National Clinical Guideline Centre

.

Multiple sclerosis

Management of multiple sclerosis in primary and secondary care

Clinical guideline 186

Methods, evidence and recommendations

October 2014

Final

Commissioned by the National Institute for Health and Care Excellence

Funding
National Institute for Health and Care Excellence

Contents

Guideline Development Group members

Name	Role
Noreen Barker	MS Clinical Nurse Specialist, The National Hospital for Neurology and Neurosurgery, London
Pamela Bostock	Consultant Occupational Therapist, Neurology, Staffordshire and Stoke On Trent Partnership NHS Trust
Peter Brex	Consultant Neurologist, Kings College Hospital NHS Foundation Trust, London
Jeremy Chataway	Consultant Neurologist, National Hospital for Neurology and Neurosurgery, UCLH NHS Foundation Trust, London
Paul Cooper (Chair)	Consultant Neurologist, Greater Manchester Neuroscience Centre, Salford Royal Foundation Trust
Aleks de Gromoboy	Patient Member
Wendy Hendrie	Specialist Physiotherapist in MS, MS Centre, Norwich
Ann Hodgson	Patient Member
Susan Hourihan	Clinical Specialist Occupational Therapist, National Hospital for Neurology and Neurosurgery, UCLH NHS Foundation Trust, London
David Kernick	General Practitioner, St Thomas Health Centre, Exeter
Emma Rowe	Patient Member
Richard Warner	MS Nurse Consultant, Gloucestershire Hospitals, NHS Foundation Trust

NCGC technical team members

Name	Role
Lola Adedokun	Health Economist (until January 2013)
Krishna Chinthapalli	Clinical Fellow
Elisabetta Fenu	Health Economics Lead (until December 2013)
Lina Gulhane	Joint Head of Information Science
Amy Kelsey	Project Manager
Sophia Kemmis Betty	Health Economist (from May 2013)
Kate Lovibond	Health Economics Lead (from January 2014)
Norma O'Flynn	Clinical Director and Guideline Lead
Mark Perry	Senior Research Fellow
Sharon Swain	Senior Research Fellow

Co-optees

Name	Role
Sarah Gillanders	MS Specialist Clinical Neuropsychologist, Astley Ainslie Hospital, Edinburgh
Paul Riordan-Eva	Consultant Ophthalmologist, Kings College Hospital, London

Acknowledgements

The development of this guideline was greatly assisted by the following people:

- Katharina Dworzynski (NCGC Senior Research Fellow)
- Jaimella Espley (NICE Senior Medical Editor)
- Caroline Keir (NICE Guideline Commissioning Manager)
- Clifford Middleton (NICE Guideline Commissioning Manager)
- Gill Ritchie (NCGC Operations Director)
- David Wonderling (NCGC Head of Health Economics)

1 Introduction

Multiple sclerosis (MS) is an acquired chronic immune-mediated inflammatory condition of the central nervous system (CNS), affecting both the brain and spinal cord. It affects approximately 100,000 people in the UK. It is the commonest cause of serious physical disability in adults of working age.

People with MS typically develop symptoms in their late 20s, experiencing visual and sensory disturbances, limb weakness, gait problems, and bladder and bowel symptoms. They may initially have partial recovery, but over time develop progressive disability. The most common pattern of disease is relapsing–remitting MS (RRMS) where periods of stability (remission) are followed by episodes when there are exacerbations of symptoms (relapses). About 85 out of 100 people with MS have RRMS at onset. Around two-thirds of people who start with RRMS may develop secondary progressive MS: this occurs when relapses are initially associated with progressively less complete recovery, then subsequently individuals gradually develop worsening symptoms without any clear remissions. Also about 10 to 15 out of 100 people with MS have primary progressive MS where symptoms gradually develop and worsen over time from the start, without ever experiencing relapses and remissions.

The cause of MS is unknown. It is believed that an abnormal immune response to environmental triggers in people who are genetically predisposed, results in immune-mediated acute, and then chronic inflammation. The initial phase of inflammation is followed by a phase of progressive degeneration of the affected cells in the nervous system. MS is a potentially highly disabling disorder with considerable personal, social and economic consequences. People with MS live for many years after diagnosis with significant impact on their ability to work, as well as an adverse and often highly debilitating effect on their quality of life and that of their families.

This guideline replaces NICE clinical guideline 8 (2003) and covers diagnosis, information and support, treatment of relapse and management of MS-related symptoms. The guideline does not address all symptoms and problems associated with MS. Some areas are addressed in other NICE guidance for example urinary symptoms and swallowing, and these are referenced where appropriate. Many of the interventions used in a rehabilitation setting to alleviate symptoms such as tremor, weakness, cardiorespiratory fitness, sensory loss, visual problems (apart from oscillopsia), and secondary complications of immobility such as deconditioning and contractures have not been covered because these are beyond the scope of the guideline. Many of these problems are complex and need individual assessment and management strategies. These assessments and treatments need to be carried out by healthcare professionals with appropriate expertise in rehabilitation and MS.

The guideline does not address the use of disease-modifying treatments; there are NICE technology appraisals about these treatments.

The guideline is aimed primarily at services provided in primary and secondary care. It does not map out a model of service delivery. Many people with MS may also attend specialised tertiary services, often established particularly to provide and monitor disease-modifying therapies.

2 Development of the guideline

2.1 What is a NICE clinical guideline?

NICE clinical guidelines are recommendations for the care of individuals in specific clinical conditions or circumstances within the NHS – from prevention and self-care through primary and secondary care to more specialised services. We base our clinical guidelines on the best available research evidence, with the aim of improving the quality of health care. We use predetermined and systematic methods to identify and evaluate the evidence relating to specific review questions.

NICE clinical guidelines can:

- provide recommendations for the treatment and care of people by health professionals
- be used to develop standards to assess the clinical practice of individual health professionals
- be used in the education and training of health professionals
- help patients to make informed decisions
- improve communication between patient and health professional.

While guidelines assist the practice of healthcare professionals, they do not replace their knowledge and skills.

We produce our guidelines using the following steps:

- Guideline topic is referred to NICE from the Department of Health.
- Stakeholders register an interest in the guideline and are consulted throughout the development process.
- The scope is prepared by the National Clinical Guideline Centre (NCGC).
- The NCGC establishes a Guideline Development Group.
- A draft guideline is produced after the group assesses the available evidence and makes recommendations.
- There is a consultation on the draft guideline.
- The final guideline is produced.

The NCGC and NICE produce a number of versions of this guideline:

- the 'full guideline' contains all the recommendations, plus details of the methods used and the underpinning evidence
- the 'NICE guideline' lists the recommendations
- 'information for the public' is written using suitable language for people without specialist medical knowledge
- NICE Pathways brings together all connected NICE guidance.

This version is the full version. The other versions can be downloaded from NICE at www.nice.org.uk.

2.2 Remit

NICE received the remit for this guideline from the Department of Health. They commissioned the NCGC to produce the guideline.

This guideline is a full replacement for multiple sclerosis (NICE clinical guideline 8).

2.3 Who developed this guideline?

A multidisciplinary Guideline Development Group (GDG) comprising health professionals and researches as well as lay members developed this guideline (see the list of Guideline Development Group members and the acknowledgements).

The National Institute for Health and Care Excellence (NICE) funds the National Clinical Guideline Centre (NCGC) and thus supported the development of this guideline. The GDG was convened by the NCGC and chaired by Dr Paul Cooper in accordance with guidance from NICE.

The group met every 6 weeks during the development of the guideline. At the start of the guideline development process all GDG members declared interests including consultancies, fee-paid work, share-holdings, fellowships and support from the healthcare industry. At all subsequent GDG meetings, members declared arising conflicts of interest.

Members were either required to withdraw completely or for part of the discussion if their declared interest made it appropriate. The details of declared interests and the actions taken are shown in Appendix B.

Staff from the NCGC provided methodological support and guidance for the development process. The team working on the guideline included a project manager, systematic reviewers, health economists and information scientists. They undertook systematic searches of the literature, appraised the evidence, conducted meta-analysis and cost effectiveness analysis where appropriate and drafted the guideline in collaboration with the GDG.

(a) What this guideline covers

Groups that will be covered

Adults who have a diagnosis of MS or possible MS or are being investigated for MS.

Key clinical issues that will be covered

- Diagnosis, assessment and information
- Disability management and rehabilitation
- Other treatments

For further details please refer to the scope in Appendix A and review questions in Section 3.1.

(b) What this guideline does not cover

Groups that will not be covered

Children and young people under the age of 18 years who have a diagnosis of MS or possible MS or are being investigated for MS.

Key clinical issues that will not be covered

- Treatment of contractures at joints
- Disease-modifying therapies

(c) Relationships between the guideline and other NICE guidance

Related NICE Health Technology Appraisals:

- Guidance on the use of computerised cognitive behavioural therapy for anxiety and depression. NICE technology appraisal 51 (2002).
- Guidance on beta interferon and glatiramer acetate for the treatment of multiple sclerosis. NICE technology appraisal 32 (2002).
- Natalizumab for the treatment of adults with highly active relapsing-remitting multiple sclerosis. NICE technology appraisal 127 (2007).

Related NICE Interventional Procedures:

- Functional electrical stimulation for drop foot of central neurological origin. NICE interventional procedure guidance 278 (2009).
- Deep brain stimulation for tremor and dystonia (excluding Parkinson's disease). NICE interventional procedure 188 (2006).
- Percutaneous venoplasty for chronic cerebrospinal venous insufficiency (CCSVI) for multiple sclerosis. NICE interventional procedure guidance 420 (2012).

Related NICE Clinical Guidelines:

- Osteoporosis: assessing the risk of fragility fracture. NICE clinical guideline 146 (2012).
- Patient experience in adult NHS services. NICE clinical guideline 138 (2012).
- Generalised anxiety disorder and panic disorder (with or without agoraphobia) in adults. NICE clinical guideline 113 (2011).
- End of life care for adults. NICE Quality Standard 13 (2011).
- Neuropathic pain – pharmacological management NICE clinical guideline 173 (2013)
- Medicines adherence. NICE clinical guideline 76 (2009).
- Depression in adults. NICE clinical guideline 90 (2009).
- The treatment and management of depression in adults with chronic physical health problems. NICE clinical guideline 91 (2009).
- Faecal incontinence. NICE clinical guideline 49 (2007).
- Nutrition support in adults. NICE clinical guideline 32 (2006).
- Infection control. NICE clinical guideline 139 (2012)
- Pressure relieving devices. NICE clinical guideline 7 (2003).
- Urinary incontinence in neurological disease. NICE clinical guideline 148 (2012)

Related NICE public health guidance:

- Behaviour change: individual approaches. NICE public health guidance 49 (2014)

Related NICE guidance currently in development:

- Pressure ulcers in primary and secondary care (update). Publication expected May 2014.

3 Methods

This chapter sets out in detail the methods used to review the evidence and to generate the recommendations that are presented in subsequent chapters. This guidance was developed in accordance with the methods outlined in the NICE guidelines manual 2012[164].

3.1 Developing the review questions and outcomes

Review questions were developed with a protocol in a PICO framework (patient, intervention, comparison and outcome) for intervention reviews, with a framework of population, prognostic factor and outcomes for prognostic reviews, and with a framework of key themes and population for qualitative reviews. This was to guide the literature searching process, critical appraisal and synthesis of evidence, and to facilitate the development of recommendations by the guideline development group (GDG). They were drafted by the NCGC technical team and refined and validated by the GDG. The questions were based on the key clinical areas identified in the scope (Appendix A).

A total of 18 review questions were identified.

Full literature searches, critical appraisals and evidence reviews were completed for all the specified review questions.

The review questions below in Table 1 are listed in chapter order.

Table 1: Review questions

Chapter	Type of review	Review questions	Health related outcomes
5	Not applicable	What are the key diagnostic criteria for the following: • Multiple sclerosis • Possible multiple sclerosis • Neuromyelitis optica • Clinically isolated syndrome	This question was not approached via a systematic review, so there were no applicable outcomes.
6	Qualitative	For adults with MS and their carers what information, education and support would they find useful?	Any information gained qualitatively from patients and carers.
7	Prognostic	Do the modifiable risk factors of exercise, vaccinations, stress, pregnancy and smoking influence progression of Multiple sclerosis?	Health related quality of life, relapse rates, patient reported outcomes, impact on carers, functional scales, cognitive function
8	Interventional	Does the use of structured assessment(s) compared with non-structured assessment(s) improve patient and carer outcomes for young people and adults with MS? What is the optimal timing of a structured assessment? What should be the frequency of a structured assessment?	Health related quality of life, patient reported outcomes, impact on carers, measures of mobility, cognitive function, psychological symptoms, hospitalisations, outpatients appointments, relapse rates, functional scales, adverse events
9	Interventional	For adults with MS and their carers what process of care has been proposed to improve coordination of care and other related health outcomes?	Health related quality of life, patient reported outcomes, impact on carers, treatment adherence, patient/carer

			satisfaction, relapse rates, relapse management, hospital admissions, length of hospital admissions, outpatient/GP attendance functional scales
10	Interventio nal	a) For adults with MS what is the clinical evidence of pharmacological management of acute relapse with steroids compared to placebo? b) If steroids are more effective than placebo, is there a difference in efficacy between IV and oral steroids? c) Is there a difference in efficacy and cost-effectiveness between steroids given at inpatients, outpatients (include day case), community or home?	Health related quality of life, patient reported outcomes, impact on carers, relapse outcomes, functional scales, cognitive function, psychological scales, adverse events
11	Interventio nal	For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of spasticity?	Health related quality of life, patient reported outcomes, impact on carers, measures of spasticity, functional scales, adverse events
11	Interventio nal	For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of mobility with fampridine?	Health related quality of life, patient reported outcomes, impact on carers, measures of mobility, Functional scales, cognitive function, psychological measures, adverse events
11	Interventio nal	For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of oscillopsia?	Health related quality of life, patient reported outcomes, impact on carers, nystagmus rating scale, nystagmus physiological measures, adverse events, relapse rates
11	Interventio nal	For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological management of emotionalism?	Health related quality of life, patient reported outcomes, impact on carers, psychological symptoms, cognitive function, adverse events
11	Interventio nal	For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of ataxia and tremor?	Health related quality of life, patient reported outcomes, impact on carers, measures of ataxia/tremor, Functional scales, cognitive function, psychological measures, relapse rates, adverse events
11	Interventio nal	For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of fatigue?	Health related quality of life, patient reported outcomes, impact on carers, measures of fatigue, functional scales, cognitive function,

			psychological symptoms, adverse events
12	Interventional	For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological management of memory and cognitive problems with neuropsychological rehabilitation?	Health related quality of life, patient reported outcomes, impact on carers, cognitive function, mood, adverse events
12	Interventional	For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological programmes (including self-management programmes) for: • Fatigue • Spasticity • Mobility • Pain • Ataxia • tremor	Measures or symptoms of fatigue, spasticity, mobility, pain, ataxia or tremor. If treatment was specifically directed at any of the six preceding outcomes, then health related quality of life, impact on carers, functional scales, and adverse events were also included.
12	Interventional	For adults with MS, what is the clinical evidence and cost effectiveness of rehabilitation provided in different settings?	Health related quality of life, impact on carers, functional scales, adverse events
13	Interventional	For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological management with vitamin D?	Health related quality of life, patient reported outcomes, impact on carers, functional scales, cognitive function, relapse rates, adverse events
13	Interventional	For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological management with complementary and alternative therapies (omega-3 fatty acids, omega-6 fatty acids, acupuncture)	Health related quality of life, patient reported outcomes, impact on carers, measures of mobility, functional scales, cognitive function, psychological symptoms, relapse rates, adverse events

3.2 Searching for evidence

3.2.1 Clinical literature search

The aim of the literature search was to systematically identify all published clinical evidence relevant to the review questions. Searches were undertaken according to the parameters stipulated within the NICE Guidelines Manual [2009]. Databases were searched using medical subject headings and free-text terms. Foreign language studies were not reviewed and, where possible, searches were restricted to articles published in the English language. All searches were conducted in MEDLINE, Embase, and the Cochrane Library, and were updated for the final time on **3rd February 2014**. No papers after this date were considered.

Search strategies were quality assured by cross-checking reference lists of highly relevant papers, analysing search strategies in other systematic reviews, and asking GDG members to highlight any additional studies. The questions, the study types applied, the databases searched and the years covered can be found in Appendix F.

The titles and abstracts of records retrieved by the searches were sifted for relevance, with potentially significant publications obtained in full text. These were then assessed against the inclusion criteria.

3.2.2 Health economic literature search

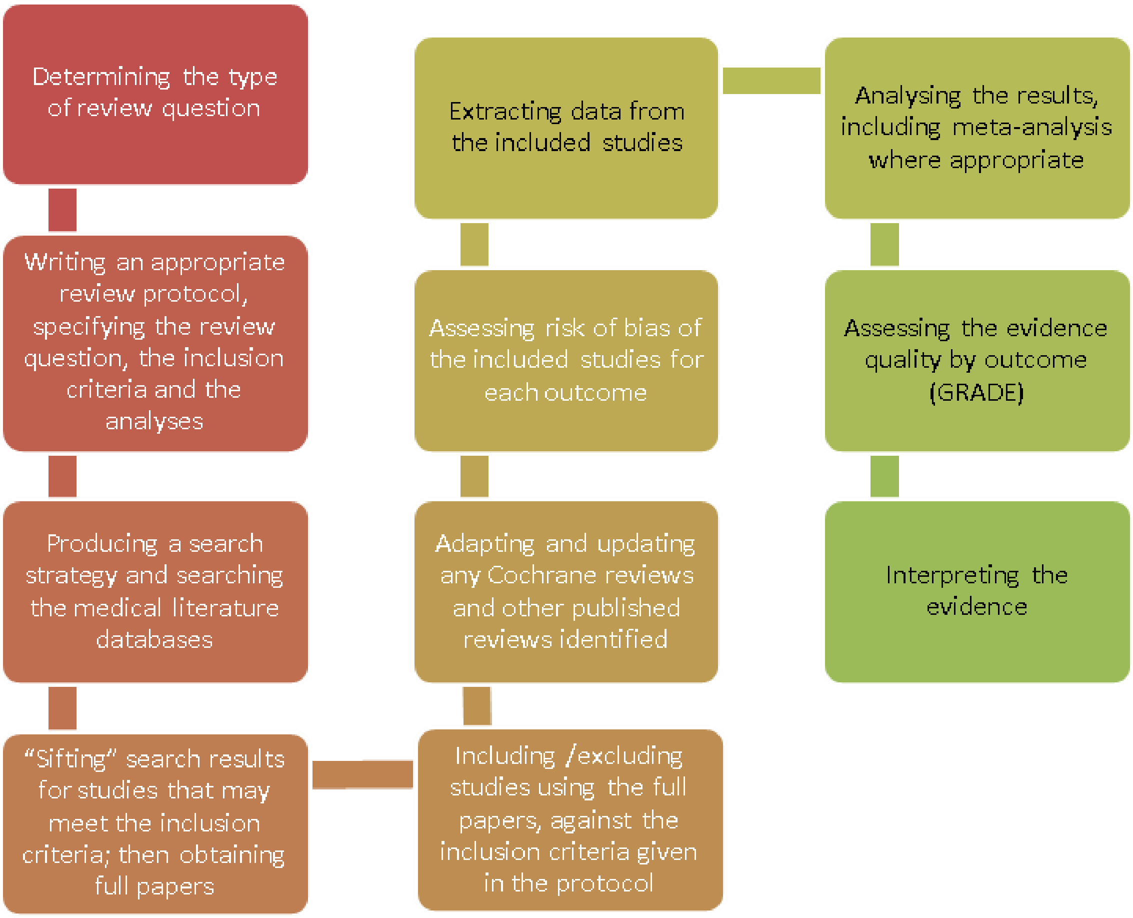
Systematic searches were undertaken to identify relevant health economic evidence within the published literature. The NHS Economic Evaluation Database (NHS EED), the Health Economic Evaluations Database (HEED) and Health Technology Assessment (HTA) database were searched using broad population terms and no date restrictions. Additionally, the search was run on MEDLINE and Embase using a specific economic filter, from 2009, to ensure recent publications that had not yet been indexed by the economic databases were identified. Where possible, searches were restricted to articles published in the English language. Economics search strategies are included in Appendix F. All searches were updated for the final time on **14th February 2014**. No papers published after this date were considered.

3.3 Evidence of effectiveness

The evidence was reviewed following the steps shown schematically in Figure 1:

- Potentially relevant studies were identified for each review question from the relevant search results by reviewing titles and abstracts. Full papers were then obtained.
- Full papers were reviewed against pre-specified inclusion and exclusion criteria to identify studies that addressed the review question in the appropriate population (review protocols are included in Appendix C). A 20% sample of the abstract lists was searched by a second reviewer to check for any potential papers that were missed. In the event of a potential missing paper being detected the entire abstract list was checked by the second reviewer.
- Relevant studies were critically appraised according to the criteria specified in the checklist in The guidelines manual.[164]
- Key information was extracted on the study's methods, PICO factors and results. These were presented in summary tables (in each review chapter) and evidence tables (in Appendix G).
- Summaries of evidence were generated by outcome (included in the relevant review chapters) and were presented in GDG meetings:
    - Randomised studies: data were meta-analysed where appropriate and reported in GRADE profiles (for intervention reviews).
    - Observational studies: data were presented as a range of values in GRADE profiles.
    - Prognostic studies: data were presented as a range of values, usually in terms of the relative effect as reported by the authors.
    - Qualitative studies: each study was summarised in a table where possible, otherwise presented in a narrative.

Figure 1: Step-by-step process of review of evidence in the guideline



3.3.1 Inclusion/exclusion criteria

The inclusion and exclusion of studies was based on the review protocols, which can be found in Appendix C. Excluded studies by review question (with the reasons for their exclusion) are listed in Appendix J. The GDG was consulted about any uncertainty regarding inclusion or exclusion.

The key population inclusion criterion was:

- Adults who have a diagnosis of MS or possible MS, or are having investigations for MS.

The key population exclusion criterion was:

- Children and young people under the age of 18 years who have a diagnosis of MS or possible MS, or are being investigated for MS.

Randomised trials, non-randomised trials, and observational studies (including diagnostic or prognostic studies) were included in the evidence reviews as appropriate.

Conference abstracts were not automatically excluded from the review but were initially assessed against the inclusion criteria and then further processed only if no other full publication was available for that review question, in which case the authors of the selected abstracts were contacted for further information.

Literature reviews, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded.

The review protocols are presented in Appendix C.

3.3.2 Methods of combining clinical studies

3.3.2.1 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted to combine the results of studies for each review question using Cochrane Review Manager (RevMan5) software.

Sometimes where a population or treatment factor (such as gender or dose) is thought to have a strong effect on the outcome of treatments, meta-analyses will be stratified from the outset for that factor. [Note that this should be differentiated from 'sub-grouping', where post-hoc meta-analyses are done separately for different strata of pre-specified factors in an attempt to reduce serious heterogeneity existing in the overall meta-analysis. This issue is dealt within the later section 'heterogeneity']. However, in this guideline, the GDG did not feel that any factor would have sufficient effect on outcome to justify prior stratification of meta-analyses.

Binary outcomes

Fixed-effects (Mantel-Haenszel) meta-analysis techniques (using an inverse variance method for pooling) were initially used to pool risk ratios (relative risk) from different studies for the binary outcomes, which included the existence/non-existence of:

- patient-assessed symptoms
- relapse
- patient satisfaction
- positive response to treatment
- subjective improvement
- adverse events

Absolute event rates were also calculated for binary outcomes with the GRADEpro software, using median event rate in the control arm of the pooled results.

For binary variables where there were zero events in either arm, Peto odds ratios, rather than risk ratios, were calculated. Peto odds ratios are more appropriate for data with a low number of events.

Where there was sufficient information provided, Hazard Ratios were calculated and/or reported for outcomes such as:

- relapse

Continuous outcomes

The continuous outcomes were meta-analysed using an inverse variance method for pooling weighted mean differences from different studies. These outcomes included:

- Heath Related Quality of Life (HRQL)
- patient assessed symptoms on a VAS or other subjective scale
- level of impact on carers
- objective measures of mobility/function/ataxia/tremor/spasticity/fatigue/pain/nystagmus
- measures of cognitive function
- psychological measures
- relapse duration

Where the studies within a single meta-analysis had different scales of measurement, standardised mean differences were used, where each different measure in each study was 'normalised' to the standard deviation value pooled between the intervention and comparator groups in that same study.

The means and standard deviations of continuous outcomes were required for meta-analysis. However, in cases where standard deviations were not reported, the standard error was calculated if the p-values or 95% confidence intervals were reported, and meta-analysis was undertaken with the mean and standard error using the generic inverse variance method in Cochrane Review Manager (RevMan5.1) software. Where p values were reported as "less than", a conservative approach was undertaken. For example, if p value was reported as "p ≤0.001", the calculations for standard deviations were based on a p value of 0.001. If these statistical measures were not available then the methods described in section 16.1.3 of the Cochrane Handbook (version 5.1.0, updated March 2011). 'Missing standard deviations' were applied as the last resort, but normally the available data would be presented in the review as 'narrative results'.

Heterogeneity

Statistical heterogeneity was assessed for the overall meta-analysis estimate by considering the chi-squared test for significance at p<0.1, or an I-squared inconsistency statistic of >50%, as indicating significant heterogeneity. Where significant heterogeneity was present, we normally carried out predefined sub-grouping of studies within the meta-analysis for:

1. type of MS: Relapsing remitting MS / Secondary progressive MS / Primary progressive MS
2. Disability: EDSS <6 / EDSS $\geq$ 6

These two strategies were applied in turn. If the 'type of MS' strategy managed to reduce heterogeneity to acceptable levels (I^2<50%) within all of the derived sub-groups, then the 'disability' strategy was not used. The latter strategy was only used if the former strategy failed to resolve heterogeneity. If either of the strategies managed to reduce I^2 to less than 50% within all the derived sub-groups, then each of the derived sub-groups were adopted as separate outcomes, pending GDG approval (for example, instead of the single outcome of '*existence of relapse*', we would now have '*existence of relapse in people with RR MS'*, '*existence of relapse in people with SP MS'* and '*existence of relapse in people with PP MS'.* Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between subgroups. Such subgroup differences were interpreted with caution since they broke randomisation and were subject to uncontrolled confounding.

For some questions different sub-grouping strategies were used, and this is documented in the individual question protocols (appendix X).

If all pre-defined strategies of sub-grouping were unable to resolve unacceptable statistical heterogeneity within each derived sub-group, then a random effects (DerSimonian and Laird) model was employed to the entire group of studies in the meta-analysis, and sub-grouping was abandoned. A random-effects model assumes a distribution of populations, rather than a single population. This leads to a widening of the confidence intervals around the overall estimate, thus providing a more realistic interpretation of the true distribution of effects across > 1 population. If, however, the GDG felt that the degree of heterogeneity was so large that meta-analysis was inappropriate, then the meta-analysis was abandoned and results were described narratively.

Special methods

Network meta-analysis was considered for the comparison of the pharmacological treatments for spasticity, but was not used because of insufficient data available for the outcomes deemed to be most relevant to clinical decision-making.

Where studies had used a cross-over design, paired continuous data were extracted where possible, and forest plots were generated in Review manager with the Generic Inverse Variance function. For cross-over study categorical data, the standard error (of the log RR) was calculated using the simplified Mantel Haenszel method for paired outcomes, when the number of subjects with an event in both interventions was known. Again, forest plots were generated in Review manager with the Generic Inverse Variance function. If paired continuous or categorical data were not available from the cross-over studies, the separate group data were analysed in the same way as data from parallel groups, on the basis that whilst this approach would tend to over-estimate CIs and thus artificially reduce study weighting, this would be a conservative effect. Where a meta-analysis contained a mixture of studies using both paired and parallel group approaches, all data were entered into Review manager using the Generic Inverse Variance function.

3.3.2.2 Data synthesis for prognostic factor reviews

Odds ratios (ORs), risk ratios (RRs) or hazard ratios (HRs), with their 95% confidence intervals (95% CIs) for the effect of the pre-specified prognostic factors were extracted from the papers. Only RCTs, pooled analysis of patient level data or prospective cohort studies were included. Retrospective cohort studies were excluded because of the likelihood that data on key confounders would not have been collected, and case-control studies were excluded because of their high risk of recall bias. Prospective cohort studies were required to have a multivariable analyses, including key confounders as identified by the GDG at the protocol stage for that outcome. Data were not combined in meta-analyses for prognostic studies.

3.3.2.3 Data synthesis for diagnostic test accuracy reviews

No diagnostic reviews were undertaken. The only review question related to diagnosis, 'what are the key diagnostic criteria for the following: multiple sclerosis, possible multiple sclerosis, neuromyelitis optica and clinically isolated syndrome?' was approached by GDG consensus rather than a formal review.

3.3.2.4 Data synthesis for qualitative reviews

Findings were synthesised narratively, often organised according to the themes discussed in the literature.

3.3.3 Type of studies

For most intervention reviews in this guideline, randomised controlled trials (RCTs) were included because they are considered the most robust type of study design that could produce an unbiased estimate of the intervention effects. If the GDG believed RCT data were not appropriate or there was limited evidence from RCTs, well-conducted non-randomised studies were included. Please refer to Appendix C for full details on the study design of studies selected for each review question. For example in the review addressing the issue of continuity of care, observational data were included because of the lack of any RCTs in the area.

For prognostic reviews, prospective and retrospective cohort studies were included. Case–control studies were not included.

Where data from observational studies were included, the GDG decided that the results for each outcome should be presented separately for each study and meta-analysis was not conducted.

3.3.4 Appraising the quality of evidence by outcomes

3.3.4.1 Interventional studies

The evidence for outcomes from the included RCT and observational studies were evaluated and presented using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group (http://www.gradeworkinggroup.org/). The software (GRADEpro) developed by the GRADE working group was used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results.

Each outcome was first examined for each of the quality elements listed and defined in Table 2.

Table 2: Description of quality elements in GRADE for intervention studies

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the patient, health care professional and assessor) and attrition bias (due to missing data causing systematic bias in the analysis).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies in the same meta-analysis.
Imprecision	Results are imprecise when studies include relatively few patients and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example a result may be consistent with both clinical benefit AND clinical harm) and thus be imprecise.
Publication bias	Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely related phenomenon is where some papers fail to report an outcome that is inconclusive, thus leading to an over-estimate of the effectiveness of that outcome.
Other issues	Sometimes randomisation may not adequately lead to group equivalence of confounders, and if so this may lead to bias, which should be taken into account. Potential conflicts of interest, often caused by excessive pharmaceutical company involvement in the publication of a study, should also be noted.

Details of how the four main quality elements (risk of bias, indirectness, inconsistency and imprecision) were appraised for each outcome are given below. Publication or other bias was only taken into consideration in the quality assessment if it was apparent.

Risk of bias

The main domains of bias for randomised controlled trials are listed in Table 3. Each outcome had its risk of bias assessed within each paper first. For each paper, if there were no risks of bias in any domain, the risk of bias was given a rating of 0. If there was risk of bias in just one domain, the risk of bias was given a 'serious' rating of -1, but if there was risk of bias in two or more domains the risk of bias was given a 'very serious' rating of -2. A weighted average score was then calculated across all studies contributing to the outcome, by taking into account the weighting of studies in the meta-analysis. For example if the heaviest-weighted studies tended to each have a score of -1 for that outcome, the overall score for that outcome would tend towards -1.

Table 3: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias –	If those enrolling patients are aware of the group to which the next enrolled patient

Limitation	Explanation
sequence generation and allocation concealment	will be allocated, either because of a non-random sequence that is predictable, or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of 1) knowledge of that participant's likely prognostic characteristics and 2) a desire for one group to do better than the other.
Performance and detection bias - Lack of patient and health care professional blinding	Patients, caregivers, those adjudicating and/or recording outcomes, and data analysts should not be aware of the arm to which patients are allocated. Knowledge of group can influence 1) the experience of the placebo effect, 2)performance in outcome measures, 3) the level of care and attention received, and 4) the methods of measurement or analysis, all of which can contribute to systematic bias.
Attrition bias	Attrition bias results from loss of data beyond a certain level (a differential of 10% between groups) which is not accounted for. Loss of data can occur when participants are compulsorily withdrawn from a group by the researchers (for example, when a per-protocol approach is used) or when participants do not attend assessment sessions. If the missing data are likely to be different from the data of those remaining in the groups, and there is a differential rate of such missing data from groups, systematic attrition bias may result.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results can also lead to bias, as this may distort the overall impression of efficacy.
Other limitations	For example: • Stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules • Use of unvalidated patient-reported outcomes • lack of washout periods to avoid carry-over effects in cross-over trials • Recruitment bias in cluster randomised trials

Indirectness

Indirectness refers to the extent to which the populations, intervention, comparisons and outcome measures are dissimilar to those defined in the inclusion criteria for the reviews. Indirectness is important when these differences are expected to contribute to a difference in effect size, or may affect the balance of harms and benefits considered for an intervention. As for risk of bias, each outcome had its indirectness assessed within each paper first. For each paper, if there were no sources of indirectness, indirectness was given a rating of 0. If there was indirectness in just one source (for example in terms of population), indirectness was given a 'serious' rating of -1, but if there was indirectness in two or more sources (for example, in terms of population and treatment) the indirectness was given a 'very serious' rating of -2. A weighted average score was then calculated across all studies contributing to the outcome, by taking into account the weights in the meta-analysis. For example if the heaviest-weighted studies tended to have an indirectness score of -1 each for that outcome, the overall score for that outcome would probably tend towards -1.

Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. When estimates of the treatment effect across studies differ widely, this suggests true differences in underlying treatment effect, which may be due to differences in populations, settings or doses. When heterogeneity existed within an outcome (Chi square $p<0.1$ or I^2 inconsistency statistic of >50%), but no plausible explanation could be found, the quality of evidence for that outcome was downgraded. Inconsistency for that outcome was given a 'serious' score of -1 if the I^2 was 50-74, and a 'very serious' score of -2 if the I^2 was 75 or more.

If inconsistency could be explained based on pre-specified subgroup analysis (that is, each sub-group had an $I^2 < 50$), the GDG took this into account and considered whether to make separate recommendations on new outcomes based on the sub-groups defined by the assumed explanatory factors. In such a situation the quality of evidence was not downgraded for those emergent outcomes.

Since the inconsistency score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

Imprecision

The criteria applied for imprecision were based on the confidence intervals for the pooled estimate of effect, and the minimal important differences (MID) for the outcome. The MIDs are the threshold for appreciable benefits and harms, separated by a zone either side of the line of no effect where there is assumed to be no clinically important effect. If either of the 95% confidence intervals of the overall estimate of effect crossed one of the MID lines, imprecision was regarded as serious and a 'serious' score of -1 was given. This was because the overall result, as represented by the span of the confidence intervals, was consistent with two interpretations as defined by the MID (for example, no clinically important effect and either clinical benefit or harm). If both MID lines were crossed by either or both of the confidence intervals then imprecision was regarded as very serious and a 'very serious' score of -2 was given. This was because the overall result was consistent with three interpretations defined by the MID (no clinically important effect and clinical benefit and clinical harm). This is illustrated in Figure 2. As for inconsistency, since the imprecision score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

The position of the MID lines is ideally determined by values as reported in the literature. "Anchor-based" methods aim to establish clinically meaningful changes in a continuous outcome variable by relating or "anchoring" them to patient-centred measures of clinical effectiveness that could be regarded as gold standards with a high level of face validity. For example, the minimum amount of change in an outcome necessary to make a patient decide that they felt their quality of life had "significantly improved" might define the MID for that outcome. MIDs in the literature may also be based on expert clinician or consensus opinion concerning the minimum amount of change in a variable deemed to affect quality of life or health. For binary variables, any MIDs reported in the literature will inevitably be based on expert consensus, as such MIDs relate to all-or-nothing population effects rather than measurable effects on an individual, as so are not amenable to patient-centred "anchor" methods.

In the absence of literature values, the alternative approach to deciding on MID levels is the "default" method, as follows:

- For categorical outcomes the MIDs are taken as RRs of 0.75 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the RR of 0.75 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit. For 'negative' outcomes such as 'bleeding', the opposite occurs, so the RR of 0.75 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm.
- For continuous outcome variables the MID is taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit will be a positive for a positive" outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a "negative" outcome (for example, a VAS pain score). Clinically significant harms will be

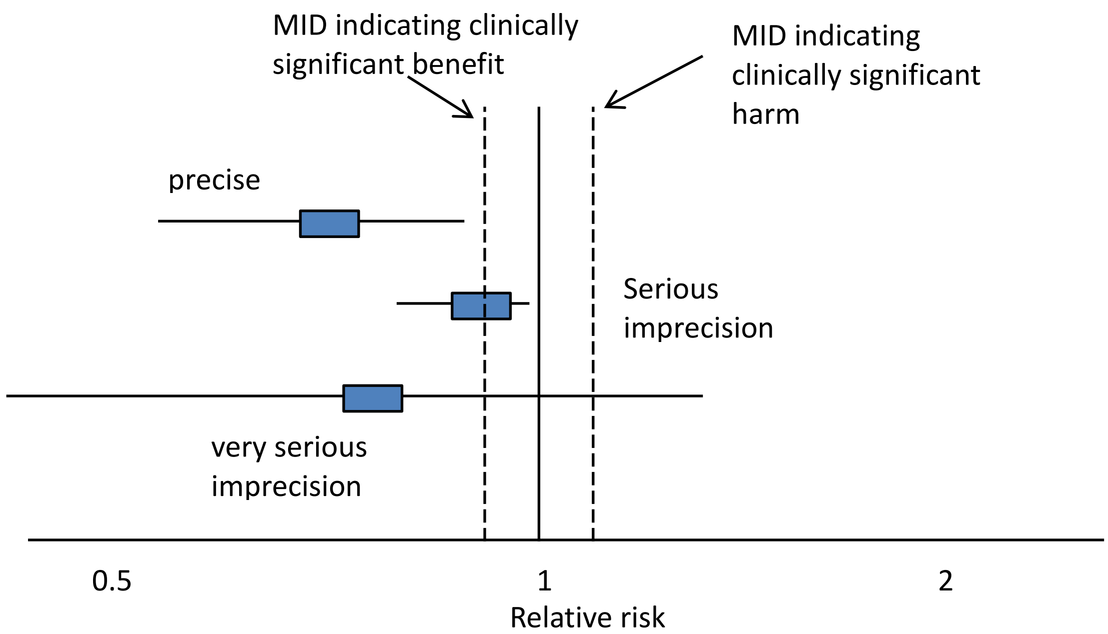
the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable will be taken as the MID.

- If standardised mean differences have been used, then the MID will be set at the absolute value of + 0.5. This follows because standardised mean differences are mean differences normalised to the pooled standard deviation of the two groups, and are thus effectively expressed in units of "numbers of standard deviation". The 0.5 MID value in this context therefore indicates half a standard deviation, the same definition of MID as used for non-standardised mean differences.

The default MID value was subject to amendment after discussion with the GDG. If the GDG decided that the MID level should be altered, after consideration of absolute as well as relative effects, this was allowed, provided that any such decision was not influenced by any bias towards making stronger or weaker recommendations for specific outcomes.

For this guideline, no appropriate MIDs for continuous or dichotomous outcomes were found in the literature, and so the default method was used.

Figure 2: Illustration of precise and imprecise outcomes based on the confidence interval of dichotomous outcomes in a forest plot. Note that all three results would be pooled estimates, and would not, in practice, be placed on the same forest plot.



Overall grading of the quality of clinical evidence

Once an outcome had been appraised for the main quality elements, as above, an overall quality grade was calculated for that outcome. The scores from each of the main quality elements (0, -1 or -

2) were summed to give a score that could be anything from 0 (the best possible) to -8 (the worst possible). However scores were capped at -3. This final score was then applied to the starting grade that had originally been applied to the outcome by default, based on study design. For example, all RCTs started as HIGH and the overall quality became MODERATE, LOW or VERY LOW if the overall score was -1, -2 or -3 points respectively. The significance of these overall ratings is explained in Table 3. The reasons or criteria used for downgrading were specified in the footnotes of the GRADE tables.

On the other hand, observational interventional studies started at LOW, and so a score of -1 would be enough to take the grade to the lowest level of VERY LOW. Observational studies could, however, be upgraded if there was: a large magnitude of effect, a dose-response gradient, and if all plausible confounding would reduce a demonstrated effect.

Table 4: Overall quality of outcome evidence in GRADE

Level	Description
High	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	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	Any estimate of effect is very uncertain

3.3.4.2 Prognostic studies

The quality of evidence for prognostic studies was evaluated according to the criteria given in Table 5.

Table 5: Description of quality elements for prospective studies

Quality element	Description of cases where the quality measure would be downgraded
Study design	If case control rather than prospective cohort
Patient recruitment	If potential for selection bias
Validity of risk factor measure(s)	If non-validated and no reasonable face validity
Validity of outcome measure	If non-validated and no reasonable face validity
Blinding	if assessors of outcome not blinded to risk factor measurement (or vice versa)
Adequate follow up (or retrospective) duration	If follow up/retrospective period inadequate to allow events to occur, or retrospective period so short that causality is in doubt because the outcome may have preceded the risk factor
Confounder consideration	If there is a lack of consideration of all reasonable confounders in a multivariable analysis
Attrition	If attrition is too high and there is no attempt to adjust for this.
Directness	If the population, risk factors or outcome differ from that in the review question.

Because prognostic reviews were not usually based on multiple outcomes per study, quality rating was assigned by study. However if there was more than one outcome involved in a study, then the quality rating of the evidence statements for each outcome was adjusted accordingly. For example, if one outcome was based on an invalidated measurement method, but another outcome in the same study wasn't, the latter outcome would be graded one grade higher than the other.

Quality rating started at HIGH for prospective studies, and each major limitation (Table 5) brought the rating down by one increment to a minimum grade of LOW, as explained for interventional studies. For prognostic studies prospective cohort studies with a multivariate analysis are regard as the gold standard because RCTs are usually inappropriate for these types of review.

3.3.4.3 Qualitative reviews

Qualitative data provides information of people's thoughts, feelings, attitudes and beliefs. As such data is necessarily subjective, there is no requirement for it to be representative of the wider population; instead it is framed in the unique context of the individual respondent. Nevertheless, these data need to be trustworthy in terms of accurately reflecting the actual opinions of the respondent.

Quality was assessed using a modified version of the NICE qualitative studies appraisal framework, which can be found at Appendix I (document pages 208-217; pdf pages: 61-70) in:

http://www.nice.org.uk/media/A67/3C/The_guidelines_manual_2009_-_All_appendices.pdf

Issues covered by this quality assessment were:

- Rigour of the research methodology
- Quality of data collection
- Clear description of role of researcher
- Clear description of context
- Trustworthy data collection methods
- Rigorous analysis methods
- Richness of data
- Trustworthy data analysis methods
- Convincing findings
- Relevance to the aims of the study

This quality assessment was carried out independently by two systematic reviewers who discussed findings to reach consensus.

3.3.5 Assessing clinical importance

The GDG assessed the evidence by outcome in order to determine if there was, or potentially was, a clinically important benefit, a clinically important harm or no clinically important difference between interventions. To facilitate this, binary outcomes were converted into absolute risk differences (ARDs) using GRADEpro software: the median control group risk across studies was used to calculate the ARD and its 95% CI from the pooled risk ratio.

The assessment of benefit, harm, or no benefit or harm was based on the point estimate of absolute effect for intervention studies which was standardised across the reviews. The GDG considered for most of the outcomes in the intervention reviews that if at least 100 participants per 1000 (10%) achieved (if positive) the outcome of interest in the intervention group compared to the comparison group then this intervention would be considered beneficial. The same point estimate but in the opposite direction would apply if the outcome was negative. For adverse events 50 participants or more per 1000 was considered to be a clinical harm.

For continuous outcomes clinical benefit, harm or no harm was based on whether the mean difference was greater than the minimally important difference.

This assessment was carried out by the GDG for each critical outcome, and an evidence summary table was produced to compile the GDG's assessments of clinical importance per outcome, alongside the evidence quality and the uncertainty in the effect estimate (imprecision).

At the protocol stage, each outcome was assigned a rating of 'critical' or 'important', to inform prioritisation of outcomes in decision making. A 'critical' outcome was defined as one that would be vital in informing a recommendation, and an 'important' outcome was defined as one that would be useful, but not vital, in informing a recommendation. The rationale for having 'important' outcomes was that sometimes critical outcomes might not be available for a particular outcome.

3.3.6 Clinical evidence statements

Clinical evidence statements are summary statements that are presented after the GRADE profiles, summarising the key features of the clinical effectiveness evidence presented. The wording of the evidence statements reflects the certainty/uncertainty in the estimate of effect. The evidence statements were presented by outcome and encompassed the following key features of the evidence:

- The number of studies and the number of participants for a particular outcome
- An indication of the direction of clinical importance (if one treatment is beneficial or harmful compared to the other or whether there is no difference between the two tested treatments).
- A description of the overall quality of evidence (GRADE overall quality).

3.4 Evidence of cost effectiveness

The GDG is required to make decisions based on the best available evidence of both clinical and cost effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost effectiveness') rather than the total implementation cost.[164] Thus, if the evidence suggests that a strategy provides significant health benefits at an acceptable cost per patient treated, it should be recommended even if it would be expensive to implement across the whole population.

Evidence on cost effectiveness related to the key clinical issues being addressed in the guideline was sought. The health economist:

- Undertook a systematic review of the published economic literature.
- Undertook new cost-effectiveness analysis in priority areas.

If both meta-analysed and narratively reported outcomes were reported, evidence statements were produced only for the meta-analysed data.

3.4.1 Literature review

The health economist:

- Identified potentially relevant studies for each review question from the economic search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against pre-specified inclusion and exclusion criteria to identify relevant studies (see below for details).

- Critically appraised relevant studies using the economic evaluations checklist as specified in the guidelines manual.[164]
- Extracted key information about the studies' methods and results into evidence tables (included in Appendix H).
- Generated summaries of the evidence in NICE economic evidence profiles (included in the relevant chapter for each review question) – see below for details.

3.4.1.1 Inclusion and exclusion criteria

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost–utility, cost-effectiveness, cost–benefit and cost–consequence analyses) and comparative costing studies that addressed the review question in the relevant population were considered potentially includable as economic evidence.

Studies that only reported cost per hospital (not per patient), or only reported average cost effectiveness without disaggregated costs and effects, were excluded. Literature reviews, abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded.

Remaining studies were prioritised for inclusion based on their relative applicability to the development of this guideline and the study limitations. For example, if a high quality, directly applicable UK analysis was available, then other less relevant studies may not have been included. Where exclusions occurred on this basis, this is noted in the relevant section.

For more details about the assessment of applicability and methodological quality see the economic evaluation checklist (Appendix F of The guidelines manual.[164] and the health economics review protocol in Appendix C).

3.4.1.2 NICE economic evidence profiles

The NICE economic evidence profile has been used to summarise cost and cost-effectiveness estimates. The economic evidence profile shows an assessment of applicability and methodological quality for each economic evaluation, with footnotes indicating the reasons for the assessment. These assessments were made by the health economist using the economic evaluation checklist from The guidelines manual.[164] It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-effectiveness ratio for the base case analysis in the evaluation, as well as information about the assessment of uncertainty in the analysis. See Table 6 for more details.

If a non-UK study was included in the profile, the results were converted into pounds sterling using the appropriate purchasing power parity.[175]

Table 6: Content of NICE economic evidence profile

Item	Description
Study	First author name, reference, date of study publication and country perspective.
Applicability	An assessment of applicability of the study to the clinical guideline, the current NHS situation and NICE decision-making[(a)]: • Directly applicable – the study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness. • Partially applicable – the study fails to meet one or more applicability criteria, and this could change the conclusions about cost effectiveness. • Not applicable – the study fails to meet one or more of the applicability criteria,

Item	Description
	and this is likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.
Limitations	An assessment of methodological quality of the study[(a)]: • Minor limitations – the study meets all quality criteria, or fails to meet one or more quality criteria, but this is unlikely to change the conclusions about cost effectiveness. • Potentially serious limitations – the study fails to meet one or more quality criteria, and this could change the conclusions about cost effectiveness. • Very serious limitations – the study fails to meet one or more quality criteria, and this is highly likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.
Other comments	Particular issues that should be considered when interpreting the study.
Incremental cost	The mean cost associated with one strategy minus the mean cost of a comparator strategy.
Incremental effects	The mean QALYs (or other selected measure of health outcome) associated with one strategy minus the mean QALYs of a comparator strategy.
Cost effectiveness	Incremental cost-effectiveness ratio (ICER): the incremental cost divided by the incremental effects.
Uncertainty	A summary of the extent of uncertainty about the ICER reflecting the results of deterministic or probabilistic sensitivity analyses, or stochastic analyses of trial data, as appropriate.

(a) Applicability and limitations were assessed using the economic evaluation checklist in Appendix G of The guidelines manual (2012)[164]

Where economic studies compare multiple strategies, results are presented in the economic evidence profiles for the pair-wise comparison specified in the review question, irrespective of whether or not that comparison was 'appropriate' within the analysis being reviewed. A comparison is 'appropriate' where an intervention is compared with the next most expensive non-dominated option – a clinical strategy is said to 'dominate' the alternatives when it is both more effective and less costly. Footnotes indicate if a comparison was 'inappropriate' in the analysis.

3.4.2 Undertaking new health economic analysis

As well as reviewing the published economic literature for each review question, as described above, new economic analysis was undertaken by the health economist in selected areas. Priority areas for new health economic analysis were agreed by the GDG after formation of the review questions and consideration of the available health economic evidence.

The GDG identified pharmacological management of mobility with fampridine as the highest priority area for original economic modelling. Fampridine is not widely used as it is a relatively new therapy. There are currently no drug alternatives to fampridine therefore the potential impact on resources would be huge if there is an increased uptake. The clinical review identified studies comparing fampridine and placebo but no published cost effectiveness were identified. Therefore an original cost utility analysis comparing fampridine to placebo was conducted.

The following general principles were adhered to in developing the cost-effectiveness analysis:

- Methods were consistent with the NICE reference case.[165]
- The GDG was involved in the design of the model, selection of inputs and interpretation of the results.

- Model inputs were based on the systematic review of the clinical literature supplemented with other published data sources where possible.
- When published data was not available GDG expert opinion was used to populate the model.
- Model inputs and assumptions were reported fully and transparently.
- The results were subject to sensitivity analysis and limitations were discussed.
- The model was peer-reviewed by another health economist at the NCGC.

Full methods for the cost-effectiveness analysis for the pharmacological management of mobility with fampridine are described in chapter 9.2.

3.4.3 Cost-effectiveness criteria

NICE's report 'Social value judgements: principles for the development of NICE guidance' sets out the principles that GDGs should consider when judging whether an intervention offers good value for money.[163] In general, an intervention was considered to be cost effective if either of the following criteria applied (given that the estimate was considered plausible):

- the intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- the intervention cost less than £20,000 per QALY gained compared with the next best strategy.

If the GDG recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in the 'Recommendations and link to evidence' section of the relevant chapter, with reference to issues regarding the plausibility of the estimate or to the factors set out in 'Social value judgements: principles for the development of NICE guidance'.[163]

3.4.4 In the absence of economic evidence

When no relevant published studies were found, and a new analysis was not prioritised, the GDG made a qualitative judgement about cost effectiveness by considering expected differences in resource use between options and relevant UK NHS unit costs, alongside the results of the clinical review of effectiveness evidence. Where feasible and deemed useful to inform consideration of cost-effectiveness, outcomes reported in the clinical review were mapped to EQ-5D using published algorithms allowing for QALYs to be estimated.

3.5 Developing recommendations

Over the course of the guideline development process, the GDG was presented with:

- Evidence tables of the clinical and economic evidence reviewed from the literature. All evidence tables are in Appendices G and H.
- Summary of clinical and economic evidence and quality (as presented in Chapters **Error! Reference source not found.**-13).
- Forest plots and summary ROC curves (Appendix I).
- A description of the methods and results of the cost-effectiveness analysis undertaken for the guideline (chapter 9.2).

Recommendations were drafted on the basis of the GDG interpretation of the available evidence, taking into account the balance of benefits, harms and costs. When clinical and economic evidence was of poor quality, conflicting or absent, the GDG drafted recommendations based on their expert opinion. The considerations for making consensus based recommendations include the balance between potential harms and benefits, economic or implications compared to the benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were done through discussions in the GDG. The GDG also considered whether the uncertainty was sufficient to justify delaying making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation (See 4.3).

The main considerations specific to each recommendation are outlined in the Evidence to Recommendation Section preceding the recommendation section.

3.5.1 Research recommendations

When areas were identified for which good evidence was lacking, the GDG considered making recommendations for future research. Decisions about inclusion were based on factors such as:

- the importance to patients or the population
- national priorities
- potential impact on the NHS and future NICE guidance
- ethical and technical feasibility.

3.5.2 Validation process

This guidance is subject to a 6-week public consultation and feedback as part of the quality assurance and peer review of the document. All comments received from registered stakeholders are responded to in turn and posted on the NICE website when the pre-publication check of the full guideline occurs.

3.5.3 Updating the guideline

A formal review of the need to update a guideline is usually undertaken by NICE after its publication. NICE will conduct a review to determine whether the evidence base has progressed significantly to alter the guideline recommendations and warrant an update.

3.5.4 Disclaimer

Health care providers need to use clinical judgement, knowledge and expertise when deciding whether it is appropriate to apply guidelines. The recommendations cited here are a guide and may not be appropriate for use in all situations. The decision to adopt any of the recommendations cited here must be made by practitioners in light of individual patient circumstances, the wishes of the patient, clinical expertise and resources.

The National Clinical Guideline Centre disclaims any responsibility for damages arising out of the use or non-use of this guideline and the literature used in support of this guideline.

3.5.5 Funding

The National Clinical Guideline Centre was commissioned by the National Institute for Health and Care Excellence to undertake the work on this guideline.

4 Guideline summary

4.1 Key priorities for implementation

From the full set of recommendations, the GDG selected 8 key priorities for implementation. The criteria used for selecting these recommendations are listed in detail in the guidelines manual.[164] The reasons that each of these recommendations were chosen are shown in the table linking the evidence to the recommendation in the relevant chapter.

Diagnosing MS

- Refer people suspected of having MS to a consultant neurologist. Speak to the consultant neurologist if you think a person needs to be seen urgently.
- Only a consultant neurologist should make the diagnosis of MS on the basis of established up-to-date criteria, such as the revised 2010 McDonald criteria[a], after:
    - assessing that episodes are consistent with an inflammatory process
    - excluding alternative diagnoses
    - establishing that lesions have developed at different times and are in different anatomical locations for a diagnosis of relapsing-remitting MS
    - establishing progressive neurological deterioration over 1 year or more for a diagnosis of primary progressive MS.
- Do not diagnose MS on the basis of MRI findings alone

Information and support

- The consultant neurologist should ensure that people with MS and, with their agreement their family members or carers, are offered oral and written information at the time of diagnosis. This should include, but not be limited to, information about:
    - what MS is
    - treatments, including disease-modifying therapies
    - symptom management
    - how support groups, local services, social services and national charities are organised and how to get in touch with them
    - legal requirements such as notifying the Driver and Vehicle Licensing Agency (DVLA) and legal rights including social care, employment rights and benefits
- Offer the person with MS a face-to-face follow-up appointment with a healthcare professional with expertise in MS to take place within 6 weeks of diagnosis

Coordination of care

- Care for people with MS using a coordinated multidisciplinary approach. Involve professionals who can best meet the needs of the person with MS and who have expertise in managing MS including:

a Polman CH, Reingold SC, Banwell B et al. (2011) Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of Neurology 69: 292–302.

- **c**onsultant neurologists
- MS nurses
- **p**hysiotherapists and occupational therapists
- **s**peech and language therapists, psychologists, dietitians, social care and continence specialists
- GPs

Non pharmacological treatment

- Consider supervised exercise programmes involving moderate progressive resistance training and aerobic exercise to treat people with MS who have mobility problems and/or fatigue.

Treating acute relapse of MS with steroids

Treating a relapse

- Offer treatment for relapse of MS with oral methylprednisolone 0.5 g daily for 5 days.

4.2 Full list of recommendations

The recommendations below are listed in chapter order. Please note that this order differs from the NICE version of the guideline.

- **Be aware that clinical presentations in multiple sclerosis (MS) include:**
  - loss or reduction of vision in 1 eye with painful eye movements
  - double vision
  - ascending sensory disturbance and/or weakness
  - problems with balance, unsteadiness or clumsiness
  - altered sensation travelling down the back and sometimes into the limbs when bending the neck forwards (Lhermitte's symptom).

- **Be aware that usually people with MS present with neurological symptoms or signs as described in recommendation 1 and:**
  - are often aged under 50 **and**
  - may have a history of previous neurological symptoms **and**
  - have symptoms that have evolved over more than 24 hours **and**
  - have symptoms that may persist over several days or weeks and then improve.

- **Do not routinely suspect MS if a person's main symptoms are fatigue, depression or dizziness unless they have a history or evidence of focal neurological symptoms or signs.**

- **Before referring a person suspected of having MS to a neurologist, exclude alternative diagnoses by performing blood tests including:**
  - full blood count

- o inflammatory markers for example erythrocyte sedimentation rate, C-reactive protein
- o liver function tests
- o renal function tests
- o calcium
- o glucose
- o thyroid function tests
- o vitamin B_{12}
- o HIV serology.

- **Do not diagnose MS on the basis of MRI findings alone.**

- **Refer people suspected of having MS to a consultant neurologist. Speak to the consultant neurologist if you think a person needs to be seen urgently.**

- **Only a consultant neurologist should make the diagnosis of MS on the basis of established up-to-date criteria, such as the revised 2010 McDonald criteria[b], after**
- o assessing that episodes are consistent with an inflammatory process
- o excluding alternative diagnoses
- o establishing that lesions have developed at different times and are in different anatomical locations for a diagnosis of relapsing-remitting MS
- o establishing progressive neurological deterioration over 1 year or more for a diagnosis of primary progressive MS.

- **If a person is suspected[c] of having MS but does not fulfil the diagnostic criteria, plan a review. Discuss the timing of the review with the person and ensure they know who to contact for advice if they develop further neurological symptoms or if current symptoms worsen.**

- **Offer people suspected of having MS information about support groups and national charities.**

- **If a person has an episode of isolated optic neuritis, confirmed by an ophthalmologist, refer them to a consultant neurologist for further assessment.**

- **Diagnosis of neuromyelitis optica should be made by an appropriate specialist based on established up-to-date criteria.**

- **NICE has produced guidance on the components of good patient experience in adult NHS services. This includes recommendations on communication, information and coordination**

---

b Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011 Feb;69(2):292-302

c Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011 Feb;69(2):292-302

of care. Follow the recommendations in Patient experience in adult NHS services (NICE clinical guideline 138).

- **The consultant neurologist should ensure that people with MS and, with their agreement their family members or carers, are offered oral and written information at the time of diagnosis. This should include, but not be limited to, information about:**
  - what MS is
  - treatments, including disease-modifying therapies
  - symptom management
  - how support groups, local services, social services and national charities are organised and how to get in touch with them
  - legal requirements such as notifying the Driver and Vehicle Licensing Agency (DVLA) and legal rights including social care, employment rights and benefits.

- **Discuss with the person with MS and their family members or carers whether they have social care needs and if so refer them to social services for assessment. Ensure the needs of children of people with MS are addressed.**

- **Offer the person with MS a face-to-face follow-up appointment with a healthcare professional with expertise in MS to take place within 6 weeks of diagnosis.**

- **Review information, support and social care needs regularly. Continue to offer information and support to people with MS or their family members or carers even if this has been declined previously.**

- **Ensure people with MS and their family members or carers have a management plan that includes who to contact if their symptoms change significantly.**

- **Explain to people with MS that the possible causes of symptom changes include:**
  - another illness such as an infection
  - further relapse
  - change of disease status (for example progression)

- **Talk to people with MS and their family members or carers about the possibility that the condition might lead to cognitive problems.**

- **When appropriate, explain to the person with MS (and their family members or carers if the person wishes) about advance care planning and power of attorney.**

- **Care for people with MS using a coordinated multidisciplinary approach. Involve professionals who can best meet the needs of the person with MS and who have expertise in managing MS including:**
  - **c**onsultant neurologists
  - MS nurses

- o physiotherapists and occupational therapists
- o speech and language therapists, psychologists, dietitians, social careand continence specialists
- o GPs.

- **Offer the person with MS an appropriate single point of contact to coordinate care and help them access services.**

- **Encourage people with MS to exercise. Advise them that regular exercise may have beneficial effects on their MS and does not have any harmful effects on their MS.**

- **Be aware that live vaccinations may be contraindicated in people with MS who are being treated with disease-modifying therapies.**

- **Discuss with the person with MS:**
- o the possible benefits of flu vaccination **and**
- o the possible risk of relapse after flu vaccination if they have relapsing–remitting MS.

- **Offer flu vaccinations to people with MS in accordance with national guidelines, which recommend an individualised approach according to the person's needs[d] .**

- **Explain to women of childbearing age with MS that:**
- o relapse rates may reduce during pregnancy and may increase 3-6 months after childbirth before returning to pre-pregnancy rates
- o pregnancy does not increase the risk of progression of disease.

- **If a person with MS is thinking about pregnancy, give them the opportunity to talk with a healthcare professional with knowledge of MS about:**
- o fertility
- o the risk of the child developing MS
- o use of vitamin D before conception and during pregnancy
- o medication use in pregnancy
- o pain relief during delivery (including epidurals)
- o care of the child
- o breastfeeding

- **Advise people with MS not to smoke and explain that it may increase the progression of disability (see Smoking cessation services NICE public health guideline 10).**

- **Determine how often the person with MS will need to be seen based on:**

d 'Chronic neurological disease: conditions in which respiratory function may be compromised, due to neurological disease (e.g. polio syndrome sufferers). Clinicians should consider on an individual basis the clinical needs of patients including individuals with cerebral palsy, multiple sclerosis and related or other similar conditions; or hereditary and degenerative disease of the nervous system of muscles; or severe disability' (Department of Health 2013).

- o their needs, and those of their family and carers **and**
- o the frequency of visits needed for different types of treatment (such as review of disease-modifiying therapies, rehabilitation and symptom management)

- **Assess and offer treatment to people with MS who have fatigue for anxiety, depression, difficulty in sleeping, and any potential medical problems such as anaemia or thyroid disease.**

- **Explain that MS-related fatigue may be precipitated by heat, overexertion and stress or may be related to the time of day.**

- **Offer amantadine[e] to treat fatigue in people with MS.**

- **Consider mindfulness-based training, cognitive behavioural therapy or fatigue management for treating MS-related fatigue.**

- **Advise people that aerobic, balance and stretching exercises including yoga may be helpful in treating MS-related fatigue.**

- **Do not use vitamin B_{12} injections to treat fatigue in people with MS.**

- **Consider a comprehensive programme of aerobic and moderate progressive resistance activity combined with cognitive behavioural techniques for fatigue in people with MS with moderately impaired mobility (an EDSS[f] score of greater than or equal to 4).**

- **Ensure people with MS and mobility problems have access to an assessment to establish individual goals and discuss ways in which to achieve them. This would usually involve rehabilitation specialists and physiotherapists with expertise in MS.**

- **Do not use fampridine to treat lack of mobility in people with MS because it is not a cost effective treatment[g].**

- **Consider supervised exercise programmes involving moderate progressive resistance training and aerobic exercise to treat people with MS who have mobility problems and/or fatigue.**

- **Consider vestibular rehabilitation for people with MS who have fatigue or mobility problems associated with limited standing balance.**

e At the time of publication (October 2014), amantadine did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

f Expanded Disability Status Scale.

g This recommendation does not apply to people who have already started treatment with fampridine in the NHS who should be able to continue treatment until they and their NHS clinician think it appropriate to stop

- **Encourage people with MS to keep exercising after the treatment programmes end for longer term benefits (see Behaviour change: individual approaches NICE public health guideline 49).**

- **Help the person with MS continue to exercise, for example by referring them to exercise referral schemes.**

- **If more than one of the interventions recommended for mobility or fatigue are suitable, offer treatment based on which the person prefers and whether they can continue the activity after the treatment programme ends.**

- **In people with MS, assess and offer treatment for factors that may aggravate spasticity such as constipation, urinary tract or other infections, inappropriately fitted mobility aids, pressure ulcers, posture and pain.**

- **Encourage people with MS to manage their own spasticity symptoms by explaining how doses of drugs can be adjusted within agreed limits.**

- **Ensure that the person with MS:**
  - has tried the drug at an optimal dose, or the maximum dose they can tolerate
  - stops the drug if there is no benefit at the maximum tolerated dose
  - has their drug treatment reviewed at least annually once the optimal dose has been reached.

- **Consider baclofen or gabapentin[h] as a first-line drug to treat spasticity in MS depending on contraindications and the person's comorbidities and preferences. If the person with MS cannot tolerate one of these drugs consider switching to the other.**

- **Consider a combination of baclofen and gabapentin[ij] for people with MS if:**
  - individual drugs do not provide adequate relief **or**
  - side effects from individual drugs prohibit the dose being increased.

- **Consider tizanidine or dantrolene as a second-line option to treat spasticity in people with MS.**

---

h At the time of publication (October 2014), gabapentin did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

i At the time of publication (October 2014), gabapentin did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

j Use caution when using gabapentin and baclofen in combination. For more information on cautions for these drugs see the summary of product characteristics for gabapentin and baclofen and the British National Formulary.

- **Consider benzodiazepines as a third-line option to treat spasticity in MS and be aware of their potential benefit in treating nocturnal spasms.**

- **Do not offer Sativex[k] to treat spasticity in people with MS because it is not a cost effective treatment.**

- **If spasticity cannot be managed with any of the above pharmacological treatments, refer the person to specialist spasticity services.**

- **Consider gabapentin[l] as a first line drug to treat oscillopsia in people with MS.**

- **Consider memantine[m] as the second-line treatment for oscillopsia in people with MS.**

- **Refer the person with MS for specialist advice if there is no improvement of oscillopsia after treatment with gabapentin and memantine or side effects prevent continued use.**

- **Consider amitriptyline[n] to treat emotional lability[o] in people with MS.**

- **Treat neuropathic pain in people with MS according to Neuropathic pain – pharmacological management (NICE clinical guideline 173) and refer to pain services if appropriate.**

- **Be aware that musculoskeletal pain is common in people with MS and is usually secondary to problems with mobility and posture. Assess musculoskeletal pain, offer treatment to the person and refer them as appropriate.**

- **Be aware that the symptoms of MS can include cognitive problems, including memory problems that the person may not immediately recognise or associate with their MS.**

- **Be aware that anxiety, depression, difficulty in sleeping and fatigue can impact on cognitive problems. If a person with MS experiences these symptoms and has problems with memory and cognition, offer them an assessment and treatment.**

---

k This recommendation does not apply to people who have already started treatment with Sativex in the NHS who should be able to continue treatment until they and their NHS clinician think it appropriate to stop.

l At the time of publication (October 2014), gabapentin did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

m At the time of publication (October 2014), memantine did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

n At the time of publication (October 2014), amitriptyline did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

o Involuntary laughing and crying related to a brain stem lesion.

- **Consider referring people with MS and persisting memory or cognitive problems to both an occupational therapist and a neuropsychologist to assess and manage these symptoms.**

- **Ensure all people with MS have a comprehensive review of all aspects of their care at least once a year.**

- **Ensure the comprehensive review is carried out by healthcare professionals with expertise in MS and its complications. Involve different healthcare professionals with expertise in specific areas of the review if needed.**

- **Tailor the comprehensive review to the needs of the person with MS assessing:**

<u>MS symptoms</u>

- mobility and balance including falls
- need for mobility aids including wheelchair assessment
- use of arms and hands
- muscle spasms and stiffness
- tremor bladder (see Urinary incontinence in neurological disease NICE clinical guideline 148), bowel (see Faecal incontinence NICE clinical guideline 49) and sexual function
- sensory symptoms and pain
- speech and swallowing (see Nutrition support in adults NICE clinical guideline 32)
- vision
- cognitive symptoms
- fatigue
- depression (see Depression in adults with chronic physical health problems NICE clinical guideline 91) and anxiety (see Generalised anxiety disorder and panic disorder NICE clinical guideline 113)
- sleep
- respiratory function.

<u>MS disease course</u>

- relapses in last year.

<u>General health</u>

- weight
- smoking, alcohol and recreational drugs
- exercise
- access to routine health screening and contraception
- care of other chronic conditions.

<u>Social activity and participation</u>

- family and social circumstances
- driving and access to transport
- employment
- access to daily activities and leisure.

<u>Care and carers</u>

- personal care needs

- o social care needs
- o access to adaptations and equipment at home.

- **Refer any issues identified during the comprehensive review of the person with MS to members of the MS multidisciplinary team and other appropriate teams so that they can be managed.**

- **Ensure people with MS are offered a medication review in line with Medicines adherence (NICE clinical guideline 76).**

- **Ensure people with MS have their bone health regularly assessed and reviewed in line with Osteoporosis: assessing the risk of fragility fracture (NICE clinical guideline 146).**

- **Check people with MS and severely reduced mobility at every contact for areas at risk of pressure ulcers (see Pressure ulcers NICE clinical guideline 179).**

- **Discuss the care provided by carers and care workers as part of the person's care plan. Ensure carers know about their right to a local authority carer's assessment and how to apply for one.**

- **Refer people with MS to palliative care services for symptom control and for end of life care when appropriate.**

- **Develop local guidance and pathways for timely treatment of relapses of MS. Ensure follow-up is included in the guidance and pathway.**

- **Non-specialists should discuss a person's diagnosis of relapse and whether to offer steroids with a healthcare professional with expertise in MS because not all relapses need treating with steroids.**

- **Diagnose a relapse of MS if the person:**
- o develops new symptoms **or**
- o has worsening of existing symptoms

  and these last for more than 24 hours in the absence of infection or any other cause after a stable period of at least 1 month.

- **Before diagnosing a relapse of MS:**
- o rule out infection – particularly urinary tract and respiratory infections **and**
- o discriminate between the relapse and fluctuations in disease or progression.

- **Assess and offer treatment for relapses of MS, that affect the person's ability to perform their usual tasks, as early as possible and within 14 days of onset of symptoms**

- **Do not routinely diagnose a relapse of MS if symptoms are present for more than 3 months.**

- **Offer treatment for relapse of MS with oral methylprednisolone 0.5 g daily for 5 days.**

- **Consider intravenous methylprednisolone 1 g daily for 3-5 days as an alternative for people with MS:**
  - in whom oral steroids have failed or not been tolerated **or**
  - who need admitting to hospital for a severe relapse or monitoring of medical or psychological conditions such as diabetes or depression.

- **Do not prescribe steroids at lower doses than methylprednisolone 0.5 g daily for 5 days to treat an acute relapse of MS.**

- **Do not give people with MS a supply of steroids to self-administer at home for future relapses.**

- **Discuss the benefits and risks of steroids with the person with MS, taking into account the effect of the relapse on the person's ability to perform their usual tasks and their wellbeing.**

- **Explain the potential complications of high-dose steroids, for example temporary effects on mental health (such as depression, confusion and agitation) and worsening of blood glucose control in people with diabetes.**

- **Give the person with MS and their family members or carers (as appropriate) information that they can take away about side effects of high-dose steroids in a format that is appropriate for them.**

- **Ensure that the MS multidisciplinary team is told that the person is having a relapse, because relapse frequency may influence which disease-modifying therapies are chosen and whether they need to be changed.**

- **Identify whether the person having a relapse of MS or their family members or carers have social care needs and if so refer them to social services for assessment.**

- **Offer inpatient treatment to the person having a relapse of MS if their relapse is severe or if it is difficult to meet their medical and social care needs at home.**

- **Explain that a relapse of MS may have short-term effects on cognitive function.**

- **Identify whether the person with MS having a relapse or exacerbation needs additional symptom management or rehabilitation.**

- **Do not offer vitamin D solely for the purpose of treating MS.**

- **Do not offer omega-3 or omega-6 fatty acid compounds to treat MS. Explain that there is no evidence that they affect relapse frequency or progression of MS.**

4.3 Key research recommendations

Cognitive rehabilitation

What is the clinical and cost effectiveness of cognitive rehabilitation for people with MS?

Continued relapses

Is intravenous methylprednisolone more clinically and cost effective than oral methylprednisolone in people with relapsing–remitting MS and people with secondary progressive MS with continued relapses?

Mobility

What is the optimal frequency, intensity and form of rehabilitation for mobility problems in people with MS?

Spasticity

What non-pharmacological interventions are effective in reducing spasticity in people with MS?

Vitamin D

Can vitamin D slow down the progression of disability in MS?

5 Diagnosing MS

5.1 Introduction

Whilst MS can present with very characteristic symptoms and signs, such as optic neuritis or Lhermitte's phenomena, it can also manifest itself through much less specific symptoms such as paraesthesia or bladder disturbance. The temporal course of the symptoms and the age at onset can help guide the diagnosis but are not typical in all cases.

The variability in the severity and nature of symptoms and the spontaneous remissions that are usual in early relapsing-remitting MS can lead to a delay in diagnosis. An early diagnosis is important as it enables patients to receive an explanation for their symptoms, to access information and gain an understanding about this chronic disease and for them to be promptly assessed for treatment, either aimed at ameliorating their symptoms or modifying the disease course.

There is no single test that can diagnose MS and so diagnostic criteria have been developed over the years, initially based on purely clinical criteria but in later years incorporating results of investigations. Whilst the primary reason behind developing such criteria is for the purposes of research they have also been incorporated into clinical practice.

Charcot raised the awareness of MS and developed Charcot's triad which was made up of poor balance, slurred speech and double vision. The criteria developed by Schumacker and published in 1965 remain the basis for current criteria. These criteria were objective evidence for disease affecting two or more white matter parts of the central nervous system, occurring in episodes lasting more than 24 hours separated by more than one month or with progression over 6 months in a person aged 10 – 50 years at onset and with no better explanation. The requirement underlying these criteria is evidence of dissemination in time (DIT) and dissemination in space (DIS).

International panels have met since 1982 to agree criteria for diagnosis of MS. What are known as the McDonald criteria have been developed since 2001 by a series of international panels who review existing criteria in light of developing technologies and research evidence on use of criteria. The aim is to simplify diagnosis without loss of sensitivity and specificity. The use of radiological techniques for example can allow for a diagnosis at first clinical presentation if older lesions are present in different areas of white matter.

The most recent revisions to the McDonald criteria were the 2010 revisions to the McDonald criteria [Polman CH *et al*. Ann Neurol 2011;69:292-302]. While the criteria have helped to standardise diagnosis clinical judgement is still required in the interpretation by healthcare professionals with experience with MS who are also able to consider and exclude alternative diagnoses.

5.2 Review question: What are the key diagnostic criteria for the following: multiple sclerosis; possible multiple sclerosis; neuromyelitis optica and clinically isolated syndrome

For full details see review protocol in Appendix C.

The GDG agreed that the recommendations for the diagnosis of multiple sclerosis should be based on McDonald criteria. These criteria are well established and accepted across the multiple sclerosis community. GDG consensus opinion was used to word these are recommendations that would be useful for clinicians in practice (by informal consensus methods).

A Medline citation search of the following paper was carried out. This paper describes the development of the revised McDonald criteria. Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A. J., Waubant, E., Weinshenker, B. and Wolinsky, J. S. (2011), Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. Ann Neurol., 69: 292–302. doi: 10.1002/ana.22366

5.3 Recommendations and link to evidence

Recommendations	**1. Be aware that clinical presentations in multiple sclerosis (MS) include:** o loss or reduction of vision in 1 eye with painful eye movements o double vision o ascending sensory disturbance and/or weakness o problems with balance, unsteadiness or clumsiness o altered sensation travelling down the back and sometimes into the limbs when bending the neck forwards (Lhermitte's symptom). **2. Be aware that usually people with MS present with neurological symptoms or signs as described in recommendation 1 and:** o are often aged under 50 **and** o may have a history of previous neurological symptoms **and** o have symptoms that have evolved over more than 24 hours **and** o have symptoms that may persist over several days or weeks and then improve. **3. Do not routinely suspect MS if a person's main symptoms are fatigue, depression or dizziness unless they have a history or evidence of focal neurological symptoms or signs.**
Relative values of different outcomes	An accurate diagnosis of multiple sclerosis will help direct appropriate management and treatment.
Trade off between clinical benefits and harms	Clinical harms include delay in diagnosis and misdiagnosis. If non-specialists have a clearer idea of the clinical presentation of MS they may refer at an earlier stage. Providing information as to which patients are unlikely to have MS is also of benefit to non-specialists and people with symptoms. .
Economic considerations	Considering specific characteristics for the diagnosis of multiple sclerosis does not have any economic implications.
Quality of evidence	The recommendations for diagnosis are based on agreed international criteria for diagnosis of Multiple Sclerosis. The GDG used informal consensus to agree the wording of the recommendations, adapting the McDonald criteria for use by non-MS specialists.
Other considerations	The GDG considered that the diagnosis of MS is complex but diagnostic criteria are a guide to who should be referred to a specialist. MS occurs primarily in people between ages of 20 and 50 years. The pathology of MS is of an inflammatory process and the time course can help differentiate symptoms from those caused e.g. by TIA or stroke where the symptoms occur suddenly or over a time course of minutes to hours. The GDG considered it useful to identify common patterns of presentation but the list is not exhaustive.

	Fatigue, depression and dizziness are non-specific symptoms and would not usually suggest a diagnosis of MS if a person does not have accompanying neurological symptoms and signs.

Recommendations	**4. Before referring a person suspected of having MS to a neurologist, exclude alternative diagnoses by performing blood tests including:** o full blood count o inflammatory markers for example erythrocyte sedimentation rate, C-reactive protein o liver function tests o renal function tests o calcium o glucose o thyroid function tests o vitamin B_{12} o HIV serology.
Relative values of different outcomes	It is important to exclude disorders that may mimic MS symptoms to ensure that the correct diagnosis is made.
Trade off between clinical benefits and harms	The tests suggested by the GDG can be carried out by means of routine blood tests. The GDG considered the benefit of a correct diagnosis, and the institution of appropriate treatment outweighed any harms.
Economic considerations	There are some costs associated with performing routine blood tests. Two GDG members reported laboratory costs from their respective hospitals. The costs for each test varied between the two hospitals (£1.40 to £32 per test). Based on these costs, the total laboratory cost for these tests would be £39 to £141, excluding nursing time to take the blood sample. All of these tests are routine tests that are carried out in primary care. The GDG considered the benefits of a correct diagnosis justify the cost of performing these tests.
Quality of evidence	The criteria for diagnosis of MS include the exclusion of other possible causes of symptoms. The GDG used informal consensus to agree a list of tests that they considered might be important to rule out the diagnosis of multiple sclerosis.
Other considerations	The GDG considered that some routine blood tests should be performed before referral to a neurologist. These should not delay urgent referral if that is required on clinical grounds. The tests listed are not an exhaustive list but were those considered most likely to inform the necessity and route of referral. Depending on the clinical presentation other tests might be appropriate. The GDG considered a full blood count necessary as it might uncover significant anaemia or alert healthcare professional to the presence of vitamin B12 or other deficiencies. An elevated C reactive protein or ESR might suggest that an alternative infectious or inflammatory process is responsible for the patient's symptoms. Glucose, calcium, and thyroid function tests can help exclude diabetic peripheral neuropathy, hypocalcaemia, and hypothyroidism, all of which can cause sensory symptoms similar to MS. Alternatively, calcium may be raised in

	mimicking disorders such as sarcoidosis. A person with symptoms suggestive of MS is likely to need an MRI scan with contrast – the use of contrast requires consideration of renal function and the availability of renal function tests when seeing a specialist would be helpful in planning investigation. Vitamin B12 deficiency can cause neurological deficits which may be attributed to other disorders unless the vitamin B12 level is checked. HIV serology is included in the list as HIV itself may mimic MS, such as in transverse myelitis. The GDG considered that missing a diagnosis of HIV would be a significant issue. HIV testing is now a routine procedure in amny areas and does not carry the stigma it once did. The GDG discussed whether testing for syphilis should be included in the list of tests to be performed. . However the GDG thought this should be based on clinical judgment in individual cases rather than on a universal recommendation. The GDG considered that neurosyphilis rates are increasing, but many cases occur in the presence of HIV infection. Therefore a positive HIV test might point to a potential need for syphilis testing. Serum autoimmune screening tests were considered but the GDG considered these should not be universally recommended. It is recognised that conditions such as cerebral vasculitis or systemic lupus erythematosus may mimic MS. Unless the non-specialist has significant suspicion that the patient has e.g. a connective tissue disorder, these tests can be conducted by the specialist, who can also interpret this fully.

Recommendations	**5. Refer people suspected of having MS to a consultant neurologist. Speak to the consultant neurologist if you think a person needs to be seen urgently.** **6. Only a consultant neurologist should make the diagnosis of MS on the basis of established up-to-date criteria, such as the revised 2010 McDonald criteria[p], after** o assessing that episodes are consistent with an inflammatory process o excluding alternative diagnoses o establishing that lesions have developed at different times and are in different anatomical locations for a diagnosis of relapsing-remitting MS o establishing progressive neurological deterioration over 1 year or more for a diagnosis of primary progressive MS. **7. Do not diagnose MS on the basis of MRI findings alone.**
Relative values of different outcomes	A diagnosis of MS has significant implications for patients. Ideally a test with high sensitivity would ensure people correctly receive treatments that delay progression and reduce relapses. A test with high specificity can reduce emotional harm from wrongly communicating a diagnosis of MS and clinical

p Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011 Feb;69(2):292-302

	harm from inappropriate investigation and treatment. Assessing the evidence base for diagnostic criteria in MS is outside the scope of this guideline. There is no single test for MS and criteria for diagnosis are agreed internationally and currently reviewed every 5 years.
Trade off between clinical benefits and harms	The GDG considered there were no harms from recommending that the diagnosis of MS be made by specialists with appropriate expertise and basing this on up to date criteria.
Economic considerations	There are costs associated with referral to a neurologist (£205 for an initial face-to-face consultation with a neurologist[52]; however the GDG considered the benefits of specialist diagnosis of MS justify the cost. Basing a diagnosis of multiple sclerosis on established up to date criteria does not have any economic implications.
Quality of evidence	The recommendations were based on established international criteria and GDG consensus.
Other considerations	The GDG considered that a full review of diagnostic criteria were beyond the scope of this guideline. Current criteria for diagnosis are reviewed regularly and it is important that neurologist use the most up to date criteria. These are the criteria used in clinical trials and it is important that criteria for diagnosis are aligned with criteria for treatment. The experience of the GDG is that delay in diagnosis can occur but that people may also be told that they have MS on the basis of isolated findings on MRI without appropriate clinical assessment by a neurologist. The GDG considered that a diagnosis of MS should be made by a neurologist and therefore people with suspected MS should be referred to a neurologist. The GDG discussed whether it was possible to recommend how soon someone with MS should be seen by a consultant but considered it not possible to make a recommendation about this given the variety of ways in which people with MS may present. They did add to the recommendation that a healthcare professional should seek advice if they thought a patient should be seen urgently and normal referral processes might not allow this. The diagnosis has significant implications for occupational, social and other aspects of an individual's life. The application of the current diagnostic criteria, the findings of tests and the symptoms and signs presented by people require interpretation. The criteria include the exclusion of alternative diagnoses which also requires expertise. Of importance also is the potential use of current treatments for MS e.g. immunosuppressive therapy, which can have significant adverse effects. The GDG considered it important to stress that MS cannot be diagnosed on the basis of MRI alone and made a recommendation to stress this. The GDG reviewed the 2010 revised McDonald criteria and used it to guide recommendations. The GDG did not wish to restate the criteria but to highlight aspects of the recommendations that would be useful particularly for non-specialists and patients reading the guideline. The GDG stressed the importance of using established, up-to-date criteria for diagnosis and it is likely that the current criteria will be refined over time. An MRI scan will probably be necessary for diagnosis in most cases, but the GDG acknowledged that the current revised McDonald criteria allow for diagnosis without an MRI scan. Primary progressive MS is a rarer form of MS, and the GDG was concerned that it may be missed. The principle underlying diagnosis of primary progressive MS is therefore included in the recommendation. It is particularly important that there is a plan for review in these patients, because diagnosis requires progression over at least a year.

People suspected of having MS

Sometimes there will be high levels of suspicion that a person has MS, but she/he does not fulfil the McDonald criteria for a diagnosis of MS. These people should not be diagnosed as having MS, but, if appropriate, the possibility of MS should be honestly explained.

Recommendations	**8. If a person is suspected[q] of having MS but does not fulfil the diagnostic criteria, plan a review. Discuss the timing of the review with the person and ensure they know who to contact for advice if they develop further neurological symptoms or if current symptoms worsen.** **9. Offer people suspected of having MS information about support groups and national charities.**
Relative values of different outcomes	An accurate diagnosis is the most important outcome. This however may not be possible on first presentation and appropriate plans for support and follow up are also important.
Trade off between clinical benefits and harms	The GDG considered there were no harms to ensuring appropriate follow up.
Economic considerations	The follow-up of people with suspected MS is associated with some costs however the GDG considered the benefits of reviewing symptoms in these people justifies the cost.
Quality of evidence	The GDG used informal consensus to make these recommendations.
Other considerations	The GDG were aware that many patients may not fulfil the criteria for a diagnosis of MS. People in this situation can be left without a definitive diagnosis and become lost to follow up. The GDG agreed it may not be possible to make a diagnosis but that a plan for follow up was required and the timing of this should be agreed with the patient. The patient should also be advised who they should contact if they have further symptoms or a change in their symptoms. This might be their GP or the specialist service they have seen depending on local circumstances and the patient's choice. The GDG also considered that people with suspected MS might benefit from information about national groups from whom they may be able to get information. People who are suspected of having MS are given the label 'possible MS in the McDonald criteria.

Optic Neuritis

Optic neuritis involves the inflammation and demyelination of the optic nerve in one eye, leading to complete or partial loss of vision, blurring or a change in colour perception. It is often accompanied by pain on eye movements. This is a common 'first event' presentation in about 20-30% of people with MS, but for some people optic neuritis will not be related to MS, and it will probably be an isolated event.

q Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011 Feb;69(2):292-302

Recommendations	10. **If a person has an episode of isolated optic neuritis, confirmed by an ophthalmologist, refer them to a consultant neurologist for further assessment.**
Relative values of different outcomes	The diagnosis of MS has significant implications for patients and it is important that this is an accurate diagnosis informed by up to date criteria.
Trade off between clinical benefits and harms	The GDG considered there were no harms from recommending that the diagnosis of MS be made by specialists with appropriate expertise.
Economic considerations	There are costs associated with referral to a neurologist (£205 for an initial face-to-face consultation with a neurologist[51] however the GDG considered the benefits of specialist diagnosis of MS justify the cost.
Quality of evidence	The GDG used informal consensus to make these recommendations.
Other considerations	This recommendation was developed by the GDG using their experience and their view of the importance of specialist diagnosis of MS. For optic neuritis, an ophthalmologist is best placed to make the diagnosis and to exclude alternative eye conditions. The GDG were aware of people who had an MRI scan performed as part of an assessment for optic neuritis to be told they had MS on the basis of this MRI alone. The GDG considered that optic neuritis is the most common presentation of MS to be seen within another speciality and that appropriate pathways should be in place to ensure a person diagnosed with optic neuritis can be seen by a neurologist to discuss further assessment, likelihood of developing MS following one episode of optic neuritis and appropriate follow up.

Neuromyelitis optica

Neuromyelitis optica is a rare disorder that is often misdiagnosed as MS. It is an antibody-meditated disease involving demyelination of the optic nerve and spinal cord. It has a high mortality rate if not diagnosed and treated appropriately.

Recommendations	11. **Diagnosis of neuromyelitis optica should be made by an appropriate specialist based on established up-to-date criteria.**
Relative values of different outcomes	An accurate diagnosis of multiple sclerosis will help direct appropriate management and treatment.
Trade off between clinical benefits and harms	The GDG considered there were no harms from recommending that the diagnosis of MS be made by specialists with appropriate expertise.
Economic considerations	An appropriate specialist may be a neurologist. There are costs associated with referral to a neurologist (£205 for an initial face-to-face consultation with a neurologist [51]); however the GDG considered the benefits of specialist diagnosis of MS justify the cost. Basing a diagnosis of neuromyelitis optica on established up-to-date criteria does not have any economic implications.
Quality of evidence	The GDG used informal consensus to make these recommendations.
Other considerations	Neuromyelitis optica is a rare disorder that is often misdiagnosed as MS. It is

	an anti-body meditated disease which responds to immunosuppressive treatments. It has a high mortality rate if not diagnosed and treated appropriately. The diagnostic criteria according to Wingerchuk et al., 2006 are A. Optic neuritis B. Acute myelitis And at least two of three supportive criteria: 1. Contiguous spinal cord MRI lesion extending over 3 vertebral segments 2. Brain MRI not meeting diagnostic criteria for multiple sclerosis 3. NMO-IgG (AQP4-Ab) seropositive status While the diagnosis may be suspected by neurologists there is currently National Specialist Services funding for a National Diagnostic and Advisory Service for Neuromyelitis Optica based at two centres - the Walton Centre NHS Foundation Trust and Oxford University Hospital NHS Trust. These are commissioned to offer a rapid access diagnostic service, patient and clinical advice, supervision of clinical management in collaboration with the local referrer and in-patient treatment for severe and acute cases.

6 Providing information and support

6.1 Introduction

People with MS have complex information needs. They have to be able to make informed decisions for both the short and long term. Their family and carers are also faced with uncertainty about the future and may also seek additional information and support. The exact level and type of information offered will depend on a number of factors such as the time since diagnosis, the nature of diagnosis, disease progression, and the person receiving the information. There are no up-to-date systematic reviews concerning the information desired by people with MS and their carers, and this chapter aims to provide a summary of qualitative research in this area.

6.2 Review question: For adults with MS and their carers what information, education and support would they find useful?

For full details see review protocol in Appendix C.

Table 7: Characteristics of the review question

Population	• Adults
Aim	• To collate and synthesise the qualitative information available on the information, education and support that people with MS would like to receive.
The review strategy	• Qualitative studies addressing the views of MS patients and their carers with respect to their information/education/support • Include studies with mixed diagnosis
Analysis	• Narrative analysis • Pooling of common themes across studies

6.3 Clinical evidence

Thirteen qualitative studies that met the eligibility criteria for this review question were included.[7,13,25,45,106,114,129,137,139,146,235,246,264] Most focused on the perceptions of adults with Multiple sclerosis,[7,13,106,129,137,146,235,246] three studies elicited the views of carers,[25,45,139] and two covered the views of both patients and carers.[114,264] Only four were wholly or partially UK-based[25,106,129,139], and some were published 10 or more years ago.[13,106,114,146,235] The relevance of some of these findings to current UK practice may therefore be limited. No papers included information needs related to end of life care and advanced decision making. All included studies are summarised in Table 8.

Limitations of each study in terms of quality criteria (see chapter 3) are described in Table 8.

Table 8: Summary of studies included in the review

Study	Population	Methods	Limitations
Andreassen 2005[7]	Adults with MS or stroke from Norway. 4 had MS, 3 of whom were women, and all aged 25-66. Only information from MS patients	Semi-structured interviews; content analysis; triangulation of findings.	Relationship of researcher to participants unclear.

	is included in the review.		
Baker 1998[13]	Adults with MS from USA having an exacerbation; 10 women and 3 men aged 32-56.	Micro-moment time-line interview type.	Data analysis insufficiently described; methods to ensure trustworthiness insufficiently described.
Bowen 2011[25]	Visiting relatives of people with advanced MS in the UK; 15 women and 10 men aged 45-64.	Interviews; grounded theory; triangulation of researchers and member checking.	
Courts 2005[45]	Spouses of people with MS in the USA. 4 women and 8 men aged 31-67.	Focus groups x2 (gender specific). Thematic analysis. Triangulation of findings.	Data analysis insufficiently described.
Johnson 2003[106]	Adults with MS from UK; 14 women and 10 men aged 34-67.	Interviews though structure unclear; member checking.	Data collection inadequately described; relationship of researcher to participants unclear; data analysis insufficiently described; methods to ensure trustworthiness insufficiently described.
Koopman 2003[114]	Adults with MS from Canada (n=10) and significant others (n=5). Average age of 43.5 years.	Focus groups; content analysis.	Relationship of researcher to participants unclear; findings superficially described.
Malcomson 2008[129]	Adults from Northern Ireland with MS duration >5 years who felt their ability to cope was good. 9 women and 4 men aged 40-67.	Focus groups x 2.Thematic analysis. Triangulation of findings and member checking.	
Matuska 2008[137]	Adults with MS from USA; 13 women and 0 men aged 29-60.	Focus groups x 2; grounded theory; member checking.	Relationship of researcher to participants unclear.
McKeown 2004[139]	MS carers from Northern Ireland and Eire. 11 women and 6 men.	Focus groups x 4.Triangulation of findings and member checking.	
Miller 1997[146]	Adults with MS from USA; 7 women and 3 men aged 40-59.	Interviews; hermeneutic phenomenologic methodology; member checking.	Data analysis insufficiently described; methods to ensure trustworthiness insufficiently described.
Solani 2007[232]	Adults with MS from Italy; 16 women and 7 men aged 23-70.	Focus groups x 2; Framework analysis; member checking.	
Thorne 2004[246]	Adults with MS from Canada; >5 years MS duration; 10 women and 2 men aged 33-54.	Loosely structured interviews or focus groups (if people could not attend interviews); thematic analysis.	Methods to ensure trustworthiness insufficiently described.

Wollin et al. 2006[264]	Australian adults with MS and their families. 6 women and 7 men aged 23-55 (patients). Carers were aged 30-60.	Semi-structured interviews; content analysis used. Triangulation of findings.	

Narrative review

There were five main categories that emerged from the review of the literature:

- Content of information desired by patients and carers
- Form and delivery of information desired by patients and carers
- Content of support required by carers
- Form and delivery of support desired by carers
- Carer support may initially be rejected, but should continue to be offered

Each category, except the last, was then sub-divided into themes (Figure 3).

Categories and themes derived from each included study are described below. Study findings for each theme have been synthesised, and this summary is followed by a description of the findings from each study. Findings have been separated into those derived from people with MS (PwMS) and those derived from carers.

No themes specifically on the subject of 'Education' emerged.

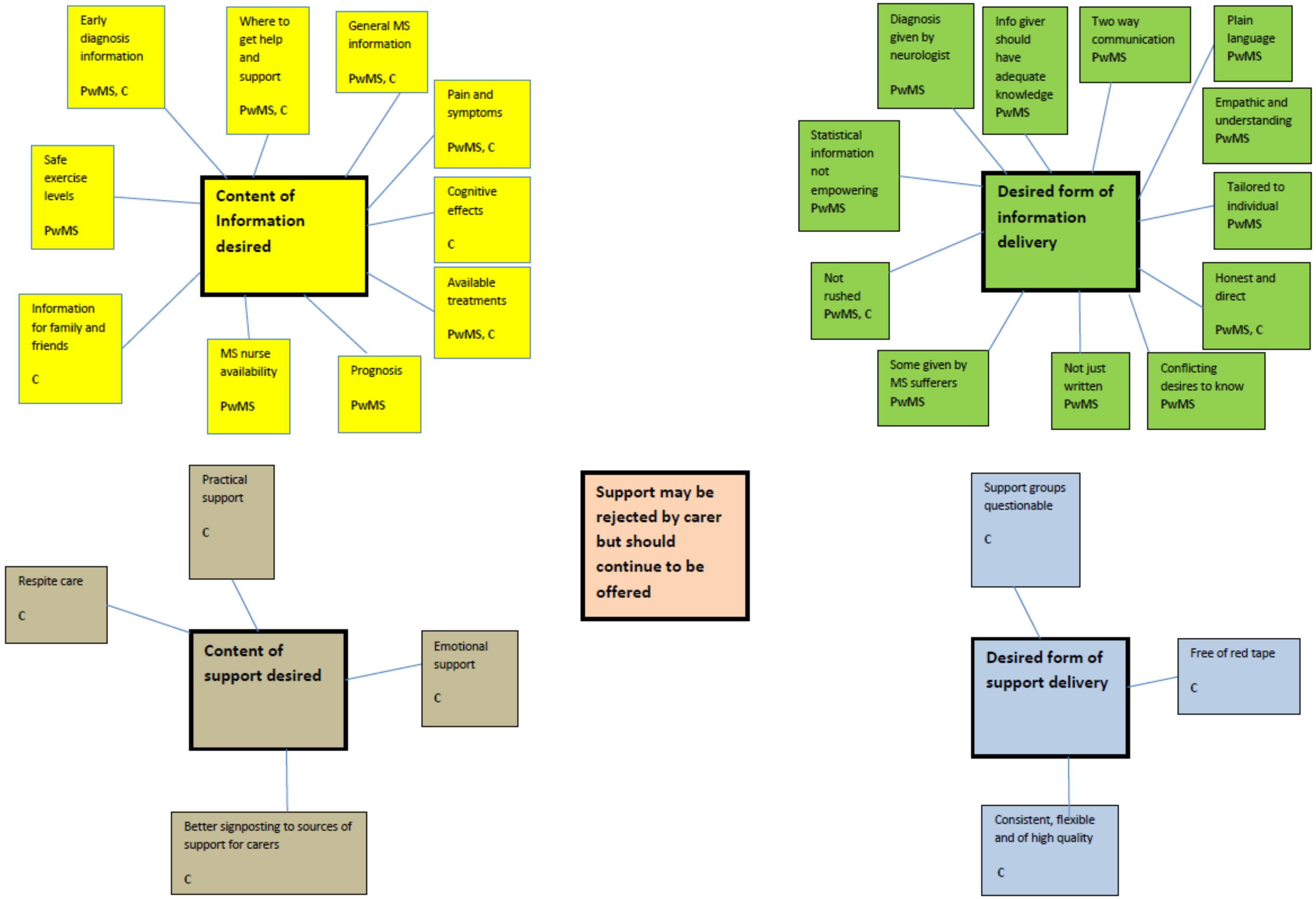


Figure 3: Categories and themes derived from the 14 qualitative studies. PwMS= data derived from people with MS; C=data derived from carers for people with MS

1. Content of Information desired by patients and carers

a. Early diagnosis information

Main findings

Although some studies contained outdated data,[146] a consistent finding was that diagnostic information was highly desired, [45,106,129,146,232,246] and should be provided as soon as possible to reduce anxiety[106,129,146,232,246]. Only two [106,129] were UK-based studies.

Detailed findings

Studies of people with multiple sclerosis (PwMS)

Malcomson 2008[129]

Information on diagnosis was felt to be delayed, with some participants waiting 25 years, which caused frustration. Reluctance by clinicians to give a definitive diagnosis was assumed to be the cause by some.

> *"When I asked them why did you not tell me, 'cause some people never have another relapse, you could be 20 years free and why worry?'. So they waited two years and the reason they told me then was the relapse."*

> *"I had the lumbar puncture 9 years earlier, which obviously showed I had MS, and they didn't tell me! They were all reluctant to give a definitive diagnosis" (nods of agreement).*

One participant praised their HCP for providing good support, giving information in book-form.

Solani 2007[232]

All participants felt that diagnostic information should be conveyed as soon as possible.

Miller 1997[146]

Many participants were relieved when finally, after many years, they were given the information of their diagnosis. Many had feared a terminal disease such as brain cancer. One participant was previously told she had 'demyelination', but the term was not explained. Some were told that their clinician had known all along about their MS but had not informed the participant due to a sense of having no power to improve the situation.

Thorne 2004[246]

Prior to diagnosis, a quest for knowledge was often to assuage fear of what the symptoms might mean, and also to understand what MS is.

Johnson 2003[106]

Prior to diagnosis the lack of information was a source of great worry as many feared they had a terminal illness. For some the diagnosis of MS was a relief.

Studies of carers

Courts 2005[45]

One carer stated that diagnosis was associated with a strong need for as much information as possible.

b. Where to get further help and support

Main findings

Information that does more than merely convey facts, but that also directs the patients and carers to practical sources of support, was a common wish. [13,45,106,129,146,232,264] Several participants in one UK-based study [129] felt that the right practical support could enhance independence.

More detailed findings

Studies of people with multiple sclerosis (PwMS)

Malcomson 2008[129]

People with MS needed information on where they could go to get help with various aspects of their disease, which was also felt to be a means by which independence and autonomy could be gained.

Wollin et al. 2006[264]

Participants expressed a need to search out help – a need to track down services and information. This was felt to be a frustrating search as help was not felt to be at hand. In particular, frustration was felt when trying to find out the services that could be accessed. Participants reported gratitude when locating a HCP who helped them access information and services.

Johnson 2003[106]

Knowledge of the sources of help was a key desire of participants.

Solani 2007[232]

Good sources of information were regarded as MS society booklets, medical institutions and pharmaceutical companies, but the internet was regarded as unreliable. However web pages of reputable sources were considered good. Charity or participant association advertisements for donations were seen as negative sources of information as they gave a very pessimistic view.

Miller 1997[146]

The MS society was seen as a valuable source of information by one participant, but a barrier to learning about MS due to inflexibility and misinformation by another.

Baker 1998[13]

Libraries were cited as a possible place to obtain information, but one participant was sceptical of the quality of the information that could be found there. Another participant expressed the view that libraries were not always accessible in a wheelchair.

Studies of carers studies

Courts 2005[45]

The need for information on resources was expressed:

> *"MS is a whole life situation….and there is so much that isn't addressed…. Resources, I would have appreciated it….. Here is the emotional thing, we need some help…. We didn't get enough information to make empowering choices ….. dealing with quality of life things"*

Carers stated their desire for information about support for themselves and their spouses.

> *"It is difficult to find a reliable source of information".*

Information from the MS society was limited to brochures, which provided superficial information that was perceived as available anywhere, without information on where support and further information could be gathered.

c. General MS information

Main findings

Four studies reported on a widespread desire by patients and carers for general information about Multiple Sclerosis [7,45,106,129]. In contrast, one older study [13] reported the views of one participant (experiencing an exacerbation) who questioned the value of such information.

More detailed findings

<u>*Studies of people with multiple sclerosis (PwMS)*</u>

Malcomson 2008[129]

Many found that they had to resort to personal research to gain a better understanding of the disease. Main sources were books, magazines and the internet. Befriending other people with MS was also used to gain information.

In general, participants wanted more information to help to adjust their lifestyles, cope with fatigue and depression, and dispel common worries and concerns such as ending up in a wheelchair.

Andreassen 2005[7]

The 4 MS participants in the study all expressed a need for education and increased knowledge about their illness.

Johnson 2003[106]

Knowledge of the disease was desired by all participants

Baker 1998[13]

One participant stated that information on MS was not perceived to be of use:

> *'I don't see how. I mean what good is information going to do? It's not going to cure anything or change anything'*.

<u>*Studies of carers*</u>

Courts 2005[45]

Carers expressed a wish for general information about MS, from the internet and MS society.

d. Pain and symptoms

Main findings

The need for information on how to deal with daily pain and symptoms[45], and how to know what symptoms were related to MS [114] were expressed. One participant with MS who was experiencing an exacerbation described how information about symptoms affirmed that she was not alone, which provided relief [13].

More detailed findings

Studies of people with multiple sclerosis (PwMS)

Baker 1998[13]

One participant found that information in the form of giving a label to experienced symptoms helped her, as it confirmed that others knew of her real problem. Information showing that symptoms were common also helped:

> *'just knowing it is common, I think, eases my mind, knowing that other people are dealing with it too…'*

Information that prepared participants for what to expect was regarded as helpful. Confirmation that symptoms were part of an exacerbation was also desired.

Koopman 2003[114]

One carer identified the usefulness of knowledge concerning what symptoms may or may not be related to MS:

> *'I'd like the information… Help us make connections to what is MS-related and what is not".*

Studies of carers

Courts 2005[45]

Information on how to cope with pain and symptoms on a daily basis was regarded as useful by carers.

e. Cognitive and personality effects

Main findings

One recent, high quality, UK-based study[25] reported that information on cognitive and personality changes is important to relatives of people with advanced MS.

More detailed findings

Studies of carers

Bowen 2011[25]

Information from health care professionals that MS was not just a physical illness, but had cognitive aspects too, was often missing, and contributed to relatives' under-preparedness for personality change and cognitive decline.

f. Available treatments

Main findings

Three studies [13,129 45] reported on the desire for treatment information. Both patients[13] and carers[45] reported a need to know more about standard treatments, specifically dosing, side effects and long term effects [13]. However two studies also implied a need for information on complementary therapies. [129 45]

More detailed findings

Studies of people with multiple sclerosis (PwMS)

Malcomson 2008[129]

Participants reported they needed information on complementary therapies.

Baker 1998[13]

Participants wanted to know about the side effects, as well as the long-term effects, of various drugs, to allow informed choices to be made. Participants also desired information on how to adapt dosing when responses to drugs were unexpected was requested.

Studies of carers

Courts 2005[45]

Information about treatments and medications was greatly desired by carers. Spouses also suggested a need for information about the availability and use of complementary therapies, which were often sought. Husband spouses shared complementary interventions that they perceived had worked for their spouses.

g. Prognosis

Main findings

A desire for prognostic information was expressed in only one study [7], but this issue was raised by more than one of the study's participants.

More detailed findings

Studies of people with multiple sclerosis (PwMS)

Andreassen 2005[7]

The MS participants expressed concerns about the future, the progression of the disease and what limitations they would experience. They expressed a need for more information about these issues.

h. Availability of the MS nurse

Main findings

Information on MS nurse availability was raised in only one study [106], but this was clearly an important issue as it was independently raised by 5 of the participants. Notably, all of them had been diagnosed close to the time of the study being reported.

More detailed findings

Studies of people with multiple sclerosis (PwMS)

Johnson 2003[106]

Being told of the availability of the MS nurse by their health care professional was greatly appreciated by participants.

i. General information to significant others

Main findings

In one study [45] spouse carers expressed a need for information that could be relayed to significant others to avoid resentments or disinterest arising from their lack of understanding. Again, despite this issue being raised in only one study, it was reiterated by more than one carer.

More detailed findings

Studies of carers

Courts 2005[45]

Husband carers expressed the need for family and friends to be given information about MS too, to avoid the resentment that can arise from ignorance of how the disease may make certain social behaviours difficult:

> *"We're always coming to your house. You don't ever come to ours".*

Wife carers implied a lack of support from friends in terms of their lack of interest of the disease:

> *"they say….'how are you?' and you say 'fine'; they don't really want to know that he had an exacerbation…. And he had an accident and couldn't get to the bathroom on time".*

Peer counselling from spouses was suggested as a way of getting non-family members to understand.

j. Exercise

Main findings

Although adequate exercise is part of general healthy-living advice, participants from two studies [129,137] expressed a need to know more about exercise in the context of having MS.

More detailed findings

Studies of people with multiple sclerosis (PwMS)

Malcomson 2008[129]

In general, participants wanted more information on exercise.

Matuska 2008[137]

Participants reported uncertainty about how much exercise was too much.

2. Desired form and delivery of information

a. Diagnosis should be given by the attending neurologist

Main findings

The desire for the diagnosis to be given by the neurologist was only raised in one study [232] , but this was the only study specifically focussed on how diagnosis should be communicated. There was general agreement within the focus groups on this issue.

More detailed findings

Studies of people with multiple sclerosis (PwMS)

Solani 2007[232]

Participants felt that the information on diagnosis should be given by the attending neurologist. Some felt the presence of a significant other was helpful in terms of helping to fill in gaps in understanding, but it was felt this decision should be up to the participant. Most felt that other professionals should not be present at the first information giving meeting. Often another person was present (i.e. another professional not involved in the meeting but on the phone or doing other work) which impaired confidentiality and the rendering of effective information.

b. The information giver should have adequate knowledge

Main findings

A sense that the information giver should have enough knowledge to provide useful information was raised in three studies [106,146,246].

More detailed findings

Studies of people with multiple sclerosis (PwMS) **Johnson 2003**[106]

Frustration with the knowledge of health care professional was common. This was especially when it limited the exchange of information on support services and practical help available.

Often knowledge was gained by ‘luck’, after chance encounters with a physiotherapist or through people at work providing information:

'Reading things and.. would bring snippets of information and I was gathering information from books, friends and over a period of time I probably learnt more… than any other help I was given professionally.'

Miller 1997[146]

Several participants were misinformed about their illness and found that even health care professionals did not understand the disease.

Thorne 2004[246]

Participants expressed a wish for clinicians to be knowledgeable about the disease, and comments by clinicians, such as "it's a mystery to me" or "you know more about this than I do" magnified the participants' sense of coping alone.

c. Communication should be two way and aimed at a common understanding

Main findings

Although only reported by one study[246], the desire for health care professionals who were willing to learn and listen to the views of patients was explicitly stated by several participants.

More detailed findings

<u>*Studies of People with multiple sclerosis (PwMS)*</u>

Thorne 2004[246]

Communications that sought common understandings were highly valued by participants.

d. Simple language should be used.

Main findings

An Italian focus group study[232] showed participants wanted information to be imparted using non-jargonised language.

More detailed findings

<u>*Studies of People with multiple sclerosis (PwMS)*</u>

Solani 2007[232]

Participants believed that explanations should be simple and in plain language.

e. Information should be imparted with support, empathy and understanding

Main findings

Two studies[106,129] focussed on the way diagnostic information had been conveyed, emphasising the need for a more empathic and supportive mode of providing such information. Both were UK-based studies, but their immediate relevance was diminished by much of their data relating to diagnoses received over 10 years ago. One other study indicated how general information should be given in person, with appropriate support[246].

More detailed findings

<u>*Studies of People with multiple sclerosis (PwMS)*</u>

Malcomson 2008[129]

The imparting of the diagnosis was often felt to be unsupportive, unhelpful, abrupt, and lacking in sensitivity, empathy and understanding. Diagnosis was often vague and uncommitted:

> *'He said 'well yes, MS would have to be on the list', but he would never say that I had MS. Even when he did tell me he told me it's likely to be.'*

Many felt a lack of psychosocial support, with the sense that clinicians focussed mainly on the physical symptoms. Most felt that no information was given in respect of how they might cope with the diagnosis, symptoms and potential lifestyle changes, although many of these had been diagnosed many years previously:

> *'I got no help, I just cried. I wasn't given information on any support groups, counselling, MS nurses, nothing' [diagnosed 12 years before study]*
>
> *'I was left to stew basically'.[diagnosed 20 years before study]*
>
> *'I didn't get any information from my doctor at all'.[diagnosed 21 years before study]*

Emotional support was greatly desired.

Johnson 2003[106]

Many participants expressed negative feelings associated with the time of diagnosis, such as poor information about the diagnosis, where sometimes the only information offered was that there was no cure and nothing could be done.

Abandonment and shock were common feelings at the time of diagnosis, often related to a lack of identifiable support and advice following diagnosis.

However some had more positive impressions at the time of diagnosis: being told of the implications of MS by the GP, being told of the availability of the MS nurse and being given telephone contact in the following weeks.

Thorne 2004[246]

Participants also felt the delivery of information was important. Receiving critical information indirectly or over the phone was seen as less than ideal. Being provided with difficult information and left without support was particularly hard to manage. Sometimes participants tried to access HCPs with questions and did not have their phone calls returned.

f. Advice should be personalised

Main findings

There was clear consensus across five studies[7,13,129,232,246] that information should be directed at personal needs and tailored to the individual.

More detailed findings

Studies of People with multiple sclerosis (PwMS)

Baker 1998[13]

Generic information was not seen as helpful by participants – more useful was honest and realistic information tailored to their needs. The lack of access to this kind of information was seen as robbing someone of their sense of control over the situation, forcing them to become dependent on physicians:

> *'I think that there's not a lot I can relate to in terms of all the things I have read. I think I'd have to rely a little more on the doctors for this [problem]'.*

For others this was problematical as they didn't want to bother the doctor or be seen as a hypochondriac.

Andreassen 2005[7]

Participants emphasised the need for information specifically for their personal situation. They found that group-based information, such as information found in diagnosis-specific organisations, was of limited value.

Malcomson 2008[129]

The majority stated that advice should be directed to individual needs.

Solani 2007[232]

It was felt that information given should be tailored to the individual. General information on the disease was felt useful but non-relevant information should not be given.

The sequence of issues to be raised should be fixed in advance, and should depend on the participants' history and work/family commitments. These should be known by the neurologist.

Thorne 2004[246]

Timely, specific and direct information, that included an acknowledgement of the limits of medical understanding, was most sought after.

g. Advice should be honest and direct

Main findings

Five studies [25,129,146,232,246]agreed that information should be honest, direct, non-patronising, and comprehensive. Too much information was generally regarded as better than too little, and there seemed to be consensus that information decreases fear.

More detailed findings

Studies of People with multiple sclerosis (PwMS)

Malcomson 2008[129]

Participants felt information should be realistic.

Thorne 2004[246]

Immediate access to information was crucial for allaying fear. Anticipation of HCPs in giving information that the participants did not quite know how to ask for was also appreciated.

> *'Physicians won't just offer information, some information they will, but I feel sometimes with the questions that I have asked, if I hadn't asked those questions I never would have gotten that information. I just get the feeling that it's a very paternalistic sort of attitude…"we don't need to tell her that" '.*

Some participants felt frustration when HCPs were seen to assume a low level of knowledge and explain things superficially in the simplest terms. Information on the results of tests was also desired, and the withholding of this caused frustration as well.

Participants wanted information delivered directly, as full of facts as possible, and not sugar coated. Too much information was regarded as better than too little. Often, fear was increased when clinicians appeared to avoid using the words 'MS' as though it were a dread disease.

Particular problems were receiving inaccurate or outdated information from HCPs, or being given an overly optimistic picture of what to expect.

When participants believed an HPC had withheld important information this created stress and made the participant feel betrayed or patronised.

Solani 2007[232]

Sensitive issues such as the difficulty in giving an early prognosis should be confronted and not skirted around

Miller 1997[146]

One participant said that in learning about her illness she had come to think of it as her friend.

<u>*Studies of carers*</u>

Bowen 2011[25]

The amount of information that people had about MS varied greatly, even within the same family. Often it was the children of parents with MS who had the least information, especially if there was a culture of not touching on this subject. This led to extremes of expectation – either complete surprise when deterioration occurred, or a long wait for the dreaded moment when death was expected to come.

> *'From what I seem to gather now, there's no reason Dad won't live another 10 years or so, but nobody bothered to explain that…. That made it worse because I kept waiting for it to happen when actually there was no need to.'*

h. But some people have conflicting desires about knowing

Main findings

However two studies[13,137] implied that not all participants wanted too much information until they were ready. Information thus needs to be given in a sensitive way that is geared to the psychology of the individual.

More detailed findings

Studies of People with multiple sclerosis (PwMS)

Baker 1998[13]

Internal barriers to obtaining information were seen as denial, uncertainty and fear. People did not want to accept another interruption in their lives.

> *'That part of the relapsing/remitting is difficult because you have to always go through it again. I don't want to deal with this, I don't want to accept it, and I don't want to have to wonder what this one is going to do, how long it is going to last, so it is difficult, never knowing.'*

Matuska 2008[137]

Participants talked about feeling fearful of learning about the disease yet desiring information about ways to improve their health:

> *'I'm really scared. It's all just pretty overwhelming. I didn't go get help right away and I didn't talk about it, and I didn't really want to hear about it.[now I'm] just taking the steps to learn. I'm trying to learn for the future, basically. So I can live a longer, happier life.'*

i. Information should not just be in written form

Main findings

A UK-based study[129] showed that participants wished information to be given in a variety of forms.

More detailed findings

Studies of People with multiple sclerosis (PwMS)

Malcomson 2008[129]

Participants felt that information should be via a variety of media, as written information may not be appropriate for those with visual problems.

j. Information should be imparted by MS sufferers

Main findings

A UK-based study[129] showed that several participants felt that some information being relayed by a fellow MS sufferer would be useful.

More detailed findings

Studies of People with multiple sclerosis (PwMS)

Malcomson 2008[129]

Participants felt that those imparting information should ideally have MS themselves, to enhance empathy.

> *"I think it's more helpful if it is someone who actually has MS, you can relate better." (Several individuals nod their heads in agreement).*

k. Information-giving should never be rushed, with opportunities for participant to ask questions

Main findings

Three studies showed that adults with MS[106,232] and their relatives [25] wanted information given in digestible amounts, with ample opportunity for clarification.

More detailed findings

Studies of People with multiple sclerosis (PwMS)

Solani 2007[232]

All participants felt that diagnostic information should be given in a non-rushed way. Most felt that any meetings were too short for adequate questioning and information to be gained.

They also felt that participants should be given ample opportunity to ask questions and there should be frequent efforts to ensure the participant fully understood the information. Important points should be re-stressed in different ways. At the end of a discussion the main points should be repeated, with another opportunity for participant questions. Subsequent sessions should be arranged to allow all information to be given in digestible chunks – it should not all be given at once. Full support and contact information should be given.

Johnson 2003[106]

More time spent in communication was desired by one participant:

> 'I needed time spending with me and I needed it explaining to me. What it meant, how you handled it, what there was available. I knew nothing! … I didn't know there was an MS society'

Studies of carers

Bowen 2011[25]

Participants felt that health care professionals need to understand that information does not always equal understanding and that understanding should be checked at regular intervals during the illness. This would avert crises where families are under-prepared for sudden events.

l. Statistical information is not empowering

Main findings

One study[246] showed that statistical information was not felt to be useful.

More detailed findings

Studies of People with multiple sclerosis (PwMS)

Thorne 2004[246]

Statistics tended to alarm rather than inform participants:

> *'My doctor would say, "Doesn't it make you feel better that in 7 years 70% of people are not in wheelchairs?" No, no it doesn't at all. That 30% that are scares the shit out of me!'.*

Another participant mentioned how disempowering statistical information could be:

> *It's a chronic disease that, you know, slow degeneration and then you will be blah, blah, blah. It hexes you and it sets you up… you're not disagreeing that it is factually correct information based on population statistics, but it disempowers.'*

3. Content of support required by carers

a. Practical support

Main findings

One study on carers[45], and one on both patients and carers[264] suggested that practical help was desired, particularly by men and people in rural areas.

More detailed findings

Studies of carers

Wollin et al. 2006[264]

People in rural communities had the greatest difficulty getting services and support and had to move to get them.

Participants greatly appreciated practical help:

> *"The council were very good. They organised a lady to come in and get N out of bed and make sure she got to kindergarten, and….that sort of thing then she fetched her back again and did some housework and I had a shower and they'd help me"*

Courts 2005[45]

Male spouses wanted pragmatic support to learn about the disease, and also to help maintain their wives' sense of self-worth and activity.

b. Emotional support

Main findings

Emotional support was desired by carers[25], particularly female carers[45].

More detailed findings

Studies of carers

Courts 2005[45]

Female spouses needed support in more emotional ways than male spouses. They required someone to listen, ask about and respond to their needs, and help them cope:

> *"There are times I would like to roll around on the floor and scream and tear my hair out and say 'I'm all better now' and go home. You know! With women [who] understand where I am coming from'.*

There was also a need for someone to reach out to them:

> *"Just once, somebody out there someplace come and put their arm around me and say 'you are going to be fine. Here, let's take care of the [problem].*

Spouses needed support for themselves, and they wanted support for their wives and husbands. They often felt overwhelmed, ignored and neglected.

Bowen 2011[25]

Relatives are faced with difficult decisions to make on behalf of their loved one, often with little emotional support.

c. Better signposting to sources of support for carers

Main findings
A need for clearer directions towards sources of practical and emotional help was expressed by carers of people with advanced MS in a UK-based study[25].

More detailed findings

<u>*Studies of carers*</u>

Bowen 2011[25]

Relatives need better sign-posting to services for support. Formal carer support was often a reaction to crises, rather than proactive, which increased stress and the challenge of managing and coping with the situation, compounded by the stress of external carers entering the home.

d. Respite care

Main findings
One UK based study[25] showed that relatives often needed reassurance that taking time out from caring was both possible and an acceptable thing to do.

More detailed findings

<u>*Studies of carers*</u>

Bowen 2011[25]

Family members reflected that it was vital that professionals let relatives know they are able to take time out from caring and that they know about the provision of respite care before they need it.

"Offloading would have been good earlier on, but whether I would have done it (accessed professional support) I don't know, but that would have been helpful really if it could be put in such a way that was acceptable"

4. Desired form and delivery of support

a. Support groups are of questionable benefit

Main findings

The notion that support groups were unhelpful was raised by most carer respondents in a recent UK study[25].

More detailed findings

Studies of carers

Bowen 2011[25]

The majority of relatives felt support groups had questionable benefits.

b. Support needs to be free of red tape

Main findings

A partially UK-based study[139] showed carers were particularly frustrated by bureaucracy.

More detailed findings

Studies of carers

McKeown 2004[139]

All caregivers in this study found it a struggle to obtain support from the formal sources. Barriers included a lack of information about sources of support, protracted waiting times for services, red tape and bureaucracy, leading to inflexible and unresponsive support services:

'I think if you are not worrying the professionals they'll not come near you and I think they should be offering us more. But the professions the doctors the physiotherapist they won't offer you…. We have a social worker who comes very six months … but she has never in the 5 years she has been coming said "are you getting help with this? Are you getting help with that" '.

Available supports were found mainly via informal channels such as carers they'd met at carer support meetings, or via the mass media. Charitable organisations were often perceived as unresponsive.

c. Support needs to be consistent, flexible and of high quality

Main findings

Many carers in a partially UK-based study[139] expressed disappointment with the quality of support offered.

More detailed findings

Studies of carers

McKeown 2004[139]

Although some carers felt support services were useful many felt support services were not ideal. Often carers said that services didn't meet their needs or those of the recipient. Services were seen as inconsistent, inflexible and of such poor quality that they caused the carer and participant distress and angst.

> *'It's the inconsistency and the inconsistency of people coming into your house, the way they treat you, the way they treat your belongings, your property. They came in some of them and said ' I can't move your wife'. I say 'why not have you had no training to use a hoist' and I show them how to use a hoist.'*

Long waits for services were discussed:

> *'It took about a year and two months for her to get one [wheelchair]'.*

5. Support may initially be rejected by the carer, but should continue to be offered

Main findings

Two mainly UK-based studies showed carers found support hard to accept [25,139], but that eventually the overwhelming need for support dominated[139]. This suggests that support should be continually offered despite initial resistance.

More detailed findings

Studies of carers

Bowen 2011[25]

Family members reflected that they were in great need of support at times but admitted they would have found it very hard to accept.

The fact that relatives often don't associate themselves with the word 'carer' makes them less likely to access whatever carer support there is. Hence there was often a feeling of relief when the participant was taken into advanced care.

> *'so that's why I never really accessed any… support I think is for carers, whereas I wouldn't class myself as a carer cos I've never been the one to get him into bed, or wash him'.*

McKeown 2004[139]

This qualitative study showed that caregivers' experience of support with care giving occurred in 4 evolutionary stages. These were: rejecting support, resisting support, seeking support and accepting support. These will be described in turn below.

Rejecting support

Rejecting support appeared to be related to a desire to protect themselves, the recipient and their family from the reality of MS.

First, rejecting support helped some carers maintain their 'ostrich-like' desire to avoid confronting facts about the disease:

> *'Well I am sort of half ostrich in the fact that I really don't want to know'.*

Second, rejection of support was seen as a way of protecting the recipient. The recipients often wanted to maintain their independence and did not want care from anyone other than their relative (and often not from the relative either).

> *'With my son…. If you go to try to help him he wants to try to do everything on his own, and it's just impossible for him to do everything because of his tremors, along with his poor balance, he can't stand or make a dinner or anything….'*

Some recipients were embarrassed by their disabilities, giving rise to the resistance to outside help.

> *'To tell you the truth he would hate anyone coming in [referring to care assistants], I know he would dear, dear, he's just so proud'.*

Finally caregivers may reject support to try to maintain normality and thus protect the family, which was particularly applicable to those with children living at home.

> *'You try to keep the family normal for the sake of the children'.*

Resisting asking for support.

Carers felt that as the illness progressed they began to feel a growing need for help due to the rising list of responsibilities and because caregiving became a norm that therefore lacked a sense of gratification.

However at the same time they still felt unable to ask for help for several reasons:

- They felt that care was their sole responsibility
- They derived great pride from their role as sole caregiver, which often led to a desire to put on a façade that all was well
- They felt nobody else could do the job as well as them because nobody knew the recipient as well
- They felt it was unfair to burden family and friends
- They also felt that many people including family and friends were ignorant about the disease and so were unaware of the levels of support provided. Also some had stopped offering support as it had been rejected so often in the past.

Nevertheless some felt angry that close relatives did not offer their support.

Seeking support with caregiving

Eventually a 'crisis' would precipitate the carer's decision to seek help. Often this crisis would involve the inability to transfer the participant on their own. The carer would then tend to approach formal sources of support (i.e. Health and social services) rather than family and friends, as the latter had often withdrawn their offers of help by this time (see section above).

Accepting support services

If sought out services were perceived as suitable the carers reached the stage of accepting the support service. However sometimes support was rejected completely because of the red tape and poor quality of support, and carers decided to cope alone without support.

6.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

6.5 Evidence statements

6.5.1 Clinical

Content of Information desired by patients and carers

Six studies comprising 94 participants showed that information on the diagnosis was highly desired, and should be provided as soon as possible to reduce anxiety. Information that does more than merely convey facts, but that also directs the patients and carers to practical sources of support, was a common wish.

Four studies comprising 53 participants reported on a widespread desire by patients and carers for general information about Multiple Sclerosis.

One study comprising 12 participants showed a need for information on how to deal with daily pain and symptoms.

One study comprising 15 participants showed a need to know what symptoms were related to MS.

One study comprising 25 participants reported that information on cognitive and personality changes is important to relatives of people with advanced MS.

Three studies comprising 38 participants reported on the desire for treatment information (including complementary therapies), specifically dosing, side effects and long term effects.

One study comprising 4 participants showed a common desire for prognostic information.

One study comprising 24 participants showed information on MS nurse availability was also desired.

One study comprising 12 participants suggested a need for information that could be relayed to significant others to avoid resentments or disinterest arising from their lack of understanding.

Two studies comprising 26 participants showed a need to know more about the relationship between exercise and MS.

Desired form and delivery of information

One study comprising 23 participants highlighted the desire for the diagnosis to be given by the neurologist.

Three studies comprising 46 participants suggested a sense that the information giver should have enough knowledge to provide useful information.

One study comprising 12 participants highlighted the desire for health care professionals who were willing to learn and listen to the views of patients.

One study comprising 23 participants showed participants wanted information to be imparted using non-jargonised language.

Two studies comprising 37 participants emphasised the need for a more empathic and supportive mode of providing such information.

One study comprising 12 participants indicated how general information should be given in person, with appropriate support.

Five studies comprising 75 participants agreed that information should be directed at personal needs and tailored to the individual.

Five studies comprising 83 participants agreed that information should be honest, direct, non-patronising, and comprehensive. Too much information was generally regarded as better than too little, and there seemed to be consensus that information decreases fear.

Two studies comprising 26 participants implied that not all participants wanted too much information until they were ready. Information thus needs to be given in a sensitive way that is geared to the psychology of the individual.

One study comprising 13 participants showed that participants wished information to be given in a variety of forms.

One study comprising 13 participants showed that several participants felt that some information being relayed by a fellow MS sufferer would be useful.

Three studies comprising 72 participants showed that adults with MS and their relatives wanted information given in digestible amounts, with ample opportunity for clarification.

One study comprising 12 participants showed that statistical information was not felt to be useful.

Content of support required by carers

Two studies comprising 25 participants suggested that practical help was desired, particularly by men and people in rural areas.

Two studies comprising 37 participants suggested emotional support was desired by carers, particularly female carers.

One study comprising 25 participants indicated a desire for clearer directions towards sources of practical and emotional help.

One study comprising 25 participants showed that relatives often needed reassurance that taking time out from caring was both possible and an acceptable thing to do.

One study comprising 25 participants highlighted the notion that support groups were unhelpful.

One study comprising 17 participants showed carers were particularly frustrated by bureaucracy.

One study comprising 17 participants showed many carers expressed disappointment with the quality of support offered.

Support may initially be rejected by the carer, but should continue to be offered

Two studies comprising 42 participants showed carers found support hard to accept but that eventually the overwhelming need for support dominated. This suggests that support should be continually offered despite initial resistance.

6.5.2 Economic

No relevant economic evaluations were identified.

6.6 Recommendations and link to evidence

Recommendations	**12. NICE has produced guidance on the components of good patient experience in adult NHS services. This includes recommendations on communication, information and coordination of care. Follow the recommendations in Patient experience in adult NHS services (NICE clinical guideline 138).** <u>**Information at time of diagnosis**</u> **13. The consultant neurologist should ensure that people with MS and, with their agreement their family members or carers, are offered oral and written information at the time of diagnosis. This should include, but not be limited to, information about:** o what MS is o treatments, including disease-modifying therapies o symptom management o how support groups, local services, social services and national charities are organised and how to get in touch with them o legal requirements such as notifying the Driver and Vehicle Licensing Agency (DVLA) and legal rights including social care, employment rights and benefits.

	14. Discuss with the person with MS and their family members or carers whether they have social care needs and if so refer them to social services for assessment. Ensure the needs of children of people with MS are addressed.
Relative values of different outcomes	People with MS are entitled to information about their condition and their care.
Trade off between clinical benefits and harms	Some people with MS may not wish to be given certain information at certain times, but provided that the clinician takes individual wishes and circumstances into account, there are unlikely to be clinical harms from offering information and support. Clinical benefits may include an increased sense of control over the person's own life, a greater ability to make appropriate self-management decisions, and reduced anxiety.
Economic considerations	No relevant economic evaluations were identified. The GDG considered that while some of these recommendations had potential cost implications, for example in terms of additional staff time for the provision of information outside of normal consultations, these are fundamental aspects of good patient care. The content of the information provided to patients and carers does not have any economic implication.
Quality of evidence	Thirteen qualitative studies were used to gather evidence. Their quality was independently assessed by two research fellows, who agreed that all of the studies were acceptable for inclusion to the study. All had minor limitations such as only one method of data collection, or unclear reporting of analytic strategies, but all contained direct evidence in terms of population and outcome.
Other considerations	The evidence reviewed in this guideline suggests that information should be delivered in the following ways: • given by someone with adequate knowledge to give appropriate information and answer questions satisfactorily • individually tailored • in plain language and presented in a variety of media • honest and direct, if appropriate for the patient. The GDG considered that there were a number of areas of information likely to be required by people at the time of diagnosis. These included information about the MS and its treatments. There are legal requirements such as informing DVLA and car insurance companies. People may also benefit from information about national charities. Healthcare professionals should also ask about social care needs and make a referral to social services for assessment if social care needs are likely. Although the evidence review did not suggest issues with the care of children, the GDG added that attention should be given to the needs of children of people with MS. MS affects people when they may have young children and this can be a significant issue. Young children may also be in the position of care-givers and this is often not recognised and has implications for the the children's own development.

Recommendations	**15. Offer the person with MS a face-to-face follow-up appointment with a healthcare professional with expertise in MS to take place within 6 weeks of diagnosis.**

Relative values of different outcomes	Appropriate follow up of people following diagnosis is important.
Trade off between clinical benefits and harms	The GDG considered that there were no harms from offering appropriate follow up.
Economic considerations	No relevant economic evaluations were identified. The costs of a face-to-face appointment with a neurologist is £148.[51] In some circumstances the follow-up may be conducted by a MS specialist nurse. The cost of a face-to-face appointment with a specialist nurse is £60.[51] The GDG considered this should occur within 6 weeks as a balance between resource use and health benefit to the patient although recognised that this may need to be varied according to patient need.
Quality of evidence	Thirteen qualitative studies were used to gather evidence. Their quality was independently assessed by two research fellows, who agreed that all of the studies were acceptable for inclusion to the study. All had minor limitations such as only one method of data collection, or unclear reporting of analytic strategies, but all contained direct evidence in terms of population and outcome.
Other considerations	The GDG reported that people newly diagnosed with MS will require a follow up appointment. The timing of this will vary – people may be having follow up as they are being assessed for or are starting DMDs, people may be undergoing physiotherapy or other treatment. The GDG considered that the neurologist responsible for the diagnosis of the patient should ensure that formal follow up is arranged following diagnosis and agreed that within 6 weeks was a reasonable period based on expert opinion and experience. Individual patients may need to be seen sooner but the experience of the GDG was that many people need time to come to terms with the diagnosis. The GDG recognised that this follow up may be with a neurologist or another healthcare professional such as an MS nurse depending on local service organisation.

Recommendations	<u>**Ongoing information and support**</u> **16. Review information, support and social care needs regularly. Continue to offer information and support to people with MS or their family members or carers even if this has been declined previously.**
Relative values of different outcomes	The outcome considered here is information required by people with MS and their carers. It is regarded as of high importance.
Trade off between clinical benefits and harms	Some people with MS may not wish to be given certain information at certain times, but provided that the clinician takes individual wishes and circumstances into account, there are unlikely to be real clinical harms from offering information and support. Clinical benefits may include an increased locus of control, a greater ability to make appropriate self-management decisions, and reduced anxiety.
Economic considerations	No relevant economic evaluations were identified. The GDG considered that reviewing of information and support needs would be part of regular appointments for the management of multiple sclerosis and should not incur additional costs.
Quality of evidence	Thirteen qualitative studies were used to gather evidence. Their quality was independently assessed by two research fellows, who agreed that all of the studies were acceptable for inclusion to the study. All had minor limitations

	such as only one method of data collection, or unclear reporting of analytic strategies, but all contained direct evidence in terms of population and outcome.
Other considerations	The GDG used the evidence review and their experience to make this recommendation. The review indicated that people and carers may initially refuse information or support. People vary in their requirement for information and what they wish to know at different times so the GDG considered it important that healthcare professionals regularly check people's information and support needs. These are also likely to change over time. Supprt needs could include vocational support to continue education or work, access to equipment or adaptations to property, and support to continue in a caring role for dependent family members.

Recommendations	**17. Ensure people with MS and their family members or carers have a management plan that includes who to contact if their symptoms change significantly.** **18. Explain to people with MS that the possible causes of symptom changes include:** o another illness such as an infection o further relapse o change of disease status (for example progression).
Relative values of different outcomes	The outcome considered here is information required by people with MS and their carers. It is regarded as of high importance.
Trade off between clinical benefits and harms	Information about how to interpret symptoms was considered a benefit and not to have harms.
Economic considerations	No relevant economic evaluations were identified. The GDG considered that provision of such information would be part of regular appointments for the management of multiple sclerosis and should not incur additional costs. Furthermore, providing this information would lead to health benefits as changes could be dealt with more promptly.
Quality of evidence	Thirteen qualitative studies were used to gather evidence. Their quality was independently assessed by two research fellows, who agreed that all of the studies were acceptable for inclusion to the study. All had minor limitations such as only one method of data collection, or unclear reporting of analytic strategies, but all contained direct evidence in terms of population and outcome.
Other considerations	The GDG used the evidence review and their experience to develop these recommendations. Their experience is that people may experience fluctuations in their clinical condition as a result of intercurrent illness such as infection, particularly urinary tract infections. This information helps people with MS interpret and manage their symptoms. People also need to be aware when they should seek medical help.

Recommendations	**19. Talk to people with MS and their family members or carers about the possibility that the condition might lead to cognitive**

	problems. **20. When appropriate, explain to the person with MS (and their family members or carers if the person wishes) about advance care planning and power of attorney.**
Relative values of different outcomes	People with MS are entitled to information about their condition and its likely impact.
Trade off between clinical benefits and harms	Some people with MS may not wish to be given certain information at certain times, but provided that the clinician takes individual wishes and circumstances into account, there are unlikely to be harms from offering information and support. Although no evidence was found for advance care planning/end of life care the GDG felt this was an important area to make a recommendation. Clinical benefits may include an increased locus of control, a greater ability to make appropriate self-management decisions, and reduced anxiety.
Economic considerations	No relevant economic evaluations were identified. The GDG considered that such discussions would be part of regular appointments for the management of multiple sclerosis and should not incur additional costs.
Quality of evidence	Fourteen qualitative studies were used to gather evidence. Their quality was independently assessed by two research fellows, who agreed that all of the studies were acceptable for inclusion to the study. All had minor limitations such as only one method of data collection, or unclear reporting of analytic strategies, but all contained direct evidence in terms of population and outcome.
Other considerations	This recommendation was informed by the experience of GDG. The GDG considered that people with MS may develop cognitive problems and this is rarely discussed. Significant problems are more likely to happen later in the disease course and sensitivity is required in discussing this with people with MS and their carers. The GDG did not consider this had to be discussed at the time of diagnosis but agreed that the issue could be very important for patients to be aware of. Advance notice allows people to institute steps such as advanced care planning and power of attorney when they are well enough to do so.

7 Coordination of care

7.1 Introduction

People with MS are faced with a varied array of symptoms and disabilities that may arise unpredictably and suddenly. Hence their physical, emotional and social needs may frequently require the action of more than one category of health professional at any particular time. Because the person with MS is usually based in the community it is often difficult for them to get the necessary treatment, advice and support from the right people when they need it most. Provision of care should therefore be adequately co-ordinated, to allow it to be timely, appropriate and comprehensive.

7.2 Review question: For adults with MS and their carers what process of care has been proposed to improve coordination of care and other related health outcomes?

Table 9: PICO characteristics of review question

Population	• Adults
Intervention/s	Any intervention aimed at improving coordination of care and other health outcomes, such as: • MS nurse • Regular review • Use of a key worker • Multidisciplinary team working • Centralised records • Electronic patient records • Established routines for handovers and exchange of information • Proactive follow-up of patients after significant life events or health events
Comparison/s	• Usual treatment
Outcomes	• Health-related Quality of Life. • Patient-reported outcomes, for example symptoms, activities. • Impact on carers. • Functional scales that quantify level of disability. • Treatment adherence • Patient and carer satisfaction • Relapse rates • Relapse management • Hospital admissions • Length of hospital admission • Outpatient/GP attendance
Study design	• Systematic reviews, RCTs, cohort studies, surveys, qualitative studies

7.3 Clinical evidence

We searched for randomised controlled trials (RCTs), non-randomised trials, surveys and qualitative studies that evaluated the use of methods to achieve co-ordination of care for adults with MS. The GDG wished to keep the genre of evidence very broad as they were not aware of any RCT or action research studies that had treated this issue as a complex intervention. Different study types that could inform consideration of the topic were therefore included.

9 studies[67,105,107,113,116,147,185,260,263] were included, which fell into 4 main groups, although there was some overlap. These were:

- those investigating the use of an MS nurse[66,67,107,113,260,263]
- those investigating the use of a multidisciplinary team[105,185]
- those investigating the use of self-assessment and management[147,148]
- qualitative studies looking at the experiences and opinions of adults with MS concerning co-ordinated care[116]

The principal aim of the review was to examine methods to improve co-ordination of care for people with MS. Evaluation of such methods ideally requires objective outcomes that directly measure co-ordination of care, but to our knowledge no such measures exist. The next best options are subjective measures, where adults with MS or clinicians give an opinion about whether the co-ordination of care has been realised.

Three quantitative studies using such subjective outcomes were found [105,260,263]. The other five quantitative studies did not contain any explicit objective or subjective measures of co-ordination of care. All contained an intervention that *could* be used to augment co-ordination of care, but it was clear that this intervention could also improve health via some other mechanism (such as more comprehensive rehabilitation). Hence improvements in generic (non-co-ordination of care) outcomes did not necessarily imply that the intervention had improved co-ordination of care. Nevertheless, these four studies were included, on the basis that there were reasonable grounds to suppose that the interventions would have had effects on health that were at least partially mediated by improvements in continuity of care. These four studies were, however, downgraded twice for indirectness.

The GDG were aware that MS nurses and other healthcare professionals involved in co-ordination of care will have other roles (for example, e.g. reviewing disease modifying therapy) and these other roles need to be taken into account when considering skill mix in services providing treatment to people with MS.

Meta-analysis of results was not carried out because there was too much heterogeneity between treatments, comparators and populations, even within the four treatment groups. Results have therefore been presented in narrative form

Details of these studies, including quality assessment, are summarised in Table 10.

Table 10: Characteristics of the included studies

Group	Study	Population	Methodology	Process of care	Grade
MS nurse	Kirker 1995[113]	Adults from UK with MS who were newly diagnosed or experiencing	'Before and after' interview questionnaire survey of adults with MS, carers	The use of a single MS liaison nurse. Compared to the situation prior to the nurse's introduction to	VERY LOW Non RCT. No independent comparator group, and serious

		MS-related problems	and GPs, and a retrospective examination of nurse records [n=82]	the service	indirectness
	Wilson 1998[263]	Adults from UK with MS	'Before and after' survey of adults with MS in two regions of England.[n=68]	The use of a MS nurse working in both community and hospital settings. No comparison group.	VERY LOW Non RCT. No independent comparator group, and serious indirectness
	Warner 2005[260]	Adults from UK with MS experiencing a relapse and indicated for IV methylprednis olone (MP)	Mixed methods quantitative/ qualitative study. [n=46 for initial audit; otherwise unclear]	Specialist nurse route to IV MP. Patients encouraged to telephone nurse if having symptoms of a relapse. Nurse would do assessment, get neurologist confirmation and administer the IV MP. Compared to GP route to IVMP.	VERY LOW Non RCT. Unclearly reported comparison group – unclear if parallel cohort or historical (pre specialist nurse) data.
	Forbes 2006[67]	People from UK with MS aged >16; mostly progressive MS	Non randomised controlled trial [n=616]	Normal service at 4 MS centres, each involving an MS nurse. Compared to normal service at 2 centres not involving an MS specialist nurse.	VERY LOW Non RCT, and serious indirectness
	Johnson 2001[106,108]	Adults with MS from UK	Before and after study [n=89]	Setting up of an MS nurse specialist post	VERY LOW Non RCT, poorly reported results
Multi-disciplinary (MD) team	Jansen 2006[105]	Adults with MS from Holland aged (intervention / comparator) 51/45 years	Prospective cohort [n=173]	Transmural care model – MD care protocol where a nurse is the case manager and biannual assessments are made by a MD team, leading to an integrated care pathway being formulated. Compared to 'traditional' care.	LOW Non-randomised comparison study.
	Pozzilli 2002[185]	Adults from Italy with MS; mostly secondary progressive	RCT [n=201]	Home based multidisciplinary rehabilitation care. Compared to usual treatment in MS referral centres.	VERY LOW Unclear allocation concealment, confounding through the multidisciplinary group having more treatments, and serious indirectness

Self-assessment	Miller 2011[147]	Adults from USA with clinically definite MS	RCT [n=167]	Online self-monitoring and self-management system as part of an internet based system of communication between the adult with MS and clinicians. Compared with internet based system of communication between the adult with MS and clinicians without self-monitoring or self-management system	VERY LOW Unclear allocation concealment and serious indirectness
Qualitative	Kroll 2003[116]	Adults from USA with MS, CP and SCI	Qualitative using semi-structured interviews. [n=30]	Aim was to find out about people's experiences of co-ordination of healthcare.	High quality qualitative study.

7.3.1 Clinical evidence concerning use of the MS nurse for fostering co-ordination of care

Kirker 1995[113]

This 'before and after' study assessed the effect of a liaison nurse on service quality, with no comparator. The liaison nurse was judged helpful or very helpful by 59/67 adults with MS and 45/51 carers. However, only 24/71 adults with MS felt the liaison nurse was a main source of support and reassurance and only 13/71 adults with MS saw the liaison nurse as someone to contact if needed. Furthermore, only 16/67 adults with MS felt the presence of the liaison nurse reduced GP visits, with the same proportion feeling that the liaison nurse reduced hospital visits. 66/101 GPs felt the nurse was helpful, but only 16/101 GPs felt their referrals had been reduced as a result.

Wilson 1998[263]

This 'before and after' study also assessed the effect of a liaison nurse on service quality, with no comparator. 90% of adults with MS in one region of England, and 100% of those in another region found MS nurse referrals helpful. Likewise, 79% and 84% felt contact with the MS nurse reduced the need to see the GP about MS. Finally, 96% and 94% found it supportive to have the MS nurse's contact number.

The most common qualitative issues raised in both regions concerned

- Satisfaction with the link with the neurologist provided by the MS specialist nurse
- The patient's desires for more follow up: "I'd like to be contacted every six months by the nurse whether I need it or not. It makes me feel someone is keeping an eye on my progress".

Warner 2005[260]

This study compared aspects of IV methylprednisolone treatment for adults with MS when led by a nurse service and when led by a GP service. Although this was not designed as a primarily quantitative study, some data were presented on access to treatment:

- The nurse service led to patient reporting of symptoms within a mean of 10 days whereas the GP service led to patient reporting of symptoms in a mean of 51 days.

- Organising a neurologist appointment took a mean of 6 days in the nurse service but 13.8 days in the GP service.
- The mean time from appointment to treatment commencement was 4.8 days for the nurse service and 6.2 days for the GP service.

Details of the GP service group were unclear and variance measures were not provided, so interpretation of these data is difficult.

48% of adults with MS also required physiotherapy and 8% required specialist continence advice. Although no data were presented, it was reported that, “these additional interventions, by the multidisciplinary team, were effectively organised by liaison between the specialist nurses and the day case clinical staff”.

Patient interviews suggested that, in the context of co-ordination of care, patients valued:

- Contact with a specialist nurse during relapse
- A close working relationship between specialist nurses and the neurologist
- Explanation of the nature of a relapse
- The specialist nurses’ ability to effectively organise appointments with consultant neurologists and treatment sessions
- Easy physical access to the clinical area where IV Methylprednisolone was provided
- Clinical staff who were knowledgeable and had the ability and time to discuss issues
- Treatment within 1 week of reporting symptoms.

Forbes 2006[67]

This study compared normal clinical service at 4 MS centres, each involving an MS nurse, to normal clinical service at 2 centres not involving an MS specialist nurse.

There was no significant reduction in the hospital admission rate in the past 12 months in the groups with MS nurses relative to the groups without MS nurses (repeated analysis chi square 2.6, p=0.26). However, the data suggested a very weak trend, with admission ranging between 12.35 to 15.6% in the intervention group compared to 18.9% to 25.2% in the control group (over 3 observation periods).

Adults with MS in the groups where an MS nurse was available were more likely (compared to adults with MS in groups where an MS nurse was not available) to report availability of a contact person at follow up, after adjustment for baseline availability of contact persons (group x time interaction p<0.001). Similarly, after baseline adjustment there were greater reports of help in an emergency (group x time interaction p=0.1), and help with urinary problems (group x time interaction p=0.11) in the groups with MS nurses. There were no clear group x time effects for help with fatigue (p=0.28), bowel problems (p=0.23), employment problems (p=0.57), depression, pressure sores (p=0.31) or relationship problems (p=0.53).

Quality of life and function at 24 months were generally poorer in the MS nurse group than the groups without an MS nurse after adjustment for baseline values. The uncertainty of the direction of the effect was high, except for SF36 general health and SF 36 energy vitality, where a clear effect favouring the group without MS nurses was observed. Table 20 and Table 12 summarise this information:

Table 11: Difference in quality of life between adults with MS in groups involving an MS nurse and adults with MS in groups not involving an MS nurse

Quality of life	Mean difference (intervention – control at 24 month follow up, adjusted for baseline values). Negative values indicate a worse outcome for the MS nurse groups
SF36 physical function	-2.81 (- 5.45 to 10.1)
SF36 role physical	-2.21(-5.8 to 1.4)
SF36 mental health	1.32 (-1.2 to 3.8)
SF36 social functioning	-1.61(-6.3 to 1.6)
SF36 bodily pain	-4.09(-7.2 to 0.9)
SF36 general health	-5.35(-8.1 to -2.5)
SF36 energy vitality	-2.82 (-5.5 to -0.1)

Table 12: Difference in function between adults with MS in groups involving an MS nurse and adults with MS in groups not involving an MS nurse

Function	MD (95% CIs) [Intervention – control] at follow up – adjusted for baseline inequality. Negative values indicate a worse outcome for the intervention group.
MSIS psychological	-2.38(-5.2 to 0.4)
MSIS physical	-1.83(-4.2 to 0.5)

Johnson 2001[106,108]

This retrospective before and after review of medical records, assessed the effects of the setting up of an MS specialist nurse post in West Berkshire in 1998. Results suggest no clear effect on patterns of attendance at emergency departments, hospital or day care. However, there did appear to be a trend for reduced length of first hospital stay (Table 13).

Table 13: Patterns of NHS attendance 6 months before and 6 months after the setting up of an MS nurse post

	6 months before MS post set up	6 months after MS Post set up	Statistics
Inpatient episodes	52	60	
Mean episodes per patient	1.5	1.447	
Emergency admissions	35/46 (76%)	33/50 (66%)	P=0.28
Length of hospital stay			
Episode 1	26.5 days (n=28)	14 days (N=33)	P<0.1
Episode 2	21.3 (n=12)	10.7 (n=13)	P=0.1
Episode 3	43.7 (n=6)	22.5 (n=4)	NS
Day case episodes	4 patients spent 12 days	9 patients spent 27 days	

Medical outpatient appts (primary diagnosis of MS)	1985	2048	

7.3.2 Clinical evidence for the use of the Multidisciplinary team (MDT) in fostering co-ordination of care.

Jansen 2006[105]

This prospective cohort study compared a multidisciplinary MS care protocol to 'traditional' care. Data were not shown but both groups were reported to have similar judgements of co-ordination of care at follow up.

In terms of healthcare use, there were differences at baseline between groups for use of rehab specialist, nurse specialist and physical therapist, so it is possible that 10 month findings were confounded by these baseline differences. No adjustments were made for baseline differences. However for other healthcare professional variables the groups were not significantly different at baseline (Table 14).

Table 14: Healthcare use in the multidisciplinary and traditional care groups

Healthcare professional	Multidisciplinary group at 10 months	Control group at 10 months	Between group p	Baseline equivalence?
Neurologist	64/80	47/96	<0.001	Y
GP	59/80	51/96	0.01	Y
Rehab specialist	17/80	11/96	NS	N – strongly favouring study group
Nurse specialist	40/80	29/96	0.01	N – favouring comparison group [*NB the baseline bias goes against the 10 month effect direction so the direction of effect favouring study group at 10 months can be taken as valid*]
Physical therapist	45/80	37/96	0.02	N – favouring study group
Occupational therapist	15/80	9/96	NS	Y
Social worker	12/80	8/96	NS	Y

Multidisciplinary care group participant's experienced better quality of life at 10 months in terms of feeling more energetic and vital, and showing fewer changes in general health. It is unclear, however, whether these changes in general health were adverse changes or not. This analysis was adjusted for baseline differences in quality of life (Table 15).

Table 15: Quality of life in the multidisciplinary and traditional care groups

Quality of life variable	Standardised regression co-efficient (95% confidence interval). This co-efficient, adjusted for baseline values, refers to the increase in the SF36 variable in the multidisciplinary group compared to the traditional care group. Hence a positive value indicates a benefit for the multidisciplinary group.	p
SF36 Physical functioning	**-1.662 (-6.099 to 2.856)**	**0.476**
SF36 Social function	2.532(-3.836 to 8.901)	0.434
SF36 role limitations (physical)	6.053(-4.283 to 16.389)	0.249
SF36 role limitations (emotional)	7.602(-4.426 to 19.632)	0.214
SF36 Mental health	-0.037(-4.313 to 4.239)	0.986
SF36 Energy and vitality	4.698(0.423 to 8.973)	0.031
SF36 Bodily pain	0.497(-5.869 to 6.863)	0.878
SF36 General health	-0.537(-5.094 to 4.019)	0.816
Reported health transition	7.678(1.886 to 13.470)	0.01

Pozzilli 2002[185]

This study compared home-based multidisciplinary care to usual care in MS referral centres. Most quality of life indices showed a greater improvement over the 12 months of the study in the multidisciplinary group (Table 16). Results from this study may have been confounded by the multidisciplinary group having greater quantities of treatment.

Table 16: Quality of life improvements (positive values indicate improvement) in the multidisciplinary and standard care groups

Measure	Mean difference (intervention group improvements minus control group improvements	95% CIs	p
SF36- Phys function (higher better)	0.27	-0.53 to 1.06	0.55
SF36- role physical (higher better)	3.67	-1.19 to 8.53	0.09
SF36- pain (higher better)	3.46	2.38 to 4.54	0.0001
SF36- general health (higher better)	5.02	4.50 to 5.51	0.0001
SF36- energy and vitality (higher better)	0.28	-0.38 to 0.94	0.41
SF36- social function (higher better)	1.09	0.51 to 1.67	0.001
SF36- general health(higher better)	12.39	9.85 to 14.93	0.0001
SF36- mental health (higher better)	-0.10	-0.25 to 0.05	0.19
Physical component score	**1.19**	**1.04 to 1.34**	**0.0001**
Mental component score	**0.75**	**0.58 to 0.91**	**0.0001**

In terms of function, no data for most results were given. It was merely stated that "no significant differences between intervention and control groups were detected for outcome measures including EDSS, FIM, MMSE, CDQ, FSS, STAI and STAXI. There was a trend in favour of the intervention group for changes in depression as measured by the CDQ score. A decrease in CDQ score was seen in the intervention group (-7.8%) while it was slightly increased (+0.7%) in the control group (p=0.11)".

7.3.3 Clinical evidence for use of the self-assessment and management in fostering co-ordination of care

Miller 2011[147]

This RCT assessed the effects of an online self-monitoring and self-management system as part of an internet based system of communication between the adult with MS and clinicians. This was compared with an internet based system of communication between the adult with MS and clinicians without any self-monitoring or self-management system.

There were no differences in function (Table 17) between the groups at 6 months. The EQ-5D index did not differ between groups but the EQ-5D VAS was significantly higher in the control group (**Table 18**). In terms of healthcare usage there were trends for a greater number of emergency room visits, a greater number of home health visits and a lower number of prescriptions in the intervention group (**Table 19**).

Table 17: Function in the intervention and control groups

Measure	Intervention (adjusted mean(SE)[n])	Control (adjusted mean(SE)[n])	p
Sickness Impact profile (higher worse)	22.4(1.8)[75]	21.7(2)[76]	0.77
MS functional composite (higher better)	-0.63(0.22)[84]	-0.8(0.24)[81]	0.51
MS self-efficacy scale (higher better)	62.5(2.6)[77]	64.5(2.8)[77]	0.50
Seniors general satisfaction and physician quality of care – general satisfaction with medical care (higher better)	23.2(0.67)[77]	23.3(0.72)[72]	0.96
Seniors general satisfaction and physician quality of care – perception of physician quality (higher better)	33.7(0.43)[77]	33.2(0.47)[77]	0.30

Table 18: Quality of life in the intervention and control groups

Measure	Intervention (adjusted mean(SE)[n])	Control (adjusted mean(SE)[n])	p
EQ-5D index (higher better)	0.756(0.023)[75]	0.757(0.025)[75]	0.96
EQ-5D VAS (higher better)	70.2(2.4)[74]	76.3(2.6)[75]	0.04

Table 19: Healthcare use in the intervention and control groups

Measure	Intervention (adjusted mean(SE)[n])	Control (adjusted mean(SE)[n])	p
Ever hospitalised	2.86%(1.87%)[83]	1.24%(2.04%)[80]	0.46
Ever admitted ER	12.4%(3.8)[83]	3.4%(4.1%)[80]	0.08
Number of medical office visits	8.53(1.31)[77]	7.36(1.59)[70]	0.46
Number home health visits	0.79(0.34)[77]	1.58(0.42)[70]	0.058
Number of prescriptions	10.5(1)[82]	11(1.1)[80]	0.068

7.3.4 Clinical evidence for patient views on co-ordination of care

Kroll 2003[116]

This qualitative study was carried out in the USA, on a sample of 30 adults with CP, MS and SCI, with mean age 44.8(8.3) years. The study used semi-structured telephone interviews, aimed to find out about people's experiences of co-ordination of health care. In the study, co-ordination of care was described thus: "co-ordination takes place when a single health care professional is knowledgeable about all of your healthcare needs and helps to manage how your needs are being met".

Person providing co-ordination of care

15/30 adults with MS reported that they had a health professional who co-ordinated their health care. 8 out of these 15 identified their care co-ordinator as a GP, primary care physician or family physician. When asked whom they ideally would want as their care co-ordinator, most reported they would want a specialist and not their primary care physician.

Problems with co-ordination of care

These fitted the following 3 themes – *disability-specific knowledge and understanding*, *time and effort* and *communication among providers.*

Disability-specific knowledge

A lack of such knowledge was often cited as a barrier to timely and effective co-ordination of services. Providers were perceived as knowing little about disabilities. One adult with MS said: "Not everything that happens is because of the MS, and having someone to identify that and recognise it as something they can handle and doesn't need a specialist, or recognise it as it when it is something specific that could need a specialist, and to make that kind of referral and so forth".

Time and effort

Some respondents felt that insufficient time and effort was invested in the co-ordination of their care by providers.

Communication among providers

Several respondents noted that lack of communication between providers involved in their care was an obstacle to co-ordinated care. One woman with MS did not experience care co-ordination because some of her previous providers did not "speak with those physicians…. Interface with those physicians, and that's what's critical".

7.4 Economic evidence

Published literature

One study was included for the use of the multidisciplinary team (MDT) in fostering co-ordination of care [185]. This study is summarised in the economic evidence profile below (Table 20). See also the study selection flow chart in Appendix C and study evidence table in Appendix F.

In addition, two UK studies were found reporting on the value of the MS specialist nurse [108,162]. The first by Johnson 2001 reported on the cost impact of the hire of an MS nurse on total hospital admissions. This study had serious limitations. The cost implications were per hospital as opposed to per patient. The analysis assumed that the reduction in total bed days was a direct consequence of the hire of a MS specialist nurse, for which there is no evidence. The costing was simplistic and did not account for other hidden cost factors that may have be involved such as the impact on other health resource use. Finally, the impact of an MS specialist nurse on patient quality of life was not assessed.

The second by Mynors 2012 was a non-comparative study which only reported the potential cost saving associated with implementing a MS nurse post. This study also had some serious limitations. The estimate of savings in resource use was illustrative and not based on evidence. The case load of each nurse was not clarified and therefore it was not possible to put the estimated resource savings into context. The national tariff as opposed to NHS reference costs were used to quantify the resource costs. When estimating costs from an NHS perspective, as we do in NICE Guidelines, NHS reference costs are preferred to national tariff as these represent actual costs without incentives. Finally, the cost of implementing a MS specialist nurse post was not explicitly compared to any other alternatives and we have assumed the comparator was ‘no MS specialist nurse’.

Although the results of Johnson 2001 and Mynors 2012 are discussed in the economic considerations, they have not been added to the inclusion list nor reported in a tabulated format as they do not meet the applicability and or methodological criteria as they are not economic evaluations but cost impact analyses.

See also the economic article selection flow chart in Appendix E.

Table 20: Economic evidence profile: Home-based multidisciplinary care versus usual care in MS referral centres

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Pozzilli 2002[185] (Italy)	Partially applicable (a)	Minor limitations (b)	Within trial cost consequence analysis (health outcome QoL using SF-36). RCT reported in clinical review [185]. Follow-up = 1 year. Cost-utility analysis conducted by NCGC by mapping SF-36 scores to EQ-5D.	saves £655 (c)	0.0207 QALYs (d)	dominant (£ per QALY)	Multivariate analysis conducted for costs in order to generate a best/worst case scenario. NCGC used the best/worst case scenario costs to analyse impact on ICER. Using the best case scenario home-based multidisciplinary care remained dominant. Using the worst case scenario the ICER was £9,015 per QALY, therefore home-based multidisciplinary care was cost-effective at the NICE threshold of £20,000 per QALY.

(a) Non-UK study. The study comparators make it difficult to distinguish if the incremental effects and cost savings observed are a result of the setting (home vs. outpatient) or the multidisciplinary approach.

(b) Cost of pharmaceuticals and aids for daily activities not included. SF-36 score used as opposed to EQ-5D.

(c) 1999 Euros, presented here as 1999 UK pounds. Euros converted using 1999 purchasing power parities[175]. Costs incorporated are: Inpatient care; medical and non-medical outpatient, home care and telephone service; and home care programme.

(d) NCGC estimated quality of life values from within trial SF-36 scores mapped to EQ-5D values using algorithm by Ara and Brazier (2008) [8] and baseline SF-36 scores from a study of people with MS in UK [193].

Economic considerations

The first paper, Johnson 2001[106,108], provided an analysis of the cost impact of hiring a MS specialist nurse in the Royal Berkshire and Battle NHS Hospital Trust on the number of MS patient admissions (bed days). The cost of MS patient admissions were calculated by multiplying cost per day of elective and non-elective admissions (reported by the Trust) by the total number of bed days over a one year period prior to and after the hire of a MS specialist nurse. The annual cost of employing a MS specialist nurse was calculated based on salary (grade H), pension contribution, national insurance, travel and office costs and was estimated to be £39,719. Overall, a net saving of £64,611 to the Trust was estimated (see Table 21).

Table 21: Cost implications of MS nurse hire on MS patient admissions (a)

	Total bed days (non-elective; elective)	Cost per bed day (non-elective; elective) (b) (£, 1999-2000)	Total cost of MS patient admissions (£, 1999-2000)
Year before MS nurse hire	1,274 (1,122; 152)	256; 180	271,573
Year after MS nurse hire	820 (523; 297)	256; 180	167,244
Total saving			**104,330**
Net saving			**64,611**

(a) Source [106,108]
(b) The authors assumed that patients admitted non-electively would spend half their stay in beds at the higher cost rate before being transferred to beds at the lower cost rate.

The second paper, a report by Mynors 2012 [162], provided illustrative costs and savings associated with a new MS specialist nurse post using the MS Society cost calculator (2011 edition). This tool is an excel spread sheet which can be used to calculate the actual cost of employing a MS specialist nurse against the cost savings from avoided admissions and other attendances, based on national tariff. The total annual cost (excluding cost savings) for one MS specialist nurse post was estimated to be £63,980. The cost components considered in the analysis were salary, overheads, telephone, mileage, computer, shared clinic receptionist, clinic room and secretarial support.

The report considered also the potential cost savings in terms of saved outpatient appointments and emergency admissions associated with one MS specialist nurse post. These were illustrative figures which assumed a saving of 300 outpatient appointments and 40 emergency admissions were attributed to the creation of one MS specialist nurse post. Using the national tariff, the authors calculated the expected cost savings to commissioners. The total estimated cost saving was £54,000 for each post (see Table 22). In addition, the report calculated that to breakeven, one MS specialist nurse would need to save 199 outpatient appointments and 21 emergency admissions (see Table 23). Two case studies estimating savings associated with an MS specialist nurse were presented in the report but the figures differed from the cost analysis above.

Table 22: Illustrative cash releasing savings (a)

	Number saved	Unit cost (£, 2011)	Total saving (£, 2011)
Neurology follow up outpatient appointment	300	91	27,300
Neurology emergency admissions	30	2,331	69,930
Other emergency admissions (e.g. UTI)	10	2,056	20,560
TOTAL SAVING			**117,790**
NET CASH RELEASING SAVING TO COMMISSIONER			**53,810**

(a) Source: [162]

Table 23: Breakeven assumptions (a)

	Number saved	Unit cost (£, 2011)	Total saving (£, 2011)
Neurology follow up outpatient appointment	199	91	18,109
Neurology emergency admissions	10	2,331	23,310
Other emergency admissions (e.g. UTI)	11	2,056	22,616
TOTAL SAVING			**64,035**

(a) Source: [162]

Finally, in the clinical review, one paper [66,67] reported quality of life differences between groups with an MS nurse compared to groups without an MS nurse. Quality of life and function at 24 months were generally poorer in the MS nurse groups than the groups without an MS nurse after adjustment for baseline values. Furthermore, there was no significant reduction in the hospital admission rate in the past 12 months in the groups with MS nurses relative to the groups without MS nurses. This contradicts the findings from the papers above. Together, the decreased quality of life and lack of reduction in hospital admissions would indicate that an MS nurse would not be cost-effective.

7.5 Evidence statements

7.5.1 Clinical

Use of the MS nurse for fostering co-ordination of care

Very low quality evidence from five qualitative/quantitative studies comprising 901 participants showed that MS nurses were generally regarded as helpful by patients. Patients particularly valued the link they provided to the neurologist, the fact that they were a reliable and knowledgeable contact in times of need, and the fact that treatment was expedited. Some quantitative data supported this, with clinically important reductions in time to treatment when MS nurses were used. However, some data showed that MS nurses were not perceived to reduce GP visits, and there was little evidence suggesting a positive effect on quality of life.

Use of the Multidisciplinary team (MDT) in fostering co-ordination of care

Low to very low quality evidence from two quantitative (cohort and randomised) studies comprising 374 participants suggested an improvement in quality of life for patients when a multidisciplinary approach was used. However no clear positive effects on function were observed.

Use of self-assessment and management in fostering co-ordination of care

Very low quality evidence from one non-randomised quantitative study comprising 167 participants showed little difference in terms of quality of life, function and healthcare use between an online self-assessment and management approach, including communication with clinicians, and an online approach restricted to communication with clinicians.

Patient views on co-ordination of care

High quality evidence from one qualitative study comprising 30 participants showed that this sample of people with MS:

- Wanted a specialist rather than generalist as their care-co-ordinator
- Felt their care co-ordinator should know more about disability
- Wanted more time and effort invested by their care co-ordinators
- Wanted better and timelier communication between members of the care team to enhance co-ordination of care.

7.5.2 Economic

One cost–utility analysis found that in people with MS home-based multidisciplinary care was dominant (less costly and more effective) compared to usual care. This analysis was assessed as partially applicable with minor limitations.

7.6 Recommendations and link to evidence

Recommendations	**21. Care for people with MS using a coordinated multidisciplinary approach. Involve professionals who can best meet the needs of the person with MS and who have expertise in managing MS including:** o **c**onsultant neurologists o MS nurses o physiotherapists and occupational therapists o speech and language therapists, psychologists, dietitians, social care and continence specialists o GPs. **22. Offer the person with MS an appropriate single point of contact to coordinate care and help them access services.**
Relative values of different outcomes	Co-ordination of care is not an end in itself but a means to improve care for a patient. It should be an integrated approach (interdisciplinary), rather than different professions working in paralell. The GDG recognised that co-ordination of care is difficult to measure. There were no objective outcome measures that directly looked at care co-ordination itself. The GDG however considered that the subjective responses on the quality of care co-ordination by clinicians and patients were of great importance. The GDG noted that the quality of life measures were also important in this area but that they might not be sensitive to changes in coordination of care. The majority of quantitative studies measured generic outcomes, such as well-being or number of hospital admissions - these may reflect other changes in care rather than co-ordination of care. These were downgraded as indirectly relevant studies, and cautiously interpreted. The GDG also found that many of the studies, including the economic evidence, did not address the perceived benefits of multidisciplinary care or an MS nurse, for example faster access or continuity of care. One study reported an increase in day case episodes in patients under the care of an MS nurse. Studies, especially the one qualitative study, suggested that patients valued: a) communication between the healthcare staff involved in their care, b) access to a knowledgeable specialist and c) speed of referral, specialist review and treatment.
Trade off between clinical benefits and harms	No harms were identified from the provision of a multi-disciplinary team or an MS nurse. It was considered that they would be of clinical benefit and that the main issue would be cost-effectiveness.
Economic considerations	One cost–consequence analysis comparing home based multi-disciplinary care to usual care in MS referral centres was identified where outcomes included quality of life measured using SF-36. The NCGC estimated quality of life values by mapping the SF-36 scores to EQ-5D values using an algorithm by Ara and Brazier (2008) [8] and baseline SF-36 scores from a study of people with MS in UK [193], thus allowing the study to be presented as a cost–utility analysis to the GDG. This study found that home-based multi-disciplinary care was dominant (both less costly and more effective) compared to usual care in MS referral centres. Of note, the study compared both the setting (home or outpatient) and a multidisciplinary approach (presence or absence) together, which make

	it difficult to identify if the incremental effects and savings observed are due to the treatment approach, the setting or both. Despite this applicability issue, the GDG felt it was cost-effective to deliver care using a coordinated multidisciplinary team. Two further studies on the value of the MS nurse were considered by the GDG however they were not included in the formal review as they did not meet the methodological and applicability criteria. Both studies were UK cost impact analyses which suggested savings to the NHS, as a result of hiring an MS nurse, in terms of hospital and emergency admissions and outpatient appointments. Both studies had serious limitations including no data on the MS nurse case loads and a lack of evidence base for their impact on admission rates.Members of the GDG considered that admission for management of MS was not common and while an MS nurse might free up time for a neurologist to see patients, additional associate specialists might see people with MS and other neurology patients. Therefore GDG felt that they did not provide sufficient evidence to recommend an MS nurse on economic grounds. The GDG felt it was impossible to generalise the role of the MS nurse in any setting, each MS nurse and each setting works and functions differently. The GDG noted that in many cases people with MS will be cared for by an MS nurse for the management of their treatment with disease modifying drugs.
Quality of evidence	The GDG included different study types- surveys, qualitative studies, and non-randomised trials- as well as randomised controlled trials, because in their experience there was comparatively little research into co-ordination of care. Despite this, only nine studies were identified – eight quantitative and one qualitative. Some studies were conducted in Europe where the healthcare team structure is known to be different. Also, the role of an MS specialist nurse and multidisciplinary teams has been evolving in the last decade. Therefore the evidence from older studies may be less relevant. The biggest drawback related to the type of study. Most studies were not randomised, controlled or blinded. There were inherent biases and confounders as a result. Studies also looked at indirect interventions or outcomes. The only study on self-management in MS looked at online self-management/self-monitoring as an intervention at a USA centre. However both groups were using internet-based communication already, which was not the usual method of self-management in the UK. The intervention group reported worse quality of life or no difference in outcomes. The economic evidence on home-based multi-disciplinary care was rated as being partially applicable with minor limitations. No economic evidence was formally included on the cost-effectiveness of MS nurses.
Other considerations	The GDG considered the importance of co-ordination of care for people with MS. They considered that a single point of contact agreed with the patient was of importance. This single point of contact essentially allows self-referral that will ensure a person gets referred appropriately. The GDG acknowledged that MS nurses may be perceived to be critical in providing co-ordination of care but considered that other ways of achieving this exist. There can be difficulty in discussing MS nurses as there is no specific qualification for this role and MS nurses may have different roles and job descriptions in different organisations. MS nurses have traditionally been associated with the delivery of disease modifying drugs.

	Studies looking at MS nurses suggested that patients and clinicians thought they were helpful, but were less clear about whether this reduced the need for admission or to see GPs and specialists. Patients appeared to value the MS nurse as a knowledgeable point of contact and a way of rapidly accessing the specialist team. For example, two studies found high satisfaction with the link to a neurologist that was provided by the MS nurse One study (Forbes, 2006) found that quality of life was worse in centres with an MS nurse present – the patient groups were adjusted for baseline differences (eg. age, time since diagnosis), but this study was observational and there was the possibility that quality of life was unrelated to the provision of an MS nurse. Multidisciplinary teams (MDTs) did not improve co-ordination of care when patients were asked directly or when using objective health outcome measures, such as EDSS. However, they did appear to increase healthcare use with a trend towards better quality of life and less depression. The two studies here suffered from flaws such as lack of randomisation and the interventional (MDT) group receiving more care or rehabilitation than the control group. The GDG considered that co-ordination of care was best seen from a patient perspective. The evidence and GDG experience indicated that people with MS want a point of contact, ideally someone with knowledge of them and of MS, and for timely communication to occur between the professionals involved in their care. The GDG considered that it was not appropriate both on the current evidence base and on their knowledge of differing service organisation, to recommend one model for co-ordination of care. They did consider that due to the complexity and low prevalence of MS, every person with the disease should be able to access healthcare professionals who are knowledgeable. The GDG did not think there was evidence that first point of contact and co-ordination professional had to be carried out by an MS nurse. One member noted that some centres employ an MS physiotherapist or MS occupational therapist instead. The GDG considered that clarity about organisation of care and how it was being co-ordinated and delivered was vital. The GDG considered that while it might be possible to define a core multi-disciplinary team of people who are involved in patient care e.g. a neurologist, MS nurse, physiotherapist and occupational therapist, individual patients might have more need of management from other health care professionals such as continence nurse, a rehabilitation physician, a speech and language therapist or a psychologist, or from social care. A multi-disciplinary team approach should encompass all these perspectives as well as those of patient and family.

8 Modifiable risk factors for relapse or progression of MS

8.1 Introduction

Many lifestyle factors are suspected of affecting the clinical course of multiple sclerosis, but the evidence for their influence is unclear. It is vital that people with multiple sclerosis and their carers are given clear and accurate advice about such factors. This is to avoid unnecessary avoidance of potentially beneficial activities, and to promote appropriate vigilance when undertaking activities for which evidence suggests a possible harm. The GDG were asked to evaluate which, if any, lifestyle factors outside of specific medication for multiple sclerosis (MS) could affect the course of the disease in terms of relapse rate and/or progression. The GDG identified the following specific areas to be reviewed: exercise, vaccinations, stress, pregnancy and smoking. These were factors felt to be those that were:

1. modifiable
2. relevant to a large number of people with MS
3. controversial in terms of their effects
4. important in terms of the potential impact their encouragement or discouragement would have.

8.2 Review question: Do the modifiable risk factors of exercise, vaccinations, stress, pregnancy and smoking influence progression of Multiple sclerosis?

For full details see review protocol in Appendix C.

Table 24: PICO characteristics of review question

Population	• Adults with MS
Risk factors for progression to be considered	• Exercise/ physical activity levels • Vaccinations • Stress • Pregnancy • Smoking
Outcomes	• Effects of the risk factor on progression of MS, as defined by adverse changes in the following: ○ Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale. ○ Relapse rates/ severity of relapses / relapse durations ○ Patient-reported outcomes, for example symptoms, activities. ○ Functional scales that quantify level of disability, such as the Expanded Disability Status Scale (EDSS), the Multiple Sclerosis Functional Composite (MSFC), the Cambridge Multiple Sclerosis Basic Score (CAMBS), the Functional Assessment of Multiple Sclerosis (FAMS), MS walking scale (MSWS-12) or the National Fatigue Index

	(NFI). ○ Cognitive functions, such as memory and concentration, and physical symptoms including fatigue, spasticity, spasms, assessed by validated and disease-specific scales, questionnaires or similar instruments, for instance the Scripps Neurologic Rating scale (SNRS) or the Krupp Fatigue Severity Scale (FSS).
Exclusion	• Children younger than 18 years • Studies conducting only univariate analyses
The review strategy	• RCTs • Pooled analysis of patient level data • Epidemiological studies – prospective cohort • The GDG did not wish to identify any specific confounders that needed to be considered in the studies but the analysis had to be adjusted for appropriate confounders identified on a study by study basis.

8.3 Clinical evidence

Seventeen prospective cohort studies[1,14,29-31,95,151,152,159,181,184,204,206,209,216,239,265] and two RCTs [145,161] were found addressing the review protocol. Their methodologies and results have been presented below in the following sections, categorised by risk factor:

- Exercise/activity levels
- Vaccinations
- Stress
- Pregnancy
- smoking

8.3.1 Clinical evidence for the prognostic effects of exercise/activity levels on MS progression

Two prospective studies were found (Motl 2011[159] Stuifbergen 2006[239]) that fulfilled the inclusion criteria.

Table 25: Summary of studies included in the review

Study	Population	Risk factor	Outcome	Quality
Motl 2011[159,159]	Men and women with MS (246 with RRMS and 46 with PPMS or SPMS) of mean age 37.7 (10.1) years and mean duration of MS 10.3 (7.9) years.	Physical activity at baseline measured by an accelerometer over 7 days.	Disability progression measured by the Patient determined Disease Steps (PDDS) scale.	MODERATE Poor reporting of results. Overly short follow up of 6 months. No assessor blinding reported, but unlikely to have been a problem as the outcome was objective. Used a “panel analysis” approach that adjusted for confounding and drop-out.
Stuifbergen 2006[239]	611 men and women with MS (41% with	Exercise behaviours -	QoL change over course of study) –	MODERATE Poor reporting of

	RRMS and 18% PPMS, 17% SPMD, 11% progressive relapsing and 3% benign or unknown classification). Aged 21-80 years.	using the 8-item exercise/physical activity subscale of the Health Promoting Life-style Profile II (HPLP-II).	using Quality of Life Index – MS version. Functional limitations change over course of study – using Incapacity status scale (ISS).	results. No assessor blinding but unlikely to be a problem as data was generated by self-report questionnaires. Attrition low and catered for by the analysis.

Narrative review for the prognostic effects of exercise/activity levels

Effects of exercise on later progression of disability

Motl 2011[159] assessed the effect of baseline activity levels on the progression of disability 6 months later. On un-adjusted analysis, baseline physical activity did not predict a change in disability from baseline to 6 months, as measured by the PDDS (path co-efficient= 0.0, p=0.49). Path co-efficients represent the amount of expected change in the dependent variable (change in disability) with a unit change of the risk factor (change in activity), thus zero represents no relationship. After adjustment for confounders (sex, age of MS onset, clinical MS course and occurrence of a relapse) very similar results were obtained, but these were not reported. [MODERATE QUALITY]

Stuifbergen 2006[239] produced evidence to suggest that the level of exercise at baseline was significantly and negatively associated with the rate of deterioration of functional performance over time. In other words, higher exercise levels at baseline led to slower deterioration in function. However this was a weak effect (r=-0.17). [MODERATE QUALITY]

Effects of exercise on later changes in quality of life

Stuifbergen 2006[239] produced evidence to suggest that the level of exercise at baseline was not associated with changes in QoL over the study period (r did not differ statistically from zero). This was largely a function of no change in QoL over the study period (annual rate of change of 0.032). [MODERATE QUALITY]

8.3.2 Clinical evidence for the prognostic effects of vaccinations on MS progression

Three RCTs were found ([14,145,148,153,161]) that fulfilled the inclusion criteria (Table 26). Results of the comparison between influenza vaccine and placebo are presented in Table 27.

Table 26: Summary of studies included in the review

Study	Population	Risk factor	Outcome	Quality
Miller, 1997[145]	Clinically definite RRMS, with EDSS <6.5. No treatment with immunosuppressive medications, interferon β or co-polymer 1 within the previous 6 months.	Standard influenza vaccination, compared to a placebo injection in an RCT methodology.	Exacerbations of MS, characterised by objective change on the neurologic examination resulting in an increase of at least 0.5 in EDSS, or an increase in one grade on the Kurtze FSS, with symptoms lasting at least 24 hours, without fever.	LOW No reports of healthcare professional blinding. Possible selection bias due to no reports of allocation concealment. No attrition.
Myers, 1977[161]	Adults with MS; duration of disease 13-17 years; 55-61% RR.	Mixed influenza virus vaccine , compared to a placebo injection in an RCT methodology	Exacerbation of symptoms. Frequency of relapses.	MODERATE Unclear allocation concealment.
Mokhtarian, 1997[153]	Clinically definite RRMS, with EDSS <6.5. No treatment with immunosuppressive medications, interferon β or co-polymer 1 within the previous 6 months.	Standard influenza vaccination, compared to a placebo injection in an RCT methodology.	Exacerbations of MS	LOW No reporting of allocation concealment or assessor blinding.

Vaccine versus placebo

Table 27: Clinical evidence profile: vaccine versus placebo

Quality assessment							Proportion with the event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vaccine	placebo	Relative risk (95% CI)	Absolute effect (risk difference)		
Exacerbations at 21-28 days												
Miller, 1997 Myers, 1977 Mokhtarian 1997	randomised trials	very erious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	6/91 (6.6%)	7/95 (7.4%)	RR 0.87 (0.31 to 2.48)	16 fewer per 1000 (from 83 fewer to 179 more)	VERY LOW	CRITICAL
Exacerbations at 3-6 months												
Miller, 1997 Myers, 1977 Mokhtarian, 1997	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	Serious[B]	none	18/91 (19.8%)	12/95 (12.6%)	RR 1.55 (0.80 to 3.02)	67 more per 1000 (from 24 fewer to 244 more)	VERY LOW	CRITICAL
Numbers worsening over follow up												
Miller, 1997	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	8/49 (16.3%)	10/54 (18.5%)	RR 0.88 (0.38 to 2.05)	22 fewer per 1000 (from 115 fewer to 194 more)	VERY LOW	CRITICAL

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

8.3.3 Clinical evidence for the prognostic effects of stress on MS progression

Six prospective studies were found [1,29-31,151,152,184,216] that fulfilled the inclusion criteria (**Table 28**). Two papers [29,30] used the same sample of participants but looked at slightly different confounding variables.

Table 28: Summary of studies included in the review

Study	Population	Risk factor	Outcome	Quality*
Schwartz 1999[216]	MS patients aged 45.7(11.3); 39% RR, 19% chronic progressive, 42% chronic stable; EDSS 4.1; 14.2 years since symptoms began.	Stressful life events, as measured by the Holmes and Rahe stressful event checklist. Measured every 6 months and possibly summated over the full 6 year follow up, but unclear.	Overall deterioration in Functional Systems Scales at 6 years.	VERY LOW Assessor blinding not mentioned. Limited range of confounders considered in the analysis, despite heterogeneous sample. Unclear reporting of analysis. Very high attrition over the 6 years (70%).
Buljevac 2003[31]	MS patients aged 19-55; All RR; EDSS median 2(0-6); 4 years since symptoms began.	Self-reported stressful events, written up on a weekly basis in a diary.	Exacerbations of RR MS at a mean of 74 weeks.	VERY LOW Assessor blinding reported, but unclear. Only infection included as a confounder in the analysis. Unclear reporting of analysis. 18% attrition, but those lost still continued for a range of 8-112 weeks.
Brown 2006A[29]	MS patients aged 42.6 (10.7) years; 75% RR, with the rest SPMS; EDSS mean 3.6(2.2); 8.3 years since symptoms began.	Self-reported stressful life situations, derived from a telephone interview conducted by trained interviewers using the Life Events and Difficulties Schedule (LEDS).	Exacerbations at a mean of 2 years.	LOW Assessor blinding carried out. Good range of confounders included in analysis. Poor reporting of analytical methods. Attrition of 49.5% by 2 years.
Brown 2006[30]	MS patients aged 42.6 (10.7) years; 75% RR, with the rest SPMS; EDSS mean 3.6(2.2); 8.3 years since symptoms began.	Self-reported stressful life situations, derived from a telephone interview conducted by trained interviewers using the Life Events and	Exacerbations at a mean of 2 years.	LOW Assessor blinding carried out. Good range of confounders included in analysis. Poor reporting of analytical methods. Attrition of 49.5% by 2 years.

Study	Population	Risk factor	Outcome	Quality*
		Difficulties Schedule.		
Ackerman 2002[1]	MS patients aged 28-57; all female; 78% RRMS, 22% SPMS; most on DMDs; 7.9 years since symptoms began.	Stressful life events, collected via the Psychiatric Epidemiologic Research interview and LEDS.	Exacerbations at a mean of 1 year.	LOW Assessor blinding. No attrition reported. Only one confounder considered.
Potagas 2008[184]	MS patients aged 20-52; all female; All RRMS; EDSS median 0.47(0-3); 3.6 years since symptoms began.	Anxiety, measured with the Hamilton rating scale on a monthly basis Also, stressful life events recorded via diaries.	Exacerbations at a mean of 1 year.	VERY LOW Assessor blinding unclear. 14% attrition by 1 year; A very limited range of confounders considered.

*Quality rating started at HIGH, and lower grades were MODERATE, LOW and VERY LOW. Each successive limitation led to a single downgrade, except a low range of confounders in the analysis, which, because of its potential for causing severe bias, led to a double downgrade.

Narrative review for the prognostic effects of stress

All evidence was classed as LOW or VERY LOW, as documented in Table 28.

<u>Effect of prior stressful life events on exacerbations of RRMS</u>

Buljevac 2003 produced evidence to suggest that participants with a stressful event in the previous 4 weeks had a RR (95% CI) of 2.2 (1.2-4) for a first exacerbation during the following week, compared to those with no stressful event in the past 4 weeks. For those participants who had already had a first exacerbation, a stressful event in the past 4 weeks led to an even higher RR of 2.7(1.2-6) for a second exacerbation over the next week, compared to those with no stressful event in the past 4 weeks. For those participants who had already had a second exacerbation, a stressful event in the past 4 weeks led to a RR of 1(0.4-2.5) for a third exacerbation over the next week, compared to those with no stressful event in the past 4 weeks. The only confounder considered, and for which adjustments were made, was infection. The unit of analysis was a participant-week. [VERY LOW QUALITY]

In agreement with Buljevac 2003, Ackerman 2002 produced evidence to suggest 36% of control dates were preceded by a negative life event within the previous 6 weeks, compared to 88% of selected exacerbations, although any statistical difference was not reported. Control dates averaged 32.8(2.8) days from the most recent stressor, but the mean time from stressor to exacerbations was 14(2.9) days ($p<0.0001$). This was not affected by MS sub-type, as shown by the stratified sub-group analyses, but other confounders were not considered. [LOW QUALITY]

<u>Effect of the frequency of prior stressful life events on exacerbations of RRMS</u>

Brown 2006A produced evidence to suggest that the frequency of acute stressors was associated with exacerbation [OR for exacerbation: 1.3(95% CIs: 1.1-1.5)] after adjustment for disability, sex, smoking, age, memory, positive stressors, time, relationship status, country of origin, recruitment site and education. Although unclearly reported, it seems likely that the OR was in relation to an increment increase in stressor frequency. [LOW QUALITY]

Using the same sample as Brown 2006A, but a different set of covariates, Brown 2006 produced evidence to suggest that the frequency of acute stressor events was again associated with exacerbation [OR for exacerbation: 1.3(95% CIs: 1.1-1.4)], after adjustment for age, relationship status, country of origin, sex, recruitment site, smoking status and education, physical fatigue, depression, seeking social support and self-controlling. Again, it seems likely that the OR was in relation to an increment increase in stressor frequency, though this is unclearly reported. [LOW QUALITY]

Ackerman 2002 also produced evidence to suggest that an increase in the rate of life events was also associated with an increased likelihood of developing an exacerbation. The HR for exacerbations was 13.18 (95% CIs: 1.67-104.39), indicating the increase in relative hazard of an exacerbation for a one unit change in the rate of life events. However, it is unclear if any adjustments were made for confounders. [LOW QUALITY]

Potagas 2008 assessed the effect of ≥3 stressful events (in a 4 week period) on the time to event for first, second and third relapses. After adjustment for only two confounders (disease duration and episode of infection), ≥3 stressful events was associated with a RR for first relapse of 8.9(3.4-23.5) within a 4 week period, a RR for second relapse of 18.1(2.8-115.4) and a RR for third relapse of 3.6(0.5-26.6), compared to <3 stressful events. [VERY LOW QUALITY]

Effect of the severity of prior stressful life events on exacerbations of RRMS

Neither Brown 2006 nor Brown 2006A reported any relationship between severity of stressors and exacerbations, although severity of stressors was measured as a potential risk factor. [LOW QUALITY]

Effect of anxiety levels and other psychological stress indicators on MS exacerbations

Potagas 2008 assessed the relationship between the level of anxiety (HAM-A ≥18) and first, second and third relapses. After adjustment for only two confounders (disease duration and episode of infection), level of anxiety (HAM-A ≥18) was associated with a RR for first relapse of 3(1.3-7.4) within a 4 week period (compared to HAM-A<18), a RR for second relapse of 7.2(2.0-26.8) within a 4 week period (compared to HAM-A<18) and a RR for third relapse of 1.8(0.2-18.8) within a 4 week period (compared to HAM-A<18). The high imprecision of the third relapse analysis probably arises from the very small sample size of 12. [VERY LOW QUALITY]

Effect of stressful life events on disease progression

Schwartz 1999 produced evidence to suggest an increased risk of disease progression, as measured by the functional systems scale, with greater levels of stress (>1 life event in past 6 months), at an OR of 1.13, p<0.0003, after adjustment for age, gender and education. [VERY LOW QUALITY]

Summary of evidence on stressful life events and effects on MS

LOW and VERY LOW quality evidence suggests that stressful life events may be associated with subsequent exacerbations and functional deterioration. In terms of exacerbation, the frequency of stressful life events may be more important than the severity. High anxiety levels may also be related to exacerbations. Finally, increased conflict and disruption in routine may be related to the development of new brain lesions.

8.3.4 Clinical evidence for the prognostic effects of pregnancy on MS progression

Four prospective studies[204,206,209,265]) were included in this review. Their populations and methodologies are summarised in Table 29.

Table 29: Summary of studies included in the review

Study	Population	Controls	Length of follow-up/Outcomes	Quality
Sadovnick 1994[209]	Women with a diagnosis of MS according to recognised criteria who attended the MS Clinic during 1982 through 1986 and subsequently became pregnant during the period 1982 through 1989. Canada N=42 (58 births).	Non pregnant women with MS matched for gender, year of birth ± 3 yrs), age of MS onset (± 2 yrs), ms type, ms course and initial symptoms N=64.	6 months postpartum. Relapse rates.	MODERATE Groups well matched for potential confounders Short follow-up period.
Worthington 1994[265]	Patients attending a MS unit. These patients were seen routinely 6 to 12 monthly for neurological interview and examination and attended for regular physiotherapy and nutritional advice. UK. N=15. N=10 relapsing remitting. N=4 secondarily progressive. N=1 primary progressive.	Pregnant women matched with non- pregnant women for age, duration of disease and Expanded Disability Status Score (EDSS) N=22. N=17 relapsing remitting. N=5 secondarily progressive.	3 yrs Relapse rates Severity of relapse.	MODERATE Groups matched on a limited number of potential confounders Adequate follow-up period.
Roullet 1993[204]	Females with MS according to Schumacher's criteria (before 1983) and Posers criteria (after 1983). For inclusion, needed to have the relapsing remitting form of MS and been followed up for one year or more. Two pregnancy groups – those with at least one pregnancy during follow up [prior pregnancies before follow up did not prohibit their inclusion in this group] (n=33), and those pregnant after MS	Females with MS according to Schumacher's criteria (before 1983) and Posers criteria (after 1983). For inclusion, needed to have the relapsing remitting form of MS and been followed up for one year or more. Controls were defined as those	10 years Relapse rates. Proportion progressing to progressive MS.	VERY LOW Inadequate consideration of potential confounders. No assessor blinding. Those withdrawing early were not included in the analysis – potential for attrition bias.

	onset but not during follow up (n=17).	who had no prior pregnancy (either before study or during FU). (n=75).		
Runmarker 1995[206]	Women with relapsing remitting MS (MS definite or probable) who had become pregnant after MS onset. N=28 [24 of these also contributed to the control group prior to their first pregnancy].	Women with relapsing remitting MS (MS definite or probable) who had not become pregnant. N=55.	25 years Progression to progressive disease Reaching level 6 on the Disability Status Scale (DSS).	VERY LOW Inadequate consideration of potential confounders. No assessor blinding reported. Unclearly reported analysis.

Narrative review for the prognostic effects of pregnancy

Sadovnick 1994[209]

There was no significant difference between the cases and controls for the number of relapses experienced, except for a lower than expected rate in the cases compared to controls in the third trimester (Table 30). [MODERATE QUALITY]

Table 30: Comparisons of observed and expected relapses: Cases compared with matched-control (N=42) (matched controls could not be found for 5 cases)

	Observed No. of relapses	Expected No.	P value
Trimester			
First (T1)	10	10.4	ns
Second (T2)	6	10.4	ns
Third (T3)	2	10.4	.014
Months after delivery			
Up to 3 (P1)	13	10.4	ns
4-6 (P2)	7	10.4	ns

Worthington 1994[265]

There was no significant difference between the pregnant and nulliparous group with respect to total number of relapses or severity of relapse (Table 31). [MODERATE QUALITY]

Table 31: No. and severity of relapse for the pregnant and nulliparous group

	Pregnant group N=15	Nulliparous group (N=22)	P value
Number of relapses	25	40	ns
Total no of yrs	51.75	80.67	

Relapse rate	0.48	0.50	
No of relapses according to age of onset < 25 yrs > 25 yrs	 19 6	 26 14	 ns
Total relapse points	1008	1610	ns

Roullet 1993[204]

The group having no history of pregnancy had a strong trend for a higher relapse rate than the two pregnancy groups. It should be noted that for the group with pregnancy *during* follow up, it is possible that some relapses occurring during follow up may have occurred *prior* to pregnancy, and so the validity of this group for the evaluation of the effects of pregnancy on relapse rate is questionable. There were no differences in the proportion of patients progressing to the progressive form of the disease in the three groups. [VERY LOW QUALITY]

Table 32: Relapse rates and proportion progressing to progressive MS in the two pregnancy and no pregnancy groups.

	Pregnant during FU	Pregnancies after MS onset but not during FU	No pregnancies after onset	P value
Relapse rate/year mean(sem)*	0.54(0.13)	0.55(0.20)	0.86(0.09)	0.07
Transition to progressive form	8/33 (24%)	4/17 (23%)	13/75 (17%)	NS

*Adjusted for age at MS onset and duration of disease at study onset.

Runmarker 1995 [206]

There was a significantly lower risk of onset of a progressive course in the state 'pregnancy after onset' compared with the state 'before pregnancy' (p=0.0239).

For each year of observation the risk of entering a progressive course was 3.2 times higher in the non-pregnant state as compared with that after pregnancy (95% CIs 1.1-10.3)

There was a strong trend towards a higher risk of reaching DSS 6 in the state before pregnancy (p=0.07). [VERY LOW QUALITY]

8.3.5 Clinical evidence for the prognostic effects of smoking on MS progression

Two prospective studies were found (Healy 2009[95], Pittas 2009[181]) that fulfilled the inclusion criteria (Table 33).

Table 33: Summary of studies included in the review

Study	Population	Risk factor	Outcome	Quality

Healy 2009[95]	Clinically definite MS. 257 current smokers, 428 ex-smokers and 780 never smokers. 69.6% with RRMS.	Smoking status (current smoker versus ex-smoker versus never smoker).	Conversion from RRMS to SPMS. Progression of EDSS. MRI evidence of brain atrophy.	LOW No assessor blinding. Adjusted for some likely confounders, but not DMD treatment, which showed a trend for differences between groups at baseline. Low attrition rate. Follow up appropriate (mean 3.6 years).
Pittas 2009[181]	Definite MS by Poser criteria. 75% with RRMS. N=198.	Smoking status and daily smoking amount.	Change in MS type. Change in disability (change in MSSS or EDSS). Number of relapses.	MODERATE Assessor blinding. Adjusted for likely confounders. Low attrition rate. Follow up appropriate (3 years). Poor reporting of results (some reference categories unclear for analysis concerning number of relapses).

Narrative review for the prognostic effects of smoking

Effect of baseline smoking status on progression of RRMS to SPMS

Healy 2009[95] examined the effect of smoking on change in type of MS at a mean follow-up of 3.6 years. This analysis was on 891 patients (154 current smokers, 237 ex-smokers, 500 never-smokers). Conversion from RRMS to SPMS occurred faster in current smokers compared to never smokers [adjusted HR 2.5 (1.42-4.41)], but similar in ex-smokers and never smokers [adjusted HR: 1.05(0.59-1.84)]. Adjusting for baseline EDSS, similar results were obtained: Conversion from RRMS to SPMS occurred faster in current smokers compared to never smokers [adjusted HR 2.08 (1.15-3.77)], but similar in ex-smokers and never smokers [adjusted HR: 0.95(0.54-1.68)]. It is not reported if other potential confounders were included in this analysis. [LOW QUALITY]

Pittas 2009[181] also examined the effect of smoking on change in type of MS at a follow-up of 3 years. The OR for a change from RRMS to SPMS was 1.02 (1.00 to 1.05) per increment increase in pack years smoked by entry. Adjustments were only reportedly made for disease duration. [MODERATE QUALITY]

Effect of baseline smoking status on progression of EDSS and/or MSSS

Although EDSS and MSSS are ordinal variables, and thus inappropriate for regression analyses, they were analysed using a regression approach in the literature reviewed. Although this methodology lacks validity, the results have been presented.

Healy 2009[95] examined the effect of smoking on progression of EDSS at a mean follow-up of 3.6 years. At 2 years, the percentage of participants in whom EDSS worsened was 23.3% in smokers, 30.8% in ex-smokers and 26% in never smokers. After adjustment for baseline age, sex, disease duration and treatment, these group differences were not significant (p=0.57). At 5 years, similar results were reported but no data was given, although the p value was reported as 0.53. Results were very similar in analyses restricted to participants with RRMS at baseline, indicating confounding from type was unlikely. [LOW QUALITY]

Pittas 2009[181] also examined the effect of smoking on progression of EDSS/MSSS at a follow-up of 3 years. The amount smoked at study inception had no clear effects on MSSS, but moderate smoking (10-19 cigarettes a day) at inception led to a worsening of EDSS by almost 1 point. However, there was not a linear dose-response pattern, with heavy smoking not causing clear increases on EDSS. For the amount smoked after cohort entry, there were clearer patterns of greater smoking leading to greater increases in both MSSS and EDSS. Table 34 summarises these results. [MODERATE QUALITY]

Table 34: Effect of smoking status on progression of EDSS and MSSS

Smoking variable	Adjusted* average increase (worsening) in MSSS (relative to no smoking)	Adjusted** average increase (worsening) in EDSS (relative to no smoking)
Amount smoked daily at cohort entry (cig/day)		
0	REF	REF
1-4	-0.84(-2.13, 0.44)	-0.03(-0.67, 0.62)
5-9	-1.19(-3.11, 0.74)	**0.01(0.08, 1.75)***
10-19	-0.59(-2.45, 1.30)	**0.91(0.08, 1.75)**
20-39	-0.86(-2.68, 0.95)	0.22(-0.52, 0.96)
40 or more	-1.38(-3.41, 0.65)	0.08(-0.91, 1.09)
Amount smoked after cohort entry (cumulative pack years)		
0	REF	REF
0.1-1	**0.34(0.28, 0.66)**	**0.36(0.13, 0.60)**
1.1-2	0.41(-0.03, 0.85)	**0.55(0.22, 0.88)**
>2	**0.99(0.41, 1.58)**	**0.96(0.52, 1.40)**

***** adjusted for MSSS at entry, follow up time, gender, age at entry, immunomodulatory treatment, education level, month of review

** adjusted for EDSS at entry, follow up time, gender, age at entry, immunomodulatory treatment, education level, month of review, disease duration

***this is the reported result, although clearly incorrect (point estimate lies outside the boundaries of the 95% confidence intervals)

Effect of baseline smoking status on time to relapse

Pittas 2009[181] examined the effect of a variety of smoking variables on the hazard of relapses. Unfortunately, the reference category was not clear for all variables. There were no clear effects, but a reasonably convincing trend for the total pack years from MS onset to entry to increase the hazard of relapses by 1.70 per pack year (Table 35). [MODERATE QUALITY]

Table 35: Effect of smoking status on time to relapse

Smoking variable	HR (95% CIs)*	Reference category
Smoker ever	0.86(0.56, 1.32)	All others?
Total pack years smoked prior to MS	0.85 (0.63, 1.15)	Per pack year
Total pack years from MS onset to entry review	1.70 (0.80, 3.62)	Per pack year
Total pack years to entry review	0.94(0.76, 1.15)	Per pack year
Current smoking	0.98 (0.58, 1.64)	All others?
Cumulative pack years smoked after cohort entry	0.94 (0.69, 1.26)	Per pack year

8.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

8.5 Evidence statements

8.5.1 Clinical

Exercise

Moderate quality evidence from two prognostic studies comprising 903 participants showed that exercise levels at baseline did not lead to clinically important effects on later MS progression or quality of life.

Vaccinations

Very low quality evidence from three RCTs comprising 186 participants showed that there was no difference between the influenza vaccine and placebo in terms of exacerbations at 28 days, with very serious imprecision.

Very low quality evidence from three RCTs comprising 186 participants showed that the influenza vaccine was clinically harmful compared to placebo in terms of a greater rate of MS exacerbations at 6 months with serious imprecision.

Very low quality evidence from one RCT comprising 103 participants showed that there was no difference between the influenza vaccine and placebo in terms of MS "worsening", with very serious imprecision.

Stress

Low to very low quality evidence from eight prognostic studies comprising 399 participants showed that stress increased rates of relapse and hasten functional deterioration.

Pregnancy

Moderate to very low quality evidence from four prognostic studies comprising 125 participants showed that pregnancy was associated with some protection from relapses.

Smoking

Low to moderate quality evidence from two prognostic studies comprising 883 participants showed that smoking had a strong association with deterioration in function and conversion of relapsing remitting MS to progressive MS.

8.5.2 Economic

No relevant economic evaluations were identified.

8.6 Recommendations and link to evidence

Recommendations	<u>**Exercise**</u> **23. Encourage people with MS to exercise. Advise them that regular exercise may have beneficial effects on their MS and does not have any harmful effects on their MS.**
Relative values of different outcomes	Important outcomes were relapse rate, progression from RRMS to secondary progressive MS and deterioration in function. The outcome 'effects on deterioration in function' measures the degree of recovery from relapses, or the degree of progression in primary or secondary progressive disease.
Trade off between clinical benefits and harms	No harmful effects from exercise were observed.
Economic considerations	No relevant economic evaluations were identified. The cost of the time spent by healthcare professionals in providing advice to patients with MS on exercise is likely to be minimal. Although no clear benefits on MS disease course were found, no harmful effects from exercise were observed and general health benefits of exercise are anticipated, therefore the provision of advice on exercise is likely to be cost effective.
Quality of evidence	Evidence was of moderate quality, but only two eligible papers were found. The main limitation in both studies was poor reporting of results, making it difficult to judge the appropriateness of analysis. Furthermore, one study was limited by an overly short follow-up of 6 months.
Other considerations	The recommendations were informed by the review on modifiable risk factors

	and also by the reviews on rehabilitation for management of MS symptoms (see Chapter 10.4). The GDG reported that many people stop exercising when diagnosed with MS and develop secondary problems from lack of exercise. The GDG considered that this review indicated no evidence for detrimental effect of exercise on MS. The reviews on symptom management indicated benefit from different types of exercise. The GDG considered it important that patients with MS be encouraged to participate in exercise for the general health benefits associated with this and that they could be reassured that exercise would not result in deterioration in MS. The type and frequency of exercise should be appropriate to general health and abilities of the patient.

Recommendations	**Vaccinations** **24. Be aware that live vaccinations may be contraindicated in people with MS who are being treated with disease-modifying therapies.** **25. Discuss with the person with MS:** o the possible benefits of flu vaccination **and** o the possible risk of relapse after flu vaccination if they have relapsing–remitting MS. **26. Offer flu vaccinations to people with MS in accordance with national guidelines, which recommend an individualised approach according to the person's needs[r].**
Relative values of different outcomes	Important outcomes were relapse rate, progression from RRMS to secondary progressive MS and deterioration in function. Relapse rates are important as for relapsing remitting MS the onset of a relapse is the main cause of morbidity. Progression from relapsing remitting MS to secondary progressive MS also represents an escalation of morbidity The outcome of effects on deterioration in function measures the degree of recovery from relapses, or the degree of progression in primary or secondary progressive disease.
Trade off between clinical benefits and harms	The influenza vaccine had no clear harmful effects, although there was a trend for an increase in exacerbation risk at 3- 6 months. Given that the outcome represents harm, and thus there are greater risks associated with a false negative than false positive result, this trend could be considered as evidence that the influenza vaccination may cause relapse.
Economic considerations	No relevant economic evaluations were identified. The GDG considered that for people with MS and respiratory conditions, the benefit of influenza vaccination (preventing influenza) outweighs the possible harms (risk of relapse).
Quality of evidence	Three RCTs were found, which evaluated the influenza vaccine. The RCT evidence was graded as very low, due to no allocation concealment, no healthcare professional blinding and serious imprecision.

r Chronic neurological disease: conditions in which respiratory function may be compromised, due to neurological disease (e.g. polio syndrome sufferers). Clinicians should consider on an individual basis the clinical needs of patients including individuals with cerebral palsy, multiple sclerosis and related or other similar conditions; or hereditary and degenerative disease of the nervous system of muscles; or severe disability (Department of Health 2013).

Other considerations	No evidence was found for vaccines other than influenza vaccine. The GDG discussed whether the evidence review was of sufficient quality to make recommendations about the use of vaccination against influenza. They considered that the evidence suggested that there might be an increase in relapse rate and this may be important for people to be aware of. While seasonal influenza vaccination is recommended for people with chronic neurological disease and the guidance documentation suggests that this is assessed on an individual basis for people with conditions such as multiple sclerosis the GDG were aware that it is common practice to offer influenza vaccination to all people with MS. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207008/130613_Flu_Letter_v_29_Gateway_GW_signed.pdf The GDG considered that the possible increase in relapse should be discussed with people with MS. The use of disease modifying drugs (DMDs) is outside the scope of the guideline but the GDG thought it important to remind non-specialists that it might not be appropriate to give live vaccines to people on DMDs.

Recommendations	**No recommendation was made on stress**
Relative values of different outcomes	Important outcomes were relapse rate, progression from RRMS to secondary progressive MS and deterioration in function. Relapse rates are important as for relapsing remitting MS the onset of a relapse is the main cause of morbidity. Progression from relapsing remitting MS to secondary progressive MS also represents an escalation of morbidity The outcome of effects on deterioration in function measures the degree of recovery from relapses, or the degree of progression in primary or secondary progressive disease.
Trade off between clinical benefits and harms	Stress was associated with higher relapse rates and functional deterioration.
Economic considerations	No relevant economic evaluations were identified.
Quality of evidence	Despite all eight studies adopting the gold standard prospective cohort methodology, evidence was low or very low. The main reason was that few studies included an appropriate array of plausible potential confounders, and thus causality was not completely clear. However the studies by Brown performed a rigorous multivariable analysis, including most plausible confounding factors, and therefore the clear association they demonstrated between more frequent stressful events and relapse may indicate a causal effect. Analytical strategies were sophisticated, allowing for the repeated nature of both risk factors and outcomes. Stress was defined in different ways. Most studies used a self-report method of stress, which included any stressors deemed important to the participant. Other studies relied on the participant only being able to report stressors that were present on a list. Whilst the former method may be less objective, it reflects the highly subjective nature of stress.
Other considerations	The GDG considered that care was required in giving advice to people with MS about stress and did not want to make recommendations about a poorly defined concept such as stress. Some amount of stress is inevitable and some causes of stress are outside an individual's control. What is stressful to one person may not be stressful to another. The GDG considered that each individual could consider what was stressful for them and how they managed

	their stress but this was not different from advice one would give to people without MS.

Recommendations	**<u>Pregnancy</u>** **27. Explain to women of childbearing age with MS that:** o relapse rates may reduce during pregnancy and may increase 3-6 months after childbirth before returning to pre-pregnancy rates o pregnancy does not increase the risk of progression of disease. **28. If a person with MS is thinking about pregnancy, give them the opportunity to talk with a healthcare professional with knowledge of MS about:** o fertility o the risk of the child developing MS o use of vitamin D before conception and during pregnancy o medication use in pregnancy o pain relief during delivery (including epidurals) o care of the child o breastfeeding.
Relative values of different outcomes	Important outcomes were relapse rates, progression from RRMS to secondary progressive MS and deterioration in function. Relapse rates are important as for relapsing remitting MS the onset of a relapse is the main cause of morbidity. Progression from relapsing remitting MS to secondary progressive MSalso represents an escalation of morbidity The outcome of effects on deterioration in function was also important as it measures the degree of recovery from relapses, or the degree of progression in primary or secondary progressive disease.
Trade off between clinical benefits and harms	Pregnancy appears to provide some protection from relapses. No harms were identified.
Economic considerations	No relevant economic evaluations were identified. The time spent by healthcare professionals in discussing pregnancy with people with MS and their partners is not expected to have a considerable economic impact.
Quality of evidence	The four studies considered were prospective cohort studies, but graded as low to moderate. Although none adjusted for a range of plausible confounding effects in the analysis, three used a form of matching between pregnant and non-pregnant groups, thus allowing for some confounding effects.
Other considerations	The GDG used the evidence review and experience to develop these recommendations. People can develop MS when before they have children or when they consider their families are not complete. The GDG considered that people with MS (male and female) who are considering having children are likely to need information about a number of aspects of MS. The GDG used their experience to develop a list of the topics that are commonly of concern to people with MS in this position. These include any effect of MS on fertility and pregnancy and delivery itself. There is some concern but limited evidence about the potential effect of IVF treatment on relapse in relapsing-remitting MS.

	There is an association between MS and family history and the risk of a child developing MS are frequently asked questions. The GDG considered it important to include attention to future care of children to be included in a discussion if one of the parents is physically disabled and the disability is likely to be progressive. The list is not intended to issues that have to be discussed but to alert healthcare professionals to those issues that might need to be discussed with patients. Vitamin D supplements during pregnancy are subject to guidance by thedepartment of health but the associations between Vitamin D and MS mean that some healthcare professionals consider added Vitamin D appropriate before conception and during pregnancy.

Recommendations	**Smoking** **29. Advise people with MS not to smoke and explain that it may increase the progression of disability (See Smoking cessation services NICE public health guideline 10).**
Relative values of different outcomes	An important outcome was relapse rates, as for relapsing remitting MS the onset of a relapse is the main cause of morbidity. Another important outcome was progression from relapsing remitting MS to secondary progressive MS, as this may represent an escalation of morbidity The outcome of effects on deterioration in function was also important as it measures the degree of recovery from relapses, or the degree of progression in primary or secondary progressive disease.
Trade off between clinical benefits and harms	There were no clinical benefits from smoking. Smoking appeared to have a strong association with deterioration in function and conversion of relapsing remitting MS to progressive MS. This association may have been causal as analyses were adjusted for plausible confounders.
Economic considerations	No relevant economic evaluations were identified. The cost of the time spent by healthcare professionals in providing advice to patients with MS on not smoking is likely to be minimal and offset by the benefits of not smoking in those that adhere to the advice.
Quality of evidence	Only two eligible studies were found. Quality of outcomes ranged from low to moderate. In one study no assessor blinding was carried out, and DMD treatment, which differed between smoking and non-smoking groups, was not adjusted for.
Other considerations	

9 Pharmacological management of MS symptoms

9.1 Pharmacological management of spasticity

9.1.1 Introduction

Spasticity is the term generally used to cover symptoms of stiffness and muscle spasm. These are commonly experienced by people with MS and may have a significant effect on function and mobility. Individual patients may be aware of triggers for spasms. There are a number of treatments available for spasticity. This chapter examines evidence for pharmacological management of spasticity.

9.1.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of spasticity?

For full details see review protocol in Appendix C.

Table 36: PICO characteristics of review question

Population	• Adults with MS • Move from wholly MS population to mixed populations that include MS (other acquired neurological conditions in adults) if <1 RCT for any comparison.
Intervention/s	• Oral baclofen • Tizanidine (Zanaflex) • Gabapentin (Neurontin) • Pregabalin (Lyrica) • Benzodiazepines (diazepam, clonazepam) • Dantrolene sodium (Dantrium) • Sativex (nabiximol) • Botulinum toxin (Azzalure, Bocouture, Botox, Dysport, Vistabel, Xeomin) • Intrathecal baclofen • phenol
Comparison/s	• Best medical management • Placebo • Oral baclofen • Tizanidine (Zanaflex) • Gabapentin (Neurontin) • Pregabalin (Lyrica) • Benzodiazepines (Diazepam, clonazepam) • Dantrolene sodium (Dantrium) • Sativex • Botulinum toxin (Azzalure, Bocouture, Botox, Dysport, Vistabel, Xeomin) • Intrathecal baclofen • phenol
Outcomes	• Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale. • Patient-reported outcomes, for example symptoms of spasticity or pain.

	• Ashworth scale, or other objective spasticity scales
	• Spasms: Penn scale
	• Functional scales that quantify level of disability
	• Mobility – for example walking speed
	• Adverse effects of treatment
	  ○ Adverse events leading to withdrawal
	  ○ Drowsiness
	  ○ Weakness
	  ○ Nausea

Study design	• Systematic reviews, RCTs. Include cross-over studies. • If no RCTs in mixed population, then move to observational.

9.1.3 Clinical evidence

Thirty three studies were included in the review.[9,27,43,44,48,59,69,77,86,99,100,103,104,118,126,141-143,143,170,172,174,176,196,205,208,213,215,228,230,238,244,247,258] A Cochrane review[220] was also found, but because this looked at different comparisons to those chosen for our review protocol, contained non-published studies, and also only contained studies up to 2003, we decided to extract and analyse from the primary sources only. The study characteristics are summarised in Table 37**.**

Twelve different comparisons were covered in this review. Eleven concerned orally-administered drugs, and one concerned intrathecal baclofen. The studies were:

- Oral baclofen v placebo[27,176,208,213]
- Tizanidine v placebo[118,228,244]
- Tizanidine v oral baclofen[59,99,230,238]
- Diazepam v oral baclofen[69,205]
- Tizanidine v diazepam[196]
- Dantrolene v diazepam[215]
- Dantrolene v placebo[77,247]
- Gabapentin v placebo[48]
- Sativex v placebo[9,43,44,258]
- Sativex responders v placebo[170-172]
- Botulinum v placebo[86,103,104]
- Intrathecal baclofen v placebo[100,126,141-143,174]

As stated in the protocol, all comparisons were made on a population with Multiple sclerosis, with the exception of the intrathecal baclofen evidence. The population in this study were a mixed population of acquired adult neurological disease. The decision to include a mixed population was made by the Guideline Development Group on the grounds that 1) there were no studies in a pure MS population, 2) intrathecal baclofen was a potentially important intervention that should be assessed, and 3) there were no good physiological reasons why the alternative neurological diagnoses should unduly influence the effects of the drug on spasticity.

Study populations - sativex

Two studies examining sativex [170-172] reported on selected populations. Novotna 2011 initially carried out a single-group 4 week trial of sativex to identify responders (>20% decrease in spasticity NRS). These responders were then randomised into the sativex and placebo group for the further trial, the results of which are presented in this review. Another sativex study [170,171] consisted of patients that had been on long term sativex and had already shown a benefit, and thus randomisation was to continuation or withdrawal. Because these two studies involve a different population, they have been analysed separately to the other studies. All other studies examining use of sativex were of samples who had not previously been treated with Sativex.[9,43,44,258].

Study design and analysis

Evidence from all comparisons are summarised in the clinical GRADE evidence profiles below (Table 38 to Table 45). See also the study selection flow chart in Appendix D, forest plots in Appendix I, study evidence tables in Appendix G and exclusion list in Appendix J.

Some outcomes were not appropriate for meta-analysis as they consisted of ordinal rather than interval scales. Others were analysed with non-parametric methods. These have been reported in a separate narrative section in 0.

Summary of included studies

Table 37: Summary of studies included in the review

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	n	Analysis
Orsnes2000[176]	Oral baclofen v placebo	Median Ashworth 0.8 (range 0-2) Median EDSS 5	14	Cross-over
Brar1991[27]		Mild to moderate spasticity EDSS 5.5 or less	38	Cross-over
Sawa1979[213]		Ashworth 3 / 3	21	Cross-over
Sachais1997[208]		Duration of disease 11/ 11 years	166	Parallel
UKTTG1994[244]	Tizanidine v placebo	Moderate or severe spasticity: 61% / 53% Disease duration 12.7 / 13.1 years	187	Parallel
Smith1994[228]		% scoring 4 on Ashworth 22% / 23% Disease duration 10.8 / 11.2 years	256	Parallel
LaPierre1987[118]		At least "moderate" spasticity EDSS 5.07 / 5.07	66	Parallel
Hoogstraten1988[99]	Tizanidine v oral baclofen	EDSS 4-7	16	Cross-over
Eyssette1988[59]		Mean duration of MS 10.8 / 13.4 years Duration of signs 17.3 / 26.6 years	100	Parallel
Bass1988		Moderate or severe spasticity: 91% / 87%	66	Cross-over
Stien1987[238]		Moderate or severe spasticity: 78% / 90% Disease duration 14 / 13 years	40	Parallel
Smolenski1981[230]		Severe spasticity 36% / 60%	21	Parallel
Roussan1997[205]	Diazepam v baclofen	Duration of spasticity 10.8 years	6	Cross-over
From1975[69]		Duration of MS 17.5 years (range 3 – 40)	17	Parallel
Rinne1980[196]	Tizanidine v diazepam	Moderate or severe spasticity: 93% / 93% MS duration 7 / 12 years	30	Parallel
Schmidt1976[215]	Dantrolene v diazepam	Moderate or severe spasticity	46	Cross-over

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	n	Analysis
Gelenberg1973[77]	Dantrolene v placebo	Moderate to severe spasticity 70% able to ambulate but with difficulty	20	Cross-over
Tolosa1975[247]		No data reported	23	Parallel
Cutter2000[48]	Gabapentin v placebo	Clinical evidence of spasticity	22	Cross-over
Collin2007[43]	Sativex v placebo	NRS spasticity 5.49 / 5.39 Disease duration 13.6 / 12.2 years	189	parallel
Collin2010[44]		NRS spasticity 6.77 / 6.48 EDSS 6.0 Disease duration 14.4/16 years	337	parallel
Wade2004[258]		Ashworth 5 / 4.6	160	parallel
Aragona 2009[9]		Significant spasticity in at least 2 muscle groups EDSS 6.1 Disease duration 20.76 years	17	Cross-over
Novotna2011[172]	Sativex responders v placebo	NRS spasticity 6.8 / 7 EDSS 6.0 Disease duration 12.3/12.6 years	572	Parallel
Notcutt2012[170,171]		NRS spasticity 3.6 / 4.1 EDSS 6.75/6.92 Disease duration 12.1/12.6 years	36	parallel
Hyman2000[103,104]	Botulinum v placebo	Modified Ashworth 8.5 – 16 EDSS > 7 Duration of MS 16.6 – 22.9 years	74	Parallel
Gusev2008[86]		Duration of MS 12.9 / 13.9 years	106	Parallel
Middel 1997[143]	Intrathecal baclofen v placebo	59% with MS, 41% had spinal cord injury; no other details available	22	Parallel
Meythaler 2001[142]		All with CVA, and intractable spastic hypertonia	22	Parallel
Loubser 1991[126]		All with spinal cord injury, with intractable spasticity	9	Cross-over
Hugenholtz 1992[100]		2/6 MS; others SCI. All with intractable spasticity	6	Cross-over
Ordia 1996[174]		Not reported for the subset in the RCT, but probably MS or SCI. All with	9	Parallel

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	n	Analysis
		intractable spasticity		
Meythaler 1996[141]		Brain injury patients, with intractable spasticity	11	Cross-over

Table 38: Clinical evidence profile: baclofen versus placebo

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Baclofen	Placebo	Relative (95% CI)	Absolute(95% CI)		
Self-evaluation of gait improvement (higher better)												
Orsenes2000	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	5/13 (38.5%)	4/13 (30.8%)	RR 1.25 (0.43 to 3.63)	77 more per 1000 (from 175 fewer to 809 more)	VERY LOW	IMPORTANT
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Patients showing improvement in Ashworth scale (higher better)												
Brar1991	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	9/30 (30%)	6/30 (20%)	RR 1.5 (0.61 to 3.69)	100 more per 1000 (from 78 fewer to 538 more)	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Baclofen	Placebo	Relative (95% CI)	Absolute(95% CI)		
Detectable improvement in spasticity assessed by investigator												
Sawa 1979	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	13/18 (72.2%)	0/18 (0%)	Peto OR: 20.98 (5.49 to 80.21)	720 more per 1000 (from 510 more to 940 more)	MOD	CRITICAL
Physician assessment of clinical change in overall spastic state (higher better)												
Sachais 1997	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	3.02(1.03)[52]	2.37(1.03)[52]	-	MD: 0.65 more (from 0.25 more to 1.05 more)	VERY LOW	CRITICAL
Physician assessment of clinical change in daytime spasms (higher better)												
Sachais 1997	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	2.88(1.35)[43]	2.23(1.35)[44]	-	MD: 0.65 more (from 0.08 more to 1.22 more)	VERY LOW	IMPORTANT
Physician assessment of clinical change in night-time spasms (higher better)												
Sachais 1997	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	2.85(1.14)[40]	2.29(1.14)[45]	-	MD: 0.56 more (from 0.07 more to 1.05 more)	VERY LOW	IMPORTANT
Adverse events leading to treatment withdrawal												
Sawa1979	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	1/21 (4.8%)	0/18 (0%)	Peto OR 6.41 (0.13 to 326.59)	50 more per 1000 (from 80 less to 180 more)	VERY LOW	CRITICAL
Adverse events - somnolence												
Sachais1997	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	66/106 (62.3%)	29/102 (28.4%)	RR 2.15 (1.56 to	206 more per 1000 (from 100	LOW	IMPORTANT

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Baclofen	Placebo	Relative (95% CI)	Absolute(95% CI)		
Sawa1979								17.9%	2.98)	more to 354 more)		
Adverse events - weakness												
Sachais1997 Sawa1979	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	Serious[B]	none	20/106 (18.9%)	9/102 (8.8%) 5.6%	RR 2.07 (1.01 to 4.24)	60 more per 1000 (from 1 more to 181 more)	VERY LOW	IMPORTANT
Adverse events – nausea												
Sachais1997 Sawa1979	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	19/106 (17.9%)	5/102 (4.9%) 3.1%	RR 3.41 (1.38 to 8.44)	75 more per 1000 (from 12 more to 231 more)	LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 39: Clinical evidence profile: tizanidine versus placebo

Quality assessment	Mean (sd) [n] – if parallel group data	Effect	Quality	Importance

							OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)					
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Tizanidine**	**Placebo**	**Relative (95% CI)**	**Absolute(95% CI)**		
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Patient assessment of efficacy - good or very good												
UKTTG1994	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	25/89 (28.1%)	13/93 (14%)	RR 2.01 (1.1 to 3.68)	141 more per 1000 (from 14 more to 375 more)	VERY LOW	CRITICAL
Patient assessment of tolerability - good or very good												
UKTTG1994	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36/89 (40.4%)	79/93 (84.9%)	RR 0.48 (0.36 to 0.62)	442 fewer per 1000 (from 323 fewer to 544 fewer)	LOW	CRITICAL
Ashworth improved												
Smith1994 UKTTG1994	randomised trials	very serious[A]	Very serious[C]	no serious indirectness	serious[B]	none	131/205 (63.9%)	112/202 (55%)	Random RR 1.16 (0.8 to 1.69)	88 more per 1000 (from 110 fewer to 380 more)	VERY LOW	CRITICAL
Patients discontinuing because of adverse events												
UKTTG1994	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	12/94 (12.8%)	5/93 (5.4%)	RR 2.37 (0.87 to 6.47)	74 more per 1000 (from 7 fewer to 294 more)	VERY LOW	CRITICAL
Numbers with improved upper limb function (higher better)												

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Tizanidine	Placebo	Relative (95% CI)	Absolute(95% CI)		
UKTTG1994	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	5/87 (5.7%)	4/88 (4.5%)	RR 1.26 (0.35 to 4.55)	12 more per 1000 (from 30 fewer to 161 more)	VERY LOW	IMPORTANT

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*

[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

[C] *Outcomes were downgraded by one increment for serious inconsistency, as shown by the I squared value being between 50 and 74%. A double downgrade was applied for very serious inconsistency if I squared was >75%. A random effects model was used for any inconsistent outcomes. No subgrouping was applied, as all outcomes with inconsistency did not have >2 studies (and thus sub-grouping would always lead to one in each sub-group, which would inevitably reduce inconsistency to zero in each sub-group, thus making any sub-grouping non-informative).*

Table 40: Clinical evidence profile: tizanidine versus baclofen

Quality assessment	Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Tizanidine	Baclofen	Relative (95% CI)	Absolute(95% CI)		
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Spasticity worse or no better												
Hoogstraten1988	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	Ln[RR](SE): -0.223(0.387)		RR 0.80 (0.37 to 1.71)	Not available	VERY LOW	CRITICAL
Spasms worse or no better												
Hoogstraten1988	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	Ln[RR](SE): -0.693(0.527)		RR 0.50 (0.18 to 1.40)	Not available	VERY LOW	IMPORTANT
Mobility worse or no better												
Hoogstraten1988	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	Ln[RR](SE): -0.201(0.142)		RR 1.22 (0.93 to 1.61)	Not available	LOW	IMPORTANT
Overall evaluation of tolerability - patients stating treatment was poorly tolerated												
Eyssette1988	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	6/50 (12%)	4/50 (8%)	RR 1.5 (0.45 to 4.99)	40 more per 1000 (from 44 fewer to 319 more)	VERY LOW	CRITICAL
Discontinuation due to adverse events												
Bass1988 Eyssette1988 Stien1987	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	11/102 (10.8%)	16/100 (16%) 8%	RR 0.66 (0.33 to 1.35)	27 fewer per 1000 (from 54 fewer to 28 more)	VERY LOW	CRITICAL
Overall assessment of patient of the efficacy (moderate/poor or "ineffective at end of study")												
Bass1988 Smolenski1981 Stien1987 Eyssette 1988	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	72/133 (54.1%)	59/131 (45%) 45.4%	RR 1.21 (0.97 to 1.49)	95 more per 1000 (from 14 fewer to 222 more)	LOW	CRITICAL
Adverse events - somnolence												
Bass1988	randomised	serious[A]	no serious	no serious	serious[B]	none	28/57	13/54	RR 2.01	289 more per 1000	LOW	IMPORTANT

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Tizanidine	Baclofen	Relative (95% CI)	Absolute(95% CI)		
Hoogstraten1988 Smolenski1981	trials		inconsistency	indirectness			(49.1%)	(24.1%) 28.6%	(1.18 to 3.42)	(from 51 more to 692 more)		
Adverse events - nausea												
Hoogstraten1988 Smolenski1981	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	2/25 (8%)	4/24 (16.7%) 15.7%	RR 0.54 (0.13 to 2.26)	72 fewer per 1000 (from 137 fewer to 198 more)	VERY LOW	IMPORTANT
Adverse events - weakness												
Bass1988 Smolenski1981	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	13/43 (30.2%)	20/47 (42.6%) 37.2%	RR 0.66 (0.38 to 1.13)	126 fewer per 1000 (from 231 fewer to 48 more)	LOW	IMPORTANT

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*

[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 41: Clinical evidence profile: diazepam versus baclofen

Quality assessment	Mean (sd) [n] – if	Effect	Quality	Importance

							parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)					
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Diazepam	baclofen	Relative (95% CI)	Absolute(95% CI)		
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Spasticity outcomes												
No evidence available												
Better patient rated global response												
Roussan1997	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	3/6 (50%)	1/6 (16.7%)	RR 3 (0.42 to 21.3)	333 more per 1000 (from 97 fewer to 1000 more)	VERY LOW	CRITICAL
Adverse events - weakness												
From1975	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	2/16 (12.5%)	3/16 (18.8%)	RR 0.67 (0.13 to 3.47)	62 fewer per 1000 (from 163 fewer to 463 more)	VERY LOW	IMPORTANT
Adverse events- somnolence												
From1975 Roussan1997	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	No serious imprecision	none	RR: 4.45(1.45 to 13.65)		RR: 4.45(1.45 to 13.65)	Not available	LOW	IMPORTANT
Adverse events – nausea												
From1975	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	0/16 (0%)	2/16 (12.5%)	RR 0.2 (0.01 to 3.86)	100 fewer per 1000 (from 124 fewer to 357 more)	VERY LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 42: Clinical evidence profile: tinazidine versus diazepam

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Tinazidine	diazepam	Relative (95% CI)	Absolute		
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Patient reported outcomes outcomes												
No evidence available												
Numbers with improvement in spasticity (higher better)												
Rinne1980	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	9/15 (60%)	9/15 (60%)	RR 1 (0.56 to 1.79)	0 fewer per 1000 (from 264 fewer to 474 more)	VERY LOW	CRITICAL
AEs												
No evidence available												

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 43: Clinical evidence profile: dantrolene versus diazepam

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Dantrolene	diazepam	Relative (95% CI)	Absolute(95% CI)		
Quality of life												
No evidence available												
Functional outcomes												
No evidence available												
Spasticity outcomes												
No evidence available												
Improvement in cramps or spasms over treatment												
Schmidt1976	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[A]	none	RR: 1.19 (0.89 to 1.60)		RR: 1.19 (0.89 to 1.60)	-	MODERATE	IMPORTANT
Improvement in stiffness over treatment												
Schmidt1976	randomised	no	no serious	no serious	serious[A]	none	RR: 0.80 (0.52 to		RR: 0.80	-		IMPORTANT

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Dantrolene	diazepam	Relative (95% CI)	Absolute(95% CI)		
	trials	serious risk of bias	inconsistency	indirectness			1.24)		(0.52 to 1.24)		MODERATE	
Improvements in gait over treatment												
Schmidt1976	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	Very serious[A]	none	RR: 1.17 (0.47 to 2.89)		RR: 1.17 (0.47 to 2.89)	-	LOW	IMPORTANT
Drug preference (higher better)												
Schmidt1976	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[A]	none	22/42 (52.4%)	13/42 (31%)	RR 1.69 (0.99 to 2.89)	214 more per 1000 (from 3 fewer to 586 more)	MODERATE	CRITICAL
AEs												
No evidence available												

[A] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 44: Clinical evidence profile: dantrolene versus placebo

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Dantrolene	Placebo	Relative (95% CI)	Absolute(95% CI)		
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Patient preference												
Gelenberg1973	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[B]	none	7/20 (35%)	4/20 (20%)	RR 1.75 (0.61 to 5.05)	150 more per 1000 (from 78 fewer to 810 more)	LOW	CRITICAL
Reduction in spasticity												
Tolosa1975	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	5/12 (41.7%)	3/11 (27.3%)	RR 1.53 (0.47 to 4.94)	145 more per 1000 (from 145 fewer to 1000 more)	VERY LOW	CRITICAL
Adverse events leading to treatment discontinuation												
Tolosa1975	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	2/12 (16.7%)	0/11 (0%)	Peto OR 7.45 (0.44	170 more per 1000 (from 80 fewer to 410	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Dantrolene	Placebo	Relative (95% CI)	Absolute(95% CI)		
									to 127.44)	more)		
Adverse events - weakness												
Gelenberg1973 Tolosa1975	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	21/32 (65.6%)	1/31 (3.2%) 4.6%	RR 13.76 (2.84 to 66.56)	587 more per 1000 (from 85 more to 1000 more)	LOW	IMPORTANT
Adverse events - nausea												
Gelenberg1973	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	7/20 (35%)	0/20 (0%)	Peto OR 10.63 (2.12 to 53.21)	350 more per 1000 (from 130 more to 570 more)	HIGH	IMPORTANT
Adverse events - somnolence												
Gelenberg1973	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	Serious[B]	none	3/20 (15%)	0/20 (0%)	Peto OR 8.23 (0.81 to 84.07)	150 more per 1000 (from 20 less to 320 more)	MODERATE	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two

increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 45: Clinical evidence profile: Gabapentin versus placebo

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Gabapentin	Placebo	Relative (95% CI)	Absolute(95% CI)		
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Existence of moderate or severe spasms at follow up (lower better)												
Cutter2000	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	3/21 (14.3%)	14/21 (66.7%)	RR 0.21 (0.07 to 0.64)	527 fewer per 1000 (from 240 fewer to 620 fewer)	HIGH	CRITICAL
Spasm freq >1 time per hour at follow up (lower better)												
Cutter2000	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[A]	none	1/21 (4.8%)	7/21 (33.3%)	RR 0.14 (0.02 to 1.06)	287 fewer per 1000 (from 327 fewer to 20 more)	MODERATE	IMPORTANT
Spasticity worse or unchanged at follow up (lower better)												

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Gabapentin	Placebo	Relative (95% CI)	Absolute(95% CI)		
Cutter2000	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[A]	none	6/21 (28.6%)	16/21 (76.2%)	RR 0.38 (0.18 to 0.77)	472 fewer per 1000 (from 175 fewer to 625 fewer)	MODERATE	
Modified Ashworth score >4 at follow up (lower better)												
Cutter2000	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[A]	none	3/21 (14.3%)	10/21 (47.6%)	RR 0.3 (0.1 to 0.94)	333 fewer per 1000 (from 29 fewer to 429 fewer)	MODERATE	CRITICAL
Spasticity making function difficult or impossible at follow up (lower better)												
Cutter2000	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[A]	none	11/21 (52.4%)	17/21 (81%)	RR 0.65 (0.41 to 1.02)	283 fewer per 1000 (from 478 fewer to 16 more)	MODERATE	CRITICAL
AEs												
No evidence available												

[A] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 46: Clinical evidence profile: Sativex versus placebo

Quality assessment	Mean (sd) [n] – if parallel group	Pooled effect	Quality	Importance

							data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)					
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Sativex**	**Placebo**	**Relative (95% CI)**	**Absolute(95% CI)**		
Quality of life												
No evidence available												
Timed 10m walk if ambulatory (lower better)												
Wade2004	randomised trials	No serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision	none	-2.35(1.41)[140]		-	MD 2.35 lower (5.16 lower to 0.46 higher)	MODERATE	IMPORTANT
Responders (at least 30% improvement in NRS)												
Collin2007 Collin2010	randomised trials	very serious[A]	serious[B]	no serious indirectness	serious[C]	none	133/286 (46.5%)	85/233 (36.5%)	Random RR 1.4 (0.95 to 2.05)	146 more per 1000 (from 18 fewer to 383 more)	VERY LOW	CRITICAL
								31.9%		128 more per 1000 (from 16 fewer to 335 more)		
EQ-5D health state index (higher better) (Better indicated by lower values)												
Collin 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	None	none	0.03(0.135)[166]	0.01(0.135)[169]	-	MD 0.02 higher (0.01 lower to 0.05 higher)	LOW	CRITICAL
EQ-5D health status VAS (higher better) (Better indicated by lower values)												
Collin 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	None	none	4.29(21.08)[166]	2.87(21.08)[169]	-	MD 1.42 higher (3.09 lower to 5.93 higher)	LOW	CRITICAL
MSQoL54 phys health (higher better) (Better indicated by lower values)												

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Pooled effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sativex	Placebo	Relative (95% CI)	Absolute(95% CI)		
Collin2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	None	none	5.1(23.03)[166]	6.61(23.03)[169]	-	MD 1.51 lower (6.44 lower to 3.42 higher)	LOW	CRITICAL
MSQoL mental health (higher better) (Better indicated by lower values)												
Collin2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	None	none	-0.05(27.9)[166]	3.04(27.9)[169]	-	MD 3.09 lower (9.07 lower to 2.89 higher)	LOW	CRITICAL
Subject global impression of improvement												
Collin 2007	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[C]	none	Log OR 0.2625 (0.31)		OR: 1.30 (0.71 to 2.39)	-	VERY LOW	CRITICAL
Adverse events leading to withdrawal												
Collin2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[C]	none	9/167 (5.4%)	5/170 (2.9%)	RR 1.83 (0.63 to 5.35)	24 more per 1000 (from 11 fewer to 128 more)	VERY LOW	CRITICAL
Adverse event - nausea												
Wade2004 Aragona2009	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[C]	none	9/97 (9.3%)	6/97 (6.2%) 6.1%	RR 1.5 (0.56 to 4.05)	31 more per 1000 (from 27 fewer to 186 more)	LOW	IMPORTANT
Adverse event - somnolence												
Wade2004 Aragona2009	randomised trials	no serious	no serious inconsistency	no serious indirectness	serious[C]	none	18/97 (18.6%)	3/97 (3.1%)	RR 6 (1.92 to	325 more per 1000 (from 60	MODERATE	IMPORTANT

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Pooled effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sativex	Placebo	Relative (95% CI)	Absolute(95% CI)		
		risk of bias						6.5%	18.74)	more to 1000 more)		
Adverse event - weakness												
Aragona2009	randomised trials	Serious[A]	no serious inconsistency	no serious indirectness	serious[C]	none	3/17 (6%)	0/17 (0%)	Peto OR 8.41 (0.81 to 86.84)	180 more per 1000 (from 20 fewer to 370 more)	LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B]Outcomes were downgraded by one increment for serious inconsistency, as shown by the I squared value being between 50 and 74%. A double downgrade was applied for very serious inconsistency if I squared was >75%. A random effects model was used for any inconsistent outcomes. No subgrouping was applied, as all outcomes with inconsistency did not have >2 studies (and thus sub-grouping would always lead to one in each sub-group, which would inevitably reduce inconsistency to zero in each sub-group, thus making any sub-grouping non-informative).
[C] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 47: Clinical evidence profile: sativex responders versus placebo

Quality assessment	Mean (sd) [n] – if parallel group	Effect	Quality	Importance

							data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)					
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Sativex RESPONDERS**	**Placebo**	**Relative (95% CI)**	**Absolute(95 % CI)**		
Responders (at least 30% improvement in NRS)												
Novotna2011	randomised trials	No serious risk of bias	no serious inconsistency	no serious indirectness	serious[B]	none	92/124 (74.2%)	60/117 (51.3%)	RR 1.45 (1.18 to 1.78)	231 more per 1000 (from 92 more to 400 more)	MODERATE	CRITICAL
Timed 10m walk (lower better) (Better indicated by lower values)												
Novotna2011 Nottcutt2012	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0.13(14.23)[124] 3.46(19.5)[18]	3.22(14.23)[117] 5.24(19.5)[18]	-	MD 3.23 lower (6.69 lower to 0.23 higher)	MOD	IMPORTANT
Subject perception of global improvement (higher better)												
Novotna 2011 Nottcutt2012	randomised trials	No serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	LnOR(SE):1.515(0.62) LnOR(SE):0.5306(0.23)		OR: 1.92(1.25, 2.95)	-	HIGH	CRITICAL
Carer perception of global improvement in ease of transfer (higher better)												
Novotna 2011	randomised trials	No serious risk of bias	no serious inconsistency	no serious indirectness	serious[B]	none	LnOR(SE):0.58(0.31)		OR: 1.79(0.97, 3.30)	-	MODERATE	CRITICAL
Carer perception of global improvement in impression of function (higher better)												
Novotna 2011	randomised trials	No serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	LnOR(SE):0.876(0.0.467)		OR: 2.4(1.29, 4.44)	-	HIGH	CRITICAL
Physician perception of global improvement (higher better)												

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sativex RESPONDERS	Placebo	Relative (95% CI)	Absolute(95 % CI)		
Novotna 2011	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	LnOR(SE):0.67(0.24)		OR: 1.96(1.23 , 3.11)	-	LOW	CRITICAL
Improvement in Barthel Index												
Novotna 2011	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	LnOR(SE):0.71(0.26)		OR: 2.04(1.22 , 3.41)	-	LOW	CRITICAL
EQ-5D health state index (higher better) (Better indicated by lower values)												
Novotna2011	randomised trials	No serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	-0.03(0.145)[124]	-0.05(0.145)[117]	-	MD 0.02 higher (0.02 lower to 0.06 higher)	HIGH	CRITICAL
EQ-5D health status VAS (higher better) (Better indicated by lower values)												
Novotna2011	randomised trials	vNo serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	-1.99(16.79)[124]	-3.24(16.79)[117]	-	MD 1.25 higher (2.99 lower to 5.49 higher)	HIGH	CRITICAL
SF36 Phys Function (higher better) (Better indicated by lower values)												
Novotna2011	randomised trials	No serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	0.3(12.88)[124]	0.76(12.88)[117]	-	MD 0.46 lower (3.71 lower to 2.79 higher)	HIGH	CRITICAL
SF36 mental health (higher better) (Better indicated by lower values)												
Novotna2011	randomised trials	No serious	no serious inconsistency	no serious indirectness	no serious imprecision	none	-2.2(14.04)[124]	-2.94(14.0)[117]	-	MD 0.74 higher (2.81	HIGH	CRITICAL

Quality assessment							Mean (sd) [n] – if parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Sativex RESPONDERS	Placebo	Relative (95% CI)	Absolute(95 % CI)		
		risk of bias								lower to 4.29 higher)		
Adverse event - nausea												
Novotna2011	randomised trials	serious [A]	no serious inconsistency	no serious indirectness	very serious [B]	none	5/124 (4%)	2/117 (1.7%)	RR 2.36 (0.47 to 11.92)	23 more per 1000 (from 9 fewer to 187 more)	LOW	IMPORTANT
Adverse event – somnolence												
Novotna2011	randomised trials	serious [A]	no serious inconsistency	no serious indirectness	very serious [B]	none	4/124 (3.2%)	1/117 (0.9%)	RR 3.77 (0.43 to 33.28)	24 more per 1000 (from 5 fewer to 276 more)	MOD	IMPORTANT

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations in these two studies were a lack of any reporting of assessor blinding.*

[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 48: Clinical evidence profile: Botulinum versus placebo

Quality assessment	Mean (sd) [n] – if	Effect	Quality	Importance

							parallel group data OR Mean difference (SE) [n] – if one paired value OR Proportions with event (%)					
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Botulinum A	Placebo	Relative (95% CI)	Absolute(95% CI)		
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Patient positive response - low dose (500 units)												
Hyman2000	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	13/21 (61.9%)	7/16 (43.8%)	RR 1.41 (0.74 to 2.71)	180 more per 1000 (from 114 fewer to 749 more)	VERY LOW	CRITICAL
Patient positive response - medium dose (1000 units)												
Hyman2000	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	10/21 (47.6%)	7/16 (43.8%)	RR 1.09 (0.53 to 2.22)	39 more per 1000 (from 206 fewer to 534 more)	VERY LOW	CRITICAL
Patient positive response - high dose (1500 units)												
Hyman2000	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	8/17 (47.1%)	7/16 (43.8%)	RR 1.08 (0.51 to 2.28)	35 more per 1000 (from 214 fewer to 560 more)	VERY LOW	CRITICAL
Adverse events - weakness												
Gusev2008	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[B]	none	12/55 (21.8%)	3/51 (5.9%)	RR 3.71 (1.11 to 12.39)	160 more per 1000 (from 6 more to 672 more)	MODERATE	IMPORTANT

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation*

concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 49: Clinical evidence profile: Intrathecal baclofen versus placebo

Quality assessment							Proportions with event (%) Mantel Haenszel test for paired categories used		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intrathecal baclofen	Placebo	Relative (95% CI)	Absolute(95% CI)		
Quality of life												
No evidence available												
Functional/mobility outcomes												
No evidence available												
Numbers with improvement in Ashworth scale (lower limb)												
Loubser 1991 Hugenholtz 1992	randomised trials	very serious risk of bias[A]	no serious inconsistency	Serious indirectness[B]	serious imprecision[C]	none	3/9 with event ONLY in baclofen gp, 6/9 with event in both gps, and 0/9 with event ONLY in placebo gp. 2/6 with event ONLY in baclofen gp, 4/6 with event in both gps, and 0/6 with event ONLY in placebo gp.		RR: 1.50 (1.05 to 2.15)	–	VERY LOW	CRITICAL
Numbers with improvement in reflex score (lower limb)												
Loubser 1991 Hugenholtz 1992	randomised trials	very serious risk of bias	no serious inconsistency	Serious indirectness[B]	serious imprecision[C]	none	2/9 with event ONLY in baclofen gp, 7/9 with event in both groups, and 0/9 with event ONLY in placebo gp. 3/6 with event ONLY in baclofen gp, 1/6 with event in both groups, and 0/6 with event ONLY in placebo gp.		RR: 1.35 (0.96 to 1.89)	–	VERY LOW	CRITICAL

Quality assessment							Proportions with event (%) Mantel Haenszel test for paired categories used		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intrathecal baclofen	Placebo	Relative (95% CI)	Absolute(95% CI)		
Improvement in spasm score (lower limb)												
Hugenholtz 1992	randomised trials	serious risk of bias[A]	no serious inconsistency	Serious indirectness[B]	serious imprecision[C]	none	4/6 with event ONLY in baclofen gp, 2/6 with event in both groups, and 0/6 with event ONLY in placebo gp		RR: 3.0 (0.97 to 9.30)	_	VERY LOW	CRITICAL
Improvement in disability (questionnaire)												
Hugenholtz 1992	randomised trials	serious risk of bias[A]	no serious inconsistency	Serious indirectness[B]	serious imprecision[C]	none	3/6 with event ONLY in baclofen gp, 2/6 with event in both groups, and 0/6 with event ONLY in placebo gp		RR: 2.5 (0.85 to 7.32)	_	VERY LOW	CRITICAL

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*
[B] *Outcomes were downgraded for indirectness because the population was a mixed population, includng people who did not have MS.*
[C] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Narrative review for outcomes not appropriate for meta-analysis

Four comparisons had outcome data that were not appropriate for meta-analysis, and so these are described narratively as follows.

Tizanidine versus placebo

Upper extremity index score (lower better)

One study[118]assessed the effects of tizanidine and placebo on arm function, as measured by the upper extremity function score. It reported its results using parametric statistics, although this was inappropriate given the ordinal nature of this measure. Its data suggested no clear effect [Tizanidine 0.48 (0.74), placebo 0.52(0.77)] although the validity of this finding is suspect in view of the inappropriate analysis.

Botulinum versus placebo

Improvement in muscle tone

No data were presented, but it was stated that: "At week 8 the difference in the proportion of patients who had an improvement of ≥ 1 point on the MAS for leg adductor muscle tone approached significance (p=0.067)".

Sativex versus placebo (general population)

FSS (lower better)

Aragona 2009[9] compared the effects of sativex and placebo on fatigue, as measured by the Fatigue Severity Scale (FSS). The data were skewed and so non parametric analysis was used, although means and standard deviations were presented. The sativex group had a mean (sd) of 5.58 (1.5) on the FSS compared to 5.89(0.93) for the placebo group (p=0.88)

VAS QoL (higher better)

Aragona 2009[9] compared the effects of sativex and placebo on the visual analogue scale on health-related quality of life (VASQoL). The data were skewed and so non parametric analysis was used, although means and standard deviations were presented. The sativex group had a mean (sd) of 3.65 (2.29) on the VASQoL compared to 4.00(2.00) for the placebo group (p=0.31)

MSIS physical and psychological (lower better)

Aragona 2009[9] compared the effects of sativex and placebo on the Multiple Sclerosis Impact Scale (MSIS) physical and psychological scales. The data were skewed and so non parametric analysis was used, although means and standard deviations were presented. The sativex group had a mean (sd) of 63 (16.8) on the MSIS physical scale compared to 62.3(13.1) for the placebo group (p=0.57). The sativex group had a mean (sd) of 47.8 (17.8) on the MSIS physical scale compared to 46.3(15.9) for the placebo group (p=0.64).

Spasticity – Ashworth scale (lower better)

Three studies [43,44,258] assessed the effects of sativex and placebo on spasticity, as measured by the Ashworth scale. All reported their results using parametric statistics, although this was inappropriate given the ordinal nature of this measure. Taken together, their data suggested no clear effect [Collin 2007: mean difference -0.11(-0.29, 0.07); Collin 2010: mean difference -0.16(-1.9, 1.58); Wade 2004:

mean difference 0.22(-0.5, 0.94)], although the validity of this finding is suspect in view of the inappropriate analysis.

Spasticity NRS score (lower better)

Two studies [43,44] assessed the effects of sativex and placebo on spasticity, as measured by the NRS score. All reported their results using parametric statistics, although this was inappropriate given the ordinal nature of this measure. Taken together, their parametric data analyses suggested a possible benefit for sativex [Collin 2007: mean difference -0.52(-1.0, -0.04); Collin 2010: mean difference -0.23(-3.21, 2.75)], although the validity of this finding is suspect in view of the inappropriate analysis.

Motricity index –arm and leg (higher better)

One study [43] assessed the effects of sativex and placebo on motor function spasticity, as measured by the Motricity index. They reported their results using parametric statistics, although this was inappropriate given the ordinal nature of these measures. Their data suggested no clear effect for arm function [mean difference 1.30 (-7.47, 10.07)] and a trend for an effect favouring sativex for leg function [mean difference 3.86(-0.06, 7.78)], although the validity of these findings is suspect in view of the inappropriate analyses.

Barthel index (higher better)

Two studies [43,44,258] assessed the effects of sativex and placebo on function in activities of daily living, as measured by the Barthel Index (BI). Both reported their results using parametric statistics, although this was inappropriate given the ordinal nature of this measure. Taken together, their parametric data analyses suggested possible harm from sativex [Wade 2004: mean difference -0.47(-1.0 to 0.06); Collin 2010: mean difference -0.15 (-2.01, 1.71)], although the validity of this finding is suspect in view of the inappropriate analysis.

Spasm severity NRS (lower better)

Two studies [43,44] assessed the effects of sativex and placebo on function spasm severity, as measured by the spasm severity NRS. Both reported their results using parametric statistics, although this was inappropriate given the ordinal nature of this measure. Taken together, their parametric data analyses suggested no effect [Collin 2007: mean difference -0.17(-0.39 to 0.05); Collin 2010: mean difference -0.01 (-32.5, 32.48)], although the validity of this finding is suspect in view of the inappropriate analysis.

Guys neurological disability scale (lower better)

One study [258]assessed the effects of sativex and placebo on disability, as measured by the Guys neurological disability scale. They reported their results using parametric statistics, although this was inappropriate given the ordinal nature of this measure. Their data suggested significant harm from sativex [mean difference 1.81 (0.03, 3.59)], although the validity of these findings is suspect in view of the inappropriate analyses.

Sativex responders versus placebo

For the following outcomes the sponsors informed us that the distribution of data were skewed, but that subsequent non-parametric analyses showed a similar or more pronounced effect.

Ashworth scale (lower better)

Two studies[170-172] assessed the effects of sativex and placebo on spasticity, as measured by the Ashworth scale. Taken together, their data suggested little effect for Sativex [Novotna 2011:mean difference -1.75(-3.79 to +0.29); Nottcutt 2012: mean difference -0.53(-6.51, 5.45)].

Spasticity NRS score (lower better)

Two studies[172 170,171] assessed the effects of sativex and placebo on spasticity, as measured by the spasticity NRS. Taken together, their data suggested a weak effect benefitting Sativex [Novotna 2011: mean difference -0.83(-1.26, -0.40); Nottcutt 2012: mean difference -0.21(-1.35, 0.93)].

Motricity index (higher better)

One study [172] assessed the effects of sativex and placebo on motor function spasticity change, as measured by the Motricity index. Their data suggested no clear effect for arm [sativex -10.5 (30.9), placebo -8.58 (30.9)] or leg [sativex -3.24(9.7), placebo -4.21 (9.71)] function.

Barthel index (higher better)

One study[172] assessed the effects of sativex and placebo on function in everyday activities, as measured by the Barthel Index (BI). Their data suggested a clear and strong effect benefitting Sativex, with a mean difference of +2.04 (SE 0.75)].

Spasm frequency NRS (lower better)

One study[172] assessed the effects of sativex and placebo on change in spasm frequency, as measured by the spasm frequency NRS. Their data suggested a clear harm for Sativex [sativex -0.03 (6.85), placebo -2.53(6.85)].

Intrathecal baclofen versus placebo

One study[143] evaluated the effects of intrathecal baclofen and intrathecal saline placebo on spasm, spasticity, pain and two measures of quality of life: sickness impact profile (SIP) and Hopkins Symptom Check List (HSCL). As the groups differed at baseline for spasm, spasticity and pain, a non-parametric Cohen estimate of between-group effect sizes was carried out (Table 50).

Table 50: Clinical evidence profile: intrathecal baclofen versus placebo

	Baclofen (n=10) mean(sd)	Placebo (n=12) mean(sd)	Cohen effect sizes, estimating the group difference in the magnitude of the change between baseline and 3 months	U Wilcoxon p value
spasm at 3 months (lower better)	1.65(1.1)	1.81(0.76)	0.2 (weakly favours baclofen)	<0.05
Ashworth scale at 3 months (lower better)	1.51(1.2)	2.87(0.57)	1.40 (strongly favours baclofen)	<0.01
Self-reported pain score at 3 months (lower better)	2.75(3.22)	5.94(3.57)	0.94 (strongly favours baclofen)	<0.05
Overall SIP at 3 months (lower better)	27.79(5.32)	28.98(8.83)	No effect size given	NS
Overall HSCL at 3 months (lower better)	20.67(11.78)	28.22(18.43)	No effect size given	NS

One study[141,142] demonstrated that intrathecal baclofen led to significantly ($p<0.01$ for all) greater improvements than placebo in both upper and lower limb Ashworth scale, spasm scale and reflex scale 6 hours after a bolus injection. No data were provided for the placebo group, so only the direction of effect is possible to report.

In a similar study on a different neurological disease population [141] intrathecal baclofen led to significantly ($p<0.01$ for all) greater improvements than placebo in both upper and lower limb Ashworth scale, spasm scale and reflex scale 6 hours after a bolus injection. No data were provided for the placebo group, so only the direction of effect is possible to report.

One study[174] showed that a group of spinal cord injured patients all improved with a bolus injection of intrathecal baclofen but that no improvements were seen in the placebo group. Improvement was denoted by a reduction in the mean Ashworth score or the mean spasm score of 2 or more points for at least 4 hours.

One cross-over study[100]assessed the effects of intrathecal baclofen and placebo on the proportion of people with improvements upper limb Ashworth scale, spasm and reflexes. It was not possible to calculate Mantel-Haenszel risk ratios for paired categorical outcomes as there were insufficient people with the event.

For the Ashworth scale, one patient showed an improvement in both treatments, but no patients showed an improvement in just one of the treatments. This indicates no difference in effect, though the uncertainty of this effect is unknown. For spasm score, no patients showed an improvement in both or just one of the treatments. This also indicates no difference in effect, though the uncertainty of this effect is unknown. For reflex score, no patients showed an improvement in both treatments, but one patient showed an improvement in just the baclofen treatment. This indicates a slight effect in favour of intrathecal baclofen, though the uncertainty of this effect is unknown.

9.1.4 Economic evidence

Published literature

Three studies were included which assessed the cost effectiveness of pharmacological treatments for spasticity. The first compared oral tizanidine with oral baclofen,[207] the second compared Sativex plus oral antispasticity agents with oral antispasticity agents alone[127] and the third compared Sativex plus standard of care treatment versus standard of care treatment alone.[227] These are summarised in the economic evidence profiles below (Table 51 and Table 52). See also the economic article selection flow chart in Appendix E and study evidence tables in Appendix H.

Three studies were excluded due to absence of an explicit comparator.[90,211,212] One study was selectively excluded as it was a duplicate of the study by Slof 2012.[64] This is summarised in Appendix K, with reasons for exclusion given.

Table 51: Economic evidence profile: tizanidine versus baclofen

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Rushton 2002[207] (UK NHS)	Partially applicable (a)	Potentially serious limitations (b)	Deterministic decision analytic model of patients with spasticity caused by MS or spinal injury whose symptoms have not been resolved through physiotherapy alone. Time horizon = 1 year. Two comparisons: Comparison 1 = tizanidine (16mg) vs baclofen (40mg) as first line therapy. Comparison 2 = tizanidine (16mg) vs lower dose baclofen (20mg) as second line therapy due to adverse event of muscle weakness.	1) £30,385 2) £34,973 (c)	1) 2,903 STDs 2) 4,132 STDs (d)	1) £10.47 per STD 2) £8.46 per STD	One way sensitivity analysis. The key determinants of the results at first line (comparison 1) were: the effectiveness of tizanidine and baclofen, the rate of muscle weakness with tizanidine and baclofen, and the non-drug cost of managing spasticity. The key determinants of the results at second line (comparison 2) were the same as those for first line and in addition, the time horizon of the study and the definition of treatment success used.

(a) Health effects not expressed as QALYs.

(b) Does not include all relevant health outcomes. Specifically the study focuses on adverse events related to the comparator drug baclofen. There are serious concerns about how the measure of effectiveness for the model is calculated and the impact this might have on introducing bias into the results.

(c) 2000 UK pounds. Costs incorporated are daily costs: gen. mgt. of spasticity, baclofen, baclofen low dose, tizanidine, tizanidine low dose, third line therapies. Consultations: liver function test, GP visit, neurologist, physio and specialist nurse.

(d) Effectiveness measure STDs (successfully treated days). STD was defined as a day when the patient experienced improvement and when patient reported no adverse event or muscle weakness. STDs were calculated from pooled trial data as follows: percentage of patients reporting improvement multiplied by percentage of patients reporting no muscle weakness. The results were referred to as number of patients experiencing adequate relief. The study notes 'The figures for adequate relief were lower than success rates reported in clinical trials as we have assumed that some patients who experience 'improvement' may also experience muscle weakness.'

Table 52: Economic evidence profile: Sativex versus placebo

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Lu 2012[127] (UK NHS)	Directly applicable	Minor limitations (a)	Markov model included adults with moderate to severe spasticity due to MS who did not respond adequately to other oral anti-spasticity medication. Time horizon = 5 years with 4-week cycles. Concomitant oral anti-spasticity agents.	£7,627 (b)	0.1549 QALYs (c)	£49,238 per QALY (d)	Probability cost-effective (£30,000 per QALY gained threshold): 10.2%. Threshold analyses were conducted to identify what changes in costs and effects would be required for the ICER to be below £20,000 per QALY. The analyses found that the cost of Sativex would need to reduce by 61% or the difference in utilities would need to be above 0.23 (difference in the base case was 0.09). A scenario analysis was also conducted. When it was assumed that patients were to gain benefits with 4 sprays per day that are similar to those gained with 8 sprays per day the ICER would be £25,324 per QALY.
Slof 2012[227] (Spain and Germany healthcare system)	Partially applicable (e)	Potential serious limitations (f)	Markov model included adults with moderate to severe spasticity due to MS who did not respond adequately to other oral anti-spasticity medication. Time horizon = 5 years with 28 day cycles. Two perspectives = Spanish and German healthcare system. Concomitant standard of care treatment (SoC). (g)	**Spain** Saves £3,236 **Germany** £2,960 (h)	**Spain** 0.3252 QALY **Germany** 0.3207 QALY (i)	**Spain** Sativex plus SoC dominant (£ per QALY) **Germany** £9,230 per QALY	One-way sensitivity analysis. For both the Spanish and German analyses, ICER was most sensitive to +/-20% change in cost of Sativex. Deviation in ICER remained below £20,000/30,000 per QALY gained threshold. Dosing assumptions challenged by assuming a dose 8.3 sprays per day for the duration of the modelled period. This analysis produced an ICER of £2,185 per QALY gained in Spain and £24,082 per QALY gained in Germany. SoC resource utilisation rates from a German retrospective study were used instead of the Delphi study panel generated rates. This analysis resulted in a small deviation of the ICER, with the

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
							ICER increasing to £11,886 per QALY gained in Germany.

(a) *The study extrapolated effectiveness data from a 16 week trial over 5 years. The model assumed that there were no withdrawals from Sativex after 16 weeks. EQ-5D tariff not stated. Costs and disutilities associated with side-effects of drugs not included in analysis.*

(b) *2009 UK pounds. Costs incorporated are: cost of Sativex and clinic visits.*

(c) *Sativex efficacy data (cycle 1-4) was based on results of the European RCT which is included in the clinical review (Novotna, 2011). Average QALYs gained for the cohort treated with Sativex plus oral medicines and the cohort receiving oral medicines alone were estimated from the EQ-5D data collected from a clinical trial of Sativex (Montalban, 2009).Study reports incremental QALY of 0.1548 in study table.*

(d) *Study reports ICER as £49,257 in study table.*

(e) *Non-UK study*

(f) *The study extrapolated effectiveness data from a 16 week trial over 5 years. No description of population baseline characteristics used in model provided such as proportion with moderate-to-severe spasticity, age or gender. Baseline health outcomes based on observational data. EQ-5D tariff not stated. Mean QALY per patient for Germany not reported. Resource use for SoC based on Delphi survey. Costs of side-effects of drugs not included in analysis. No probabilistic sensitivity analysis undertaken.*

(g) *SoC treatment, described by the Delphi study panel, included specialist and physiotherapy visits and the following drug therapy: oral baclofen, intrathecal baclofen, tizanidine, diazepam, gabapentin, dantrolene sodium, botulinum toxin.*

(h) *2010 Euros, presented here as 2010 UK pounds. Euros converted using 2010 purchasing power parities [175]. Costs incorporated are: drugs, drug administration surgery, healthcare visits, (homecare worker, GP, nurse, physical therapist, occupational therapist, social worker, hospital emergency and routine) and tests (MRI and lab tests).*

(i) *SoC risks (baseline health outcomes) were taken from a Spanish retrospective observational study. Sativex efficacy data (cycle 1-4) was based on results of the European RCT by Novotna 2011 which is included in the clinical review. Sativex discontinuation rates from cycle 5-56 were from a long-term open label UK study (Wade, 2006). EQ-5D (from patients in Novotna 2011 study).*

The decision analytic study by Rushton 2002 was funded by Elan Pharmaceuticals Ltd, the manufacturer of tizanidine. The chosen measure of effectiveness, STD (successfully treated day) appears to introduce bias into the cost effectiveness results to the detriment of the comparator drug baclofen. Results from the eight trials included in the model showed that there was no statistically significant difference in effectiveness between baclofen and tizanidine. In the model, STDs (or days on which a patient experienced adequate relief) were calculated by multiplying percentage of patients reporting improvement in the clinical trial by percentage of patients *not* reporting the adverse event muscle weakness. However, patients receiving baclofen were more likely to report muscle weakness, while the most commonly reported adverse event with tizanidine was drowsiness/somnolence. The model justified its definition of STD by making the assumption that the likelihood of muscle weakness was independent of whether or not a patient's spasticity was perceived to have improved. The effect of this assumption, along with the inclusion of only one type of adverse event (muscle weakness) in the calculation of STDs was to systematically underestimate the effectiveness of baclofen compared to tizanidine. The study noted this by stating 'the figures for adequate relief were lower than success rates reported in clinical trials as we have assumed that some patients who experience 'improvement' may also experience muscle weakness.' However, what they have not stated is that this impacts more on success rates for baclofen than on tizanidine.

Lu 2012 and Slof 2012 both assessed Sativex using Markov models. A number of key differences exist between the two models which may account for the conflicting results. Differences included the model structure, perspective, comparators, costs and utilities. The model structure in Slof 2012 incorporated health states for different severities of spasticity, whereas the Lu 2012 model only separated the health states as responders and non-responders. As a result the utility levels varied between the two studies. Slof 2012 used EQ-5D scores of 0.6112, 0.5589 and 0.4321 for mild, moderate and severe spasticity respectively. Lu 2012 used EQ-5D scores of 0.48 for non-responders and 0.57 for responders to Sativex. The perspective in Slof 2012 was German and Spanish healthcare systems, not UK NHS. Lu 2012 limited the comparison to oral antispasticity drugs whereas Slof 2012 included other standard of care treatment for spasticity (pharmacological and non-pharmacological treatment). Finally, Lu 2012 did not include the cost of the oral antispasticity drugs in the analysis. Of note, the study by Slof 2012 was funded by SA Almirall, the manufacturer of Sativex.

Unit costs

Relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

9.1.5 Evidence statements

9.1.5.1 Clinical

Baclofen versus placebo

Very low quality evidence from one RCT comprising 13 participants showed that there was no difference in clinical effectiveness between baclofen and placebo in terms of self-evaluated gait improvement, with very serious imprecision.

Very low quality evidence from one RCT comprising 30 participants showed that baclofen was clinically effective compared to placebo in terms of numbers with improvement in the Ashworth scale, with very serious imprecision.

Moderate quality evidence from one RCT comprising 18 participants showed that baclofen was clinically effective compared to placebo in terms of numbers with detectable improvements in investigator assessed spasticity, with no imprecision.

Very low quality evidence from one RCT comprising 104 participants showed that baclofen was clinically effective compared to placebo in terms of improvements in physician assessed spasticity, with serious imprecision.

Very low quality evidence from one RCT comprising 77 participants showed that baclofen was clinically effective compared to placebo in terms of improvements in physician assessed daytime spasms, with serious imprecision.

Very low quality evidence from one RCT comprising 85 participants showed that baclofen was clinically effective compared to placebo in terms of improvements in physician assessed daytime spasms, with serious imprecision.

Very low quality evidence from one RCT comprising 39 participants showed that baclofen was clinically harmful compared to placebo in terms of numbers with adverse events leading to withdrawal, with very serious imprecision.

Low quality evidence from two RCTs comprising 208 participants showed that baclofen was clinically harmful compared to placebo in terms of numbers with somnolence, with no imprecision.

Very low quality evidence from two RCTs comprising 208 participants showed that baclofen was clinically harmful compared to placebo in terms of numbers with weakness, with serious imprecision.

Low quality evidence from two RCTs comprising 208 participants showed that baclofen was clinically harmful compared to placebo in terms of numbers with nausea, with no imprecision.

Tizanidine versus placebo

Very low quality evidence from one RCT comprising 182 participants showed that tizanidine was clinically effective compared to placebo in terms of numbers with good or very good self-assessment of efficacy, with serious imprecision.

Low quality evidence from one RCT comprising 182 participants showed that tizanidine was clinically harmful compared to placebo in terms of numbers with good or very good self-assessment of tolerability, with no imprecision.

Very low quality evidence from two RCTs comprising 407 participants showed that there was no difference in clinical effectiveness between tizanidine and placebo in terms of numbers with improvement in Ashworth score, with serious imprecision.

Very low quality evidence from one RCT comprising 187 participants showed that tizanidine was clinically harmful compared to placebo in terms of numbers with adverse events leading to withdrawal, with serious imprecision.

Very low quality evidence from one RCT comprising 175 participants showed that there was no difference in clinical effectiveness between tizanidine and placebo in terms of numbers with improved upper limb function, with very serious imprecision.

Tizanidine versus baclofen

Very low quality evidence from one RCT comprising 14 participants showed that there was no difference in clinical effectiveness between tizanidine and baclofen in terms of numbers with worse or no better spasticity, with very serious imprecision.

Very low quality evidence from one RCT comprising 14 participants showed that tizanidine was clinically effective compared to baclofen in terms of numbers with worse or no better spasms, with very serious imprecision.

Low quality evidence from one RCT comprising 14 participants showed that there was no difference in clinical effectiveness between tizanidine and baclofen in terms of numbers with worse or no better spasticity, with serious imprecision.

Very low quality evidence from one RCT comprising 100 participants showed that there was no difference in clinical effectiveness between tizanidine and baclofen in terms of numbers stating treatment was poorly tolerated, with very serious imprecision.

Very low quality evidence from three RCTs comprising 202 participants showed that there was no difference in clinical effectiveness between tizanidine and baclofen in terms of numbers withdrawing due to adverse events, with very serious imprecision.

Low quality evidence from four RCTs comprising 264 participants showed that there was no difference in clinical effectiveness between tizanidine and baclofen in terms of numbers perceiving efficacy as moderate/poor or ineffective, with very serious imprecision.

Low quality evidence from three RCTs comprising 111 participants showed a clinical harm for tizanidine compared to baclofen in terms of numbers with somnolence, with serious imprecision.

Very low quality evidence from two RCTs comprising 49 participants showed that tizanidine was clinically effective compared to baclofen in terms of numbers with nausea, with very serious imprecision.

Low quality evidence from two RCTs comprising 90 participants showed that tizanidine was clinically effective compared to baclofen in terms of numbers with somnolence, with serious imprecision.

Diazepam versus baclofen

Very low quality evidence from one RCT comprising 6 participants showed that diazepam was clinically effective compared to baclofen in terms of the numbers with improvements in patient related global response, with very serious imprecision.

Very low quality evidence from one RCT comprising 16 participants showed that diazepam was clinically effective compared to baclofen in terms of the numbers with weakness, with very serious imprecision.

Low quality evidence from two RCTs comprising 22 participants showed that diazepam was clinically harmful compared to baclofen in terms of numbers with somnolence, with no imprecision.

Very low quality evidence from one RCT comprising 16 participants showed that diazepam was clinically effective compared to baclofen in terms of the numbers with nausea, with very serious imprecision.

Tizanidine versus diazepam

Very low quality evidence from one RCT comprising 30 participants showed that there was no difference in clinical effectiveness between tizanidine and diazepam in terms of the numbers with improvements in spasticity, with very serious imprecision.

Dantrolene versus diazepam

Moderate quality evidence from one RCT comprising 42 participants showed that there was no difference in clinical effectiveness between dantrolene and diazepam in terms of the numbers with improvements in spasticity, with serious imprecision.

Moderate quality evidence from one RCT comprising 42 participants showed that there was no difference in clinical effectiveness between dantrolene and diazepam in terms of the numbers with improvements in stiffness, with serious imprecision.

Low quality evidence from one RCT comprising 42 participants showed that there was no difference in clinical effectiveness between dantrolene and diazepam in terms of the numbers with improvements in gait, with very serious imprecision.

Moderate quality evidence from one RCT comprising 42 participants showed that dantrolene was clinically effective compared to diazepam in terms of the proportion of patients preferring it, with serious imprecision.

Dantrolene versus placebo

Low quality evidence from one RCT comprising 20 participants showed that dantrolene was clinically effective compared to placebo in terms of the proportion of patients preferring it, with very serious imprecision.

Very low quality evidence from one RCT comprising 23 participants showed that dantrolene was clinically effective compared to placebo in terms of the proportion with reduction in spasticity, with very serious imprecision.

Very low quality evidence from one RCT comprising 23 participants showed that dantrolene was clinically harmful compared to placebo in terms of numbers with adverse events leading to withdrawal, with very serious imprecision.

Low quality evidence from two RCTs comprising 63 participants showed that dantrolene was clinically harmful compared to placebo in terms of numbers with weakness, with no imprecision.

High quality evidence from one RCT comprising 20 participants showed that dantrolene was clinically harmful compared to placebo in terms of numbers with nausea, with no imprecision.

Moderate quality evidence from one RCT comprising 20 participants showed that dantrolene was clinically harmful compared to placebo in terms of numbers with somnolence, with serious imprecision.

Gabapentin versus placebo

High quality evidence from one RCT comprising 21 participants showed that gabapentin was clinically effective compared to placebo in terms of the proportion with moderate or severe spasms at follow up, with no imprecision.

Moderate quality evidence from one RCT comprising 21 participants showed that gabapentin was clinically effective compared to placebo in terms of numbers with spasm frequency >once per hour, with serious imprecision.

Moderate quality evidence from one RCT comprising 21 participants showed that gabapentin was clinically effective compared to placebo in terms of numbers with worse or unchanged spasticity, with serious imprecision.

Moderate quality evidence from one RCT comprising 21 participants showed that gabapentin was clinically effective compared to placebo in terms of numbers with modified Ashworth score >4, with serious imprecision.

Moderate quality evidence from one RCT comprising 21 participants showed that gabapentin was clinically effective compared to placebo in terms of numbers with spasticity making function difficult or impossible, with serious imprecision.

Sativex versus placebo

Moderate quality evidence from one RCT comprising 140 participants showed that sativex was clinically effective compared to placebo in terms of timed 10m walk (if ambulatory) , with serious imprecision.

Very low quality evidence from two RCTs comprising 519 participants showed that sativex was clinically effective compared to placebo in terms of numbers with at least 30% improvement in NRS, with serious imprecision.

Low quality evidence from 1 RCT comprising 335 participants showed that there was no difference in clinical effectiveness between sativex and placebo in terms of EQ-5D health status index, with no imprecision.

Low quality evidence from 1 RCT comprising 335 participants showed that there was no difference in clinical effectiveness between sativex and placebo in terms of EQ-5D health status VAS, with no imprecision.

Low quality evidence from 1 RCT comprising 335 participants showed that there was no difference in clinical effectiveness between sativex and placebo in terms of MSQoL54 physical health, with no imprecision.

Low quality evidence from 1 RCT comprising 335 participants showed that there was no difference in clinical effectiveness between sativex and placebo in terms of MSQoL54 mental health, with no imprecision.

Very low quality evidence from 1 RCT comprising 184 participants showed that sativex was clinically effective compared to placebo in terms of numbers with subjective global impression of improvement, with very serious imprecision.

Very low quality evidence from 1 RCT comprising 337 participants showed that there was no difference in clinical effectiveness between sativex and placebo in terms of numbers with adverse events leading to withdrawal, with very serious imprecision.

Low quality evidence from 2 RCTs comprising 194 participants showed that there was no difference in clinical effectiveness between sativex and placebo in terms of numbers with nausea, with very serious imprecision.

Moderate quality evidence from 2 RCTs comprising 194 participants showed that sativex was clinically harmful compared to placebo in terms of numbers with somnolence, with serious imprecision.

Low quality evidence from 1 RCT comprising 34 participants showed that sativex was clinically harmful compared to placebo in terms of numbers with weakness, with serious imprecision.

Sativex responders versus placebo

Moderate quality evidence from one RCT comprising 241 participants showed that sativex used in a population of positive responders was clinically effective compared to placebo in terms of numbers with at least 30% improvement in NRS, with no imprecision.

Moderate quality evidence from two RCTs comprising 277 participants showed that sativex used in a population of positive responders was clinically effective compared to placebo in terms of timed 10m walk (if ambulatory), with serious imprecision.

High quality evidence from two RCTs comprising 277 participants showed that sativex used in a population of positive responders was clinically effective compared to placebo in terms of patient global impression of improvement, with no imprecision.

Moderate quality evidence from one RCT comprising 241 participants showed that sativex used in a population of positive responders was clinically effective compared to placebo in terms carer global impression of improvement in ease of transfer, with serious imprecision.

High quality evidence from one RCT comprising 241 participants showed that sativex used in a population of positive responders was clinically effective compared to placebo in terms of carer global impression of improvement of function, with no imprecision.

Low quality evidence from one RCT comprising 241 participants showed that sativex used in a population of positive responders was clinically effective compared to placebo in terms of physician global impression of improvement, with serious imprecision.

Low quality evidence from one RCT comprising 241 participants showed that sativex used in a population of positive responders was clinically effective compared to placebo in terms of improvement in Barthel index, with serious imprecision.

High quality evidence from 1 RCT comprising 241 participants showed that there was no difference in clinical effectiveness between sativex used in a population of positive responders and placebo in terms of EQ-5D health status index, with no imprecision.

High quality evidence from 1 RCT comprising 241 participants showed that there was no difference in clinical effectiveness between sativex used in a population of positive responders and placebo in terms of EQ-5D health status VAS, with no imprecision.

High quality evidence from 1 RCT comprising 241 participants showed that there was no difference in clinical effectiveness between sativex used in a population of positive responders and placebo in terms of SF36 physical function, with no imprecision.

High quality evidence from 1 RCT comprising 241 participants showed that there was no difference in clinical effectiveness between sativex used in a population of positive responders and placebo in terms of SF36 mental function, with no imprecision.

Low quality evidence from 1 RCT comprising 241 participants showed that there was no difference in clinical harm between sativex used in a population of positive responders and placebo in terms of numbers with nausea, with very serious imprecision.

Low quality evidence from 1 RCT comprising 241 participants showed that there was no difference in clinical harm between sativex used in a population of positive responders and placebo in terms ofnumbers with somnolence, with very serious imprecision.

Botulinum versus placebo

Very low quality evidence from 1 RCT comprising 37 participants showed that a low dose of botulinum was clinically effective compared to placebo in terms of a patient positive response, with very serious imprecision.

Very low quality evidence from 1 RCT comprising 37 participants showed that there was no difference in clinical effectiveness between a medium dose of botulinum and placebo in terms of a patient positive response, with very serious imprecision.

Very low quality evidence from 1 RCT comprising 37 participants showed that there was no difference in clinical effectiveness between a high dose of botulinum and placebo in terms of a patient positive response, with very serious imprecision.

Moderate quality evidence from 1 RCT comprising 106 participants showed that botulinum was clinically harmful compared to placebo in terms of numbers with weakness, with serious imprecision.

Intrathecal baclofen versus placebo

Very low quality evidence from 2 RCTs comprising 15 participants showed that intrathecal baclofen was clinically effective compared to placebo in terms of numbers with an improvement in Ashworth scale (lower limb), with serious imprecision.

Very low quality evidence from 2 RCTs comprising 15 participants showed that intrathecal baclofen was clinically effective compared to placebo in terms of numbers with an improvement in reflex score (lower limb), with serious imprecision.

Very low quality evidence from 1 RCT comprising 6 participants showed that intrathecal baclofen was clinically effective compared to placebo in terms of numbers with an improvement in spasm score (lower limb), with serious imprecision.

Very low quality evidence from 1 RCT comprising 6 participants showed that intrathecal baclofen was clinically effective compared to placebo in terms of numbers with an improvement in function, with serious imprecision.

9.1.5.2 Economic

One cost-effectiveness analysis found that tizanadine was more costly and more effective than baclofen for the treatment of spasticity (ICER: £10.47 per STD gained for 1st line, £8.46 per STD gained for second line). This analysis was assessed as partially applicable with potentially serious limitations.

One cost–utility analysis found that Sativex was not cost-effective compared to placebo for the treatment of spasticity (ICER: £49,238 per QALY gained). This analysis was assessed as directly applicable with minor limitations.

One cost–utility analysis found Sativex was dominant (less costly and more effective) compared to placebo for the treatment of spasticity in a Spanish context. This analysis was assessed as partially applicable with potential serious limitations.

9.1.6 Recommendations and link to evidence

Recommendations	**30. In people with MS sssess and offer treatment for factors that may aggravate spasticity such as constipation, urinary tract or other infections, inappropriately fitted mobility aids, pressure ulcers, posture and pain.** **31. Encourage people with MS to manage their own spasticity symptoms by explaining how doses of drugs can be adjusted within agreed limits.** **32. Ensure that the person with MS:** o has tried the drug at an optimal dose, or the maximum dose they can tolerate. o stops the drug if there is no benefit at the maximum tolerated dose o has their drug treatment reviewed at least annually once the optimal dose has been reached. **33. Consider baclofen or gabapentin[s] as a first-line drug to treat spasticity in MS depending on contraindications and the person's comorbidities and preferences. If the person with MS cannot tolerate one of these drugs consider switching to the other.** **34. Consider a combination of baclofen and gabapentin[tu] for people with MS if:** o individual drugs do not provide adequate relief **or** o side effects from individual drugs prohibit the dose being increased. **35. Consider tizanidine or dantrolene as a second-line option to treat spasticity in people with MS.** **36. Consider benzodiazepines as a third-line option to treat spasticity in MS and be aware of their potential benefit in treating nocturnal spasms.**

s At the time of publication (October 2014), gabapentin did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

t At the time of publication (October2014), gabapentin did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

u Use caution when using gabapentin and baclofen in combination. For more information on cautions for these drugs see the summary of product characteristics for gabapentin and baclofen and the British National Formulary.

	37. If spasticity cannot be managed with any of the above pharmacological treatments, refer the person to specialist spasticty services.
Relative values of different outcomes	Quality of life is usually regarded as the most important outcome, and this may show change if a reduction in spasticity or spasms reduces pain or improves activities of daily living or carer burden. The most commonly used outcomes were those evaluating changes in spasticity, such as the Ashworth scale or patient-reported spasticity outcomes which ranged from global satisfaction to rating scales for spasms and stiffness. The Ashworth and modified Ashworth scale for spasticity however, are known to have serious limitations. Functional improvements were also regarded as important sensitive indicators of improvement, as even small changes in spasticity can have a major impact on functioning.
Trade off between clinical benefits and harms	Gabapentin had the clearest clinical benefits, followed by baclofen, tizanidine and dantrolene. These benefits were often highly uncertain. Most treatments for spasticity had adverse effects, especially tizanidine, dantrolene and gabapentin. The GDG felt that these adverse effects were not sufficiently severe to counter the potential benefits of these drugs. For example, although successful treatment of spasticity often results in muscle weakness, this is often clinically justified by the benefits. Intrathecal baclofen showed benefits over placebo in some underpowered trials. Adverse events were not reported by any studies so it is not possible to comment on potential harms of intrathecal baclofen.
Economic considerations	One cost–effectiveness study was identified which found that oral tizanidine was more costly and more effective than oral baclofen, where effectiveness was measured in terms of successfully treated day. The chosen measure appears to introduce bias into the cost effectiveness results to the detriment of the comparator drug baclofen, as this measure only takes account of one adverse event which is more common with baclofen (muscle weakness). The GDG acknowledging this important limitation did not have confidence in the conclusion of this study which is also in conflict with the conclusion of our clinical review, where there appeared to be no difference in the majority of outcomes between baclofen and tizanidine. Furthermore, the unit costs of the individual pharmacological treatments were presented to the GDG and they showed that on average treatment with oral baclofen is the cheapest among the available drug therapies for spasticity. The annual cost of the drugs varied depending on the prescribed dose and was between £11–46 for oral baclofen, £49–157 for gabapentin, £53–665 for tizanidine, £62–629 for dantrolene and £11–77 for diazepam. Gabapentin was shown to be more effective than baclofen; however it is also more expensive. Based on these considerations on costs and clinical effectiveness, the GDG considered that gabapentin or baclofen should be offered as a first-line treatment for spasticity. Three studies of intrathecal baclofen were identified but were excluded due to the lack of explicit comparator. The first study reported the cost of intrathecal baclofen (£2,500–3,000 a year); the second indicated that the mean quality of life gain after the intervention was 0.42 and the third study reported a cost per QALY for intrathecal baclofen of £6,900–12,800. The GDG considered these studies with caution as they did not have an explicit comparator. The annual cost of the intrathecal baclofen (excluding staff costs and other consumables) varied depending on the prescribed dose and was between £39–2,484. In addition, a test dose would be required which would be between £1–2 depending on the prescribed dose. The GDG agreed that there was insufficient

	evidence to make a recommendation.
Quality of evidence	The quality of the evidence from the 33 RCTs was generally low or very low, with the main methodological limitations being a lack of allocation concealment, insufficient blinding and inadequate handling of drop-outs in the analyses. Many trials had limited numbers of participants, leading to possible type II errors. A network meta-analysis was not possible due to the differing populations and the lack of common outcomes across studies. The economic evidence for oral tizanidine compared with oral baclofen was assessed as partially applicable with potential serious limitations.
Other considerations	Limited evidence was available for treatment of spasticity and this applied more to older established drugs than newer drugs. The recommendations were therefore informed by the experience of the GDG. Spasticity is a very common symptom in MS and a high proportion of MS patients take drug treatment for this. Baclofen was chosen as first-line therapy in view of cost, tolerability and effectiveness seen in placebo-controlled trials and comparative trials. It is currently the first choice of treatment for spasticity and there is considerable experience among patients and professionals in using baclofen. Gabapentin is also used for neuropathic pain and it may be a better option to be tried first in people with spasticity and neuropathic pain. Benzodiazepines are commonly used for treatment of spasticity in all musculoskeletal conditions but their use can be limited by drowsiness. This can be a particular problem for people with MS who may need relatively high doses to treat their spasticity. This effect can however be used positively by using benzodiazepines for spasticity at night when they can also help with sleep. It is important that any drug is tried at an adequate but tolerated dose before it is judged to be ineffective. Involving patients in these decisions will improve treatment adherence and symptom control. The GDG considered that there were some important principles in how anti-spasticity treatments are used. They considered that people with MS need to be empowered in their use of drugs to treat spasticity. Spasticity is a symptom and the aim of treatment should be to help the patient manage their symptom in the best possible way. The experience of spasticity may differ between individuals. Environmental factors can affect spasticity and the appropriate timing of treatment of spasticity may vary according to each individual's lifestyle, commitments and management of activities of daily living. The GDG wished to emphasise that some patients may require permission and encouragement to adjust their dose according to their needs. Clinical experience is that use of drugs may be limited by side effects and people may need to take more than one drug at doses they can tolerate. The GDG has experience both of people not being given adequate doses of drugs and also of remaining on drugs that they did not find useful for prolonged periods of time. The GDG considered that more research is required into the efficacy and tolerability of spasticity treatments with a particular emphasis on functional and patient reported outcomes. The GDG did not make a recommendation on the use of botulinum or on intrathecal baclofen. There is a poor evidence base for these drugs, they are not commonly used and expertise and specialist services are required for their delivery. The GDG considered that they may have a place in specialist services for people with severe spasticity and complications from their spasticity. A research recommendation on the use of intrathecal baclofen and botulinum toxin has been made by the GDG. A specialist spasticity service consists of a multidisciplinary outpatient service, usually based in secondary care or in a specialist neurological rehabilitation centre. Members of the team typically include a neurologist or consultant in

	rehabilitation medicine, a neurophysiotherapist and a neuro-occupational therapist or neuro-specialist nurse. The service typically provides comprehensive assessment and management for people with complex spasticity or dystonias due to neurological disorders, including MS, Interventions may include advice on posture and positioning, customised seating, splinting and standing as well a drug review, intramuscular botulinum toxin or phenol injections. People are referred to this service when their needs cannot be addressed by local services i.e. when they have been unresponsive to pharmacological or therapeutic interventions and when their spasticity, spasms or pain continue to cause difficulties affecting independence, carer burden or quality of life.

Recommendations	**38. Do not offer Sativex[v] to treat spasticity in people with MS because it is not a cost effective treatment.**
Relative values of different outcomes	Quality of life is usually regarded as the most important outcome, but the GDG considered that it might not be sensitive to changes in spasticity. The most important outcomes were those evaluating changes in spasticity, such as the Ashworth scale or patient-reported spasticity outcomes which ranged from global satisfaction to rating scales for spasms and stiffness. Functional improvements such as improved ambulation were also regarded as important sensitive indicators of improvement, as even small changes in spasticity can have a major impact on functioning.
Trade off between clinical benefits and harms	In the studies with non-enriched study designs, clinically important benefits for Sativex were seen for spasticity, patient satisfaction and ambulation ability, but there was high uncertainty in the magnitude and direction of effect, and no benefits were observed in terms of quality of life. Because of this uncertainty, it was unclear if the adverse effects in the form of drowsiness and weakness were outweighed by the potential clinical benefits. We noted that Sativex is known to be more effective in specific groups of people and so more weight was placed on the studies conducted in people who had shown a previous positive response to sativex. In the two enriched studies, there was clear evidence of the clinical efficacy of sativex in terms of reduction in spasticity, carer, patient and physician global impression of improvement, and improvement in ambulation ability, with good precision in the magnitude and direction of effect overall. No clinically important benefits were observed for quality of life, but, as explained in the section above, this was regarded as a less important outcome for decision making in this context. No clinically important adverse effects were observed in the enriched studies, and so the clinical benefits were unopposed.
Economic considerations	Two cost–utility analyses were identified which compared sativex with placebo. One study took a UK NHS perspective and found that sativex was more costly and more effective than placebo, with an ICER of £49,238 per QALY, which is considered to not be cost effective at £20,000 per QALY. The second study found that, with a Spanish healthcare system perspective, sativex was dominant (both less costly and more effective than placebo). With a German healthcare system perspective, this study found that sativex was more costly and more effective than placebo (ICER: £9,230 per QALY). This second study was funded by the manufacturer of sativex. The two studies had key differences which may account for their conflicting results. The GDG considered the two studies and agreed that the UK study was both more applicable and had fewer limitations than the Spanish and German study.

v This recommendation does not apply to people who have already started treatment with Sativex in the NHS who should be able to continue treatment until they and their NHS clinician think it appropriate to stop.

	Therefore they agreed not to recommend offering sativex sativex for the treatment of spasticity in people with MS.
Quality of evidence	The evidence was very low to moderate in the non-enriched studies. Main risks of bias were a lack of reporting of allocation concealment, insufficient blinding and inadequate handling of drop-outs in the analyses, and many outcomes were seriously or very seriously imprecise. Evidence in the enriched studies was better quality, rated at low to high. The 3 outcomes rated as low were regarded by the GDG as less important in terms of decision-making. Risk of bias was due to a lack of reporting of assessor blinding (where this was relevant) and some outcomes had serious imprecision. The sativex UK cost–utility analysis was assessed as directly applicable with minor limitations and the Spanish/German study was assessed as partially applicable with potential serious limitations.
Other considerations	The GDG felt that it was more appropriate to place more weight on the evidence from the enriched studies, as Sativex is known to work best in a small proportion of the population.The GDG discussed the use of cannabis by people with MS as a means of managing their symptoms. They acknowledged that while sativex was seen as a potential development in the management of MS symptoms, and that there were clear clinically important benefits in the enriched studies, the cost effectiveness evidence did not support its use.

9.2 Pharmacological management of mobility

9.2.1 Introduction

Problems with walking are a significant problem for people with MS. This has significant effect on activities of daily living and vocational and recreational activities. Treatment is mainly using non-pharmacological methods and these are reviewed in chapter 10.4. This chapter examines the evidence for the use of fampridine for treatment of mobility problems.

9.2.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of mobility with fampridine?

Table 53: PICO characteristics of review question

Population	• Adults
Intervention/s	• Fampridine (use of disease modifying drugs is permissible as an accompaniment to this treatment)
Comparison/s	• Usual treatment (use of disease modifying drugs is permissible as part of usual treatment)
Outcomes	• Quality of life [critical] • Changes in disability or impairment scales (validated) assessing ○ Motor function ○ Fatigue ○ Spasticity ○ Walking speed [critical]

	• Incidence of adverse events [secondary] [timescale for all outcomes was end of treatment]
Study design	• RCTs

9.2.3 Clinical evidence

We searched for randomised controlled trials (RCTs) comparing fampridine and usual treatment for mobility. 8 RCTs were found, four of which were parallel group trials[84 81 83 82] and 4 of which were cross-over trials.[22,201,217,254] No minimum period of washout was necessary as an inclusion criterion as the half-life of fampridine is low (2-3 hours). We excluded studies that evaluated 3,4 diaminopyridine. A Cochrane review was found but it was not appropriate for use as its results were in narrative form. Its reference list was checked for appropriate studies.

Details of these studies are summarised in Table 54.

Table 54: Characteristics of the included studies

Study	population (as far as known)	Interventions	Methodology
Goodman 2007[84]	n=36; approx 65% women; approx mean duration 156 months; Relapsing remitting approx 20%, mostly secondary progressive; EDSS:5.3	Fampridine 40mg twice daily for 8 weeks vs identical placebo. The dose started at 10mg and was increased in 5mg increments per week. Downward titration to 10mg occurred in the final week in 2 steps.	Parallel group study
Goodman 2008[81]	n=206; Mostly secondary progressive; EDSS:5.8	Fampridine 10mg, 15mg or 20mg twice daily for 15 weeks vs identical placebo. 12 weeks were at a stable dose; the first 2 weeks involved escalation of doses for the higher dose groups, and the final week involved downward titration as appropriate.	Parallel group study
Goodman 2009[83]	n=301; approx 65% women; approx mean duration 156 months; Relapsing remitting 28%, mostly secondary progressive; EDSS:5.8	Fampridine 10mg twice daily for 14 weeks vs identical placebo	Parallel group study. This study also performed an additional analysis, splitting the fampridine group into “responders” and “non-responders” according to whether or not the patient improved their 25m timed walk consistently over the follow up visits. They then also compared the “responders” to the original placebo group for all outcomes. The results from these additional analyses are not reported in this review because 1) for ambulation outcomes the fampridine group would, by virtue of being selected for their positive ambulation response, have

Study	population (as far as known)	Interventions	Methodology
			inevitably done better than the placebo group, where no such selection took place, 2) For the non-ambulation outcomes the possibility of correlations between timed walk performance and other outcomes may also have led to a general overestimation of effect for fampridine, and 3) the sub-analysis did not address the review question.
Goodman 2010[82]	n=239; approx 70% female; mean duration approx 14 years; relapsing remitting approx 34%, mostly secondary progressive; EDSS:5.8.	Fampridine 10mg twice daily for 9 weeks vs identical placebo	Parallel group study. This study also performed an additional sub-analysis exactly as for Goodman 2009 (above). Again, those results have not been reported.
Rossini 2001[201]	n=54; 59% female; mean duration 13.2 years; 6 primary progressive, but most secondary progressive; EDSS: 6.2	Fampridine 8mg 4 times per day for 6 months vs identical placebo. No washout period "due to short half-life of drug".	Cross-over study
Schwid 1997[217]	n=10; 60% female; mean duration of 13.5 years; EDSS: 6-7.5; MS type unclear.	17.5 mg twice daily for 1 week vs identical placebo. 1 week washout period.	Cross-over study
Bever 1994A[22]	n=8; 50% women; Disease duration 2-30 years; 2 relapsing remitting but most characterised as chronic progressive; EDSS: 6.	Low dose (30-59 ng/ml serum) and high dose (60-100 ng/ml serum). Duration unclear but could be as short as 30 hours. Existence of washout period unclear.	Cross-over study
van Dieman 1992[254]	n=70; 61% women; mean disease duration 86 months; Relapsing remitting 25.7%, but mostly chronic progressive; EDSS: 5.	Fampridine 0.5mg/kg body weight for 12 weeks vs identical placebo. No washout period given.	Cross-over study

Fampridine versus placebo.

Table 55: Clinical evidence profile: Fampridine versus placebo

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Fampridine**	**placebo**	**Relative (95% CI)**	**Absolute (Fampridine compared to placebo)**		
Quality of life												
No papers covered this critical outcome												
positive **response to treatment – numbers with improvement of >20% in walking speed throughout study or numbers with 75% of walking tests during treatment better than pre-treatment tests.**												
Goodman 2008 Goodman 2009 Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	183/491 (37.3%) median control event rate	22/247 (8.9%) 8.8%	RR 4.34 (2.85 to 6.62)	294 more per 1000 (from 163 more to 495 more)	LOW	CRITICAL
Time to walk 8m (seconds) (better indicated by lower values)												
Schwid 1997	randomised trials	none	no serious inconsistency	no serious indirectness	serious[B]	none	n=8 Cross-over study Paired data used		-	Generic Inverse Variance MD 7.22 lower (0.36 lower to 14.08 lower)	MODERATE	CRITICAL
Percentage change from baseline in gait speed (better indicated by higher values)												
Schwid 1997	randomised trials	very serious[A]	serious[B]	no serious indirectness	no serious imprecision	none	n=245 Cross-over study and parallel		-	Random effects Generic	VERY LOW	CRITICAL

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
Goodman 2010							data combined using generic inverse variance			Inverse Variance MD 9.96 higher (1.02 less to 20.93 higher)		
MSWS-12 score change from baseline - low dose (10 mg) (Better indicated by lower values)												
Goodman 2008 Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	-5.33(16.15)[51] -2.62(10.8)[119]	-3.56(14.55)[46] 0.73(10.8)[118]	-	MD 3.08 lower (5.59 lower to 0.58 lower)	VERY LOW	CRITICAL
MSWS-12 score change from baseline - medium dose (15 mg) (Better indicated by lower values)												
Goodman 2008	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	-7.32(16.29)[49]	-3.56(14.55)[46]	-	MD 3.76 lower (9.96 lower to 2.44 higher)	LOW	CRITICAL
MSWS-12 score change from baseline - high dose (20 mg) (Better indicated by lower values)												
Goodman 2008	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	-5.76(15.3)[52]	-3.56(14.55)[46]	-	MD 2.2 lower (8.11 lower to 3.71 higher)	LOW	CRITICAL
Strength score – sum of MRC gradings for 4 lower limb muscles - low dose 30-59 ng/ml (Better indicated by higher values)												

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
Bever 1994A	randomised trials	none	no serious inconsistency	no serious indirectness	serious[B]	none	n=8 Cross-over study Paired data used		-	Generic Inverse Variance MD 1.38 higher (1.6 lower to 4.36 higher)	MOD	CRITICAL
Strength score – sum of MRC gradings for 4 lower limb muscles - high dose 60-100 ng/ml (Better indicated by higher values)												
Bever 1994A	randomised trials	none	no serious inconsistency	no serious indirectness	no serious imprecision	none	n=7 Cross-over study Paired data used		-	Generic Inverse Variance MD 3.28 higher (1.75 to 4.83 higher)	HIGH	CRITICAL
Fatigue												
No data for this outcome suitable for Grade – but see narrative review.												
Any adverse events												
Goodman 2007 Goodman 2008 Goodman 2009 Goodman 2010 Rossini 2001 van Diemen 1992	randomised trials	very serious[A]	very serious[C]	no serious indirectness	serious[B]	none	NA	NA	Random effects generic inverse variance RR: 1.36(1.10 to 1.68)	Not available	VERY LOW	IMPORTANT
adverse events – fall												
Goodman 2008	randomised	very serious[A]	no serious	no serious	very serious[B]	none	75/507	36/248	RR 1.01	1 more per	VERY LOW	IMPORTA

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
Goodman 2009 Goodman 2010	trials		inconsistency	indirectness			(14.8%)	(14.5%)	(0.68 to 1.49)	1000 (from 46 fewer to 71 more)		NT
							median control event rate	15.3%		2 more per 1000 (from 49 fewer to 75 more)		
adverse events – UTI												
Goodman 2008 Goodman 2009 Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	72/507 (14.2%)	22/248 (8.9%)	RR 1.64 (1.05 to 2.59)	57 more per 1000 (from 4 more to 141 more)	VERY LOW	IMPORTANT
							median control event rate	8.4%		54 more per 1000 (from 4 more to 134 more)		
adverse events – dizziness												
Goodman 2007 Goodman 2008 Goodman 2009 Goodman 2010 Van Dieman 1992	randomised trials	very serious[A]	serious[C]	no serious indirectness	no serious imprecision	none	93/601 (15.5%)	16/328 (4.9%)	Random effects RR 2.94 (1.20 to 7.19)	95 more per 1000 (from 10 more to 302 more)	VERY LOW	IMPORTANT
							median control event rate	5.8%		113 more per 1000 (from 12 more to 359 more)		
adverse events – insomnia												

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
Goodman 2007 Goodman 2008 Goodman 2009 Goodman 2010	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	62/532 (11.7%)	10/259 (3.9%)	RR 2.76 (1.47 to 5.19)	68 more per 1000 (from 18 more to 162 more)	MODERATE	IMPORTANT
							median control event rate	5.6%		99 more per 1000 (from 26 more to 235 more)		
adverse events – fatigue												
Goodman 2008 Goodman 2009	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	34/387 (8.8%)	7/129 (5.4%)	RR 1.66 (0.76 to 3.64)	36 more per 1000 (from 13 fewer to 143 more)	MODERATE	IMPORTANT
							median control event rate	5.8%		38 more per 1000 (from 14 fewer to 153 more)		
adverse events – nausea												
Goodman 2007 Goodman 2008 Goodman 2009 Goodman 2010 Van Dieman 1992	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	55/601 (9.2%)	7/328 (2.1%)	RR 3.69 (1.83 to 7.45)	57 more per 1000 (from 18 more to 138 more)	LOW	IMPORTANT
							median control event rate	3.5%		94 more per 1000 (from 29 more to 226 more)		

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
adverse events – URTI												
Goodman 2008 Goodman 2009 Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	30/507 (5.9%)	16/248 (6.5%)	RR 0.92 (0.5 to 1.68)	5 fewer per 1000 (from 32 fewer to 44 more)	VERY LOW	IMPORTANT
							median control event rate	6.7%		5 fewer per 1000 (from 34 fewer to 46 more)		
adverse events – asthenia												
Goodman 2007 Goodman 2008 Goodman 2009 Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	52/532 (9.8%)	11/259 (4.2%)	RR 2.3 (1.2 to 4.4)	55 more per 1000 (from 8 more to 144 more)	VERY LOW	IMPORTANT
							median control event rate	4.9%		64 more per 1000 (from 10 more to 167 more)		
adverse events - back pain												
Goodman 2009 Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	20/348 (5.7%)	3/191 (1.6%)	RR 3.58 (1.05 to 12.16)	41 more per 1000 (from 1 more to 175 more)	VERY LOW	IMPORTANT
							median control event rate	1.3%		34 more per 1000 (from 1 more to 145		

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
										more)		
adverse events - balance disorders												
3Goodman 2008 Goodman 2009 Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32/507 (6.3%)	4/248 (1.6%)	RR 3.43 (1.27 to 9.26)	39 more per 1000 (from 4 more to 133 more)	LOW	IMPORTANT
							median control event rate	1.7%		41 more per 1000 (from 5 more to 140 more)		
adverse events – headache												
Goodman 2007 Goodman 2008 Goodman 2009 Goodman 2010 Van Dieman 1992	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	51/601 (8.5%)	11/328 (3.4%)	RR 2.05 (1.12 to 3.74)	35 more per 1000 (from 4 more to 92 more)	VERY LOW	IMPORTANT
							median control event rate	5.6%		59 more per 1000 (from 7 more to 153 more)		
adverse events – arthralgia												
Goodman 2008 Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	15/279 (5.4%)	8/176 (4.5%)	RR 1.14 (0.48 to 2.68)	6 more per 1000 (from 24 fewer to 76 more)	VERY LOW	IMPORTANT
							median control event rate	4.7%		7 more per 1000 (from 24		

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
										fewer to 79 more)		
adverse events – nasopharyngitis												
Goodman 2010	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	6/120 (5%)	5/119 (4.2%)	RR 1.19 (0.37 to 3.79)	8 more per 1000 (from 26 fewer to 117 more)	VERY LOW	IMPORTANT
							median control event rate	4.2%		8 more per 1000 (from 26 fewer to 117 more)		
adverse events – paraesthesia												
Goodman 2007 Goodman 2008 Goodman 2010 Van Dieman 1992	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	44/373 (11.8%)	16/256 (6.3%)	RR 1.9 (1.09 to 3.31)	56 more per 1000 (from 6 more to 144 more)	VERY LOW	IMPORTANT
							median control event rate	7.2%		65 more per 1000 (from 6 more to 166 more)		
adverse events - accidental injury												
Goodman 2007	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	4/25 (16%)	3/11 (27.3%)	RR 0.59 (0.16 to 2.19)	112 fewer per 1000 (from 229 fewer to 325 more)	VERY LOW	IMPORTANT

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
							median control event rate	27.3%		112 fewer per 1000 (from 229 fewer to 325 more)		
adverse events – tremor												
Goodman 2007	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	6/25 (24%)	0/11 (0%)	Peto OR: 5.37 (0.82 to 35.03))	240 more per 1000 (from 40 more to 440 more)	VERY LOW	IMPORTANT
							median control event rate	0%		240 more per 1000 (from 40 more to 440 more)		
adverse events – oedema												
Goodman 2008	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	13/159 (8.2%)	3/57 (5.3%)	RR 1.55 (0.46 to 5.25)	29 more per 1000 (from 28 fewer to 224 more)	VERY LOW	IMPORTANT
							median control event rate	5.3%		29 more per 1000 (from 29 fewer to 225 more)		
adverse events - muscle spasm												
Goodman 2008	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	9/159 (5.7%)	3/57 (5.3%)	RR 1.08 (0.3 to	4 more per 1000 (from 37 fewer to 149	VERY LOW	IMPORTANT

Quality assessment							Proportion of participants with event OR mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fampridine	placebo	Relative (95% CI)	Absolute (Fampridine compared to placebo)		
									3.83)	more)		
							median control event rate	5.3%		4 more per 1000 (from 37 fewer to 150 more)		
exacerbation of MS												
Goodman 2008	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	12/159 (7.5%)	0/57 (0%)	Peto OR: 4.19 (1.12 to 15.64)	80 more per 1000 (from 30 more to 120 more)	VERY LOW	IMPORTANT
							median control event rate	0%		-		
Discontinuation due to adverse events												
Goodman 2007 Goodman 2008 Goodman 2009 Goodman 2010 van Dieman 1992	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	31/596 (5.2%)	6/317 (1.9%)	RR 2.6 (1.1 to 6.15)	30 more per 1000 (from 2 more to 97 more)	VERY LOW	IMPORTANT
							median control event rate	1.5%		24 more per 1000 (from 2 more to 77 more)		

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*
[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two*

increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

[c] *Outcomes were downgraded by one increment for serious inconsistency, as shown by the I squared value being between 50 and 74%. A double downgrade was applied for very serious inconsistency if I squared was >75%. If serious or very serious inconsistency existed, and there were >2 studies, pre-defined sub-grouping (see review question protocol) was applied. If consistency within each sub-group was achieved, then the results for each sub-group were reported as separate outcomes. If this did not reduce inconsistency to acceptable levels within all sub-groups, or there were only 2 studies, then the entire group was re-analysed using a random effects model to allow for the fact that a homogeneous population was not present. In this instance, sub-grouping was applied to two outcomes with heterogeneity and >2 studies, but this did not reduce inconsistency, and so a random effects model was used.*

Narrative review

Some outcomes were not appropriate for GRADE because of the use of non-normally distributed interval data, the lack of effect-size data, or the lack of variance measures. These are presented below in narrative form

EDSS

This outcome could not be analysed in review manager and GRADE because it an ordinal scale. Four cross-over studies[201,217 22 254] assessed this outcome. Schwid 1997[217] found that for all 10 subjects, 3 showed a greater improvement for EDSS on Fampridine compared to placebo, and the other 7 showed the same improvement on both treatments. Schwid 1997[217] correctly analysed these data non-parametrically, finding a trend (p=0.16) for an effect favouring fampridine. Rossini[201] used parametric methods to compare effects of the two treatments on EDSS, and showed identical changes (-0.05) in both groups. Bever 1994A[22] reported that no changes were seen in EDSS in either group, although data are not provided. Van Diemen 1992[254] described EDSS data in each period separately. For the first period, EDSS improved by 0.18 in the fampridine group and worsened by 0.15 in the placebo group, and in the second period EDSS improved by 0.09 in the fampridine group and worsened by 0.23 in the placebo group. Variance data or p values were not given. In summary, there is little good evidence to suggest fampridine has an appreciable effect on EDSS.

Ashworth scale

This outcome could not be analysed in review manager and GRADE because it is an ordinal scale. Three parallel studies (Goodman 2008[81] , Goodman 2009[83] and Goodman 2010[82]) assessed this outcome. All analysed Ashworth scale parametrically, and hence results are inevitably misleading. Goodman 2008[81] reported that the placebo group showed greater mean improvements (-0.11) than each of the 10mg (-0.04), 15mg (-0.06) and 20 mg (0.02) fampridine doses but that these were not statistically significant. In contrast, Goodman 2009[83] reported that there was a significantly greater Ashworth scale improvement in the fampridine than the placebo group (p=0.0210). Likewise, Goodman 2010[82] found that mean improvements from baseline were significantly greater (p=0.015) in the fampridine group (-0.18) than the placebo group (-0.06). In summary, however, because of inappropriate analysis methods, it is difficult to know if fampridine affects Ashworth scale.

Average change from baseline in walking speed

Goodman 2009[83] reported that there was a greater improvement in the fampridine than the placebo group (p=0.0004) but no effect sizes were given. Goodman 2007[84] performed a repeated measures analysis for the changes in gait speed over 7 weeks, and this demonstrated a significantly greater improvement in the fampridine group (p=0.03). However, no effect sizes were provided. Goodman 2008[81] presented their results in a low resolution graph, but stated that there were no significant differences between any of the 4 groups, though all actively treated groups had larger numerical improvements than the placebo group. Overall, these findings of a generally positive effect of fampridine on walking speed appear to support those that were included in the meta-analysis.

Lower Extremity Manual Muscle Testing (LEMMT)

Four studies[84 81 83 22] derived the LEMMT score by summing Medical Research Council scores across 4 different lower limb muscle groups to derive an overall strength score. They then analysed this outcome parametrically. It is likely that such an analysis was flawed, as the MRC grading system is ordinal and not interval. Hence LEMMT measures were not included in review manager or GRADE. In addition, three studies[84 81 83] also did not include effect sizes. Results of these studies are summarised below.

Goodman 2009[83] reported that there was a greater improvement in the fampridine than the placebo group (p=0.0029) but no effect sizes were given. Goodman 2007 performed a repeated measures analysis for the changes in LEMMT over 7 weeks, and this demonstrated a significantly greater improvement in the fampridine group (p=0.01). However, no effect sizes were provided. Goodman 2008[81] presented their results in a low resolution graph , but stated that there were significant differences between placebo and each of 10mg (p=0.018) and 15mg (p=0.003) fampridine doses, but not with a 20mg fampridine dose (p>0.05). In a cross-over study, Bever 1994[22] summed MRC scores across 4 different muscle groups to derive an overall strength score, finding a weak trend in terms of improvements favouring fampridine at a lower dose of 30-59 ng/ml(MD: 1.38 (-1.6 to 4.36), but a clear effect favouring fampridine at a higher dose of 60-100 ng/ml(MD: 3.28 (1.75 to 4.83). Overall, because of the flawed analyses, it is difficult to know if fampridine affects lower limb strength.

<u>Change in Subjective Global Impression (SGI) score</u>

Goodman 2010[82] reported that the Fampridine group was favoured but that it was non- significant, and no effect sizes were given. Goodman 2007[84] measured SGI scores but did not formally analyse differences across groups. This failure to present data was obscurely explained by “the dose – exploratory nature of the study and the expectation that the global impression would relate both to potential effects and side effects”.

<u>Change in Clinician Global Impression (CGI) score</u>

Goodman 2007[84] measured CGI scores but did not formally analyse differences across groups. This failure to present data was also obscurely explained by “the dose –exploratory nature of the study and the expectation that the global impression would relate both to potential effects and side effects”.

<u>Fatigue</u>

Goodman 2008[81] reported that reductions in fatigue, as measured by the Modified Brief Fatigue inventory (BFI) score, were similar for both groups (p=0.13), but no data were provided except in a low resolution graph. Rossini 2001[201] reported that the decrements in fatigue, as measured by the fatigue severity scale (FSS) did not differ between groups (p=0.19). No data were presented except in a low resolution graph. These data appear to match the meta-analysis results.

Other data

Goodman 2008 and Goodman 2010 also performed additional post-hoc analyses, splitting the fampridine group into “responders” and “non-responders” according to whether or not the patient improved their 25m timed walk consistently over the several follow up visits. They then compared the “responders” to the “non-responders” and the original placebo group for all outcomes, and Goodman 2010 presented these as the main outcomes.

The results from these additional analyses are not reported in this review for the following reasons.

- Original randomisation, and thus group comparability at baseline, was broken by the post-hoc selection of a sub-group from the fampridine group.
- By virtue of the definition of a positive response, the “responder” group would inevitably have had better outcomes than the other groups for ambulation speed, and this is likely to have applied to other outcomes related to mobility.

- The “responder” analysis did not address the review question, which concerned the effects of fampridine on the population of adults with multiple sclerosis, not the population of adults with multiple sclerosis that responded to fampridine. It is acknowledged that some drugs may only work on a sub-set of the population, and that research designs should reflect this to provide a realistic assessment of efficacy. However, a more robust responder analysis should have randomised the responders from this RCT into fampridine and placebo groups, so that a proper randomised comparison could be made.

9.2.4 Economic evidence

Published literature

No relevant economic evaluations comparing fampridine with usual care were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

Relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

New cost-effectiveness analysis

This area was prioritised for new cost-effectiveness analysis. The summary of the results can be found in **Table 56** below and the details of the analysis can be found in the following paragraphs.

Table 56: Economic evidence profile: Fampridine vs. placebo

Study	Applicability	Limitations	Other comments	Incremental cost per year	Incremental effects (QALY)	ICER	Uncertainty
NCGC analysis	Directly applicable	Potentially serious limitations (a)	Population: patients who have responded to treatment with fampridine. Time horizon: one year. Based on an RCT included in the clinical review[82,84]	£4,719 (b)	0.029 QALY (c)	£160,884 per QALY	Threshold analysis: change in incremental EQ-5D for the ICER to decrease to £20,000/QALY is 0.236. Assuming baseline MSWS-12 scores and MSWS-12 score at 9 weeks in the placebo group are unchanged, this corresponds to a decrease in the MSWS-12 score in the fampridine responders group by 52.11 (compared to the 6.04 reported in the study).

(a) *Analysis based on a single RCT[82,84]; utilities were estimated through a mapping function which is associated with limitations. Non-responder costs and adverse event costs have not been included.*

(b) Cost of drug treatment only.

(c) Difference in QALY calculated as the incremental change in EQ-5D score between baseline and follow-up using an algorithm that mapped MSWS-12 scores to EQ-5D scores. The improvement in EQ-5D was assumed to be constant over a year.

Methods

A simple cost-utility analysis was conducted from the NHS perspective to compare treatment with fampridine with placebo for improving mobility in people with multiple sclerosis. Methods were consistent with the NICE reference case unless otherwise stated.

Firstly we decided to conduct the analysis for a population who had already had a trial with fampridine and had been categorised as responders. Responders were defined as individuals with a faster walking speed for at least 3 of the 4 visits during the treatment period as compared with the maximum speed for any of the 5 off-drug visits. It was planned that if fampridine was found to be cost-effective in this population, then a broader analysis including the overall population of fampridine-naïve individuals would be conducted. On the other hand, if fampridine was found to be not cost-effective in the responsive population, offering a four-week trial with fampridine would not be worthwhile and a formal analysis would not be necessary.

Effectiveness was expressed as quality-adjusted life-years (QALYs); this was estimated through the mapping of changes in MSWS-12 scores, obtained from our systematic review of the clinical evidence (see 9.2.3) to EQ-5D. Due to the limited follow-up time of the clinical data, a one-year time horizon was considered. It may be feasible that benefits would continue beyond this period if treatment was continued however the treatment costs would also continue; therefore it was deemed not necessary to further extrapolate beyond the clinical data to a longer time horizon. Costs and QALYs were not discounted due to the short time horizon.

Clinical effectiveness

Two RCTs included in the clinical review for the guideline (9.2.3) reported the MSWS-12 score change from baseline in the group of patients randomised to the fampridine treatment who responded to treatment and in the group of patients randomised to placebo.[81,82,84] We decided to conduct the analysis using the Goodman et al. (2010) RCT as this trial reported more favourable MSWS-12 score changes than the Goodman et al. (2008) RCT. It was planned that if fampridine was found to be cost-effective using the Goodman et al. (2010) data, then a sensitivity analysis using the Goodman et al. (2008) data would be conducted. On the other hand, if fampridine was found to be not cost-effective, using less favourable data would equally not be cost-effective and a formal analysis would not be necessary. MSWS-12 scores at baseline and follow-up for Goodman et al. (2010) are reported in Table 57.

Of note, the Goodman et al. (2008) RCT, looked at a three different doses, 10 mg, 15 mg and 20 mg twice daily. In this analysis only the 10 mg twice daily dose was considered as this is the recommended dose reported in the summary of product characteristics for fampridine.

The clinical evidence found that fampridine was associated with a greater risk of adverse events such as nausea, dizziness and insomnia compared to placebo. These adverse events are not captured in this simple model as they were considered unlikely to have a large impact on resource use or quality of life. This is a conservative approach as including the impact of adverse events may make fampridine less cost effective compared to placebo.

QALYs

In line with the NICE reference case, EQ-5D data was sought in order to estimate QALYs. Preferably, direct EQ-5D data measuring treatment effect on health-related quality of life would be used but this was not available from the systematic review of RCTs carried out for the guideline.

A systematic search of quality of life (QoL) studies was conducted and a study was found[91] which provided us with a mapping function to estimate EQ-5D scores from MSWS-12 scores. The characteristics of the Hawton et al. (2012) mapping study were considered to be similar to, and overlapping with the Goodman et al. (2010). Age and gender were similar and both studies included people with different types of MS. The baseline MSW2-12 scores were 60.1 for Hawton et al. (2012) and 70.8 for Goodman et al. (2010).

In this study, 21 regression models were estimated using MSWS-12 and EQ-5D data collected in a longitudinal cohort study of 560 individuals with multiple sclerosis in the UK followed up for 6 months. The best performing model is the model that most accurately estimates EQ-5D values for a population; this is selected by comparing the models' estimation errors which is the difference between the actual EQ-5D score for an individual and the relative EQ-5D score estimated using the model. Although the best performing model was one based on individual item scores, for practical reasons the best performing model based on aggregate data (ordinary least squares [OLS] total score and total score squared model) was selected and its algorithm is reported in Equation I.

The EQ-5D values were estimated at baseline and follow-up time using the algorithm developed by Hawton et al. (2012):[91]

I ***EQ-5D score = 0.8863602 - 0.0047809 * MSWS-12 + 0.00000325 * MSWS-12 * MSWS-12***

where MSWS-12 represents the total MSWS-12 score. Estimated scores and differences between the placebo groups and the fampridine group are reported in Table 57.

Table 57: Calculating EQ-5D from MSWS-12 scores

	MSWS-12 score at baseline (a)	MSWS-12 score at 9 weeks (a)	EQ-5D score at baseline (b)	EQ-5D score at 9 weeks (b)	Estimated change in EQ-5D at 9 weeks
Fampridine responders group (n=51)	72.1	66.06	0.5586	0.5847	0.0262
Placebo group (n=118)	67.7	68.43	0.5776	0.5744	-0.0032
EQ-5D improvement - fampridine versus placebo group					***0.029***

(a) From Goodman et al. (2010)[82,84]
(b) Calculated by substituting the MSWS-12 score in the previous columns in equation I.

QALY gain with fampridine was estimated assuming the effectiveness throughout the year is similar to the effectiveness observed at 9 weeks (i.e. the difference in MSWS-12 scores and therefore in EQ-5D between fampridine and placebo is constant).

Since the time horizon of our analysis is one year and it is assumed no one dies in that time,the QALY gain corresponds to the improvement in EQ-5D value (0.029).

During stakeholder consultation for theguideline it was highlighted that Macdonell et al. (2013)[128] reported EQ-5D data from a 48-week, open-label, single arm, multicentre Phase 4 study of fampridine. This non-randomised study was excluded from the clinical review and therefore not deemed suitable for the base case analysis in this economic model. As this data is direct EQ-5D data measuring treatment effect on health-related quality of life and includes a longer follow up of 48 weeks, it was included in a sensitivity analysis. In the study, data was collected for those receiving fampridine, which are those identified as responders at week 4, and for those who are not receiving fampridine, which are non-responders at week 4 who agreed to be allocated to the control group. The mean change in EQ-5D score from baseline at different follow-up points is reported in Table 58. Since the time horizon of the economic analysis is one year, it was assumed that the EQ-5D improvement observed at week 48 was maintained for a further 4 weeks, to week 52 for comparability to the base case analysis. Furthermore, it was assumed the improvement at week 12 occurred from the beginning of the trial. The EQ-5D data was ploted and the area under the curve was calculated to estimate the incremental QALYs.

Table 58: Change from baseline in EQ-5D-3L scores

	Week 12	Week 24	Week 36	Week 48
Fampridine responders group	0.06 (n=652)	0.05 (n=624)	0.03 (n=598)	0.04 (n=568)
Control group	0.02 (n=78)	0.01 (n=69)	0.01 (n=58)	0.00 (n=47)
EQ-5D-3L improvement – fampridine versus control group	***0.04***	***0.04***	***0.02***	***0.04***

Source: Macdonell 2013[128]

Costs

The cost of identifying responders was not included in this analysis as the population was those who had already been identified as responders following an initial four week trial.

Fampridine is available as 10 mg tablets costing £181 and £362 for a 28 and 56 pack respectively. The annual cost of treatment with fampridine is reported in Table 59 as £4,719 based on a 10 mg twice a day dose.

Table 59: Unit cost of fampridine

	Mg/ units	Units/ pack	Cost/ pack (£)	Cost/ unit (£)	Units/day	Cost/day (£)	Cost/year (£)
Fampridine – 28 pack, 10 mg twice a day	10	28	181	6.46	2	12.93	4,719
Fampridine – 56 pack, 10 mg twice a day	10	56	362	6.46	2	12.93	4,719

Source: MIMS[94]

Only drug costs were included in our analysis as the number of assessments is uncertain and could be equal to that number of visits in an untreated population. This could mean that if more visits are required for patients undergoing treatment, the cost of fampridine in our analysis is an underestimate.

Downstream costs were not included in the analysis as no data was available from the RCTs on the impact of fampridine on healthcare utilisation. The GDG considered that fampridine may result in plausible downstream savings due to delayed deterioration of mobility and accounted for this when interpreting results. The cost of specialist equipment was not included in the economic analysis as the GDG considered that those eligible to receive fampridine, people with MS and a walking disability (EDSS 4-7), are likely to have already received specialist equipment.

Cost of assessing responders

Although the cost of assessing responders was not included in this analysis, we considered what these costs would be to help inform GDG discussion and interpretation of the results of this analysis.

According to the BNF, patients started on fampridine have to be assessed for response to therapy after 2 weeks, after which treatment is continued or discontinued. A cost analysis undertaken by NHS

Regional Drug and Therapeutics Centre[169](NETAG) states that there is a scheme in place by the drug manufacturers to cover 28 days of treatment. This would cover the assessment period after which the NHS bears the treatment costs. In this scenario, in the first four weeks no drug cost would be incurred. However, even if this scheme was in place and drug costs would be null, patient assessments would need to be carried out in order to determine responders to treatment and this would generate costs to the NHS. The cost of a multi-professional neurology outpatient visit was estimated at £136. The total costs of initial treatment with fampridine are listed in **Error! Reference source not found.** for two scenarios: scenario A where the assessments undertaken are two over four weeks, and scenario B where the assessments undertaken are four over four weeks.

Table 60: Costs of initial four-week treatment with fampridine

Assessment period	Scenario A	Scenario B
Number of assessments	2	4[c]
Cost of assessments[a]	£272	£544
Drug costs[b]	-	-
Cumulative total costs	£272	£544

(a) Source: NETAG March,2012[169]
(b) For the first four weeks, the cost of drug would be covered by the drug manufacturers.
(c) Based on clinical study protocols

The results show that the total costs for the first four weeks vary between £272 and £544 depending on the number of assessments that are undertaken. These costs simply represent the costs of identifying responders to the initial fampridine treatment over the first two-four weeks.

Model validation

The model was developed in consultation with the GDG; inputs and results were presented to and discussed with the GDG for clinical validation and interpretation.

The model was systematically checked by the health economist undertaking the analysis; this included checking that results were plausible given inputs. The model was peer reviewed by a second experienced health economist from the NCGC; this included systematic checking of the model calculations.

Computations and estimation of cost effectiveness

The model was constructed in Microsoft Excel 2010 and allowed for the calculation of the incremental cost effectiveness ration (ICER). The ICER is calculated by dividing the difference in costs associated with two alternatives by the difference in QALYs. The decision rule then applied is that if the ICER falls below a given cost per QALY threshold the result is considered to be cost effective. If both costs are lower and QALYs are higher the option is said to dominate and an ICER is not calculated.

$$ICER = \frac{Costs(B) - Costs(A)}{QALYs(B) - QALYs(A)}$$

Where: Costs(A) = total costs for option A; QALYs(A) = total QALYs for option A

Cost-effective if:

- ICER < Threshold

Interpreting results

NICE's report 'Social value judgements: principles for the development of NICE guidance'[163] sets out the principles that GDGs should consider when judging whether an intervention offers good value for money. In general, an intervention was considered to be cost effective if either of the following criteria applied (given that the estimate was considered plausible):

- The intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- The intervention costs less than £20,000 per quality-adjusted life-year (QALY) gained compared with the next best strategy.

Results

The analysis was conducted deterministically and a threshold analysis was performed to determine the improvement in EQ-5D and MSWS-12 at which fampridine is considered cost-effective at a threshold of £20,000 per QALY.

Base case analysis

Fampridine was found to have an incremental cost effectiveness ratio of £160,884 per QALY gained in a population who responded to treatment (Table 61).

Table 61: Results of incremental deterministic analysis

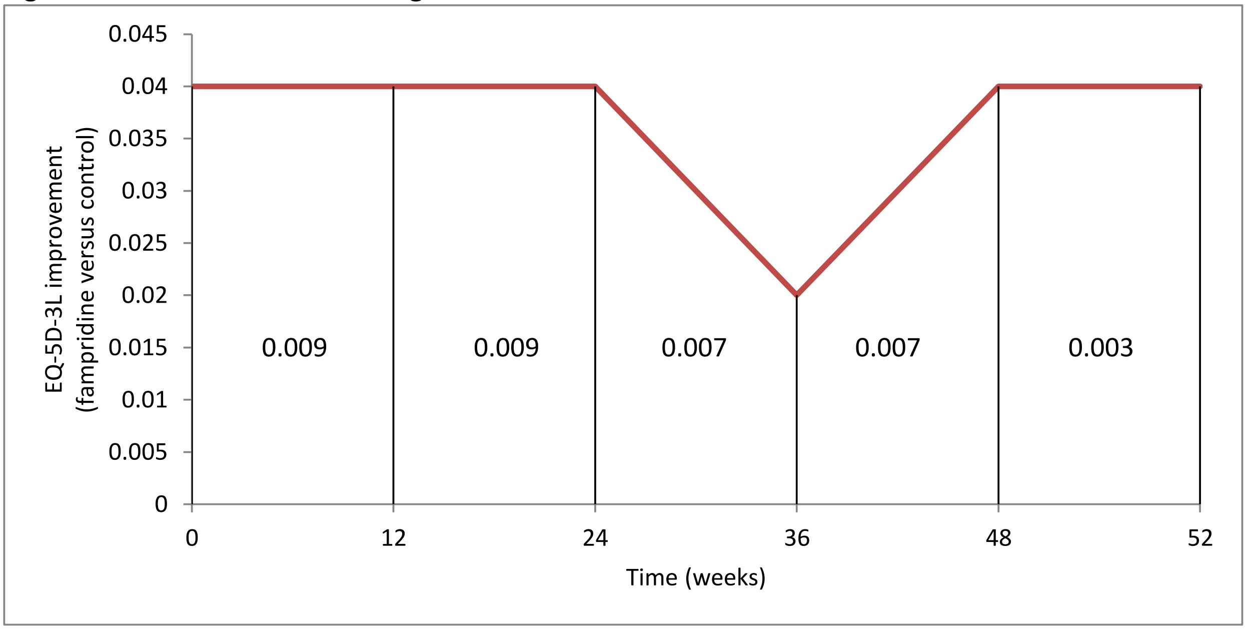
Strategy	Incremental cost	Incremental QALY	ICER (£ per QALY gained) (a)
Fampridine versus placebo	£4,719	0.029	£160,884

(a) ICER = incremental cost effectiveness ratio

Sensitivity analysis

Using the EQ-5D data from Macdonell et al (2003)[128], the EQ-5D improvement for fampridine versus control was ploted against time. This allowed for the area under the curve and therefore the incremental QALY for fampridine versus control to be calculated. The incremental QALY is the sum of each area in **Figure 4**divided by the number of weeks, which equals 0.035. Using this incremental QALY, fampridine was found to have an incremental cost effectiveness ratio of £133,361 per QALY gained in a population who responded to treatment (Table 62).

Figure 4: QALY calculations using area under the curve method



The numbers below the line are the incremental QALYs for that time period.

Table 62: Results of incremental sensitivity analysis

Strategy	Incremental cost	Incremental QALY	ICER (£ per QALY gained) (a)
Fampridine versus placebo	£4,719	0.035	£133,361

(a) ICER = incremental cost effectiveness ratio

Threshold analysis

The cost of treatment being constant, the change in incremental EQ-5D scores which is required for the ICER to decrease to £20,000 per QALY is 0.236. Assuming baseline MSWS-12 scores and MSWS-12 score at 9 weeks in the placebo group are unchanged, this corresponds to a decrease in the MSWS-12 score in the fampridine responders group by 52.11 (compared to the 6.04 reported in the study).

Given the magnitude of the QALY gained required for fampridine to be cost-effective relative to the QALY gained observed and the limited number of inputs in the model, it was deemed unnecessary to quantify uncertainty probabilistically.

Discussion

At the threshold of £20,000 per QALY, fampridine was not found to be cost-effective for improving mobility in people with multiple sclerosis who have responded to the initial trial with fampridine.

Based on these results, it was concluded that fampridine would be even less cost-effective for a group of patients who have not had the trial yet. This is because the effectiveness of the drug would be diluted in the broader group, which included non-responders as well, compared to the responders and the cost of the initial assessments would have to be added to the overall cost of the fampridine strategy. In the RCT on which we based the analysis[82,84], 57% of the individuals randomised to the fampridine group did not respond to treatment.

This analysis has some limitations: the base case relies on a single RCT with a limited number of participants and all the limitations of the clinical data also apply to the economic analysis. Utilities were estimated by mapping a condition-specific measure to a generic quality of life measure. This is associated with several limitations and uncertainty, as important domains could be lost in the mapping algorithm. As MSWS-12 only assesses mobility it may be that other treatment effects are not captured (mobility is one domain of EQ-5D, other are self-care, usual activities, pain/discomfort and anxiety and depression). Furthermore, the mapping function had not been validated. Of note, a sensitivity analysis was conducted using EQ-5D data reported directly from people receiving fampridine. This study had limitations as it was a non-randomised trial where fampridine non-responders were used as controls to fampridine responders. The incremental QALY gain using this direct data was greater then when using the mapped data, thus indicating that fampridine may have treatment effects other than improvements in mobility. Despite the greater QALY gain observed using the direct data, it was not sufficient to make fampridine cost-effective at £20,000 per QALY. Even if there had been evidence to suggest that fampridine delayed deterioration of mobility and therefore decreased healthcare utilisation, the GDG felt that it was unlikely that these downstream cost savings would offset the cost of fampridine.

Finally, fampridine was associated with a higher risk of adverse events compared to placebo; the possible impact of these on quality of life is not captured in the analysis. Incorporating this may make fampridine even less cost-effective compared to placebo.

9.2.5 Evidence statements

9.2.5.1 Clinical

Walking ability

Low quality evidence from three studies comprising 738 participants showed that fampridine was clinically effective compared to placebo in terms of a greater rate of positive response to treatment than placebo, with no imprecision.

Moderate quality evidence from one study comprising 8 participants showed that there was no difference in clinical effectiveness between fampridine and placebo in terms of time to walk 8m, with no imprecision.

Very low quality evidence from two studies comprising 245 participants showed that there was no difference in clinical effectiveness between fampridine and placebo in terms of gait speed, with no imprecision.

Very low quality evidence from two studies comprising 334 participants showed that there was no difference in clinical effectiveness between a low dose of fampridine and placebo in terms of MSWS-12 score, with no imprecision.

Low quality evidence from one study comprising 95 participants showed that there was no difference in clinical effectiveness between a medium dose of fampridine and placebo in terms of MSWS-12 score, with serious imprecision.

Low quality evidence from one study comprising 98 participants showed that there was no difference in clinical effectiveness between a high dose of fampridine and placebo in terms of MSWS-12 score, with serious imprecision.

Adverse events

Very low quality evidence from 6 studies comprising 1006 participants showed that fampridine was clinically harmful compared to placebo in terms of a higher rate of any adverse events, but with serious imprecision.

Very low quality evidence from 3 studies comprising 755 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of falls, but with very serious imprecision.

Very low quality evidence from 3 studies comprising 755 participants showed that fampridine was clinically harmful compared to placebo in terms of worse rate of UTIs, but with serious imprecision.

Very low quality evidence from 5 studies comprising 929 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of UTIs, with no imprecision.

Moderate quality evidence from 4 studies comprising 791 participants showed that fampridine was clinically harmful compared to placebo in terms of worse rate of UTIs, with no imprecision.

Moderate quality evidence from 2 studies comprising 516 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of fatigue, with serious imprecision.

Low quality evidence from 5 studies comprising 929 participants showed that fampridine was clinically harmful compared to placebo in terms of a worse rate of UTIs, with no imprecision.

Very low quality evidence from 3 studies comprising 755 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of URTIs, with very serious imprecision.

Very low quality evidence from 4 studies comprising 791 participants showed that fampridine was clinically harmful compared to placebo in terms of a worse rate of asthenia, with serious imprecision.

Very low quality evidence from 2 studies comprising 539 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of back pain, with serious imprecision.

Low quality evidence from 3 studies comprising 755 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of balance disorders, with no imprecision.

Very low quality evidence from 4 studies comprising 929 participants showed that fampridine was clinically harmful compared to placebo in terms of a worse rate of headaches, with serious imprecision.

Very low quality evidence from 2 studies comprising 455 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of arthralgia, with very serious imprecision.

Very low quality evidence from one study comprising 239 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of nasopharyngitis, with very serious imprecision.

Very low quality evidence from 4 studies comprising 629 participants showed that fampridine was clinically harmful compared to placebo in terms of a worse rate of paraesthesia, with serious imprecision.

Very low quality evidence from 1 study comprising 36 participants showed that fampridine was clinically harmful compared to placebo in terms of a lower rate of accidental injury, with very serious imprecision.

Very low quality evidence from 1 study comprising 36 participants showed that fampridine was clinically harmful compared to placebo in terms of a worse rate of tremor, with serious imprecision.

Very low quality evidence from 1 study comprising 216 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of oedema, with very serious imprecision.

Very low quality evidence from 1 study comprising 216 participants showed that there was no difference in clinical effectiveness between fampridine and placebo in terms of muscle spasm, with very serious imprecision.

Very low quality evidence from 1 study comprising 216 participants showed that fampridine was clinically harmful compared to placebo in terms of exacerbation, with serious imprecision.

Very low quality evidence from 5 studies comprising 913 participants showed that there was no difference in clinical harm between fampridine and placebo in terms of discontinuation due to adverse events, with serious imprecision.

9.2.5.2 Economic

One original cost–utility analysis found that fampridine was not cost effective compared to placebo for treating mobility problems in people with multiple sclerosis (ICER: £160,884 per QALY gained). This analysis was assessed as directly applicable with potentially serious limitations.

9.2.6 Recommendations and link to evidence

Recommendations	39. Do not use fampridine to treat lack of mobility in people with MS because it is not a cost effective treatment[w].
Relative values of different outcomes	Quality of life and walking speed were regarded as the critical outcomes. Important outcomes were motor function, spasticity, changes in disability or function scales and adverse events.
Trade off between clinical benefits and harms	Fampridine had beneficial effects on subjective improvement of walking speed and muscle strength. Although there were benefits for fampridine in terms of objectively measured walking speed these were too small to be considered clinically important by the GDG. Fampridine was associated with some adverse events, such as nausea, dizziness and insomnia, but these were not considered to be enough to outweigh any clinical benefits.
Economic considerations	The original cost–utility analysis undertaken for the guideline found that fampridine was not cost effective compared to placebo for treating mobility problems in people with MS who have had been categorised as responders to fampridine treatment following a four week trial. QALYs were estimated by mapping MSWS-12 data from the clinical review to EQ-5D utility (health-related quality of life). Fampridine cost £160,884 per QALY gained compared to placebo. In addition it was noted that fampridine would likely be even less cost effective when taking into consideration the need to establish who responds to treatment as that would mean including additional costs for the initial assessment but no additional patient benefits. Currently the manufacturer covers the drug costs of this trial but there will still likely be costs in terms of healthcare professional time. The GDG concluded that fampridine should not be offered based on the existing cost-effectiveness evidence.
Quality of evidence	Much of the evidence was graded LOW or VERY LOW. Blinding was unclearly reported by most studies, and four demonstrated incomplete outcome reporting. The two parallel studies both lacked evidence of allocation concealment (this was not regarded as an important source of bias for the cross-over studies). 3 studies had clear conflicts of interest, as they were funded by the manufacturers of fampridine. The economic evaluation was assessed as directly applicable with potentially serious limitations.
Other considerations	In two studies heavily biased responder analyses were presented, where a subgroup of patients responding to the drug were compared to the original placebo group, without any re-randomisation. The responder sub-group responded better than the placebo group but these outcomes were not considered by the GDG because of how the study was conducted.

9.3 Pharmacological management of oscillopsia

9.3.1 Introduction

Nystagmus is abnormal eye movement that is found on clinical examination. Patients may not be aware of this and may not have any symptoms related to it. Some people do notice an effect on their vision

w This recommendation does not apply to people who have already started treatment with fampridine in the NHS who should be able to continue treatment until they and their NHS clinician think it appropriate to stop.

and the name given to the symptom reported by a patient is oscillopsia. Patients' experience is that objects in their field of vision appear to move.

9.3.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of oscillopsia?

For full details see review protocol in Appendix C.

Table 63: PICO characteristics of review question

Population	• Adults with MS • Move from wholly MS population to anyone with acquired pendular nystagmus if <1 RCT for any comparison.
Intervention/s	• Gabapentin (brand names: Fanatrex, Gabarone, Neogab, Gralise, Neurontin, Nupentin) • Memantine (Ebixa) • Levetiracetam (Keppra) • Botulinum toxin • Baclofen • Clonazepam • Isoniazid • Valproate • Antimuscarinic agents (scopolamine, benztropine, trihexyphenidyl)
Comparison/s	Each other, Usual treatment or placebo

Outcomes	• Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale.
	• Patient-reported symptoms of nystagmus
	  ○ VAS
	  ○ Patient global satisfaction
	• Nystagmus rating scale
	• Nystagmus-related physiological measures (e.g., median eye speed, or distance visual acuity).
	• Adverse effects of treatment (drowsiness, unsteadiness and weight gain)

Study design	Systematic reviews, RCTs. Include cross-over and dosing studies.

9.3.3 Clinical evidence

Five studies[11,15,18,122,236] were included in the review. The study characteristics are summarised in Table 37.

Seven different comparisons were covered in this review. These were:

- Gabapentin versus vigabatrin[15]
- Memantine versus Gabapentin [236]
- Gabapentin versus baclofen [11]
- Trihexylphenidyl versus Tridihexylchloride[122]
- Scopolamine versus Benztropine [18]
- Scopolamine versus glycopyrrolate [18]
- Glycopyrrolate versus Benztropine [18]

The first two comparisons (Gabapentin versus vigabatrin and Gabapentin versus Memantine) were exclusively made on adults with multiple sclerosis who had pendular nystagmus. Because there were no other eligible studies exclusively using a multiple sclerosis population, all other comparisons were made on a mixed population of people with acquired adult pendular nystagmus: gabapentin versus baclofen contained 9/15 adults with MS, trihexyphenidyl versus tridihexylchloride contained 4/5 adults with MS and the comparisons between scopolamine, benztropine and glycopyrrolate contained 3/5 adults with MS.

All studies used a cross-over design. For cross-over study categorical data, the standard error (of the log RR) was calculated using the simplified Mantel Haenszel method for paired outcomes, when the number of subjects with an event in both interventions was known. Forest plots were generated in Review manager with the Generic Inverse Variance function. For some variables there were no subjects who had events in BOTH groups, thus making the Mantel Haenszel method for paired outcomes unsuitable, so a Peto odds ratio was calculated instead. Although this statistic assumed parallel and not paired groups, it was used on the basis that whilst this approach would tend to over-estimate CIs and thus artificially reduce study weighting, this would be a conservative effect.

Evidence from all comparisons are summarised in the clinical GRADE evidence profiles below (Table 65 to 77). See also the study selection flow chart in Appendix D, forest plots in Appendix I, study evidence tables in Appendix G and exclusion list in Appendix J.

Some outcomes were not appropriate for meta-analysis as only p values and directions of effect were reported. These have been reported in a separate narrative section in 0.

Summary of included studies

Table 64: Summary of studies included in the review

Study	Intervention *Daily dose and duration*	Comparator *Daily dose and duration*	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised/analysed	Analysis
Bandini 2001[15]	Gabapentin *300-1200mg* *21 days*	Vigabatrin *500-2000mg* *21 days*	EDSS mean 5.5; 2 RR and 3 Chronic progressive; disease duration	8/5	Cross-over *2 week wash-out*

Study	Intervention ***Daily dose and duration***	Comparator ***Daily dose and duration***	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised/analysed	Analysis
			10.5 years		
Starck 2010[236]	Gabapentin *200-1200mg* *7 days*	Memantine *10-60mg* *7 days*	EDSS mean 6.1; 2 primary progressive and 9 secondary progressive; disease duration 15.3 years	11/9	Cross over *5 days wash-out*
Averbuch-Heller 1997[11]	Gabapentin *300-900mg* *14 days*	Baclofen *10-30mg* *14 days*	No details given for MS patients. MS (9), CVA (3), cerebellar degeneration (1), hypoxic encephalopathy (1), idiopathic (1)	15/15	Cross-over *1-2 week wash-out*
Leigh 1991[122]	Trihexylphenidyl *5-20mg* *28 days*	Trihexylchloride *25-100mg* *28 days*	No details given for MS patients. Of 5 completing, MS (4), post-surgical hypoxia (1)	10/5	Cross-over *1-2 week wash-out*
Barton 1994[18]	Scopolamine *IV 0.4mg x 3 over 4+ days (unclear)* Benztropine *IV 2mg x 3 over 4+ days (unclear)*	Glycopyrrolate *IV 0.2mg x 3 over 4+ days (unclear)*	No details given for MS patients. MS (3), cerebellar degeneration (2), 1 unknown	5/5	Cross-over *wash-out not reported*

Table 65: Clinical evidence profile: gabapentin versus vigabatrin

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Gabapentin	vigabatrin	Relative (95% CI)	Absolute		
Health related quality of life												
No studies reported on this outcome												
Improved nystagmus amplitude straight vision right eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.099 (0.816)		RR 3.00(0.61-14.85)	-	VERY LOW	IMPORTANT
Improved nystagmus amplitude straight vision right eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.386 (0.866)		RR 4.00(0.73-21.83)	-	VERY LOW	IMPORTANT
Improved nystagmus amplitude eccentric vision right eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.099 (0.816)		RR 3.00(0.61-14.85)	-	VERY LOW	IMPORTANT
Improved nystagmus amplitude straight vision left eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.386 (0.866)		RR 4.00(0.73-21.83)	-	VERY LOW	IMPORTANT
Improved nystagmus frequency straight vision right eye												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Gabapentin	vigabatrin	Relative (95% CI)	Absolute		
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	2/5 (40%)	0/5 (0%)	OR 9.49 (0.5 to 179.46)	400 more per 1000 (from 50 less to 850 more)	VERY LOW	IMPORTANT
Improved nystagmus frequency straight vision left eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	2/5 (40%)	0/5 (0%)	OR 9.49 (0.5 to 179.46)	400 more per 1000 (from 50 less to 850 more)	VERY LOW	IMPORTANT
Improved nystagmus frequency eccentric vision right eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	2/5 (40%)	0/5 (0%)	OR 9.49 (0.5 to 179.46)	400 more per 1000 (from 50 less to 850 more)	VERY LOW	IMPORTANT
Improved nystagmus frequency eccentric vision left eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	2/5 (40%)	0/5 (0%)	OR 9.49 (0.5 to 179.46)	400 more per 1000 (from 50 less to 850 more)	VERY LOW	IMPORTANT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Gabapentin	vigabatrin	Relative (95% CI)	Absolute		
Improved visual acuity near straight vision right eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	2/5 (40%)	0/5 (0%)	OR 9.49 (0.5 to 179.46)	400 more per 1000 (from 50 less to 850 more)	VERY LOW	CRITICAL
Improved visual acuity near straight vision left eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Not estmable	none	0/5 (0%)	0/5 (0%)	not pooled	not pooled	Not estimable	CRITICAL
Improved visual acuity near eccentric vision left eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.386 (0.866)		RR 4.00(0.73-21.83)	-	VERY LOW	CRITICAL
Improved visual acuity near eccentric vision right eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	none	none	4/5 (80%)	0/5 (0%)	OR 20.09 (1.82 to 221.51)	800 more per 1000 (from 390 more to 1000 more)	LOW	CRITICAL
Improved visual acuity far straight vision right eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision	none	2/5 (40%)	0/5 (0%)	OR 9.49 (0.5 to 179.46)	400 more per 1000 (from 50 less to 850	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Gabapentin	vigabatrin	Relative (95% CI)	Absolute		
										more)		
Improved visual acuity far straight vision left eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	2/5 (40%)	0/5 (0%)	OR 9.49 (0.5 to 179.46)	400 more per 1000 (from 50 less to 850 more)	VERY LOW	CRITICAL
Improved visual acuity far eccentric vision right eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	serious imprecision[B]	none	3/5 (60%)	0/5 (0%)	OR 13.08 (1.01 to 170.31)	600 more per 1000 (from 150 more to 1000 more)	VERY LOW	CRITICAL
Improved visual acuity far eccentric vision left eye												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.386 (0.866)		RR 4.00(0.73-21.83)	-	VERY LOW	CRITICAL
Mild drowsiness												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	1/5 (20%)	0/5 (0%)	OR 7.39 (0.15 to 372.38)	200 more per 1000 (from 210 less to 610 more)	VERY LOW	IMPORTANT
Nausea												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Gabapentin	vigabatrin	Relative (95% CI)	Absolute		
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	1/5 (20%)	0/5 (0%)	OR 7.39 (0.15 to 372.38)	200 more per 1000 (from 210 less to 610 more)	VERY LOW	IMPORTANT
Subjective improvement in oscillopsia												
Bandini 2001	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (0.707)		RR 2.00(0.50-7.99)	-	VERY LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 66: Clinical evidence profile: memantine versus gabapentin

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	memantine	Gabapentin	Relative (95% CI)	Absolute		
Health related quality of life												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**memantine**	**Gabapentin**	**Relative (95% CI)**	**Absolute**		
No studies reported on this outcome												
Improved right eye horizontal amplitude												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	serious imprecision[B]	none	GIV: log[RR](SE) = 0.405 (0.289)		RR 1.5(0.85-2.64)	-	VERY LOW	IMPORTANT
Improved left eye horizontal amplitude												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	serious imprecision[B]	none	GIV: log[RR](SE)) = 0.182 (0.183)		RR 1.2(0.84-1.72)	-	VERY LOW	IMPORTANT
Improved right eye horizontal frequency												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (0.408)		RR 2.0(0.90-4.45)	-	VERY LOW	IMPORTANT
Improved left eye horizontal frequency												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	serious imprecision[B]	none	GIV: log[RR](SE) = 0.405 (0.289)		RR 1.5(0.85-2.64)	-	VERY LOW	IMPORTANT
Improved right eye vertical amplitude												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = -0.405 (0.408)		RR 0.67(0.30-1.48)	-	VERY LOW	IMPORTANT
Improved left eye vertical amplitude												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0 (0.283)		RR 1.0(0.57-1.74)	-	VERY LOW	IMPORTANT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**memantine**	**Gabapentin**	**Relative (95% CI)**	**Absolute**		
Improved right eye vertical frequency												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = -0.288 (0.289)		RR 0.75(0.43-1.32)	-	VERY LOW	IMPORTANT
Improved left eye vertical frequency												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	serious imprecision[B]	none	GIV: log[RR](SE) = 0.182 (0.183)		RR 1.2(0.84-1.72)	-	VERY LOW	IMPORTANT
Subjective improvement												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	serious imprecision[B]	none	GIV: log[RR](SE) = 0.405 (0.289)		RR 1.5(0.85-2.64)	-	VERY LOW	CRITICAL
fatigue												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	1/11 (9.1%)	4/11 (36.4%)	RR 0.25 (0.03 to 1.9)	273 fewer per 1000 (from 353 fewer to 327 more)	VERY LOW	IMPORTANT
dizziness												
Starck 2010	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	1/11 (9.1%)	3/11 (27.3%)	RR 0.33 (0.04 to 2.73)	183 fewer per 1000 (from 262 fewer to 472 more)	VERY LOW	IMPORTANT

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*

[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 67: Clinical evidence profile: gabapentin versus baclofen

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**gabapentin v baclofen**	**Relative (95% CI)**	**Absolute**		
Health related quality of life											
No studies reported on this outcome											
Patient desire to continue medication											
Averbuch-Heller 1997	RCT – cross-over	serious risk of bias[A]	none	none	none	none	GIV: log[RR](SE) = 1.705 (0.707)	RR 5.5(1.38-21.99)	-	MOD	IMPORTANT
Adverse events leading to withdrawal											
Averbuch-Heller 1997	RCT – cross-over	serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = -0.693 (0.707)	RR 0.5(0.13-2.00)	-	VERY LOW	IMPORTANT
Ataxia or worsened balance											
Averbuch-Heller 1997	RCT – cross-over	serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.099 (0.816)	RR 3.0(0.61-14.85)	-	VERY LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.

[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 68: Clinical evidence profile: tridihexylchloride versus trihexylphenidyl

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	tridihexylchloride	Trihexylphenidyl	Relative (95% CI)	Absolute		
Health related quality of life												
No studies reported on this outcome												
visual acuity improved												
Leigh 1991	RCT – cross-over	Very serious risk of bias[A]	none	none	none	none	4/5 (80%)	0/5 (0%)	OR 20.09 (1.82 to 221.51)	800 more per 1000 (from 390 more to 1000 more)	LOW	CRITICAL
improvement in slow phase velocity in primary eye position												
1Leigh 1991	RCT – cross-over	Very serious risk of bias[A]	none	none	serious imprecision[B]	none	3/5 (60%)	0/5 (0%)	OR 13.08 (1.01 to 170.31)	600 more per 1000 (from 150 more to 1000 more)	VERY LOW	IMPORTANT
improvement in slow phase velocity in non-primary eye position												
Leigh 1991	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.099 (0.816)		RR 3 (0.61 to 14.85)	-	VERY LOW	IMPORTANT
improvement in frequency												
Leigh 1991	RCT – cross-over	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	2/5 (40%)	0/5 (0%)	OR 9.49 (0.5 to 179.46)	400 more per 1000 (from 50 less to 850 more)	VERY LOW	IMPORTANT
unable to tolerate												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**tridihexylchloride**	**Trihexylphenidyl**	**Relative (95% CI)**	**Absolute**		
Leigh 1991	no methodology chosen	Very serious risk of bias[A]	none	none	Very serious imprecision[B]	none	2/5 (40%)	1/5 (20%)	OR 2.36 (0.18 to 30.67)	171 more per 1000 (from 157 fewer to 685 more)	VERY LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 69: Clinical evidence profile: scopolamine versus benzotropine

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Scopolamine v benzotropine**	**Relative (95% CI)**	**Absolute**		
Health related quality of life											
No studies reported on this outcome											
Improved frequency											
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = -0.693 (0.5)	RR 0.5 (0.19 to 1.33)	-	VERY LOW	IMPORTANT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Scopolamine v benzotropine		Relative (95% CI)	Absolute		
Improved acuity												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = -0.288 (0.5)		RR 0.75 (0.28 to 2.0)	-	VERY LOW	CRITICAL
Improved mean nystagmus velocity												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	5/5	5/5	RR 1(0.71 to 1.41)	0 more (from 290 fewer to 410 more)	VERY LOW	CRITICAL
Improved amplitude												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	serious imprecision[B]	none	GIV: log[RR](SE) = 1.609 (0.894)		RR 5 (0.87 to 28.82)	-	LOW	IMPORTANT
Dizziness												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = -0.693 (0.707)		RR 0.5 (0.13 to 2.0)	-	VERY LOW	IMPORTANT
Drowsiness												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (0.707)		RR 2.0 (0.50 to 7.99)	-	VERY LOW	IMPORTANT

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Scopolamine v benzotropine	Relative (95% CI)	Absolute		
Poor balance											
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (1.225)	RR 2.0 (0.18 to 22.06)	-	VERY LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 70: Clinical evidence profile: scopolamine versus glycopyrrolate

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Scopolamine v glycopyrrolate	Relative (95% CI)	Absolute		
Health related quality of life											
No studies reported on this outcome											
Improved frequency											
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (1.225)	RR 2.0 (0.18 to 22.06)	-	VERY LOW	IMPORTANT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Scopolamine v glycopyrrolate		Relative (95% CI)	Absolute		
Improved acuity												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0 (0.707)		RR 1.0 (0.25 to 4.0)	-	VERY LOW	CRITICAL
Improved mean nystagmus velocity												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	serious imprecision[B]	none	5/5	3/5	RR 1.57(0.77 to 3.22)	342 more (from 138 fewer to 1000 more)	LOW	CRITICAL
Improved amplitude												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	none	none	5/5 (100%)	0/5 (0%)	OR 36.6 (3.48 to 384.51)	1000 more per 1000 (from 690 more to 1000 more)	MOD	IMPORTANT
Dizziness												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Not estimable	none	GIV: log[RR](SE) = 0 (0)		Not estimable	-	Not estimable	IMPORTANT
Drowsiness												

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Scopolamine v glycopyrrolate	Relative (95% CI)	Absolute		
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (1.225)	RR 2.0 (0.18 to 22.06)	-	VERY LOW	IMPORTANT
Poor balance											
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (0.707)	RR 2.0 (0.50 to 7.99)	-	VERY LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 71: Clinical evidence profile: benzotropine versus glycopyrrolate

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	benzotropine v glycopyrrolate	Relative (95% CI)	Absolute		
Health related quality of life											
No studies reported on this outcome											

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	benzotropine v glycopyrrolate		Relative (95% CI)	Absolute		
Improved frequency												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 1.099 (0.816)		RR 3.0 (0.61 to 14.85)	-	VERY LOW	IMPORTANT
Improved acuity												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (0.707)		RR 2.0 (0.50 to 7.99)	-	VERY LOW	CRITICAL
Improved mean nystagmus velocity												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	serious imprecision[B]	none	5/5	3/5	RR 1.57(0.77 to 3.22)	342 more (from 138 fewer to 1000 more)	LOW	CRITICAL
Improved amplitude												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	1/5 (20%)	0/5 (0%)	OR 7.39 (0.15 to 372.38)	200 more per 1000 (from 210 less to 610 more)	LOW	IMPORTANT
Dizziness												
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (0.707)		RR 2.0 (0.50 to 7.99)	-	VERY LOW	IMPORTANT

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	benzotropine v glycopyrrolate	Relative (95% CI)	Absolute		
Drowsiness											
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0.693 (0.707)	RR 2.0 (0.50 to 7.99)	-	VERY LOW	IMPORTANT
Poor balance											
Barton 1994	RCT – cross-over	Serious risk of bias[A]	none	none	Very serious imprecision[B]	none	GIV: log[RR](SE) = 0 (1.414)	RR 1.0 (0.06 to 15.98)	-	VERY LOW	IMPORTANT

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*
[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Narrative review for outcomes not appropriate for meta-analysis

One comparison had outcome data that were not appropriate for meta-analysis, and so these are described narratively as follows.

Gabapentin v baclofen

Table 72: Gabapentin v baclofen

	Gabapentin v baclofen
Visual acuity – near	Gabapentin showed a significantly greater improvement than baclofen over treatment (p<0.001 for within gabapentin, NS for within baclofen)
Visual acuity - far	Gabapentin showed a significantly greater improvement than baclofen over treatment (p<0.006 for within gabapentin, NS for within baclofen)
Visual acuity	12 patients reported some illusory motion of the visual target before treatment and gabapentin reduced this in 6 patients. Baclofen did not have any effect on visual acuity at near or far.
Median eye speed	Median eye speed was reduced in all 3 planes by gabapentin (FAR: p<0.001 for horizontal and vertical, p<0.005 for torsional; NEAR: p<0.005 horizontal and torsional, p<0.005 vertical), during viewing of the near or far targets. The predominant frequency of oscillation was reduced by <9% by gabapentin (p<0.05). Median eye speed was reduced significantly (p<0.005) only in the vertical plane by baclofen. Baclofen caused no changes in the predominant frequency of oscillation.

9.3.4 Economic evidence

Published literature

No relevant economic evaluations comparing pharmacological treatments of oscillopsia were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided Appendix M to aid consideration of cost effectiveness.

9.3.5 Evidence statements

9.3.5.1 Clinical

Gabapentin versus vigabatrin

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus amplitude (straight vision in the right eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus amplitude (straight vision in the left eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus amplitude (eccentric vision in the right eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus amplitude (eccentric vision in the left eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus frequency (straight vision in the right eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus frequency (straight vision in the left eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus frequency (eccentric vision in the right eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus frequency (eccentric vision in the left eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus far acuity (straight vision in the right eye), with serious to very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus far acuity (straight vision in the left eye), with serious to very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus far acuity (eccentric vision in the right eye), with serious to very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus far acuity (eccentric vision in the left eye), with serious to very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus near acuity (straight vision in the right eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to vigabatrin in terms of improved nystagmus near acuity (eccentric vision in the right eye), with no imprecision.

There was no estimable evidence for the comparison between gabapentin and vigabatrin in terms of improved nystagmus near acuity (straight vision in the left eye).

Very low quality evidence from one cross-over study comprising 5 participants showed that there was no difference in clinical effectiveness between gabapentin and vigabatrin in terms of improved nystagmus near acuity (eccentric vision in the left eye), with very serious imprecision.

Gabapentin versus memantine

Very low quality evidence from one cross-over study comprising 11 participants showed that memantine was clinically effective compared to gabapentin in terms of improved nystagmus amplitude (horizontal in the right eye), with serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that there was no difference in clinical effectiveness between memantine and gabapentin in terms of improved nystagmus amplitude (horizontal in the left eye), with serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that memantine was clinically effective compared to gabapentin in terms of improved nystagmus frequency (horizontal in the right eye), with serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that memantine was clinically effective compared to gabapentin in terms of improved nystagmus frequency (horizontal in the left eye), with serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that memantine was clinically harmful compared to gabapentin in terms of improved nystagmus amplitude (vertical in the right eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that there was no difference in clinical effectiveness between memantine and gabapentin in terms of improved nystagmus amplitude (vertical in the left eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that memantine was clinically harmful compared to gabapentin in terms of improved nystagmus frequency (vertical in the right eye), with very serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that there was no difference in clinical effectiveness between memantine and gabapentin in terms of improved nystagmus frequency (vertical in the left eye), with serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that memantine was clinically effective compared to gabapentin in terms of subjective improvement, with serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that memantine was clinically effective compared to gabapentin in terms of fatigue, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 11 participants showed that memantine was clinically effective compared to gabapentin in terms of dizziness, with very serious imprecision.

Gabapentin versus baclofen

Moderate quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to baclofen in terms of patient desire to continue medication, with no imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically effective compared to baclofen in terms of adverse events leading to withdrawal, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that gabapentin was clinically harmful compared to baclofen in terms of ataxia or worsened balance, with very serious imprecision.

Trihexylchloride versus trihexylphenidyl

Low quality evidence from one cross-over study comprising 5 participants showed that trihexylchloride was clinically effective compared to trihexylphenidyl in terms of improved visual acuity, with no imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that trihexylchloride was clinically effective compared to trihexylphenidyl in terms of improvement in slow phase velocity in primary eye position, with serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that trihexylchloride was clinically harmful compared to trihexylphenidyl in terms of improvement in slow phase velocity in non-primary eye position, with very serious imprecision.

Scopolamine versus benzotropine

Very low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically harmful compared to benzotropine in terms of improved frequency, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that there was no difference between scopolamine and benzotropine in terms of improved velocity, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically harmful compared to benzotropine in terms of improved acuity, with very serious imprecision.

Low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically effective compared to benzotropine in terms of improved amplitude, with serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically effective compared to benzotropine in terms of dizziness, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically harmful compared to benzotropine in terms of drowsiness, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically harmful compared to benzotropine in terms of balance, with very serious imprecision.

Scopolamine versus glycopyrrolate

Very low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically effective compared to glycopyrrolate in terms of improved frequency, with very serious imprecision.

Low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically effective compared to glycopyrrolate in terms of improved velocity, with serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed no difference between scopolamine and glycopyrrolate in terms of improved acuity, with very serious imprecision.

Moderate quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically effective compared to glycopyrrolate in terms of improved amplitude, with no imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically harmful compared to glycopyrrolate in terms of drowsiness, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that scopolamine was clinically harmful compared to benzotropine in terms of balance, with very serious imprecision.

Benzotropine versus glycopyrrolate

Very low quality evidence from one cross-over study comprising 5 participants showed that benzotropine was clinically effective compared to glycopyrrolate in terms of improved frequency, with very serious imprecision.

Low quality evidence from one cross-over study comprising 5 participants showed that benzotropine was clinically effective compared to glycopyrrolate in terms of improved velocity, with serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that benzotropine was clinically effective compared to glycopyrrolate in terms of improved acuity, with very serious imprecision.

Low quality evidence from one cross-over study comprising 5 participants showed that benzotropine was clinically effective compared to glycopyrrolate in terms of improved amplitude, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that benzotropine was clinically harmful compared to glycopyrrolate in terms of dizziness, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that benzotropine was clinically harmful compared to glycopyrrolate in terms of drowsiness, with very serious imprecision.

Very low quality evidence from one cross-over study comprising 5 participants showed that there was no difference between benzotropine and glycopyrrolate in terms of balance, with very serious imprecision.

9.3.5.2 Economic

No relevant economic evaluations were identified.

9.3.6 Recommendations and link to evidence

Recommendations	**40. Consider gabapentin[x] as a first line drug to treat oscillopsia in people with MS.** **41. Consider memantine[y] as the second-line treatment for oscillopsia in people with MS.** **42. Refer the person with MS for specialist advice if there is no improvement of oscillopsia after treatment with gabapentin and memantine or side effects prevent continued use.**
Relative values of different outcomes	Improved visual acuity and reduction in slow phase velocity were judged to be the most important objective outcomes from treatment of oscillopsia. Subjective improvement was equally important. Visual acuity on straight ahead gaze was more important than visual acuity on eccentric gaze. Less important outcomes were frequency and amplitude of nystagmus.
Trade off between clinical benefits and harms	The GDG understood from an expert neuro-ophthalmologist that the effects of oscillopsia can range from bothersome to critical. For example, severe oscillopsia can lead to deterioration of balance requiring a wheelchair for mobility and/or remove the ability to read or watch television. It tends to occur in severe or progressive disease. Memantine appeared to be slightly more effective than gabapentin, and both were more effective than vigabatrin or baclofen. Trihexylchloride was more effective than trihexylphenidate. Hyoscine and benztropine were more effective than glycopyrrolate. Drug treatments used for oscillopsia can have significant adverse effects. Studies suggested that gabapentin causes drowsiness, nausea, fatigue and dizziness. Memantine has been reported to cause reversible neurological deterioration in multiple sclerosis. Expert opinion was that gabapentin may impair balance and botulinum toxin injections can increase disability, by requiring occlusion of one eye to overcome double vision and by impairing vestibulo-ocular reflexes.
Economic considerations	No relevant economic evaluations were identified. A threshold analysis was conducted for gabapentin and memantine. These two drugs had been

x At the time of publication (October 2014), gabapentin did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

y At the time of publication (October 2014), memantine did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

	identified in the clinical review as resulting in higher visual acuity in people with MS and oscillopsia. The aim of this analysis was to identify what gain in quality of life (utility) would be required for these drugs to be deemed cost effective compared to no treatment according to the NICE threshold (£20,000 per QALY). The annual cost of the drugs varied depending on the prescribed dose and was between £28–415 for gabapentin and £450–2,699 for memantine. The results of the threshold analysis indicated that a sustained utility gain of 0.001–0.021 was required for gabapentin to be cost effective and of 0.022–0.135 for memantine to be cost effective. The GDG considered the unit costs and the results of this threshold analysis and felt that the required improvement in quality of life for gabapentin was achievable and therefore it should be considered as a first-line agent. The GDG felt that as the required improvement in quality of life for memantine was greater it should only be considered as a second-line agent.
Quality of evidence	All five trials included were small crossover interventional trials, with the largest study including 15 participants. A significant limitation was that they were not placebo controlled. This made it difficult to assess efficacy, as relative benefit for the experimental drug over the comparator drug could simply be due to the comparator drug causing an actual worsening of the condition. In this situation, the experimental drug might have little or no efficacy. For example, it was unclear if vigabatrin worsened visual acuity or if gabapentin had improved acuity in the study by Bandini et al, 2001.
Other considerations	The GDG developed the recommendations using the evidence and the advice of a co-opted expert. The experience of the GDG was that people who experience oscillopsia are quite distressed and functionally limited by the symptom. While the research evidence base in oscillopsia is poor, neurologists have experience of using gabapentin and memantine for other conditions and a trial of treatment is appropriate for this condition. If a patient was already taking gabapentin for other indications, then it was thought to be reasonable to increase the dose as a first-line measure to control oscillopsia. People who do not respond to these two drugs should be referred to a specialist, such as a neuro-ophthalmologist according to local availability.

9.4 Pharmacological treatment and management of emotional lability

9.4.1 Introduction

Emotional lability can be a very distressing symptom for a minority of people with MS. It is also known as pseudobulbar affect (PBA) and people with PBA may laugh or cry without any apparent trigger. Laughing or crying once it starts cannot easily be controlled and may occur at inappropriate times.

9.4.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological management of emotionalism?

For full details see review protocol in Appendix C.

Table 73: PICO characteristics of review question

Population	• Adults with MS
Intervention/s	Antidepressants, such as medications in the following classes: • SSRIs • SNRIs • NaSSAs • NRIs • Tricyclics • MAOIs • Antiepileptics • Atypical antipsychotics
Comparison/s	Usual treatment or placebo
Outcomes	• Measure of emotionalism • Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale. • Psychological symptoms assessed by validated and disease-specific scales, questionnaire or similar instruments. • Impact on carers. • Cognitive functions, such as memory and concentration, and physical symptoms including fatigue, spasticity, spasms, assessed by validated and disease-specific scales, questionnaires or similar instruments, for instance the Scripps Neurologic Rating scale (SNRS) or the Krupp Fatigue Severity Scale (FSS). • Adverse effects of treatment- sedation, fatigue, dizziness or mood disturbance.
	• Systematic reviews, RCTs. Include cross-over and dosing studies.

9.4.3 Clinical evidence

Summary of included studies

1 RCTs was found, which[214] evaluated the effects of amitryptiline compared to placebo.

Table 74: Summary of study included in the review

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised /analysed	Analysis
Schiffer 1985[214,214]	Amitryptiline 30 days (no dose information provided) versus placebo	All MS; MS symptoms 3-25 years; lability 1-40 months; 12 completers all with emotional lability (2 laughing, 2 mixed and 8 weeping)	17/12	Cross-over

Table 75: Clinical evidence profile: Amitryptiline versus placebo

Quality assessment							Log[Risk Ratio](SE)	Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Amitryptiline versus placebo**	**Relative (95% CI)**	**Absolute**		
Quality of life											
No outcomes available											
Psychological symptoms											
No outcomes available											
Impact on carers											
No outcomes available											
Cognitive function											
No outcomes available											
Numbers with reduction in episodes of laughing at 30 days compared to baseline											
Schiffer 1985	randomised trials	Serious[A]	no serious inconsistency	no serious indirectness	Serious[B]	None	0.981(0.54)	2.67(0.93 to 7.69)	-	LOW	CRITICAL

[A] *The outcome was downgraded by one increment because the study had attrition bias.*
[B] *The outcome was downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variabl*

9.4.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided in Appendix M.

9.4.5 Evidence statements

9.4.5.1 Clinical

Amitryptiline versus placebo

Low quality evidence from one cross-over RCT comprising 12 participants showed that amitryptiline had a clinically important benefit in relation to placebo in terms of the proportion of people with a reduction in episodes of laughing from baseline to 30 days, with serious imprecision.

9.4.5.2 Economic

No relevant economic evaluations were identified.

9.4.6 Recommendations and link to evidence

Recommendations	**43. Consider amitriptyline[z] to treat emotional lability[aa] in people with MS.**
Relative values of different outcomes	Measures of emotionalism were regarded as the most critical outcome as they were the most directly relevant to the review question. Quality of life, psychological symptoms, impact on carers, cognitive function and adverse events were also regarded as critical outcomes.
Trade off between clinical benefits and harms	Amitryptiline had a clinically important benefit for emotionalism, and no harms were reported by the available evidence.
Economic considerations	No relevant economic evaluations were identified. The costs of pharmacological treatments used for emotionalism were presented. Amitriptyline and dextromethorphan/quinidine both showed clinical efficacy compared to placebo. The annual cost of amitriptyline was £119 based on a

z At the time of publication (October 2014), amitriptyline did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

aa Involuntary laughing and crying related to a brain stem lesion.

	25mg daily dose. A threshold analysis was conducted for amitriptyline. The aim of this analysis was to identify what gain in quality of life (utility) would be required for this drug to be deemed cost effective compared to no treatment according to the NICE threshold (£20,000 per QALY). The results of the threshold analysis indicated that a sustained utility gain of 0.006 is required for amitriptyline to be cost effective. The GDG reported that amitriptyline is the current standard of care for emotionalism and considered the cost was low compared to the likely benefits in terms of quality of life. Therefore, the GDG agreed that the use of amitriptyline for the management of emotionalism in people with MS is likely to be cost-effective.
Quality of evidence	There was very little RCT evidence available. The single available study was at very serious risk of bias due to attrition, and was based on a very small sample size with no data for adverse events.
Other considerations	The GDG discussed whether it was appropriate to make a recommendation for the treatment of emotionalism. The GDG considered that emotionalism needs to be distinguished from mood disorder which should be treated appropriately. Emotionalism can be seriously debilitating problem for those who suffer from it with a significant functional impact. The professional members of the GDG considered that there is experience of the use of amitriptyline for emotionalism and that a trial of this is worthwhile for those affected. Amitriptyline is a drug commonly used for a number of different medical problems and therefore neurologists and other healthcare professionals have a lot of experience in using it and of the adverse effects that can occur.

9.5 Pharmacological management of ataxia and tremor

9.5.1 Introduction

MS can cause ataxia and tremor which can be disabling. The prevalence is unclear with some reports suggesting that tremor can occur in up to 80% of people with MS at some stage of their disease.

9.5.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of ataxia and tremor?

For full details see review protocol in Appendix C.

Table 76: PICO characteristics of review question

Population	Adults with MS with ataxia and/or tremor
Intervention/s	• Baclofen (oral/intrathecal) • Isoniazid • Carbamazepine (plus other antiepileptics)Propranolol (all beta blockers), • Clonazepam (all benzodiazepines), • Primidone, • Ondansetron • fampridine, • Botox
Comparison/s	Usual treatment (including exercise, deep brain stimulation etc) or placebo, or drugs above
Outcomes	• Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale. • Patient-reported outcomes, for example symptoms of ataxia and tremor. • Impact on carers. • Ataxia measurement scales – ie International Cooperative Ataxia Rating Scale (ICARS) • Tremor rating scales – ie TRS, Fahn • Functional scales that quantify level of disability, such as the Expanded Disability Status Scale (EDSS), the Multiple Sclerosis Functional Composite (MSFC), the Cambridge Multiple Sclerosis Basic Score (CAMBS), the Functional Assessment of Multiple Sclerosis (FAMS) or the National Fatigue Index (NFI). + mobility scales • Adverse effects of treatment.
Review strategy	Systematic reviews, RCTs. Include cross-over studies.

9.5.3 Clinical evidence

Summary of included studies

4 cross-over RCTs[26,88,176,253] were found. Two[26,88] compared isoniazid and placebo for the treatment of ataxia/tremor in people with MS, one [176] compared baclofen to placebo and one[253] compared botulinum toxin to placebo. Their characteristics are shown in Table 77 below. Results from these studies are outlined in the GRADE evidence profiles (Table 78 to Table 80).

Table 77: Summary of studies included in the review

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised /analysed	Analysis
Bozek 1987[26]	Isoniazid versus placebo	Clinically definite MS (stable or chronically progressive); mean age 36 years; duration tremor 5 (5) years; all had postural tremor	10/8	Cross-over
Hallett 1985[88]	Isoniazid versus placebo	Advanced MS; severe postural cerebellar tremor; age 31-51	7/6	Cross-over
Orsnes 2000[176]	Baclofen versus placebo	Clinically definite MS with spasticity; age 24-57; EDSS: 3.5-6	14/13	Cross-over
Van Der Walt 2012A[253]	Botulinum versus placebo	RR and SP MS; disabling arm tremor; mean age 49.6 years; duration tremor 6.5(5.1) years; EDSS median (IQR):5.5(4-6.5)	33 limbs/variable from 22-33	Cross-over

Table 78: Clinical evidence profile: Isoniazid versus placebo

Quality assessment							No of patients with event (%) OR Mean(sd)[n] OR LnRR (SE)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Isoniazid	Placebo	Relative (95% CI)	Absolute		
Quality of life												
No data available												
Impact on carers												
No data available												
Functional scales												
No data available												
Ataxia rating scales												
No data available												
Mean tremogram acceleration measures (10^{-1} g) at 4 weeks (Better indicated by lower values)												
Bozek 1987	RCT cross-over	Very serious risk of bias[B]	No serious inconsistency	No serious indirectness	Very serious imprecision[C]	none	5(3)[8]	6.7(4.1)[8]	-	MD 1.7 lower (5.62 lower to 2.22 higher)	VERY LOW	CRITICAL
Improvements from baseline in tremor according to tremogram acceleration measures at 4 weeks												
Bozek 1987	RCT cross-over	Very serious risk of bias[B]	No serious inconsistency	No serious indirectness	Very serious imprecision[C]	none	0.511(0.365)	1.67(0.82 to 3.41)	-		VERY LOW	CRITICAL
Patient subjective improvements in tremor at 4 weeks												
Hallett 1985	RCT cross-over	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[C]	none	0.916(0.707)	2.50(0.63 to 9.99)	-		VERY LOW	CRITICAL

Quality assessment							No of patients with event (%) OR Mean(sd)[n] OR LnRR (SE)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Isoniazid	Placebo	Relative (95% CI)	Absolute		
Adverse events - Nausea												
Hallett 1985	RCT cross-over	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Serious imprecision[C]	none	0/8 (0%)	1/8 (12.5%)	Peto OR 0.14 (0 to 6.82)	130 fewer per 1000 (from 410 fewer to 160 more)	VERY LOW	CRITICAL

[A] Outcomes were downgraded by one increment because health care professional blinding was unclear. Some attrition but this was probably at random, so this did not count towards a further downgrade.
[B] Outcomes were downgraded by two increments. Health care professional and patient blinding was unclear, leading to one incremental downgrade. Overall 20% attrition, which was not at random but appeared to be related to outcome, may have also been a source of some bias. The differential attrition was unclear as the phase in which one patient was experiencing lack of efficacy was not reported.
[C] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 79: Clinical evidence profile: Botulinum versus placebo

Quality assessment							Median (IQR) change from baseline)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Botulinum	Placebo	P value for Wilcoxon signed ranks	Absolute		
Quality of life QUEST score at 3 months (higher worse)												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0(-4 to 6)	-4(-12 to 1)	0.1136	NA	LOW	CRITICAL

Quality assessment							Median (IQR) change from baseline)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Botulinum	Placebo	P value for Wilcoxon signed ranks	Absolute		
Impact on carers												
No data available												
Bain composite tremor score (0-10) at 12 weeks (higher worse)												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-2(-2 to -1)	0(-1 to 1)	0.0001	NA	MODERATE	CRITICAL
Bain writing score (0-10_ at 12 weeks (higher worse) N=22												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-1(-1 to 0)	0(0 to 0)	0.002	NA	MODERATE	CRITICAL
Bain Archimedes spiral (0-10) at 12 weeks (higher worse) N=22												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-1(-2 to 0)	0(0 to 1)	0.0007	NA	MODERATE	CRITICAL
CRST writing (0-4) at 12 weeks (higher worse) N=22												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0(-1 to 0)	0(0 to 0)	0.197	NA	LOW	CRITICAL
CRST drawing (0-4) at 12 weeks (higher worse) N=22												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-0.5(-1 to 0)	0(0 to 0)	0.024	NA	MODERATE	CRITICAL
ICARS Archimedes spiral (0-4) at 12 weeks. . Higher worse. N=22												

Quality assessment							Median (IQR) change from baseline)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Botulinum	Placebo	P value for Wilcoxon signed ranks	Absolute		
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0(-1 to 0)	0(0 to 0)	0.3351	NA	LOW	CRITICAL
CRST pouring (0-4) at 12 weeks. Higher worse. N=29												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0(-1 to 0)	0(0 to 0)	0.0628	NA	LOW	CRITICAL
Drinking from cup (0-4) at 12 weeks. Higher worse. N=29												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	0(-1 to 0)	0(0 to 0)	0.0089	NA	MODERATE	CRITICAL
9 hole peg test at 12 weeks. N=28												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-4.5(-14 to -1)	0(-6 to 4)	0.0195	NA	MODERATE	CRITICAL
Kinetic tremor severity(0-4) at 12 weeks. Higher worse.												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-1(-1 to 0)	0(0 to 1)	<0.0001	NA	MODERATE	CRITICAL
CRST action tremor arm (0-4) at 12 weeks. Higher worse												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	0(-1 to 0)	0(0 to 0)	0.021	NA	MODERATE	CRITICAL
CRST action tremor amplitude (cm) at 12 weeks. Higher worse												

Quality assessment							Median (IQR) change from baseline)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Botulinum	Placebo	P value for Wilcoxon signed ranks	Absolute		
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-1(-2 to 0)	0(0 to 0.5)	0.0012	NA	MODERATE	CRITICAL
ICARS finger-finger test at 12 weeks. Higher worse.												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0(-1 to 0)	0(0 to 0)	0.4274	NA	LOW	CRITICAL
Postural tremor severity (0-4) at 12 weeks. Higher worse.												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-1(-1 to 0)	0(0 to 0)	0.0161	NA	MODERATE	CRITICAL
Batwing position tremor (0-4) at 12 weeks. Higher worse.												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-1(-1 to 0)	0(0 to 0)	0.0268	NA	MODERATE	CRITICAL
CRST postural tremor arm (0-4) at 12 weeks. Higher worse												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	0(-1 to 0)	0(0 to 0)	0.0076	NA	MODERATE	CRITICAL
CRST postural tremor amplitude (cm) at 12 weeks. . Higher worse												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-0.5(-5 to 0)	0(0 to 0.5)	0.0077	NA	MODERATE	CRITICAL
Ataxia rating scores - SARA score change from baseline at 12 weeks. Higher worse												

Quality assessment							Median (IQR) change from baseline)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Botulinum	Placebo	P value for Wilcoxon signed ranks	Absolute		
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-2(-3 to 0)	0.5(1.5 to 2)	0.089	NA	LOW	CRITICAL
Adverse events - Muscle weakness												
Van Der Walt 2012A	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	No serious imprecision[C]	none	14/33 (42.4%)	2/33 (6.1%)	RR: 7 (1.72 to 28.41)	364 more per 1000 (from 44 more to 1000 more)	MODERATE	CRITICAL

[A] The outcome was downgraded by one increment because the study had attrition bias.
[B] Because of the lack of confidence intervals or absolute effect sizes, imprecision was based on the Wilcoxon signed ranks test. If p≤0.05 it was rated as precise and if p>0.05 as seriously imprecise.
[C] This outcome was rated as precise because neither confidence interval crossed either of the default MIDs (0.75 and 1.25)

Table 80: Clinical evidence profile: Baclofen versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Baclofen versus placebo	Control	Relative (95% CI)	Absolute		
Quality of life												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Baclofen versus placebo	Control	Relative (95% CI)	Absolute		
No data available												
Impact on carers												
No data available												
Functional scales												
No data available												
Ataxia or tremor rating scales												
No data available												
vertical unsteadiness improved at 18 days												
Orsnes 2000	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	Serious indirectness[B]	Serious imprecision[C]	none	10/13 (76.9%)	5/13 (38.5%)	RR 2 (0.95 to 4.23)	385 more per 1000 (from 19 fewer to 1000 more)	VERY LOW	CRITICAL
Adverse events - fatigue												
Orsnes 2000	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[C]	none	5/13 (38.5%)	1/13 (7.7%)	RR 5 (0.67 to 37.12)	308 more per 1000 (from 25 fewer to 1000 more)	VERY LOW	CRITICAL
Adverse events - dizziness												
Orsnes 2000	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[C]	none	3/13 (23.1%)	1/13 (7.7%)	RR 3 (0.36 to 25.21)	154 more per 1000 (from 49 fewer to 1000 more)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Baclofen versus placebo	Control	Relative (95% CI)	Absolute		
Adverse events - nausea												
Orsnes 2000	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[C]	none	1/13 (7.7%)	0/13 (0%)	PETO OR 7.39 (0.15 to 372.38)	80 more per 1000 (from 110 fewer to 270 more)	VERY LOW	CRITICAL
Adverse events - diarrhoea												
Orsnes 2000	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[C]	none	1/13 (7.7%)	1/13 (7.7%)	RR 1 (0.07 to 14.34)	0 fewer per 1000 (from 72 fewer to 1000 more)	VERY LOW	CRITICAL
Adverse events - worse incontinence												
Orsnes 2000	Cross-over RCT	Serious risk of bias[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[C]	none	1/13 (7.7%)	0/13 (0%)	PETO OR 7.39 (0.15 to 372.38)	80 more per 1000 (from 110 fewer to 270 more)	VERY LOW	CRITICAL

[A] Outcome assessor bias was not reported
[B] The outcome was an indirect measure of ataxia/tremor
[C] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID <u>or</u> the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Narrative review for outcomes not appropriate for meta-analysis

Isoniazid versus placebo

Hallet et al. (1985) [88] gave quantitative tremor measures on a patient by patient basis in low resolution graphs, so no qualitative data were available for meta-analysis. However Hallet et al. reported that there was little difference in the effects of isoniazid and placebo in terms of displacement, acceleration, spectral peak amplitudes and Fourier transformed signal power.

Baclofen versus placebo

Some quantitative data were reported by Ornes (2000) [176] for unsteadiness of gait, which is an indirect measure of ataxia/tremor, but these were not based on baclofen versus placebo data, so are not included in this review. However the paper reported that all parameters of gait unsteadiness were similar in the cross-over study between the treatments, except for vertical unsteadiness, which was better in the baclofen group (see Table 80). Function, as shown by EDSS, ambulation index, NRS, and MSIS, was also described as the same ('non-significant') between treatments, without provision of any quantitative data.

9.5.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided Appendix M to aid consideration of cost effectiveness.

9.5.5 Evidence statements

9.5.5.1 Clinical

Isoniazid versus placebo

Very low quality evidence from one study comprising 8 participants showed that there was no difference in clinical effectiveness between isoniazid and placebo in terms of mean tremogram acceleration measures, with very serious imprecision.

Very low quality evidence from one study comprising 8 participants showed that Isoniazid was clinically effective compared to placebo in terms of the number of people with objective improvements in tremor, with very serious imprecision.

Very low quality evidence from one study comprising 6 participants showed that Isoniazid was clinically effective compared to placebo in terms of the number of people with a subjective improvements in tremor, with very serious imprecision

Very low quality evidence from one study comprising 6 participants showed that there was no difference in clinical harm between isoniazid and placebo in terms of nausea, with serious imprecision.

Baclofen versus placebo

Very low quality evidence from one study comprising 26 participants showed that baclofen was clinically effective compared to placebo in terms of the number of people with improved vertical unsteadiness during gait, with serious imprecision

Very low quality evidence from one study comprising 26 participants showed that baclofen was clinically harmful compared to placebo in terms of the number of people with fatigue, with very serious imprecision

Very low quality evidence from one study comprising 26 participants showed that baclofen was clinically harmful compared to placebo in terms of the number of people with dizziness, with very serious imprecision

Very low quality evidence from one study comprising 26 participants showed that baclofen was clinically harmful compared to placebo in terms of the number of people with nausea, with very serious imprecision

Very low quality evidence from one study comprising 26 participants showed that there was no difference in clinical harm between baclofen and placebo in terms of diarrhoea, with very serious imprecision.

Very low quality evidence from one study comprising 26 participants showed that baclofen was clinically harmful compared to placebo in terms of the number of people with worsened incontinence, with very serious imprecision

Botulinum toxin versus placebo

Low quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of quality of life, with serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of Bain composite tremor score, with no serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of Bain writing score, with serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of Bain Archimedes spiral, with no serious imprecision.

Low quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of CRST writing, with serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of CRST drawing, with no serious imprecision.

Low quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of ICARS Archimedes spiral, with serious imprecision.

Low quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of CRST pouring, with no serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of drinking from a cup, with no serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of 9 hole peg test, with no serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of kinetic tremor severity, with no serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of CRST action tremor in the arm, with serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of CRST action tremor amplitude, with no serious imprecision.

Low quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of ICARS finger-finger test, with serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of postural tremor severity, with no serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of Batwing position tremor, with no serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of CRST postural tremor, with no serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of CRST postural tremor amplitude, with no serious imprecision.

Low quality evidence from one study comprising 33 participants showed that there was no difference in clinical effectiveness between botulinum and placebo in terms of SARA ataxia rating scale score, with serious imprecision.

Moderate quality evidence from one study comprising 33 participants showed that there was no difference in clinical harm between botulinum and placebo in terms of muscle weakness, with no serious imprecision.

9.5.5.2 Economic

No relevant economic evaluations were identified.

9.5.6 Recommendations and link to evidence

Recommendations	
Relative values of different outcomes	Quality of life was considered the most critical outcome, closely followed by subjective assessments of ataxia and tremor. Objective assessments of ataxia and tremor and adverse events were regarded as of lowest importance, but still critical for decision making.
Trade off between clinical benefits and harms	Isoniazid versus placebo There was evidence of a small but clinically important benefit from isoniazid, though this was not consistent across all outcomes. The only adverse event considered was nausea, and this was not reported for the isoniazid group. Hence the small benefits were not compromised. Botulinum toxin versus placebo Quality of life was made worse by the use of botulinum, and this may be at least partly due to the greater number reporting muscle weakness in the botulinum group. Although there were benefits for botulinum toxin in terms of the SARA ataxia score, the BAIN tremor score, and the 9 hole peg test, these were not regarded as clinically important effects, and other outcomes were inconclusive. Overall, these inconsistent benefits were outweighed by the adverse effects on quality of life and muscle strength. Baclofen versus placebo There was some evidence for baclofen reducing one objective measure of tremor, but this was not observed in other objective measures. The small and inconsistent benefit was probably outweighed by the harms of baclofen in terms of fatigue, dizziness, nausea, diarrhoea and incontinence.
Economic considerations	No relevant economic evaluations comparing pharmacological interventions for ataxia and / or tremor were identified. The costs of pharmacological treatments used for ataxia and tremor were presented. The annual costs of isoniazid, baclofen and botulinum toxin were: £140–179, £34 and £310, respectively. For botulinum toxin there is an additional cost for administration of the drug and nursing needs which has not been included in the unit cost estimate.
Quality of evidence	There were only 2 studies for isoniazid versus placebo and one each for the other two comparisons. The quality of evidence was very low for the isoniazid and baclofen studies and low to moderate for the botulinum study. Studies were limited by imprecision and risk of bias (mainly due to attrition bias and a lack of blinding).
Other considerations	The GDG considered both the unit costs and the clinical evidence and felt that there was insufficient evidence to recommend any of these pharmacological treatments for ataxia and / or tremor in people with MS.

9.6 Pharmacological management of fatigue

9.6.1 Introduction

Fatigue is one of the commonest symptoms of multiple sclerosis. MS related fatigue is not well understood. It can be associated with heat, can occur at particular times of the day, and appears out of proportion to activity levels. Fatigue is however also a common symptom in the population and can be caused by a variety of medical problems. Assessment of the person with fatigue should not ignore these other potential causes.

9.6.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment of fatigue?

For full details see review protocol in Appendix C.

Table 81: PICO characteristics of review question

Population	Adults with MS experiencing fatigue
Intervention/s	• Amantadine • SSRIs • Aspirin • B12
Comparison/s	• Usual treatment or placebo • Amantadine • SSRIs • Aspirin • B12 • acupuncture • rehabilitation
Outcomes	Critical: • Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale. • Patient-reported outcomes, for example symptoms of fatigue • Impact on carers. • Fatigue scales – ie Neurological Fatigue Index (NFI), fatigue Severity Scale (FSS) • Adverse effects leading to withdrawal Important: • Functional scales that quantify level of disability, such as the Expanded Disability Status Scale (EDSS), the Multiple Sclerosis Functional Composite (MSFC), the Cambridge Multiple Sclerosis Basic Score (CAMBS), or the Functional Assessment of Multiple Sclerosis (FAMS). • Cognitive functions, such as memory and concentration • Psychological symptoms assessed by validated and disease-specific scales, questionnaire or similar instruments. • Adverse effects of treatment.
Study design	RCTs

9.6.3 Clinical evidence

Eight studies were included in the review. [87; 42; 58; 76; 117; 121; 160; 223,224]. Evidence from these are summarised in the clinical GRADE evidence profile below (Table 196). See also the study selection flow chart in Appendix D, forest plots in Appendix I, study evidence tables in Appendix G and exclusion list in Appendix J.

We searched for randomised trials which compared pharmacological interventions for MS related fatigue compared with each other, non-pharmacological interventions or placebo. Six trials were identified that compared amantadine with placebo. [87]; [42]; [76]; [117]; [121]; [160]. One Cochrane review on this intervention was cross-checked for relevant references. [187]. One trial was identified comparing amantadine with aspirin[223,224] and one comparing paroxetine with placebo.[58]. Two trials comparing amantadine with placebo were excluded due to no relevant outcomes.[200]; [210]

Summary of included studies

Table 82: Summary of studies included in the review

Study	Intervention/comparison	Population	Outcomes	Comments
Amantadine versus placebo				
Anon 1987[87]	100mg 2 x per day for 3 weeks. Preceded by a two week single-blind placebo period Identical placebo. Preceded by a two week single-blind placebo period	Minimum 6 months of definite MS, according to Schumacher criteria; minimum 3 month history of chronic, persistent, moderate to severe daily fatigue. Excluded: History of depression N=86 per protocol	Mean decrease on fatigue VAS Mean decrease in the effect of fatigue on walking or standing Beck depression Inventory Patient and physician subjective assessment Adverse events	Crossover trial
Cohen 1989[42]	Amantadine hydrochloride (100mg) twice daily for 4 weeks. 2 week wash-out period. Placebo exactly as for intervention	Definite or probable diagnosis of MS made at least 6 months prior to the study; daily symptomatic fatigue for at least 3 months. Excluded: moderate/severe depression N=22 per protocol	Fatigue Drug preference Neurobehavioural measures Adverse events	Crossover trial
Geisler 1996 [76]	Amantadine 100 mg table twice a day Placebo	MS patients with severe fatigue. Inclusion criteria 18 to 50 yrs, clinically or laboratory definite MS based on Poser et al, Fatigue Severity Scale	Fatigue Severity Scale Cognition	Parallel RCT

Study	Intervention/comparison	Population	Outcomes	Comments
		(FSS) score of 4.0 or greater, Kurzke's Expanded Disability Status Scale (EDSS) score of 6.5 or less. Excluded: severe depression N=32 no drop-outs		
Krupp 1995[117]	Amantadine 200mg daily (100mg dose am and pm) for 6 weeks Placebo	Clinically definite MS with severe fatigue. 18-52 years; ambulatory; EDSS of <6; Fatigue severity scale score >4. Excluded severe depression N=66 per protocol	Fatigue Severity Scale Drug preference Adverse events	Parallel RCT
Ledinek 2013[121]	Amantadine 200mg daily for 1 month Placebo	MS; disability level ≤5.5 on the EDSS; fatigue	Modified Fatigue Impact Scale (MFIS) Quality of life (SF-36: physical component score and Mental Component Score	Parallel RCT
Murray 1985 [160]	Amantadine 100mg twice a day for 6 weeks. One week washout period Placebo	MS, with persistent fatigue >3 months; fatigue was felt to be abnormal or greater than normal N=64	Subjective improvement Drug preference Adverse events	Crossover trial
Amantadine versus aspirin				
Shaygannejad 2012[223,224]	Amantadine 100 mg orally twice daily. 2 wk washout Aspirin 500mg orally once daily	Consecutive patients with definite MS who sought treatment for MS-related fatigue. Expanded Disability Status	Fatigue Severity Scale	Parallel RCT

Study	Intervention/comparison	Population	Outcomes	Comments
		Scale score ≤ 6 and clinical evidence of fatigue as documented by a score of ≥ 4 on the Fatigue Severity Score (FSS), but no clinical exacerbations for at least 4 wks. None of the patients had been treated with a medication known to influence MS-related fatigue. Patients had received interferon-beta treatment for at least 1 yr in order to avoid the frequent occurrence of fatigue in the early stage of interferon-beta therapy. Excluded: severe depression. N=52		
Paroxetine versus placebo				
Ehde 2008[58]	Paroxetine for 12 weeks. Starting dose of 10 mg/day, titrated up to 40 mg daily as tolerated Placebo exactly as paroxetine N=36 available case analysis	Clinically definite MS; age >18 yrs; major depressive disorder (> or =16 on the CES-D).	SF-36 Modified Fatigue Impact scale. Hamilton Depression Scale Perceived Deficits Questionnaire Adverse events	Parallel RCT

Table 83: Clinical evidence profile: Amantadine versus placebo

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Amantadine versus placebo mean (SD)	Control mean (SD)	Relative (95% CI)	Absolute	Quality	Importance
Quality of life												
No data												
Fatigue Severity Scale (follow-up 6 weeks; Better indicated by lower values)												
Geisler 1996	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	5.2 (0.8)	5.4 (1.2)	-	MD 0.2 lower (0.91 lower to 0.51 higher)	LOW	CRITICAL
Overall fatigue (follow-up 4 weeks; range of scores: 1-5; Better indicated by higher values)												
Cohen 1989	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	3.18 (0.04)	2.96 (0.03)	-	MD 0.22 higher (0.2 to 0.24 higher)	MODERATE	CRITICAL
MFIS at 1 month (lower better).												
Ledinek 2013	randomised trial	Very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	31.2 (3.75)	48.5 (3.7)	-	MD 17.30 lower (19.97 lower to 14.63 lower)	LOW	CRITICAL
SF-36 Physical at 1 month (lower better).												
Ledinek 2013	randomised trial	Very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	34.4 (2.14)	40.2 (2.14)	-	MD 5.80 lower (7.33 lower to 4.27 lower)	LOW	CRITICAL
SF-36 Mental at 1 month (lower better).												
1 Ledinek 2013	randomised trial	Very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	48.8 (2.07)	40.4 (2.07)	-	MD 8.40 higher (6.92 higher to 9.88 higher)	LOW	CRITICAL
Overall improvement (follow-up 6 weeks)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Amantadine versus placebo mean (SD)	Control mean (SD)	Relative (95% CI)	Absolute		
Murray 1985	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	20/32 (62.5%)	21.9%	RR 2.86 (1.41 to 5.8)	407 more per 1000 (from 90 more to 1000 more)	MODERATE	CRITICAL
Felt better on the drug (follow-up 3-10 weeks)												
Hader 1987 Cohen 1989 Krupp 1995 Murray 1985	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	61/163 (37.4%)	13.7%	RR 2.37 (0.79 to 7.16)	188 more per 1000 (from 29 fewer to 844 more)	LOW	CRITICAL
Side effects leading to study withdrawal (follow-up 6-10 weeks)												
Krupp 1995 Murray 1985	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	3/63 (4.8%)	3.4%	RR 1.38 (0.24 to 8.06)	13 more per 1000 (from 26 fewer to 240 more)	VERY LOW	CRITICAL
Physicians rating of better response on drug (follow-up 3 weeks)												
Hader 1987	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	21/86 (24.4%)	10.5%	RR 2.33 (1.13 to 4.8)	140 more per 1000 (from 14 more to 399 more)	LOW	IMPORTANT
Selective reminding long term retrieval (follow-up 6 weeks; Better indicated by higher values)												
Geisler	randomised trials	serious[a]	no serious inconsisten	no serious indirectnes	very serious[b]	none	42.2 (17.5)	45.2 (11.4)	-	MD 3 lower (13.23 lower to	VERY	IMPORTANT

Quality assessment							No of patients		Effect			
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Amantadine versus placebo mean (SD)**	**Control mean (SD)**	**Relative (95% CI)**	**Absolute**	**Quality**	**Importance**
1996			y	s						7.23 higher)	LOW	
Selective reminding sum of recall (follow-up 6 weeks; Better indicated by higher values)												
Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	52.3 (10.1)	53.3 (6.7)	-	MD 1 lower (6.94 lower to 4.94 higher)	VERY LOW	IMPORTANT
Selective reminding delayed recall (follow-up 6 weeks; Better indicated by lower values)												
1= Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	8.9 (3.6)	8.9 (3.1)	-	MD 0 higher (2.33 lower to 2.33 higher)	VERY LOW	IMPORTANT
Trail making part A (follow-up 6 weeks; Better indicated by lower values)												
Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	30.9 (9.4)	36.2 (14.2)	-	MD 5.3 lower (13.64 lower to 3.04 higher)	LOW	IMPORTANT
Trail making part B (follow-up 6 weeks; Better indicated by higher values)												
Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	68.9 (31.2)	83.1 (29.2)	-	MD 14.2 lower (35.14 lower to 6.74 higher)	LOW	IMPORTANT
WAIS-R Digit Span (follow-up 6 weeks; Better indicated by higher values)												
Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	15.6 (2.7)	16.5 (3.5)	-	MD 0.9 lower (3.07 lower to 1.27 higher)	LOW	IMPORTANT
Benton Visual Retention (errors) (follow-up 6 weeks; Better indicated by lower values)												
Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	4.3 (2.4)	2.8 (1.8)	-	MD 1.5 higher (0.03 to 2.97 higher)	LOW	IMPORTANT
SDMT written (follow-up 6 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Amantadine versus placebo mean (SD)	Control mean (SD)	Relative (95% CI)	Absolute		
Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	48.6 (15.7)	46.6 (14.2)	-	MD 2 higher (8.37 lower to 12.37 higher)	VERY LOW	IMPORTANT
SDMT oral (follow-up 6 weeks; Better indicated by higher values)												
Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	57.8 (19.7)	58.3 (16.8)	-	MD 0.5 lower (13.19 lower to 12.19 higher)	VERY LOW	IMPORTANT
Finger tapping test (follow-up 6 weeks; Better indicated by higher values)												
Geisler 1996	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	56.6 (14.9)	57.2 (9.5)	-	MD 0.6 lower (9.26 lower to 8.06 higher)	VERY LOW	IMPORTANT
ADL physical function (follow-up 3 weeks; Better indicated by lower values)												
Hader 1987	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.76 (2.87)	11.64 (2.87)	-	MD 0.88 lower (1.74 to 0.02 lower)	LOW	IMPORTANT
ADL intellectual function (follow-up 3 weeks; Better indicated by lower values)												
Hader 1987	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	7.67 (5.56)	8.25 (5.1)	-	MD 0.58 lower (2.17 lower to 1.01 higher)	MODERATE	IMPORTANT
ADL total score (follow-up 3 weeks; Better indicated by lower values)												
Hader 1987	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	24.09 (6.86)	25.85 (6)	-	MD 1.76 lower (3.8 lower to 0.28 higher)	LOW	IMPORTANT
Beck Depression Inventory (follow-up 3 weeks; Better indicated by lower values)												
Hader 1987	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	7.34 (7.51)	7.59 (7.79)	-	MD 0.25 lower (2.54 lower to 2.04 higher)	MODERATE	IMPORTANT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Amantadine versus placebo mean (SD)	Control mean (SD)	Relative (95% CI)	Absolute		
Patients experiencing adverse events (follow-up 3-6 weeks)												
Hader 1987 Cohen 1989 Murray 1985	randomised trials	serious1	no serious inconsistency	no serious indirectness	serious[b]	none	92/169 (54.4%)	18.8%	RR 1.28 (1.03 to 1.59)	53 more per 1000 (from 6 more to 111 more)	LOW	IMPORTANT

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 84: Amantadine versus aspirin

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Amantadine	Aspirin	Relative (95% CI)	Absolute		
Quality of life												
No data												
Fatigue Severity Scale mean difference (follow-up 4 weeks; Better indicated by lower values)												
Shaygannejad	randomised trials	no serious risk of	no serious inconsistency	no serious indirectness	very serious[a]	none	1.1 (1.54)	0.8 (1.22	-	MD 0.3 higher (0.24 lower to	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Amantadine	Aspirin	Relative (95% CI)	Absolute		
2010		bias						)		0.84 higher)		
Proportion with reduction in Fatigue Severity Scale (follow-up 4 weeks)												
Shaygannejad 2010	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	Serious[a]	none	19/26 (73.1%)	57.7 %	RR 1.27 (0.85 to 1.9)	156 more per 1000 (from 87 fewer to 519 more)	MODERATE	CRITICAL

[A] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 85: Paroxetine versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Paroxetine versus placebo	Control	Relative (95% CI)	Absolute		
SF-36 physical (follow-up 12 weeks; Better indicated by higher values)C												
Edhe 2008	randomised trials	Serious[a]	no serious inconsistency	serious indirectness	very serious[b]	none	36.4 (12.3)	35.5 (13.3)	-	MD 0.9 higher (7.46 lower to 9.26 higher)	VERY LOW	CRITICAL
SF-36 mental (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	Serious[a]	no serious inconsistency	serious indirectness	very serious[b]	none	48.4 (32.3)	42.5 (9.7)	-	MD 5.9 higher (10.06 lower to 21.86 higher)	VERY LOW	CRITICAL
Modified Fatigue Impact Scale (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	Serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	39.3 (14.8)	52.1 (18.3	-	MD 12.8 lower (23.63 to 1.97	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Paroxetine versus placebo	Control	Relative (95% CI)	Absolute		
								)		lower)		
Modified Fatigue Impact Scale psychosocial (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	3.4 (1.7)	4.8 (1.9)	-	MD 1.4 lower (2.58 to 0.22 lower)	VERY LOW	CRITICAL
Modified Fatigue Impact Scale physical (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	19.5 (7.3)	23.1 (9.2)	-	MD 3.6 lower (9 lower to 1.8 higher)	VERY LOW	CRITICAL
Modified Fatigue Impact Scale cognitive (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	23.1 (9.2)	16.2 (8.8)	-	MD 6.9 higher (1 to 12.8 higher)	VERY LOW	CRITICAL
Withdrawal due to adverse events (follow-up 12 weeks)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	2/21 (9.5%)	0%	Peto OR 7.41 (0.45 to 122.78)	10 more per 1000 (from 5 less to 240 more)	VERY LOW	CRITICAL
Perceived Deficits Questionnaire (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	29.1 (13.2)	40.4 (12.6)	-	MD 11.3 lower (19.75 to 2.85 lower)	VERY LOW	IMPORTANT
Perceived Deficits Questionnaire attention concentration (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	8.1 (4.2)	11.8 (3.5)	-	MD 3.7 lower (6.27 to 1.13 lower)	VERY LOW	IMPORTANT
Perceived Deficits Questionnaire plan organise (follow-up 12 weeks; Better indicated by lower values)												
1	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	8 (3.5)	11 (3.9)	-	MD 3 lower (5.42 to 0.58 lower)	VERY LOW	IMPORTANT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Paroxetine versus placebo	Control	Relative (95% CI)	Absolute		
Perceived Deficits Questionnaire prospective memory (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	5.4 (3.2)	8 (2.4)	-	MD 2.6 lower (4.47 to 0.73 lower)	VERY LOW	IMPORTANT
Perceived Deficits Questionnaire retrospective memory (follow-up 12 weeks; Better indicated by lower values)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	7.7 (4.5)	9.7 (4.3)	-	MD 2 lower (4.88 lower to 0.88 higher)	VERY LOW	IMPORTANT
50% reduction on HAM-D (follow-up 12 weeks)												
Edhe 2008	randomised trials	serious[a]	no serious inconsistency	serious indirectness	serious[b]	none	13/17 (76.5%)	42.1%	RR 1.82 (1.01 to 3.27)	345 more per 1000 (from 4 more to 956 more)	VERY LOW	IMPORTANT

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*

[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Narrative review for outcomes not appropriate for meta analysis

Amantadine versus placebo

Table 86: Anon 1987[87]

mean decrease on fatigue VAS	Data only included on low resolution graph. Reports of a significantly greater improvement in the Amantadine group than the placebo group at 1 week (p=0.01) and trends at 2 and 3 weeks (p=0.09 and 0.11 respectively). There was also a repeated measures analysis done across all time points, showing a benefit for Amantadine (p<0.01).
mean decrease in the effect of fatigue in walking or standing	Data only included on low resolution graph. Reports of a significantly greater improvement in the Amantadine group than the placebo group at all 3 weeks (p=0.05). There was also a repeated measures analysis done across all time points, showing a benefit for Amantadine (p<0.01).

Table 87: Cohen 1989[42]

	Amantadine	Placebo	p
Grooved pegboard task R	113.72	118.0	NS
Grooved pegboard task L	134.27	137.36	NS
Trail making test A	39.45	41.73	NS
Trail making test B	91.05	94.82	NS
Symbol digit modality test score	40.45	41.77	NS
Consonant trigram test score	33.86	32.77	NS
verbal fluency task (number of words)	46.27	46.00	NS
Continuous performance task – error rate			
misses	2.55	2.72	NS
false positives	2.14	2.45	NS
Stroop test (s)			
colour naming	71.88	81.68	NS
interference	123.13	139.31	<0.05

Table 88: Krupp 1995[117]

	Amantadine	placebo	Amantadine v placebo
FSS	**No data**	**No data**	**F=1.13; p=0.327**
MS-FS	No data	No data	F=3.40; p=0.037
Worse after stopping drug	No data	No data	Chi square=3.97; p=NS
Rand Index of vitality	No data	No data	F<1.0; p=0.750
CES-D	No data	No data	F=2.00; p=0.140

9.6.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided in Appendix M.

9.6.5 Evidence statements

9.6.5.1 Clinical

Amantadine versus placebo

Low quality evidence from 1 RCT containing 32 participants showed that there was no clinical difference between amantadine and placebo in terms of Fatigue Severity Scale , with serious imprecision

Moderate quality evidence from 1 RCT containing 30 participants showed that there was no clinically important difference between amantadine and placebo in terms of overall fatigue, with no imprecision

Low quality evidence from 1 RCT containing 30 participants showed that amantadine was clinically beneficial compared to placebo in terms of Modified Fatigue Impact Scale, with no serious imprecision

Low quality evidence from 1 RCT containing 30 participants showed that amantadine was clinically beneficial compared to placebo in terms of SF-36 physical, with no imprecision

Low quality evidence from 1 RCT containing 30 participants showed that placebo was clinically beneficial compared to amantadine in terms of SF-36 mental, with no serious imprecision

Moderate quality evidence from 1 RCT containing 64 participants showed that amantadine was clinically beneficial compared to placebo in terms of overall improvement, with no imprecision

Low quality evidence from 4 RCTs containing 327 participants showed that amantadine was clinically beneficial compared to placebo in terms on ‘felt better on drug’, with serious imprecision

Very low quality evidence from 2 RCTs containing 122 participants showed that there was no clinical difference between amantadine and placebo in terms of side effects leading to study withdrawal, with very serious imprecision

Low quality evidence from 1 RCT containing 30 participants showed that amantadine was clinically beneficial compared to placebo in terms of physicians rating of better response on drug, with serious imprecision

Very low to low quality evidence from 1 RCT (per outcome) containing 32 participants showed that there was no clinical difference between amantadine and placebo in terms of selective reminding long term retrieval, selective reminding sum of recall, Trail Making part A and B, WAIS-R Digit Span, Benton Visual Retention (errors), SDMT written and oral and the finger tapping test, with serious or very serious imprecision

Low to moderate quality evidence from 1 RCT (per outcome) containing 172 participants showed that there was no clinical difference between amantadine and placebo on the ADL physical function, ADL intellectual function, ADL total or Beck Depression Inventory with no or serious imprecision

Low quality evidence from 3 RCTs containing 338 participants showed that amantadine was clinically harmful compared to placebo in terms of patients experiencing adverse events, with serious imprecision

Amantadine versus aspirin

Low quality evidence from 1 RCT comprising 52 participants showed that there was no clinical difference between amantadine and aspirin in terms of mean Fatigue Severity Score, with very serious imprecision

Moderate quality evidence from 1 RCT comprising 52 participants showed that amantadine was clinically beneficial compared to placebo in terms of the proportion of people experiencing a reduction of Fatigue Severity Score, with serious imprecision

Paroxetine versus placebo

Very low to low quality evidence from 1 RCT comprising 32 participants showed that there was no clinical difference between paroxetine and placebo in terms of SF-36 physical, Modified Fatigue Impact Scale physical or Perceived Deficits Questionnaire retrospective memory, with very serious or serious imprecision

Very low quality evidence from 1 RCT comprising 32 participants showed that paroxetine was clinically beneficial compared to placebo in terms of SF-36 mental, with very serious imprecision

Very low quality evidence from 1 RCT comprising 32 participants showed that paroxetine was clinically beneficial compared to placebo in terms of Modified Fatigue Impact Scale, psychosocial and cognitive, with serious imprecision

Very low quality evidence from 1 RCT comprising 32 participants showed that paroxetine was clinically beneficial compared to placebo in terms of Modified Fatigue Impact Scale psychosocial , with serious imprecision

Very low quality evidence from 1 RCT comprising 32 participants showed that paroxetine was clinically beneficial compared to placebo in terms of Perceived Deficits Questionnaire, attention and prospective memory, with serious imprecision

Very low quality evidence from 1 RCT comprising 32 participants showed that paroxetine was clinically beneficial compared to placebo in terms of a 50% reduction in Hamilton Depression Scale, with serious imprecision

9.6.5.2 Economic

No relevant economic evaluations were identified.

9.6.6 Recommendations and link to evidence

Recommendations	**44. Offer amantadine[bb] to treat fatigue in people with MS.** **45. Do not use vitamin B_{12} injections to treat fatigue in people with MS.**
Relative values of different outcomes	It was acknowledged that there was no good way for measuring improvement of fatigue and a number of related outcomes were used in the relevant studies. Fatigue outcomes included a subjective rating of fatigue by a patient

bb At the time of publication (October 2014), amantadine did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

	or physician, a visual analogue scale, the Fatigue Severity Scale or Modified Fatigue Impact Scale. Some studies also used a subjective rating of overall improvement and measurement of cognitive function or depression. The Fatigue Severity Scale is primarily used as a screening tool with a cut-off of >4 for those needing further assessment, however the trials reported data that could be analysed only as a continuous variable. This analysis may have been less sensitive to detecting clinically important changes in fatigue. Quality of life and adverse effects were also regarded as critical outcomes.
Trade off between clinical benefits and harms	Amantadine had clinical benefits compared to placebo in terms of reduction in fatigue (as measured by the modified fatigue scale) and improved quality of life, although the effects on fatigue were inconsistent and depended on the scale used. One study (Shaygannejad et al, 2012) found that aspirin was probably comparable to amantadine but there is no other relevant literature on the use of aspirin, and the GDG were not aware of the use of aspirin for fatigue in UK clinical practice. There was one small study on the use of paroxetine in people with multiple sclerosis and major depressive disorder. This found clinically important improvements in fatigue scores and scores on the Perceived Deficits Questionnaire. The GDG felt that these improvements may be specific to the patient population recruited in the study and therefore should not be generalised to patients without major depressive disorder. Harms were identified for all drugs under consideration. For amantadine, a meta-analysis of three studies found an increased incidence of adverse effects compared to placebo. However when looking at drug withdrawal due to adverse effects in two studies, amantadine was no worse than placebo. Overall the small benefits of amantadine were felt to outweigh the potential harms. The GDG also felt further research on aspirin was justified, given its similar effect to amantadine, despite its adverse effects.
Economic considerations	No relevant economic evaluations comparing pharmacological interventions of fatigue were identified. The costs of pharmacological treatments used for fatigue were presented. The annual costs of amantadine, aspirin and paroxetine were: £129, £25–78 and £42–154, respectively. Although aspirin cost less than amantadine, the GDG agreed that there was not sufficient clinical evidence, particularly on possible adverse events to recommend aspirin. Paroxetine was not recommended due to increased adverse events and withdrawal. Therefore, based on the clinical and economic evidence, the GDG considered amantadine should be offered for the treatment of fatigue. No evidence was identified for the use of B12 injections. The cost of hydroxocobalamin (B12) injections, including nursing time, was estimated to be £14-18 based on four injections a year. The GDG considered that with the absence of clinical evidence, the cost of these injections are wasteful and B12 injections should not be offered.
Quality of evidence	Four studies assessed overall subjective improvement on the same dose (200mg daily) of amantadine and consistently found a benefit. Combined, this moderate quality evidence was considered to be a large and meaningful effect with 407 per 1000 people finding overall improvement on amantadine when compared to placebo. There was one other study of amantadine (Geisler et al, 1996) which found no significant difference to placebo on cognitive functioning or the Fatigue Severity Scale. For aspirin and paroxetine, there was only one study for each. Indeed, the study with paroxetine included fatigue as a secondary outcome only.
Other considerations	Other suggested treatments for which there was no evidence were for vitamin B12 injections and modafinil. Modafinil should not be used to treat fatigue in multiple sclerosis. This follows a European Medicines Agency directive in 2010 that there are significant harms associated with the drug. The GDG reported that vitamin B12 injections are sometimes used in clinical practice even for

	people who are not B12 deficient. Historically vitamin B12 has been used generally for fatigue but the GDG were not aware of any rationale for this use in MS and considered that it is not now common practice. The search for evidence found no evidence for use of B12, it has cost and resource implications and in the absence of evidence, the GDG made a ‘do not offer’ recommendation for vitamin B12 for the treatment of fatigue.

10 Non-pharmacological management of MS symptoms

10.1 Non-pharmacological management of cognition and memory

10.1.1 Introduction

Cognitive problems are a common symptom of multiple sclerosis. People can experience a wide range of difficulties including attention, memory and decision making and planning and these can impact significantly on work, home and social activities. Neuropsychological rehabilitation encompasses a diverse range of strategies, techniques and programmes, including computerised training delivery.

10.1.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological management of memory and cognitive problems with neuropsychological rehabilitation?

For full details see review protocol in Appendix C.

Table 89: PICO characteristics of review question

Population	Adults with MS
Intervention/s	• Specific or non-specific cognitive retraining • Memory retraining techniques – ie, Personal Digital Assistants, Lumosity, 'Brain Trainer', Learning internal and external memory strategies, 'Brain Stim' • Neuropsychological Compensatory Training (NCT) • Story memory technique (SMT) • Executive functioning textbook exercises • Cognitive training of concentration • Computer aided (VILAT-G 1.0) training for memory • Computer aided RehaCom module 'Plan a Day' for organization and planning • Computer aided RehaCom module 'Divided Attention' for attention • Computer aided RehaCom module 'memory and Attention' • Computer aided memory retraining programme (SCRP) • Computer aided 'Cognifit Personal Coach' for cognition • Restitution – encoding and retrieval strategies, attention retraining • Compensation – external memory aids • Memory and working memory rehab tasks • Social cognition/theory of mind
Comparison/s	• Standard treatment • Pharmacological approaches • No treatment / placebo • Other neuropsychological rehabilitation treatment
Outcomes	• Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale. • Cognitive functions, such as memory, attention, executive functions • Mood • Patient-reported outcomes, for example symptoms, activities.(for example Canadian

	Occupational Performance measure) • Patient satisfaction • Impact on carers
Study design	SR or RCTs

10.1.3 Clinical evidence

18 studies were included in the review.[6]; [37]; [40]; [61]; [130]; [136]; [98]; [109]; [123]; [135,136]; [140]; [202]; [222]; [234]; [240]; [243]; [257]. One Cochrane review was identified but interventions aimed at different cognitive processes were combined [203]. Evidence from these are summarised in the clinical GRADE evidence profiles below. See also the study selection flow chart in Appendix D, forest plots in Appendix I, study evidence tables in Appendix G and exclusion list in Appendix J.

Summary of included studies

Table 90: Summary of studies included in the review

Study	Population	Intervention	Comparison	Comments
Data meta-analysed				
General cognitive rehabilitation versus control				
Mantynen 2014[130], Rosti-Otajarv 2013[202]	Patients with clinically definite relapsing remitting MS, EDSS < 6, subjective (total score of questions 1, 2 and 11 in the Multiple Sclerosis Neuropsychological Questionnaire ≥ 6) and objective Symbol Digit Modalities Test total score ≤ 50) deficits in attention and processing speed and age 18-59	Neuropsychological rehabilitation Computer-based attention and working memory retraining used for increasing awareness of attentional problems, learning strategies, psychoeducation and homework assignment connected with rehabilitation goals as well as psychological support to promote coping with cognitive impairments Once a week for 13 weeks	Control No training	
Memory and problem solving versus control				
Hildenbrandt	Patients with relapsing	Cognitive	Control	

Study	Population	Intervention	Comparison	Comments
2007[98]	remitting MS diagnosed according to the McDonald criteria. Group performance as baseline. Depending on the specific task a performance below one standard deviation or above one standard deviation of the published norms were defined as impaired. According to this criterion 28% of the control group and 41% of the treatment group showed impairments on the PASAT, 24% vs 23.4% in CVLT learning or recall, 20% vs 12% in cognitive speed, 36% vs 17% in object alteration. Taking the results of all neuropsychological tests together 48% of control group and 47% of the treatment group showed some impairment	training Compact disc with memory and working memory rehabilitation tasks (VILAT-G 1.o (Hildenbrandt, 2002). Patients were requested to train for 6 week, at least 5 days a week, for 30 minutes a day	No training	
Stuifbergen2012[240]	Clinically definite MS for at least six mths Responded 'sometimes' or more often to at least five problems on the Perceived Deficits Questionnaire	Memory and Problem Solving Skills for people with Multiple Sclerosis (MAPSS-MS) Teaches the use of compensatory skills, retraining skills (the computer component) and environmental/lifestyle support for cognitive functioning. a) Eight weekly 2-hr group sessions focused on building efficacy for use of cognitive compensatory	Waiting list	

Study	Population	Intervention	Comparison	Comments
		strategies (b) a computer-assisted cognitive training program. Enabled the participants to engage in practice sessions (minimum of 45 minutes three times per week). Translation of skills practiced to everyday issues was a focus of the group sessions.		
Learning versus control (some data not meta-analysed)				
Chiaravalloti 2005[39]	Patients with clinically definite MS (Poser criteria). 17 patients had a relapsing-remitting course, 4 primary-progressive and 7 secondary-progressive. Duration of MS 12 to 432 mths, mean 135.72 (SD 87.53) All patients were determined to have impaired verbal new learning, as documented by performance at least one standard deviation below the mean for a healthy control sample on an adaptation of the Buschke Selective Reminding Test	Rehabilitation Eight therapeutic sessions (2 x 4 wks). Participant learns the story memory technique (SMT). Within the SMT, the participant was taught two interrelated skills: 1) to use visualisation ie imagery to facilitate new learning (sessions 1-4) and 2) to utilize context to learn new information e.g a story even if information is seemingly unrelated (sessions 5-8). Each session latest	Control Met with the same therapist as did the rehabilitation gp. Sessions were held at the same frequency as the rehabilitation gp but the control gp engaged in non-training orientated tasks to control for professional contact. Training sessions for the two gps were matched for stimulus presentation, content, examiner contact, and session duration.	

Study	Population	Intervention	Comparison	Comments
		approximately 45 mins		
Executive versus control				
Fink 2010[61]	Patients with relapsing remitting MS	Cognitive intervention 6 week programme. Participants spent 25-30 minutes per day, four times per week, on textbook exercises for executive functioning and they met with a psychologist for 1.5 hrs to receive feedback and to discuss the exercises	Placebo Trained 5 days per week for 40 minutes. Patients had to respond fast and accurately to visual stimuli. They had to call the psychologist once a week to report on time having spent training. The amount of time invested in completing the exercises was comparable in both gps	
Rehacom versus active control				
Cerasa 2013[37]	Patients with relapsing remitting MS. Inclusion: No evidence of a severe cognitive impairment; predominant deficits in either attention and/or information processing speed, working memory and/or executive function	Intervention Twice weekly for one hour sessions for six weeks. Training consisted of the Rehacom software.	Control Twice weekly for one hour sessions for six weeks. Visuomotor coordination task	Data meta analysed
Mendozzi 1998[140]	Patients with a relapsing-remitting course or secondary chronic progressive course were eligible	Specific cognitive retraining programme (SCRP) 15 bi-weekly sessions lasting 45 min each average 8 weeks duration The programmes employed were part of Rehacom.	Non-specific cognitive retraining programme (NCRP) Two periods of similar duration to SCRP, one spent on a visual tracking task and the other on a reaction-time task.	

Study	Population	Intervention	Comparison	Comments
Solari 2004[234]	Patients meeting the diagnostic criteria of Posner and who complained of poor attention or memory, confirmed by a score below the 80th percentile in at least two components of the Brief Repeatable Battery of Neuropsychological Tests (BRBNT Disease course Cognitive training relapsing remitting 42.5%, relapsing progressive 50.0%, chronic progressing 7.5% Control: relapsing remitting 59.5%, relapsing progressive 40.5%	Cognitive training Individual treatment as outpatients for 45 mins, twice a week, for 8 consecutive weeks. The training program was Rehacom. The study treatment consisted of the Rehacom memory and attention retraining procedures	Control As for cognitive training except the treatment consisted of the Rehacom visuo-constructional and visuo-motor coordination retraining procedures.	
Rehacom versus control				
Mendozzi 1998[140]	Patients with a relapsing-remitting course or secondary chronic progressive course were eligible	Specific cognitive retraining programme (SCRP) 15 bi-weekly sessions lasting 45 min each average 8 weeks duration The programmes employed were part of Rehacom.	No training	
Tesar 2005[243]	Patients with MS meeting the criteria of Posner plus a positive MRI scan. Inclusion criteria: mild to moderate cognitive deficit	Rehabilitation Rehacom computer training. Direct functional training of the two cognitive areas which were most severely affected and then teaching of	Control No treatment	

Study	Population	Intervention	Comparison	Comments
		compensation strategies to everyday life. 12 sessions each last one hour. Total duration 4 wks		
Cognifit versus control				
Shatil 2010[222]	Outpatients with multiple sclerosis. Inclusion criteria: Diagnosis of relapsing remitting or relapsing progressive MS Exclusion criteria: Primary progressive MS. At baseline 15/22 completers in the training gp were classified by the program as having low or intermediate scores on general memory, visual working memory or verbal working memory	Cognitive training CogniFit Personal Coach (CPC), a home-based, computerised, individualised cognitive training program. Three times a week	Control No training	
High intensity versus distributed rehabilitation				
Vogt 2009[257]	Outpatients with clinically definite multiple sclerosis according to the McDonald criteria. 36/45 female, 36/45 relapsing remitting, 8/45 secondary progressive, 1/45 chronic progressive.	High intensity 45 mins training 4 times per week for 4 weeks BrainStim (Penner et al., 2006).	Distributed training 45 mins training 2 times per week for 8 weeks Control No training	
Data not meta-analysed				
General cognitive rehabilitation and psychotherapy				
Jonsson 1993[109]	Patients fulfilling Schumacher's diagnostic criteria of MS. Hospitalised patients. Exclusion criteria: severe visual or motor dysfunction, very severe cognitive impairment,	Cognitive training and neuropsychotherapy	Control (non-specific mental stimulation)	

Study	Population	Intervention	Comparison	Comments
	Six had relapsing remitting disease course, 25 secondary chronic progressive disease and 9 had primary chronic progressive disease Compared with a normal Danish sample all cognitive factors but one (visual perception) were significantly impaired in both treatment gps. There were no significant different on the Beck Depression Inventory and the State Trait Anxiety Inventory			
Memory and problem solving versus control				
Lincoln 2002[123]	Patients with either clinically definite, clinically probable, or laboratory supported multiple sclerosis. Inclusion criteria included being able to cooperate with assessment for 30 mins at a time. Selection criteria were based on the assumption that patients might benefit as much from being told that they had no cognitive deficit as from being identified. Therefore patients were not excluded on the basis of a cognitive screening assessment.	Rehabilitation Detailed cognitive assessment as above. Included various techniques such as diaries, lists, and visual mnemonics. Maximum 6 mth duration	Screening General cognitive testing Assessment Patients received detailed cognitive assessment taking about 3 hrs. Feedback given to healthcare professionals	
Attention versus active control				
Amato 2014[6]	Outpatients with relapsing-remitting MS aged 18-55 yrs. EDSS greater than or equal to 6, MMSE greater than equal to 26. Impairments on at least two out of seven attention tests defined as scores < 1.5 SD of normative values. Excluded patients with important impairment on other cognitive tasks, defined as performance greater than or equal to 2.0 SD of normative values.	Attention Processing Training (APT) program. Aimed as focused, sustained, selective, alternating and divided attention Twice weekly sessions for three mths. Each session lasted one hour	Nonspecific training Nonspecific exercises including text and newspaper article reading and comprehension	

Study	Population	Intervention	Comparison	Comments
Learning versus active control				
Chiaravalloti 2013[40]	Patients with clinically definite MS and 1) new learning impairment, 2) aged 30-70 yrs, 3) free of exacerbations and steroid use for 1 mth or more 4) no major mental health problem	Modified Story Memory Technique see Chiaravalloti (2005) 2 sessions every week for five weeks	Placebo Met with therapist for same time as intervention. Engaged in non training specific tasks No contact after 5 wks	
Rehacom versus control				
Flavia 2010[65]	Patients with relapsing remitting MS (Poser and Brinar criteria). Patients were included in the study if their scores in both tests fell below z=-1.5 for PASAT (either 2" or 3" interval) and T=35 for WCST in any of the following measures: total errors (WCSTte), number of perseverative errors (WCSTpe) and number of perseverative response (WCSTpr).	Rehabilitation 3 month duration. Individual sessions last for 1 hr with a frequency of three sessions per week. Sessions consisted on computer-assisted training of attention, information processing and planning exercises for executive functions. The software used, Plan a Day and Divided Attention, were part of the RehCom package (www. Schohfried.at)	Control No treatment	
Mattioli 2012[135]	Patients with relapsing remitting MS. EDSS < 4 and if their scores fell below Z= -1.5 for the PASAT and T=35 for WCST.	Intensive neuropsychological training 3 mths duration (1 hr session for three times per week)	No rehabilitation	

Study	Population	Intervention	Comparison	Comments
		Attention, information processing and planning exercises for executive functions. Plan a day and divided attention components of the RehaCom package.		

Table 91: General cognitive rehabilitation versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General cognitive rehabilitation mean (SD)[N]	Control mean (SD)[N]	Relative (95% CI)	Absolute		
Buschke Selective Reminding Test (BSRT)/Long Term Storage (LTS) (total score) (follow-up 6 months; Better indicated by higher values)												
Mantynen 2013	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	56.7 (14.7)[58]	53.9 (11.1)[40]	-	MD 2.8 higher (2.31 lower to 7.91 higher)	LOW	CRITICAL
BSRT/Consistent Long Term Retrieval (CLTR) (total score) (follow-up 6 months; Better indicated by higher values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	50.2 (18/2)[58]	45/7 (15.2)[40]	-	MD 4.5 higher (2.14 lower to 11.14 higher)	LOW	CRITICAL
BSRT (delayed recall) (follow-up 6 months; Better indicated by higher values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	1.4 (2.2)[58]	10 (1.7)[40]	-	MD 0.4 higher (0.37 lower to 1.17 higher)	LOW	CRITICAL
10/36 (total correct) (follow-up 6 months; Better indicated by higher values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	23.8 (4/5)[58]	20.9 (4.8)[40]	-	MD 2.9 higher (1.01 to 4.79 higher)	LOW	CRITICAL
10/36 (delayed recall) (follow-up 6 months; Better indicated by higher values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	8.5 (1.9)[58]	7.4 (1.9)[40]	-	MD 1.1 higher (0.33 to 1.87 higher)	LOW	CRITICAL
3 Paced Auditory Serial Additions Test (PASAT) (total correct) (follow-up 6 months; Better indicated by higher values)												
Mant	randomis	serio	no serious	no serious	serious[b]	none	46.7 (11.8)[58]	43.5	-	MD 3.2 higher		CRITICA

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General cognitive rehabilitation mean (SD)[N]	Control mean (SD)[N]	Relative (95% CI)	Absolute	Quality	Importance
ynen 2013	ed trials	us[a]	inconsistency	indirectness				(11)[40]		(1.37 lower to 7.77 higher)	LOW	L
2 PASAT (total correct) (follow-up 6 months; Better indicated by higher values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	32.9 (12.1)[58]	30.8 (10.3)[40]	-	MD 2.1 higher (2.36 lower to 6.56 higher)	LOW	CRITICAL
Controlled Oral Word Association Test (COWAT) (total correct) (follow-up 6 months; Better indicated by higher values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	25.5(7.1)[58]	24.2 (7/9)[40]	-	MD 1.3 higher (1.75 lower to 4.35 higher)	LOW	CRITICAL
Stroop (colour naming time) (follow-up 6 months; Better indicated by lower values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	73.7 (17.7)[58]	77 (17.8)[40]	-	MD 3.3 lower (10.45 lower to 3.85 higher)	LOW	CRITICAL
Stroop (colour/word interference-time) (follow-up 6 months; Better indicated by lower values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	116.2 (36.2)[58]	116 (30.3)[40]	-	MD 0.2 higher (13.03 lower to 13.43 higher)	MODERATE	CRITICAL
Trail making A (time) (follow-up 6 months; Better indicated by lower values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	32.1 (12.4)[58]	31 (9.2)[40]	-	MD 1.1 higher (3.18 lower to 5.38 higher)	LOW	CRITICAL
Trail making B (time) (follow-up 6 months; Better indicated by lower values)												
Mantynen 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	79.1 (36.4)[58]	75.4 (35.6)[40]	-	MD 3.7 higher (10.77 lower to 18.17	LOW	CRITICAL

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General cognitive rehabilitation mean (SD)[N]	Control mean (SD)[N]	Relative (95% CI)	Absolute	Quality	Importance
										higher)		
Perceived Deficits Questionnaire (follow-up 12 months; Better indicated by lower values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	27.9 (11/8)[50]	35.3 (13)[28]	-	MD 7.3 lower (13.12 to 1.48 lower)	LOW	CRITICAL
MSNQ-P, total score (follow-up 12 months; Better indicated by lower values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	22.3 (9.2)[50]	28.3 (11.6)[28]	-	MD 6 lower (11 to 1 lower)	LOW	CRITICAL
MSNQ-I, total score (follow-up 12 months; Better indicated by lower values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	18.6 (8.8)[50]	19.8 (11)[28]	-	MD 1.2 lower (5.95 lower to 3.55 higher)	LOW	CRITICAL
BDI-II (follow-up 12 months; Better indicated by lower values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.8 (7.7)[50]	29.7 (7)[28]	-	MD 1.1 higher (2.26 lower to 4.46 higher)	LOW	CRITICAL
MSIS-29 physical total score (follow-up 12 months; Better indicated by lower values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	22.9 (15.5)[50]	24.2 (14)[28]	-	MD 1.3 lower (8.03 lower to 5.43 higher)	LOW	CRITICAL
MSIS-29 psychological total score (follow-up 12 months; Better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**General cognitive rehabilitation mean (SD)[N]**	**Control mean (SD)[N]**	**Relative (95% CI)**	**Absolute**		
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	23.6 (16.8)[50]	22.5 (16.9)[28]	-	MD 1.1 higher (6.7 lower to 8.9 higher)	LOW	CRITICAL
WHOQOL-BREF S1 physical total score (follow-up 12 months; Better indicated by higher values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	14.4 (2.6)[50]	13.7 (2.4)[28]	-	MD 0.7 higher (0.44 lower to 1.84 higher)	LOW	CRITICAL
WHOQOL-BREF S2 psychological total score (follow-up 12 months; Better indicated by higher values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	14.1 (2.7)[50]	13.6 (2.5)[28]	-	MD 0.5 higher (0.69 lower to 1.69 higher)	LOW	CRITICAL
WHOQOL-BREF S3 social relationship total score (follow-up 12 months; Better indicated by higher values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	14.5 (3.7)[50]	14.4 (2.7)[28]	-	MD 0.1 higher (1.33 lower to 1.53 higher)	LOW	CRITICAL
WHOQOL-BREF S4 environment total score (follow-up 12 months; Better indicated by lower values)												
Posti-Otajarvi 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	15.3 (2.5)[50]	14.4 (2)[28]	-	MD 0.9 higher (0.11 lower to 1.91 higher)	LOW	CRITICAL

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.

[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 92: Clinical evidence profile: Memory and problem solving versus control

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General cognitive rehabilitation mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
Patient reported outcomes – no data												
Patient satisfaction – no data												
Impact on carers – no data												
SF12 Bodily Score (follow-up 6 weeks; Better indicated by higher values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	38.6 (12.1) [17]	41.1 (11.9)[25]	-	MD 2.5 lower (9.91 lower to 4.91 higher)	VERY LOW	CRITICAL
SF12 Mental score (follow-up 6 weeks; Better indicated by higher values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	48.5 (13.3) [17]	47.8 (9.7)[25]	-	MD 0.7 higher (6.68 lower to 8.08 higher)	VERY LOW	CRITICAL
Learning trials (CVLT) (follow-up 6 weeks; Better indicated by higher values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	12.29 (2.12)[17]	11.3 (1.94)[25]	-	MD 0.99 higher (0.27 lower to 2.25 higher)	VERY LOW	CRITICAL
Short delay free recall (CVLT) (follow-up 6 weeks; Better indicated by higher values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	13.18 (3.05) [17]	11.32 (3.45) [25]	-	MD 1.86 higher (0.12 lower to 3.84 higher)	VERY LOW	CRITICAL
Short delay cued recall (CVLT) (follow-up 6 weeks; Better indicated by higher values)												
Hildeb	randomis	very	no serious	no serious	serious[b]	none	13.47 (3) [17]	12.48	-	MD 0.99 higher	VERY	CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General cognitive rehabilitation mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
randt 2007	ed trials	serious[a]	inconsistency	indirectness				(2.95) [25]		(0.85 lower to 2.83 higher)	LOW	L
Long delay free recall (CVLT) (follow-up 6 weeks; Better indicated by higher values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	13.24 (3.35) [17]	12.16 (3.22) [25]	-	MD 1.08 higher (0.95 lower to 3.11 higher)	VERY LOW	CRITICAL
Long delay cued recall (CVLT) (follow-up 6 weeks; Better indicated by higher values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	13.31 (3.16) [17]	12.96 (2.69) [25]	-	MD 0.35 higher (1.49 lower to 2.19 higher)	VERY LOW	CRITICAL
Object alternation RTs (follow-up 6 weeks; Better indicated by lower values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	820 (323) [17]	744 (233) [25]	-	MD 76 higher (102.65 lower to 254.65 higher)	VERY LOW	CRITICAL
Object alternation Errors (follow-up 6 weeks; Better indicated by lower values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	1.18 (1.7) [17]	2.16 (3.04) [25]	-	MD 0.98 lower (2.42 lower to 0.46 higher)	VERY LOW	CRITICAL
Nine Hole Peg Test (follow-up 6 weeks; Better indicated by higher values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-0.134 (0.81) [17]	-0.083 (0.94) [25]	-	MD 0.05 lower (0.58 lower to 0.48 higher)	VERY LOW	CRITICAL
PASAT (MSFC) (follow-up 6 weeks; Better indicated by higher values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	0.017 (0.83) [17]	0.01 (1.09) [25]	-	MD 0.01 higher (0.57 lower to 0.59 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General cognitive rehabilitation mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Beck Depression Inventory (follow-up 6 weeks; Better indicated by lower values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	10.3 (8.5) [17]	11 (7.9) [25]	-	MD 0.7 lower (5.79 lower to 4.39 higher)	VERY LOW	CRITICAL
Fatigue Severity Scale (follow-up 6 weeks; Better indicated by lower values)												
Hildebrandt 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	37.7 (15) [17]	36.8 (14.5) [25]	-	MD 0.7 higher (8.42 lower to 9.82 higher)	VERY LOW	CRITICAL
CVLT total (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	58.4 (13.6) [34]	53.8 (14.3) [27]	-	MD 4.6 higher (2.47 lower to 11.67 higher)	VERY LOW	CRITICAL
CVLT delay (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	12.5 (4.1) [34]	11.4 (4.1) [27]	-	MD 1.1 higher (0.97 lower to 3.17 higher)	VERY LOW	CRITICAL
Brief Visuospatial Memory Test total (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	24.9 (6) [34]	24.6 (6.9) [27]	-	MD 0.3 higher (2.99 lower to 3.59 higher)	VERY LOW	CRITICAL
Brief Visuospatial Memory Test delay (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	9.3 (2.1) [34]	8.8 (2.8) [27]	-	MD 0.5 higher (0.75 lower to 1.75 higher)	VERY LOW	CRITICAL
Judgement of Line Orientation (follow-up 5 months; Better indicated by higher values)												
Stuifb	randomis	very	no serious	no serious	serious[b]	none	27.8 (3.9) [34]	27.4	-	MD 0.4 higher	VERY	CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General cognitive rehabilitation mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
ergen 2012	ed trials	serious[a]	inconsistency	indirectness				(4.2) [27]		(1.66 lower to 2.46 higher)	LOW	L
Symbol Digit Modalities Test (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	49.7 (12.7) [34]	50.6 (13.1) [27]	-	MD 0.9 lower (7.43 lower to 5.63 higher)	VERY LOW	CRITICAL
PASAT-3 (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	47.4 (9.6) [34]	47.2 (10.7) [27]	-	MD 0.2 higher (4.97 lower to 5.37 higher)	VERY LOW	CRITICAL
PASAT-2 (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	34.2 (9.8) [34]	38.1 (9.8) [27]	-	MD 3.9 lower (8.85 lower to 1.05 higher)	VERY LOW	CRITICAL
Controlled Oral Word Association Test (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	36.1 (10.7) [34]	36.4 (12) [27]	-	MD 0.3 lower (6.08 lower to 5.48 higher)	VERY LOW	CRITICAL
Delis-Kaplan Executive Function System (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	39.6 (8.7) [34]	41.7 (10.5) [27]	-	MD 2.1 lower (7.02 lower to 2.82 higher)	VERY LOW	CRITICAL
Self efficacy (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	557.72 (157.84) [34]	534.26 (201.06) [27]	-	MD 23.46 higher (69.09 lower to 116.01 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	General cognitive rehabilitation mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Memory strategy (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious2	none	43.12 (11.93) [34]	41.15 (10.65) [27]	-	MD 1.97 higher (3.71 lower to 7.65 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Neuropsychological Screening Questionnaire (follow-up 5 months; Better indicated by higher values)												
Stuifbergen 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	28.41 (11.13) [34]	26.15 (11.56) [27]	-	MD 2.26 higher (3.49 lower to 8.01 higher)	VERY LOW	CRITICAL

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 93: Clinical evidence profile: Learning versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Learning mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Health-related quality of life – no data												
Mood – no data												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Learning mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Patient reported outcomes – no data												
Patient satisfaction – no data												
Impact on carers – no data												
Hopkins Verbal Learning Test - revised week 0 to 6 (follow-up 6 weeks)												
Chiaravalloti 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	8/14 (57.1%)	35.7 %	RR 1.6 (0.69 to 3.69)	214 more per 1000 (from 111 fewer to 960 more)	VERY LOW	CRITICAL
HVLT - mean change score week 0 to 11 (follow-up 11 weeks; Better indicated by lower values)												
Chiaravalloti 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	3.07 (5.88) [14]	0.57 (4.2) [14]	-	MD 2.5 higher (1.29 lower to 6.29 higher)	VERY LOW	CRITICAL

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 94: Clinical evidence profile: Executive versus control

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Executive mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Health-related quality of life – no data												
Mood – no data												
Patient reported outcomes – no data												
Patient satisfaction – no data												
Impact on carers – no data												
PS TTC post treatment (follow-up 6 weeks; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	33.3 (19) [11]	38.8 (18.7) [14]	-	MD 5.5 lower (20.4 lower to 9.4 higher)	VERY LOW	CRITICAL
Preference Shifting Trials To Criterion 1 yr (follow-up 1 years; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	59.2 (22.5) [6]	45.7 (20.1) [8]	-	MD 13.5 higher (9.26 lower to 36.26 higher)	VERY LOW	CRITICAL
PS Reaction Time (ms) post treatment (follow-up 6 weeks; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	638 (185) [11]	598 (124) [14]	-	MD 40 higher (87.17 lower to 167.17 higher)	VERY LOW	CRITICAL
PS RT (ms) 1 yr (follow-up 1 years; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	685 (142) [6]	734 (196) [8]	-	MD 49 lower (226.08 lower to 128.08 higher)	VERY LOW	CRITICAL
Response Shifting TTC post treatment (follow-up 6 weeks; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	49.3 (23.7) [11]	49.9 (27) [14]	-	MD 0.6 lower (20.5 lower to 19.3 higher)	VERY LOW	CRITICAL
RS TTC 1 yr (follow-up 1 years; Better indicated by lower values)												
Fink	randomise	very	no serious	no serious	very	none	40.4	49.9	-	MD 9.5 lower (40.23	VERY	CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Executive mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
2010	d trials	serious[a]	inconsistency	indirectness	serious[b]		(31.6) [6]	(25.2) [8]		lower to 21.23 higher)	LOW	L
RS RT (ms) post treatment (follow-up 6 weeks; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	656 (219) [11]	676 (170) [14]	-	MD 20 lower (177.1 lower to 137.1 higher)	VERY LOW	CRITICAL
RS RT (ms) 1 yr (follow-up 1 years; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	684 (230) [6]	747 (230) [8]	-	MD 63 lower (306.46 lower to 180.46 higher)	VERY LOW	CRITICAL
2-back com post treatment (follow-up 6 weeks; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	4.2 (6.5) [11]	3.1 (1.6) [14]	-	MD 1.1 higher (2.83 lower to 5.03 higher)	VERY LOW	CRITICAL
2-back com 1 yr (follow-up 1 years; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	4.2 (5.2) [6]	2.2 (1.5) [8]	-	MD 2 higher (2.29 lower to 6.29 higher)	VERY LOW	CRITICAL
2-back om post treatment (follow-up 6 weeks; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	1.5 (0.7) [11]	1.4 (1.2) [14]	-	MD 0.1 higher (0.65 lower to 0.85 higher)	VERY LOW	CRITICAL
2-back om 1 yr (follow-up 1 years; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	1.6 (1.1) [6]	3.5 (1.5) [8]	-	MD 1.9 lower (3.26 to 0.54 lower)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Executive mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
2-back RT (ms) post treatment (follow-up 6 weeks; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	589 (146) [11]	680 (241) [14]	-	MD 91 lower (243.91 lower to 61.91 higher)	VERY LOW	CRITICAL
2-back RT (ms) 1 yr (follow-up 1 years; Better indicated by lower values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	685 (184) [6]	587 (202) [8]	-	MD 98 higher (105.15 lower to 301.15 higher)	VERY LOW	CRITICAL
CVLT learning post treatment (follow-up 6 weeks; Better indicated by higher values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	12.1 (2.1) [11]	11.5 (1.2) [14]	-	MD 0.6 higher (0.79 lower to 1.99 higher)	VERY LOW	CRITICAL
CVLT learning 1 yr (follow-up 1 years; Better indicated by higher values)												
Fink 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	12.5 (2.1) [6]	11.5 (1.1) [8]	-	MD 1 higher (0.85 lower to 2.85 higher)	VERY LOW	CRITICAL

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*
[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 95: Clinical evidence profile: Rehacom versus active control

Quality assessment	No of patients	Effect	Quality	Importa

												nce
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Rehacom mean (SD) [n]**	**Active control mean (SD) [n]**	**Relative (95% CI)**	**Absolute**		
Health-related quality of life – no data												
Mood – no data												
Patient reported outcomes – no data												
Patient satisfaction – no data												
Impact on carers – no data												
Spatial span (Corsi) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	25.4 (21.5) [20]	14.7 (23.1) [20]	-	MD 10.7 higher (3.13 lower to 24.53 higher)	VERY LOW	CRITICAL
Digit span (forward) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	17.8 (22.9) [20]	0 (17.5) [20]	-	MD 17.8 higher (5.17 to 30.43 higher)	VERY LOW	CRITICAL
Digit span (backward) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.8 (29.4) [20]	-1.25 (20) [20]	-	MD 12.05 higher (3.53 lower to 27.63 higher)	VERY LOW	CRITICAL
Paired associates (easy) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.3 (20.5) [20]	1.9 (11.1) [20]	-	MD 8.4 higher (1.82 lower to 18.62 higher)	VERY LOW	CRITICAL
Paired associates (hard) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	59 (87) [20]	21.6 (46.5) [20]	-	MD 37.46 higher (5.83 lower to 80.63 higher)	VERY LOW	CRITICAL
Short story recall % change (follow-up 14 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rehacom mean (SD) [n]	Active control mean (SD) [n]	Relative (95% CI)	Absolute		
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	37.6 (33) [20]	1.55 (23.6) [20]	-	MD 36.05 higher (18.27 to 53.83 higher)	VERY LOW	CRITICAL
Visual reproduction % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	49.1 (48.8) [20]	46.9 (77.1) [20]	-	MD 2.2 higher (37.79 lower to 42.19 higher)	VERY LOW	CRITICAL
LNNB % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	2.5 (3) [20]	0.4 (2.8) [20]	-	MD 2.1 higher (0.3 to 3.9 higher)	VERY LOW	CRITICAL
Recognition memory % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	5.5 (5.4) [20]	6.8 (13.3) [20]	-	MD 1.3 lower (7.59 lower to 4.99 higher)	VERY LOW	CRITICAL
Signal detection no. of hits % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	8.5 (17.9) [20]	3.8 (12.5) [20]	-	MD 4.7 higher (4.87 lower to 14.27 higher)	VERY LOW	CRITICAL
Signal detection reaction time % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	9.4 (10.3) [20]	4.5 (8.8) [20]	-	MD 4.9 higher (1.04 lower to 10.84 higher)	VERY LOW	CRITICAL
Selective reminding test long term storage (follow-up 6 weeks; Better indicated by higher values)												
Cerasa	randomis	very	no serious	no serious	serious[b]	none	36.9	29.9	-	MD 7 higher (2.12	VERY	CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rehacom mean (SD) [n]	Active control mean (SD) [n]	Relative (95% CI)	Absolute		
2013	ed trials	serious[a]	inconsistency	indirectness			(12.46) [12]	(9.8) [11]		lower to 16.12 higher)	LOW	L
Selective reminding test consistent long term retrieval (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	24.86 (11.05) [12]	17.1 (7.3) [11]	-	MD 7.76 higher (0.16 to 15.36 higher)	VERY LOW	CRITICAL
Selective reminding test delayed (follow-up 6 weeks; Better indicated by lower values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	7.11 (2.93) [12]	6.2 (3.02) [11]	-	MD 0.91 higher (1.53 lower to 3.35 higher)	VERY LOW	CRITICAL
Spatial recall test immediate (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	18.42 (6.22) [12]	24.3 (3.99) [11]	-	MD 5.88 lower (10.12 to 1.64 lower)	VERY LOW	CRITICAL
Spatial recall test delayed (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	5.58 (2.47) [12]	8.3 (1.89) [11]	-	MD 2.72 lower (4.51 to 0.93 lower)	VERY LOW	CRITICAL
Word list generation (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	20.8 (5.96) [12]	20.6 (5.59) [11]	-	MD 0.2 higher (4.52 lower to 4.92 higher)	VERY LOW	CRITICAL
Symbol digit modalities test (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	38.69 (9.9) [12]	37.3 (8.45) [11]	-	MD 1.39 higher (6.11 lower to 8.89 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rehacom mean (SD) [n]	Active control mean (SD) [n]	Relative (95% CI)	Absolute		
Stroop test (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	19.41 (5.14) [12]	16.5 (5.22) [11]	-	MD 2.91 higher (1.33 lower to 7.15 higher)	VERY LOW	CRITICAL
Paced auditory serial addition test - 3 (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	41.23 (12.7) [12]	41 (8.79) [11]	-	MD 0.23 higher (8.64 lower to 9.1 higher)	VERY LOW	CRITICAL
Trail making test A (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	44.83 (13.1) [12]	40.9 (13.94) [11]	-	MD 3.93 higher (7.15 lower to 15.01 higher)	VERY LOW	CRITICAL
Trail making test B (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	120.9 (37.9) [12]	121.1 (37.4) [11]	-	MD 0.2 lower (30.99 lower to 30.59 higher)	VERY LOW	CRITICAL
Trail making test B-A (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	76.08 (34.1) [12]	76.9 (30.7) [11]	-	MD 0.82 lower (27.3 lower to 25.66 higher)	VERY LOW	CRITICAL
State trait anxiety inventory-Y1 (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	36.6 (8.9) [12]	41 (11.1) [11]	-	MD 4.4 lower (12.67 lower to 3.87 higher)	VERY LOW	CRITICAL
Stait trait anxiety inventory Y2 (follow-up 6 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rehacom mean (SD) [n]	Active control mean (SD) [n]	Relative (95% CI)	Absolute		
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	35.5 (8.6)	46 (11.1)	-	MD 10.5 lower (18.67 to 2.33 lower)	VERY LOW	CRITICAL
Beck II (follow-up 6 weeks; Better indicated by higher values)												
Cerasa 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	3.94 (4.33) [12]	12.8 (13.5) [11]	-	MD 8.86 lower (17.21 to 0.51 lower)	VERY LOW	CRITICAL
Improvement greater than 20% in 2/5 BRBNT tests (follow-up 16 weeks)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	18/40 (45%)	43.2%	RR 1.04 (0.63 to 1.72)	17 more per 1000 (from 160 fewer to 311 more)	VERY LOW	CRITICAL
Consistent long term retrieval mean % change (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	160 (314) [40]	143 (284) [37]	-	MD 17 higher (116.58 lower to 150.58 higher)	LOW	CRITICAL
Delayed recall mean % change (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	9 (39.7) [40]	44.3 (97) [37]	-	MD 35.3 lower (68.89 to 1.71 lower)	LOW	CRITICAL
Symbol digit modalities mean % change (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	15 (27) [40]	17 (36) [37]	-	MD 2 lower (16.3 lower to 12.3 higher)	MODERATE	CRITICAL
PASAT 2 mean % change (follow-up 16 weeks; Better indicated by higher values)												
Solari	randomis	seriou	no serious	no serious	serious[b]	none	16 (49)	39	-	MD 23 lower	LOW	CRITICA

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rehacom mean (SD) [n]	Active control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
2004	ed trials	s[a]	inconsistency	indirectness			[40]	(101) [37]		(58.91 lower to 12.91 higher)		L
Word list generation mean % change (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	32 (49) [40]	0 (29) [37]	-	MD 32 higher (14.17 to 49.83 higher)	LOW	CRITICAL
Spatial recall immediate recall mean % change (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	17 (53) [40]	27 (67) [37]	-	MD 10 lower (37.13 lower to 17.13 higher)	LOW	CRITICAL
Spatial recall delayed recall mean % change (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	12 (63) [40]	77 (150) [37]	-	MD 65 lower (117.13 to 12.87 lower)	LOW	CRITICAL
MSQOL-54 mean improvements mental health (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	16 (47) [40]	23 (11) [37]	-	MD 7 lower (21.99 lower to 7.99 higher)	VERY LOW	CRITICAL
MSQOL-54 mean improvements cognitive (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	43 (126) [40]	56 (140) [37]	-	MD 13 lower (72.66 lower to 46.66 higher)	LOW	CRITICAL
CMDI mean % change (follow-up 16 weeks; Better indicated by higher values)												
Solari 2004	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-6 (19) [40]	-5 (21) [37]	-	MD 1 lower (9.97 lower to 7.97 higher)	MODERATE	CRITICAL

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 96: Clinical evidence profile: Rehacom versus control

Quality assessment							No of patients		Effect			
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Rehacom mean (SD) [n]**	**Control mean (SD) [n]**	**Relative (95% CI)**	**Absolute**	**Quality**	**Importance**
Health-related quality of life – no data												
Patient reported outcomes – no data												
Patient satisfaction – no data												
Impact on carers – no data												
Beck Depression Inventory (follow-up 3 months; Better indicated by lower values)												
Tesar 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	8.3 (5.8) [10]	8.3 (3.4) [9]	-	MD 0 higher (4.23 lower to 4.23 higher)	VERY LOW	CRITICAL
Fatigue Impact Scale (follow-up 3 months; Better indicated by lower values)												
Tesar 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	41.8 (15.5) [10]	31.7 (18.8) [9]	-	MD 10.1 higher (5.49 lower to 25.69 higher)	VERY LOW	CRITICAL
Card sorting correct (follow-up 3 months; Better indicated by higher values)												
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	42.1 (12.6) [10]	53.9 (21.5) [9]	-	MD 11.8 lower (27.87 lower to 4.27 higher)	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rehacom mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Card sorting incorrect (follow-up 3 months; Better indicated by lower values)												
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	14.1 (4.1) [10]	16.8 (2.2) [9]	-	MD 2.7 lower (5.62 lower to 0.22 higher)	LOW	CRITICAL
Sustained attention correct (follow-up 3 months; Better indicated by higher values)												
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	42.1 (12.6) [10]	53.9 (21.5) [9]	-	MD 11.8 lower (27.87 lower to 4.27 higher)	LOW	CRITICAL
Sustained attention incorrect (follow-up 3 months; Better indicated by lower values)												
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	46.2 (16.1) [10]	51.2 (14.2) [9]	-	MD 5 lower (18.62 lower to 8.62 higher)	VERY LOW	CRITICAL
Sustained attention reaction time (follow-up 3 months; Better indicated by lower values)												
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	42.7 (9.7) [10]	46.8 (7.5) [9]	-	MD 4.1 lower (11.86 lower to 3.66 higher)	LOW	CRITICAL
Sustained attention variation reaction time (follow-up 3 months; Better indicated by lower values)												
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	42.7 (9.7) [10]	46.8 (7.5) [9]	-	MD 5.9 lower (14.73 lower to 2.93 higher)	LOW	CRITICAL
Verbal learning test (follow-up 3 months; Better indicated by higher values)												
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	56.9 (13.1) [10]	50.4 (13.6) [9]	-	MD 6.5 higher (5.54 lower to 18.54 higher)	LOW	CRITICAL
Non-verbal learning test (follow-up 3 months; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rehacom mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	49 (14.9) [10]	48.3 (12.2) [9]	-	MD 0.7 higher (11.5 lower to 12.9 higher)	VERY LOW	CRITICAL
HAWIE-R (follow-up 3 months; Better indicated by higher values)												
Tesar 2005	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	10.6 (2.9) [10]	10.4 (2.1) [9]	-	MD 0.2 higher (2.06 lower to 2.46 higher)	VERY LOW	CRITICAL
Spatial span (Corsi) % change. (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25.4 (21.5) [20]	-1.1 (15.5) [20]	-	MD 26.5 higher (14.88 to 38.12 higher)	LOW	CRITICAL
Digit span (forward) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	17.8 (22.9) [20]	-6.35 (21.1) [20]	-	MD 24.15 higher (10.5 to 37.8 higher)	LOW	CRITICAL
Digit span (backward) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.8 (29.4) [20]	-5.75 (28.2) [20]	-	MD 16.55 higher (1.3 lower to 34.4 higher)	VERY LOW	CRITICAL
Paired associates (easy) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.3 (20.5) [20]	1.1 (10.4) [20]	-	MD 9.2 higher (0.87 lower to 19.27 higher)	VERY LOW	CRITICAL
Paired associates (hard) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendo	randomis	very	no serious	no serious	serious[b]	none	59	2.21	-	MD 56.79 higher	VERY	CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rehacom mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
zzi 1998	ed trials	serious[a]	inconsistency	indirectness			(87) [20]	(64.8) [20]		(9.25 to 104.33 higher)	LOW	L
Short story recall % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	37.6 (33) [20]	22.9 (40.4) [20]	-	MD 14.7 higher (8.16 lower to 37.56 higher)	VERY LOW	CRITICAL
Visual reproduction % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	49.1 (48.8) [20]	-0.7 (21) [20]	-	MD 49.8 higher (26.52 to 73.08 higher)	LOW	CRITICAL
LNNB memory scale % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	2.5 (3) [20]	-0.6 (2.2) [20]	-	MD 3.1 higher (1.47 to 4.73 higher)	LOW	CRITICAL
Recognition memory % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	5.5 (5.4) [20]	-0.4 (9.8) [20]	-	MD 5.9 higher (1 to 10.8 higher)	VERY LOW	CRITICAL
Signal detection (n hits) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	8.5 (17.9) [20]	6.4 (14.8) [20]	-	MD 2.1 higher (8.08 lower to 12.28 higher)	VERY LOW	CRITICAL
Signal detection, reaction times (s) % change (follow-up 14 weeks; Better indicated by higher values)												
Mendozzi 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	9.4 (10.3) [20]	1.7 (9.7) [20]	-	MD 7.7 higher (1.5 to 13.9 higher)	VERY LOW	CRITICAL

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 97 Clinical evidence profile: CogniFit versus control

Quality assessment							No of patients		Effect			
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**CogniFit Mean (SD) [n]**	**Control Mean (SD) [n]**	**Relative (95% CI)**	**Absolute**	**Quality**	**Importance**
Health-related quality of life – no data												
Mood – no data												
Patient reported outcomes – no data												
Patient satisfaction – no data												
Impact on carers – no data												
Divided attention (follow-up 12 weeks; Better indicated by higher values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	very seriousb	none	2.37 (0.78) [22]	2.41 (0.72) [24]	-	MD 0.04 lower (0.47 lower to 0.39 higher)	VERY LOW	CRITICAL
Avoiding distractions (follow-up 12 weeks; Better indicated by lower values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	-0.7 (0.47) [22]	-0.67 (0.69) [24]	-	MD 0.03 lower (0.37 lower to 0.31 higher)	VERY LOW	CRITICAL
Hand-eye coordination (follow-up 12 weeks; Better indicated by higher values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	very seriousb	none	0.26 (1.2) [22]	0.38 (0.99) [24]	-	MD 0.12 lower (0.76 lower to 0.52 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CogniFit Mean (SD) [n]	Control Mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
General memory (follow-up 12 weeks; Better indicated by higher values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	1.13 (0.82) [22]	0.56 (1.1) [24]	-	MD 0.57 higher (0.01 to 1.13 higher)	VERY LOW	CRITICAL
Naming (follow-up 12 weeks; Better indicated by lower values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	0.68 (0.56) [22]	0.54 (0.85) [24]	-	MD 0.14 higher (0.27 lower to 0.55 higher)	VERY LOW	CRITICAL
Response time (follow-up 12 weeks; Better indicated by higher values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	-0.39 (0.74) [22]	-0.51 (0.67) [24]	-	MD 0.12 higher (0.29 lower to 0.53 higher)	VERY LOW	CRITICAL
Shifting attention (follow-up 12 weeks; Better indicated by higher values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	0.37 (0.91) [22]	0.48 (0.62) [24]	-	MD 0.11 lower (0.56 lower to 0.34 higher)	VERY LOW	CRITICAL
Spatial perception (follow-up 12 weeks; Better indicated by higher values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	0.46 (0.69) [22]	0.54 (0.64) [24]	-	MD 0.08 lower (0.47 lower to 0.31 higher)	VERY LOW	CRITICAL
Time estimation (follow-up 12 weeks; Better indicated by lower values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	0.62 (0.61) [22]	0.34 (1) [24]	-	MD 0.28 higher (0.19 lower to 0.75 higher)	VERY LOW	CRITICAL
Visual working memory (follow-up 12 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CogniFit Mean (SD) [n]	Control Mean (SD) [n]	Relative (95% CI)	Absolute		
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	1.15 (0.84) [22]	0.65 (1.03) [24]	-	MD 0.5 higher (0.04 lower to 1.04 higher)	VERY LOW	CRITICAL
Visual scanning (follow-up 12 weeks; Better indicated by higher values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	very seriousb	none	-0.53 (0.74) [22]	-0.57 (0.94) [24]	-	MD 0.04 higher (0.45 lower to 0.53 higher)	VERY LOW	CRITICAL
Verbal auditory working memory (follow-up 12 weeks; Better indicated by higher values)												
Shatil 2010	randomised trials	very serious a	no serious inconsistency	no serious indirectness	seriousb	none	1.09 (0.81) [22]	0.53 (1.02) [24]	-	MD 0.56 higher (0.03 to 1.09 higher)	VERY LOW	CRITICAL

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*

[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 98: Clinical evidence profile: High intensity versus distributed rehabilitation

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	High intensity mean (SD) [n]	Distributed mean (SD) [n]	Relative (95% CI)	Absolute		
Corsi blocks backward (follow-up 4-8 weeks; Better indicated by higher values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	8.87 (2.03) [15]	9.33 (1.58) [15]	-	MD 0.46 lower (1.76 lower to 0.84 higher)	VERY LOW	CRITICAL
Digit span backward (follow-up 4-8 weeks; Better indicated by higher values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	7.87 (2.38) [15]	7.41 (1.72) [15]	-	MD 0.46 higher (1.03 lower to 1.95 higher)	VERY LOW	
2-back, number correct (follow-up 4-8 weeks; Better indicated by higher values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	55.07 (4.02) [15]	57.33 (4.06) [15]	-	MD 2.26 lower (5.15 lower to 0.63 higher)	VERY LOW	CRITICAL
2-back, omissions (follow-up 4-8 weeks; Better indicated by lower values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.4 (0.73) [15]	0.06 (0.26) [15]	-	MD 0.34 higher (0.05 lower to 0.73 higher)	VERY LOW	CRITICAL
2-back reaction time (follow-up 4-8 weeks; Better indicated by lower values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	767.66 (272.31) [15]	666.4 (191.57) [15]	-	MD 101.26 higher (67.23 lower to 269.75 higher)	VERY LOW	CRITICAL
PASAT (follow-up 4-8 weeks; Better indicated by higher values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	50.41 (7.91) [15]	53.61 (5.69) [15]	-	MD 3.2 lower (8.13 lower to 1.73 higher)	VERY LOW	CRITICAL
Corsi blocks forward (follow-up 4-8 weeks; Better indicated by higher values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	9.21 (1.93) [15]	9.22 (1.37) [15]	-	MD 0.01 lower (1.21 lower to 1.19 higher)	VERY LOW	CRITICAL
Digit span forward (follow-up 4-8 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**High intensity mean (SD) [n]**	**Distributed mean (SD) [n]**	**Relative (95% CI)**	**Absolute**		
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	7.2 (2.01) [15]	7.73 (1.94) [15]	-	MD 0.53 lower (1.94 lower to 0.88 higher)	VERY LOW	CRITICAL
Faces Symbols Test (follow-up 4-8 weeks; Better indicated by higher values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	53.87 (14.78) [15]	62.22 (16.22) [15]	-	MD 8.35 lower (19.45 lower to 2.75 higher)	VERY LOW	CRITICAL
Fatigue Scale for Motor and Cognitive Functions (follow-up 4-8 weeks; Better indicated by higher values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	61.73 (19.08) [15]	58 (22.08) [15]	-	MD 3.73 higher (11.04 lower to 18.5 higher)	VERY LOW	CRITICAL
Modified Fatigue Impact Scale (follow-up 4-8 weeks; Better indicated by lower values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	34.13 (17.34) [15]	29.61 (17.09) [15]	-	MD 4.52 higher (7.8 lower to 16.84 higher)	VERY LOW	CRITICAL
Functional Assessment of MS (follow-up 4-8 weeks; Better indicated by lower values)												
Vogt 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	118.61 (34.08) [15]	134.2 (18.57) [15]	-	MD 15.59 lower (35.23 lower to 4.05 higher)	VERY LOW	CRITICAL

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.*

[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Narrative review

General cognitive rehabilitation and psychotherapy versus control

One study[109] (N=32) compared general cognitive rehabilitation and psychotherapy versus control. The data could not be meta-analysed due to the lack of variance data. No statistically significant differences were reported except for visual perception (intervention mean 2.0 vs control 0.6 p=0.04), Beck Depression Inventory (intervention 2.4 vs control 0.0 p=0.04), visual-spatial memory (intervention 2.7 vs control 0.2 p=0.05)

Memory and problem solving versus control

One study[123] (N=149 numbers varied per outcome) compared memory and problem solving rehabilitation with control (patients received an assessment but no intervention). The results were reported as median and inter-quartile ranges. No statistical significant differences were noted.

Attention versus active control

One study[6] (N=102) compared the rehabilitation of attentional processing with an active control. No raw data was presented. A significant improvement from baseline to three month performance that was maintained at 6 mths on the PASAT 3 and 2 in both the specific and non-specific training groups

Learning versus active control

Two studies reported on a learning strategy versus contro[39,40] . No raw data was presented in either study. One study[39] (N=29) found significant improvements in favour of the learning strategy for remember things in everyday life week 0 to 6 (rehabilitation mean change 2.00 vs control -1.29 p<0.01), remember things in everyday life week 0 to 11 (rehabilitation 3.07 vs control -1.86 p<0.001), subjective assessment of ability to remember things in everyday life week 0 to 6 (p<0.01), subjective assessment of ability to remember things in everyday life week 0 to 11 (p<0.01). A second study[40] (N=88) reported significant differences on the California Verbal Learning Test (CVLT) immediate follow up (intervention 95%CI 1.67 to 2.10 vs placebo 1.26 to 1.72, p=0.0075 the treatment effect maintained at follow up) and the (objective everyday memory immediate follow up intervention 95%CI 1.382 to 1.763 vs placebo 1.050 to 1.450, p<0.0115).

Rehacom versus control

Two studies[65,135]reported on the Rehacom intervention versus control. One study[65,65] (N=20) reported median and interquartile ranges and found statistically significant differences in favour of rehabilitation for the Paced Auditory Serial Addition Test (PASAT) 2" (control change score median (lower quartile upper quartile) 0.00 (0.00 12.75) vs rehabilitation 22.00 (17.00 27.00), p=0.004), PASAT 3 change score (control 7.00 (0.00 26.50) vs rehabilitation 36.00 (24.50 44.75), p=0.023), Wisconsin Care Sorting Test total error (WCSTte) (control 45.00 (21.50 62.75) vs rehabilitation 20.00 (15.25 27.50), p=0.037), Montomery-Asberg Depression Rating Scale (MADRS) control 14.00 (8.75 22.50) vs rehabilitation 4.50 (3.00 6.50), p=0.01). One study[135,136] (N=24) reported medians and interquartile ranges and found statistically significant differences in favour of rehabilitation for the PASAT 2" (follow up 3 mths control l 0, (0 11) vs intervention 3 (14 46), p<0.05), PASAT 3" (3 mths control -8, (0 20) vs intervention 8 (17, 41), p<0.05), (9 mths control 0 (3 21) vs intervention (14 (20 30), p<0.05), WCSTpe (3 mths control -23.5 (-6 0) vs intervention -41 (-28 -13) p<0.05), (9 mths control -20.7 (-15 21) vs intervention -45 (-27 19), p<0.05), COWA/P (3 mths control -6 (-3 10) vs intervention (3 (7 17), p<0.05), COWA/S (9 mths Control -3.(5 0) vs intervention 0 (8 21), p<0.05), MADRAS (3 mths control -1.5 (1 -24.5) vs intervention -9 (-4 1), p<0.05), (9 mths control -2.5 (3 28) vs intervention -15 (-8 6), p<0.05), MNSQoL ns (9 mths control -22.5 (-13 46) vs Intervention -17 (33 104), p<0.05).

10.1.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

Relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

10.1.5 Evidence statements

10.1.5.1 Clinical

General cognitive rehabilitation versus control

Low quality evidence from one RCT comprising 98 participants showed that general cognitive rehabilitation was clinically effective compared to control in terms of 10/36 (total correct, delayed recall), with serious imprecision

Moderate to low quality evidence from one RCT (per outcome) comprising between 98 and 78 participants showed that there was no difference in clinical effectiveness between general cognitive rehabilitation and control in terms of the outcomes below, with no serious or serious imprecision:

- Buschke Selective Reminding Test (BSRT)/Long Term Storage (total score)
- BSRT/Consistent Long Term Retrieval (total score)
- BSRT (delayed recall)
- 3 Paced Auditory Serial Additions Test (PASAT) (total correct)
- 2 PASAT (total correct)
- Controlled Oral Word Association Test (total correct)
- Stroop (colour naming time)
- Stroop (colour/word interference-time)
- Trail making A (time)
- Trail making B (time)
- Perceived Deficits Questionnaire
- MSNQ-P, total score
- MSNQ-I, total score
- WHOQOL-BREF S1 physical
- WHOQOL-BREF S2 psychological
- WHOQOL-BREF S3 social relationship
- WHOQOL-BREF S4 environment

Learning and control

Very low quality evidence from one RCT comprising 13 participants showed that learning was clinically effective compared to control in terms of Hopkins verbal learning test no. improvement, with very serious imprecision

Very low quality evidence from one RCT comprising 28 participants showed that was clinically effective compared to control in terms of Hopkins verbal learning test change score, with serious imprecision

Executive versus control

Very low quality evidence from one RCT comprising 14 participants showed that executive was clinically effective compared to placebo in terms of 2-back om one year, with very serious imprecision

Very low quality evidence from one RCT comprising 14 participants showed that placebo was clinically effective compared to executive in terms of PS TTC, with serious imprecision

Very low quality evidence from one RCT comprising 25 participants showed that placebo was clinically effective compared to executive in terms of 2-back com post treatment, with very serious imprecision

Very low quality evidence from one RCT comprising 14 participants showed that placebo was clinically effective compared to executive in terms of 2-back com 1 yr, with very serious imprecision

Very low quality evidence from one RCT (per outcome) comprising 14 to 25 participants showed that there was no difference in clinical effectiveness between executive and control in terms of the outcomes below, with very serious or serious imprecision:

- PS TTC post treatment
- PS RT post treatment
- PS RT one year
- RS TTC post treatment
- RS TTC one year
- RS RT post treatment
- RS RT post treatment one year
- 2-back om post treatment
- 2-back RT one year
- 2-back RT one year
- CVLT learning post treatment
- CVLT learning one year

Rehacom versus active control

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to active control in terms of digit span forward % change, with serious imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to active control in terms of digit span backward % change, with serious imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to active control in terms of paired associates easy % change, with serious imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to active control in terms of paired associates hard % change, with very serious imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to active control in terms of short story recall % change, with very serious imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to active control in terms of LNNB % change, with serious imprecision

Very low quality evidence from one RCT comprising 23 participants showed that Rehacom was clinically effective compared to active control in terms of signal detection reaction time, with serious imprecision

Very low quality evidence from one RCT comprising 23 participants showed that Rehacom was clinically effective compared to active control in terms of selective reminding long term retrieval, with serious imprecision

Very low quality evidence from one RCT comprising 23 participants showed that Rehacom was clinically effective compared to active control in terms of stroop, with very serious imprecision

Very low quality evidence from one RCT comprising 23 participants showed that Rehacom was clinically effective compared to active control in terms of State trait anxiety inventory Y2, with serious imprecision

Very low quality evidence from one RCT comprising 23 participants showed that Rehacom was clinically effective compared to active control in terms of, Beck II with serious imprecision

Low quality evidence from one RCT comprising 97 participants showed that Rehacom was clinically effective compared to active control in terms of word list generation mean % change, with serious imprecision

Very low quality evidence from one RCT comprising 23 participants showed that active control was clinically effective compared to Rehacom in terms of spatial recall immediate, with serious imprecision

Very low quality evidence from one RCT comprising 23 participants showed that active control was clinically effective compared to Rehacom in terms of spatial recall delayed, with serious imprecision

Very low quality evidence from one RCT comprising 97 participants showed that active control was clinically effective compared to Rehacom in terms of MS QOL-54 mean improvement mental health, with very serious imprecision

Very low quality to moderate quality evidence from one RCT (per outcome) comprising 23 to 97 participants showed that there was no difference in clinical effectiveness between Rehacom and active control in terms of the outcomes below, with very serious, serious or no imprecision:

- Spatial span (Corsi) % change
- Visual reproduction % change
- Recognition memory % change
- Signal detection no. of hits % change
- Selective reminding long term storage
- Spatial recall immediate
- Word list generation
- Symbol digit modalities
- Paced auditory serial additions test -3
- Trails test A and B
- State trait anxiety Y1
- BRNT
- Consistent long term retrieval
- Delay recall % change
- Symbol digit modalities % change
- PASAT 2 % change

- Spatial recall immediate recall % change
- Spatial recall delayed recall % change
- MS QOL-54 cognitive
- CMDI % change

Rehacom versus control

Very low quality evidence from one RCT comprising 19 participants showed that Rehacom was clinically effective compared to control in terms of Fatigue Impact Scale, with serious imprecision

Low quality evidence from one RCT comprising 19 participants showed that Rehacom was clinically effective compared to control in terms of card sorting correct, with serious imprecision

Low quality evidence from one RCT comprising 19 participants showed that Rehacom was clinically effective compared to control in terms of card sorting incorrect, with serious imprecision

Low quality evidence from one RCT comprising 19 participants showed that Rehacom was clinically effective compared to control in terms of sustained attention incorrect, with serious imprecision

Low quality evidence from one RCT comprising 19 participants showed that Rehacom was clinically effective compared to control in terms of sustained attention reaction time, with serious imprecision

Very low quality evidence from one RCT comprising 19 participants showed that Rehacom was clinically effective compared to control in terms of sustained attention variation reaction time, with serious imprecision

Low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to control in terms of spatial span (Corsi) % change, with no imprecision

Low quality evidence from one RCT comprising 40participants showed that Rehacom was clinically effective compared to control in terms of digit span forward % change, with no imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to control in terms of spatial span (Corsi) % change, with serious imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to control in terms of paired associates (easy, hard) % change, with serious imprecision

Low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to control in terms of LNNB % change, with no imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to control in terms of recognition memory % change, with serious imprecision

Very low quality evidence from one RCT comprising 40 participants showed that Rehacom was clinically effective compared to control in terms of signal detection (n hits) % change, with no imprecision

Very low quality evidence from one RCT comprising 19 participants showed that Rehacom was clinically effective compared to control in terms of signal detection reaction time % change, with no imprecision

Very low quality to low quality evidence from one RCT (per outcome) comprising 19 to 40 participants showed that there was no difference in clinical effectiveness between Rehacom and control in terms of the outcomes below, with very serious or serious imprecision:

- Beck Depression Inventory
- Sustained attention incorrect
- Verbal learning test
- Non-verbal learning test
- HAWRIE
- Short story recall % change

General cognitive rehabilitation versus control

Very low quality evidence from one RCT comprising 42 participants showed that general cognitive rehabilitation was clinically effective compared to control in terms of short delay free recall, with serious imprecision

Very low quality evidence from one RCT (per outcome) comprising 42 to 61 participants showed that there was no difference in clinical effectiveness between general cognitive rehabilitation and control in terms of the outcomes below, with very serious or serious imprecision:

- SF12 bodily score
- SF12 mental score
- Learning trials
- short delay cued recall
- long delay free recall
- long delay cued recall
- object alternation RTs
- object alternation errors
- nine hole peg test
- PASAT (MSFC)
- Beck Depression Inventory
- Fatigue Severity Scale
- CVLT total, delay
- Brief visuospatial memory test total, delay
- Judgement of line orientation
- Symbol digit modalities test
- PASAT 2, 3
- Controlled oral word association test
- Delis Kaplan executive function system
- Self efficacy
- Memory strategy
- MS Neuropsychological Screening Questionnaire

CogniFit versus control

Very low quality evidence from one RCT (per outcome) comprising 46 participants showed that there was no difference in clinical effectiveness between CogniFit and control in terms of the outcomes below, with very serious or serious imprecision:

- Divided attention
- Avoiding distractions
- Hand-eye coordination

- General memory
- Naming
- Response time
- Shifting attention
- Spatial perception
- Time estimation
- Visual working memory
- Visual scanning
- Verbal auditory working memory

High intensity versus distributed rehabilitation

Very low quality evidence from one RCT comprising 30 participants showed that distributed was clinically effective compared to high intensity in terms of 2 back no correct, with serious imprecision

Very low quality evidence from one RCT comprising 30 participants showed that distributed was clinically effective compared to high intensity in terms of 2 back reaction time, with serious imprecision

Very low quality evidence from one RCT comprising 30 participants showed that distributed was clinically effective compared to high intensity in terms of PASAT, with serious imprecision

Very low quality evidence from one RCT comprising 30 participants showed that distributed was clinically effective compared to high intensity in terms of Faces symbols test, with serious imprecision

Very low quality evidence from one RCT comprising 30 participants showed that distributed was clinically effective compared to high intensity in terms of functional assessment of MS, with serious imprecision

Very low quality evidence from one RCT (per outcome) comprising 30 participants showed that there was no difference in clinical effectiveness between high intensity and distributed rehabilitation in terms of the outcomes below, with very serious or serious imprecision:

- C orsi block backward, forward
- Digit span backward
- 2 back, omissions
- Fatigue scale for motor and cognitive function
- Functional assessment of MS

10.1.5.2 Economic

No relevant economic evaluations were identified.

10.1.6 Recommendations and link to evidence

Recommendations	**46. Be aware that the symptoms of MS can include cognitive problems, including memory problems that the person may not immediately recognise or associate with their MS.** **47. Be aware that anxiety, depression, difficulty in sleeping and fatigue can impact on cognitive problems. If a person with MS experiences these symptoms and has problems with memory**

	and cognition, offer them an assessment and treatment. **48. Consider referring people with MS and persisting memory or cognitive problems to both an occupational therapist and a neuropsychologist to assess and manage these symptoms.**
Relative values of different outcomes	A wide range of cognitive functions were reported in the studies with very little commonality across the studies for each intervention. Whilst these outcomes were validated tests of cognitive function they lacked ecological validity and were not thought to reflect those that would lead to improvements in everyday life. The majority of interventions did not report any quality of life data.
Trade off between clinical benefits and harms	Neuropsychological rehabilitation is unlikely to result in any harms but it was noted that the rehabilitation interventions reported in the studies involved a considerable investment of time although a number of the interventions were delivered in the home environment.
Economic considerations	No relevant economic evaluation studies comparing non-pharmacological management of cognition (including memory) were found. One computerised programme, Rehacom, was assessed in five studies in the clinical review. The costs, based on a price list published by the manufacturer, were presented to the GDG. The software costs start at £88 per licence per procedure (module) and hardware varies between £78 and £475 depending on the chosen technology. According to the studies, a trained psychologist assists with the set-up and running of the sessions. The unit cost of a clinical psychologist is £60 per hour and £136 per patient contact hour. The GDG discussed that in practice; occupational therapists may be assisting with the set-up and running of sessions instead of psychologists. The unit cost of a hospital occupational therapist is £32 per hour. The GDG agreed that additional clinical evidence is required to justify the cost of these computer-based interventions and therefore made a recommendation for further research. There are costs associated with assessing and treating people with evidence of memory and cognitive problems for anxiety, depression, difficulty sleeping and fatigue as well as referral to a psychologist or a memory service. The GDG considered that this was standard practice for people with or without MS and they wanted to reinforce the importance of addressing these needs.
Quality of evidence	14 parallel and crossover RCTs were included in the review. These covered a wide range of interventions including strategies to improve learning, executive function techniques and computerised programmes aimed at improving a range of cognitive functions. Outcomes were graded at low or very low quality. General rehabilitation programmes were unsuccessful in remediating cognitive functions. Patients who participated in a computerised training programme, Rehacom, did improve on some outcomes including digit span and paired associates but there was uncertainty around the estimate of effect. This in conjunction with lack of data on functional outcomes led the GDG to make a research recommendation
Other considerations	The GDG discussed the importance of addressing issues such as sleep quality and fatigue, and having appropriate assessment and treatment of mood disorders as these may all affect cognitive function. This reflects current good practice and a recommendation was made to reinforce this. The GDG acknowledged that cognitive symptoms impact significantly on work, home and social activities and on family and carers. The GDG considered that the evidence available lacked functional outcomes of importance to people with MS. Activities of daily living and the ability to achieve goals are likely to be important to patients. The GDG considered that interventions to improve these are already carried out by occupational

	therapists e.g. breaking down tasks and using technology to provide reminders of tasks. The GDG agreed that people with MS who have cognitive or memory problems should have access to appropriate assessment by neuropsychologist, memory service or similar expertise according to local availability. These services should be considered part of the multidisciplinary care that may be required for people with MS and appropriate communication channels should be available. A recommendation was made to reinforce current good practice. There can also be a role for psychological input in helping people come to terms with either the diagnosis of MS and/or the distress associated with physical and cognitive disability The GDG agreed that the research recommendation should include the following: • function and patients orientated outcomes • Fewer focussed outcome measures • Adequately powered sample sizes • Include people with mild and moderate MS • It should be recurrent • Goal oriented

10.2 Non-pharmacological management of ataxia and tremor

10.2.1 Introduction

Ataxia and tremor are common symptoms in MS. Ataxia is commonly used as an umbrella term to cover both symptoms. There are a number of possible reasons by people with MS suffer from ataxia e.g. disease affecting cerebellum, sensory nerves and visual loss.

10.2.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological programmes (including self-management programmes) for ataxia and/or tremor?

For full details see review protocol in Appendix C.

Table 99: PICO characteristics of review question

Population	Adults with MS only
Intervention/s	Any non-pharmacological management programme, including self-management programmes , for example: • Multidisciplinary rehabilitation/programmes • Self-management programmes • Treatment programmes for various symptoms • FACETS programmes, energy conservation programs, mindfulness (Grossman Paul), exercise (John Saxton), Getting To Grips (MS Society), stretching, standing, splinting, gym prescription, diet, yoga, tai chi, Pilates, relaxation, lycra garments
Comparison/s	Usual treatment or placebo
Outcomes	• ataxia [symptoms or measures (ie ICARS)] • tremor [symptoms or measures Also, any of the following outcomes, provided the treatment has been directed at **ataxia or tremor**: • Quality of life • Function (i.e. EDSS, ambulation measures, MSIS, Guys scale, etc.) • carer perceptions • Incidence of adverse events

10.2.3 Clinical evidence

10.2.3.1 Tremor

Summary of included studies

No RCTs were found covering the non-pharmacological management of tremor.

10.2.3.2 Ataxia

Summary of included studies

Two RCTs[10,112] were found that covered the non-pharmacological management of ataxia in people with MS.

Table 100: Summary of studies included in the review

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised/analysed	Analysis
Armutlu 2001[10]	Conventional PNF-based neurorehabilitation combined with Johnstone Pressure Splints / Conventional PNF-based neurorehabilitation only	10/26 primary progressive; 16/26 secondary progressive; EDSS 4.53/4.88	26/26	Parallel
Keser 2013[112]	Bobath neurorehabilitation / conventional physiotherapy	EDSS 2.8/2.85; Disease duration 4.45years/8.25 years	23/20	Parallel

Table 101: Clinical evidence profile for Bobath versus conventional physiotherapy

Quality assessment							Mean (sd) [n] for change from baseline		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Bobath conventional physiotherapy	Conventional physiotherapy	Relative (95% CI)	Absolute		
ICAR total (change from baseline) (Better indicated by lower values)												
Keser 2013	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	-6.6(4.59)[10]	-6.4(4.59)[10]	-	MD 0.2 lower (4.22 lower to 3.82 higher)	VERY LOW	CRITICAL
ICAR1 – posture and stance (change from baseline) (Better indicated by lower values)												
Keser 2013	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	Serious[C]	none	-3.5(2.22)[10]	-2.4(1.71)[10]	-	MD 1.1 lower (2.84 lower to 0.64 higher)	VERY LOW	CRITICAL
ICAR2 – limb movement (change from baseline) (Better indicated by lower values)												
Keser 2013	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	-2.8(2.85)[10]	-3.3(4.21)[10]	-	MD 0.5 higher (2.65 lower to 3.65 higher)	VERY LOW	CRITICAL
ICAR3 – speech disorders (change from baseline) (Better indicated by lower values)												
Keser 2013	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	-0.5(0.84)[10]	-0.6(0.96)[10]	-	MD 0.1 higher (0.69 lower to 0.89 higher)	VERY LOW	CRITICAL
ICAR4 – oculomotor problems (change from baseline) (Better indicated by lower values)												
Keser 2013	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	0.2(1.03)[10]	-0.1(0.31)[10]	-	MD 0.3 higher (0.37 lower to 0.97 higher)	VERY LOW	CRITICAL
MSFTC (change from baseline) (Better indicated by higher values)												
Keser 2013	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	0.3(0.28)[10]	0.26(0.25)[10]	-	MD 0.04 higher (0.19 lower to 0.27 higher)	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] for change from baseline		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Bobath conventional physiotherapy	Conventional physiotherapy	Relative (95% CI)	Absolute		
Quality of life												
No evidence available												
Carer perceptions												
No evidence available												
Adverse events												
No evidence available												

[A] No reporting of allocation concealment, blinding and potential attrition bias.
[B] Both MIDs (+/-0.5 x sd of control group) were crossed by the lower and upper CIs
[C] Lower MIDs (-0.5 x sd of control group) was crossed by the lower CI

Table 102: Clinical evidence profile for Conventional physiotherapy with splinting versus conventional physiotherapy alone

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Conventional physiotherapy with splinting	Conventional physiotherapy alone	Relative (95% CI)	Absolute		
Step width – can be a proxy measure for ataxia (Better indicated by lower values)												
Armutlu 2001	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	12.9(2.06)[13]	13.4(3.3)[13]	-	MD 0.5 lower (2.61 lower to 1.61 higher)	VERY LOW	CRITICAL
Single limb stance time (right) (Better indicated by higher values)												
Armutlu 2001	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	No serious imprecision	none	36.1(18.6)[13]	17.7(12.5)[13]	-	MD 18.4 higher (6.22	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Conventional physiotherapy with splinting	Conventional physiotherapy alone	Relative (95% CI)	Absolute		
										to 30.58 higher)		
Single limb stance time (left) (Better indicated by higher values)												
Armutlu 2001	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	35.1(18.6)[13]	16.4(15.7)[13]	-	MD 18.7 higher (5.47 to 31.93 higher)	VERY LOW	CRITICAL
Time to walk 3m (s) (Better indicated by lower values)												
Armutlu 2001	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	3.42(1.23)[13]	2.88(0.74)[13]	-	MD 0.54 higher (0.24 lower to 1.32 higher)	VERY LOW	CRITICAL
Ambulation index (Better indicated by lower values)												
Armutlu 2001	randomised trials	very serious[A]	no serious inconsistency	no serious indirectness	very serious[C]	none	2.07(0.49)[13]	2.0(0.4)[13]	-	MD 0.07 higher (0.27 lower to 0.41 higher)	VERY LOW	CRITICAL
Quality of life												
No evidence available												
Carer perceptions												
No evidence available												
Adverse events												
No evidence available												

[A] *No reporting of allocation concealment or blinding, and large baseline differences in age and disease duration (favouring intervention group)*
[B] *One MIDs (+/-0.5 x sd of control group) was crossed by one CI*
[C] *Both MIDs (+/-0.5 x sd of control group) were crossed by the lower and upper CI*

Narrative review of results

Armutlu measured co-ordination, an important aspect of ataxia, with ‘equilibrium co-ordination tests’ (using footprints of the number of steps taken outside a support base of 10cm and during tandem walking) and ‘non-equilibrium co-ordination tests’ (using the knee-heel test, the dysdiadakokinesia test, and the number of pendular movements of a limb after requested movement). However, no between-group data were provided, apart from the fact that no significant differences were noted between groups.

10.2.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

Relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness

10.2.5 Evidence statements

10.2.5.1 Clinical

Bobath versus conventional physiotherapy

Very low quality evidence from one RCT comprising 20 participants showed that Bobath was clinically effective compared to conventional physiotherapy in terms of ICAR (posture and stance) score, with serious imprecision.

Very low quality evidence from one RCT comprising 20 participants showed that Bobath was clinically harmful compared to conventional physiotherapy in terms of ICAR (oculomotor) score, with very serious imprecision.

Very low quality evidence from one RCT comprising 20 participants showed that there was no difference in clinical effectiveness between Bobath and conventional physiotherapy in terms of ICAR (total) score, ICAR (limb movement) score, ICAR (speech disorders) score, MSFC score, with very serious imprecision.

Conventional physiotherapy plus splinting versus conventional physiotherapy

Low quality evidence from one RCT comprising 26 participants showed that conventional physiotherapy plus splinting was clinically effective compared to conventional physiotherapy in terms of right single limb stance time, with no serious imprecision.

Low quality evidence from one RCT comprising 26 participants showed that conventional physiotherapy plus splinting was clinically effective compared to conventional physiotherapy in terms of right single limb stance time, with serious imprecision.

Very low quality evidence from one RCT comprising 26 participants showed conventional physiotherapy plus splinting was clinically effective compared to conventional physiotherapy in terms of left single limb stance time, with serious imprecision.

Very low quality evidence from one RCT comprising 26 participants showed that conventional physiotherapy plus splinting was clinically harmful compared to conventional physiotherapy in terms of time to walk 3m, with serious imprecision.

Very low quality evidence from one RCT comprising 26 participants showed that there was no clinical difference between Conventional physiotherapy plus splinting and conventional physiotherapy in terms of step width and ambulation index, with serious to very serious imprecision.

10.2.5.2 Economic

No relevant economic evaluations were identified.

10.2.6 Recommendations and link to evidence

Recommendations	
Relative values of different	A measure of ataxia/tremor was the most critical outcome as this was the most directly relevant outcome to this question. Quality of life, function, carer

outcomes	perceptions and adverse events were also regarded as critical.
Trade off between clinical benefits and harms	Bobath versus conventional physiotherapy The only clinically significant benefit for Bobath treatment over conventional physiotherapy was observed for the posture and stance sub-scale of the ICARS. In contrast, Bobath had a clinically significant harm compared to conventional physiotherapy in terms of the oculomotor sub-scale of the ICARS. No adverse effects were reported for either intervention, but overall the relative harms of Bobath treatment appeared to nullify its benefits. Conventional physiotherapy with splinting treatment versus conventional physiotherapy alone The conventional physiotherapy with Johnstone Pressure Splints treatment did show clinically important benefits for R and L stance time over conventional physiotherapy alone, but this was regarded as too indirect a measure of ataxia to influence recommendations. For the more directly relevant measures of 'co-ordination' no benefits were shown for conventional physiotherapy with splinting compared to conventional physiotherapy alone. No harms were identified for either approach.
Economic considerations	No relevant economic evaluation studies comparing non-pharmacological treatment of ataxia and tremor were found. Relevant unit costs, based on the resource use of the interventions found in the clinical evidence were presented. The unit cost of a hospital based physiotherapist was £31 per hour. The unit cost of Johnstone pressure splints were between £115.10–120.33 depending on the size of the splint. The GDG agreed that additional clinical evidence is required to justify the cost of these interventions and therefore made a recommendation for further research.
Quality of evidence	The evidence was of very low quality in the two included studies, due to a lack of allocation concealment and adequate blinding in both. Furthermore the Bobath study had likely attrition bias and the splinting study had very unequal groups at baseline. All evidence was focussed on ataxia, and no studies looked specifically at tremor.
Other considerations	The GDG presumed that clinical and other assessment would be carried out if there was any concern that ataxia was not related to MS. People with ataxia need to be assessed by relevant healthcare professionals and individualised treatment agreed. The GDG considered that ataxia is often used as an umbrella term for disorder of coordination including tremor and it may be more appropriate to consider specific ataxias e.g. upper limb or trunk and specific deficits contributing to ataxia i.e. cerebellar damage, sensory loss, eye coordination problems. The GDG felt that there was insufficient evidence to be able to recommend specific non-pharmacological management programmes for ataxia and tremor. Many small scale studies have shown promising results with both compensatory and restorative approaches e.g. cooling, weights, lycra, wheeled walking aids. Researchers need to concentrate on specific ataxias e.g. upper limb, trunk and specify deficits contributing to ataxia and evidence is needed on dose of exercise needed to bring about benefit. The GDG developed a research recommendation for RCTs to evaluate the benefits and harms of different non-pharmacological treatments for ataxia and tremor.

10.3 Non-pharmacological management of fatigue

10.3.1 Introduction

Excessive fatigue may affect up to 80% of people with MS. The level of fatigue can be overwhelming, and is usually out of proportion to prior activity levels. Such fatigue may be a direct effect of the disease process on the central nervous system, or may be secondary to weakness, stiffness, tremor, disturbed sleep or depression. Some medications may have a beneficial effect on MS fatigue, but they do not help all people and may also have adverse effects. Non-pharmacological methods may therefore also be useful to help manage this disabling symptom. It is possible that pharmacological and non-pharmacological methods may influence different aspects of fatigue, and so their combined use may be complementary.

10.3.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological programmes (including self-management programmes) for fatigue?

For full details see review protocol in Appendix C.

Table 103: PICO characteristics of review question

Population	Adults with MS only
Intervention/s	Any non-pharmacological management programme, including self-management programmes , for example: • Multidisciplinary rehabilitation/programmes • Self-management programmes • Treatment programmes for various symptoms • Fatigue management programmes, FACETS programme, energy conservation programs, mindfulness based training(Grossman Paul), exercise (John Saxton), Getting To Grips (MS Society), stretching, standing, splinting, gym prescription, diet, yoga, tai chi, Pilates, relaxation, lycra garments
Comparison/s	Usual treatment or placebo
Outcomes	• Fatigue [symptoms or measures (i.e. FSS)] Also, any of the following outcomes, provided the treatment has been directed at **fatigue**: • Quality of life • Function (i.e. EDSS, ambulation measures, MSIS, Guys scale etc.) • carer perceptions • Incidence of adverse events
Study types	Systematic reviews, RCTs. Include cross-over studies.

10.3.3 Clinical evidence

Summary of included studies

38 published articles[2,3,24,28,34,35,49,53,54,62,63,71-75,78,85,92,96,97,102,111,115,119,120,134,138,157,158,166,179,180,183,192,194,241,242,245,252,255,256] covering 33 RCTs were found, all of which looked at non-pharmacological approaches to reduce fatigue in a population of people with MS. There were 19 categories of non-pharmacological interventions covered in these papers. 14 of these were predominantly physical approaches and 5 were cognitive or psychological approaches. Most studies compared interventions to a non-treatment control group, and there were 20 different combinations of approach and comparator. These are summarised in Table 104.

The physical approaches were:

- Resistance training
- Aerobic training
- Mixed resistance/aerobic training
- Mixed resistance/aerobic training with cognitive behavioural therapy
- Supervised resistance/balance training
- Home-based resistance/balance training
- High level resistance plus standard exercise versus exercise
- Vestibular rehabilitation training
- Yoga
- Electro-magnetic field therapy
- Individualised rehabilitation
- Neurorehabilitation
- Massage
- Wii balance board training

The cognitive or psychological approaches were:

- Cognitive behavioural therapy
- Fatigue management/energy conservation
- Mindfulness' based training, based on standard Mindfulness Based Stress Reduction (MBSR)
- 'Motivational interviewing'
- Group wellness intervention

Table 104: Summary of studies included in the review

Study	Intervention/comparison	Population characteristics*	N randomised /analysed
Dalgas 2010A[49]	Resistance training versus control	RR; EDSS 3-5.5; pyramid function score ≥ 2	39/34
Dodd 2011[54]		AI score of 2-4; 41/71 MFIS > 38	76/71
Tarakci 2013[242]		EDSS 2-6.5; FSS 39.3/39.9; mostly RR	114/95
Ahmadi 2013[2,3]	Aerobic training versus control	EDSS 1-4; DMDs allowed; disease duration 5 years	20/20
Dettmers 2009[53]		MFIS 36.8/41.8; EDSS <4.5; mostly female; mostly RR	30/30, but depended on

Study	Intervention/comparison	Population characteristics*	N randomised /analysed
			outcome
Geddes 2009[75]		EDSS 4.7; 75% female; MS>1 year	15/12
Gervasoni 2014[78]		EDSS 5/5.5; 15 years since onset; 12/30 female; RR: 37%/54.6%	30/30
Hebert 2011[97]		MFIS $\geq$ 45; ambulant >100m with/without aids;	26/26
McCullagh 2008[138]		Independently mobile without assistance; disease duration 5.4/5; MFIS 26/26.5	30/24
Mostert 2002[158]		EDSS 1-6.5; mostly relapsing progressive	37/26
Van den Berg 2006[252]		Able to walk 10m in <60 secs;	19/16
Kargarfard 2012[111]		Women only; RRMS; min 2 years since diagnosis; EDSS 2.9/3.	32/21
Rampello 2007[192]	Aerobic training versus neurorehabiliation	EDSS<7; aged 20-55	11/11
Learmonth 2012[120]	Mixed aerobic/resistance versus control	EDSS 5-6.5; MMSE >24; mostly female; years since onset 13.4/12.6	32/25
Hayes 2011[92]		18-65; ambulatory with/without assistive devices	22/19
Garrett 2013A[73,74]		Aged c50; mostly RR;	151/112
Negahban 2013[166]		EDSS 3.5/3.8; Time since diagnosis 102/87 months	
Surakka 2004[241]		EDSS 1-5.5; FSS 4.6; mostly RR; 6 years since diagnosis	110/99
Carter 2014[35]	Mixed aerobic/resistance + CBT versus usual care	EDSS 3.8; MFIS (total) 45/42.8	120/99
Cakit 2010[34]	Supervised resistance/balance versus control	EDSS$\leq$6; able to stand independently > 3 secs;	30/23
	home based resistance/balance versus control	EDSS$\leq$6; able to stand independently > 3 secs;	30/19
	Supervised versus home based resistance/balance	EDSS$\leq$6; able to stand independently > 3 secs;	30/24
Hayes 2011[92]	High resistance + standard exercise versus standard exercise	EDSS 5.2; age 49(11); 11/19 women	20/19
Garrett2013A[73,74]	Yoga versus mixed resistance/aerobic	Aged c50; mostly RR;	157/126
Garrett 2013[74,74]			157/79
Ahmadi 2013[2,3]	Yoga versus aerobic	EDSS 1-4; DMDs allowed; disease duration	21/21

Study	Intervention/comparison	Population characteristics*	N randomised /analysed
		5 years	
Ahmadi 2010, 2013[2];[3]	Yoga versus control	EDSS 1-4; DMDs allowed; disease duration 5 years	21/21
Garret 2013A[73,74]		Aged c50; mostly RR;	148/112
Velikonja 2010[256]	Resistance versus yoga	RR, PP or SP; 26-50 years, EDSS <7; $EDSS_{pyr}$ >2	10/10
Hebert 2011[97]	Vestibular rehab versus control	MFIS $\geq$ 45; ambulant >100m with/without aids;	25/25
Kargarfard 2012[111]	Hydrotherapy versus control	Women only; RRMS; min 2 years since diagnosis; EDSS 2.9/3.	32/21
Piatkowski 2009[179,180]	Electromagnetic field therapy versus placebo device	MFIS: 32/38; EDSS 3.7/3.1; mostly female	41/37
Richards 1997[194]		EDSS 5.13/4.98	30/30
Lappin 2003[119]		72% had duration MS $\geq$ 4yrs; 57% moderately disabled or worse	145/117
Brichetto 2013[28]	Wii balance board vs control	22/36 female; EDSS 3.9/4.3; Disease duration 11.2/12.3 years	36/36
Negahban 2013[166]	Massage versus massage and exercise vs exercise only vs usual care	EDSS 3.8; Time since diagnosis 87-149 months (range between groups); age 36	48/48
Plow 2009[183]	Individualised rehab versus group wellness intervention	Able to walk with or without assistive device, and physician-confirmed diagnosis of MS.	50/42
Moss-Morris 2012[157]	CBT versus usual care	FS>4; ambulant > 100m; mostly female; 21/16 years since diagnosis	45/40
Van Kessel 2008[255]		Auckland; mainly RR; EDSS$\leq$6; fatigue scale $\geq$4; DMDs allowed	72/72
Finlayson 2011[63]	Fatigue management/energy conservation versus control	FSS$\geq$4; 15 years since diagnosis; 79% female; mostly PP	190/181
Garcia 2013[71]		EDSS $\leq$ 6; FSS$\geq$4; mostly SP; MS duration 11/14.2	23/23
Hugos 2010[102]		EDSS $\leq$ 6; No DMDs allowed within 6 months before study; EDSS 4.9/5.5	41/30
Kos 2007[115]		3 or more on fatigue sub-scale of GNDS; ambulant >100m without assistance; mostly RR	51/51
Mathiowetz 2005[134]		FSS$\geq$4; independent in community; mostly RR; 83% female	169/169
Thomas 2013[245]		FSS>4; mostly RR; 73% female	164/146
Grossman 2010[85]	Mindfulness training versus control	EDSS <7; Mostly RR; allowed to be on DMDs; MFIS 35/30	150/138
Bombardier 2008[24]	Motivational interviewing versus control	EDSS<6; able to walk 90m without assistance; all types of MS	130/130

*Where group-specific data is reported by the study, data for each group is separated by a forward slash. For example, ‘age: 42/44’ indicates a mean age of 42 for the intervention group and 44 for the control group.

Table 105: Clinical evidence profile: Resistance training versus control

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Resistance training**	**Control**	**Relative (95% CI)**	**Absolute**		
FATIGUE OUTCOMES												
MFIS total change from baseline to 10-12 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-10.2(11.2)[36]	-3(14.1)[35]	-	MD 7.2 lower (13.13 lower to 1.27 lower)	VERY LOW	CRITICAL
MFIS (phys) change from baseline to 10-12 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-5.9(5.9)[36]	-1.8(6.8)[35]	-	MD 4.1 lower (7.06 lower to 1.14 lower)	VERY LOW	CRITICAL
MFIS (cog) change from baseline to 10-12 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-3.2(5.9)[36]	-1.7(6.9)[35]	-	MD 1.5 lower (4.49 lower to 1.49 higher)	VERY LOW	CRITICAL
MFIS (psychosocial)change from baseline to 10 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-1.1(1.6)[36]	-0.4(2.4)[35]	-	MD 0.7 lower (1.65 lower to 0.25 higher)	VERY LOW	CRITICAL
MFIS total change from baseline to 22 weeks (Better indicated by lower values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance training	Control	Relative (95% CI)	Absolute		
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-2.9(12.8)[36]	-4.8(12.4)[35]	-	MD 1.9 higher (3.96 lower to 7.76 higher)	VERY LOW	CRITICAL
MFIS (phys) change from baseline to 22 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-2.6(6.8)[36]	-2.1(5.4)[35]	-	MD 0.5 lower (3.35 lower to 2.35 higher)	VERY LOW	CRITICAL
MFIS (cog) change from baseline to 22 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-0.2(7)[36]	-2.1(6.3)[35]	-	MD 1.9 higher (1.2 lower to 5 higher)	VERY LOW	CRITICAL
MFIS (psychosocial)change from baseline to 22 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-0.1(2)[36]	-0.5(2.2)[35]	-	MD 0.4 higher (0.58 lower to 1.38 higher)	VERY LOW	CRITICAL
FSS change from baseline at 12 weeks(Better indicated by lower values)												
Dalgas 2010A Tarakci 2013	RCT	Very serious[A]	Very serious inconsistency[C]	No serious indirectness	No serious imprecision	none	MD(SE): -0.7(0.327) MD(SE): -11.55(3.36)			MD 0.8 lower (1.44 lower to 0.16 lower)	VERY LOW	CRITICAL
MFI-20 general fatigue change from baseline at 12 weeks(Better indicated by lower values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance training	Control	Relative (95% CI)	Absolute		
Dalgas 2010A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	MD(SE): -2.9(0.935)			MD 2.9 lower (4.73 lower to 1.07 lower)	VERY LOW	CRITICAL
MFI-20 physical fatigue change from baseline at 12 weeks(Better indicated by lower values)												
Dalgas 2010A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	MD(SE): -1.8(0.923)			MD 1.8 lower (3.61 lower to 0.01 higher)	LOW	CRITICAL
Change in MUSIQOL from baseline to 12 weeks (Better indicated by higher values)												
Tarikci 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	1.98(5)[51]	-0.4(5)[48]	-	MD 2.38 higher (0.41 to 4.35 higher)	VERY LOW	CRITICAL
QUALITY OF LIFE OUTCOMES												
SF-36 physical change from baseline at 12 weeks(Better indicated by higher values)												
Dalgas 2010A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	None	MD(SE): 4.5(1.64)			MD 4.5 higher (1.29 higher to 7.71 higher)	LOW	CRITICAL
SF-36 mental change from baseline at 12 weeks(Better indicated by higher values)												
Dalgas 2010A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	MD(SE): 4.4(2.52)			MD 4.4 higher (0.54 lower to 9.34 higher)	VERY LOW	CRITICAL
WHOQoL overall change from baseline to 10 weeks (Better indicated by higher values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Resistance training**	**Control**	**Relative (95% CI)**	**Absolute**		
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0.4(0.9)[36]	0.1(0.8)[35]	-	MD 0.3 higher (0.1 lower to 0.7 higher)	VERY LOW	CRITICAL
WHOQoL overall change from baseline to 22 weeks (Better indicated by higher values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-0.1(1.1)[36]	0.1(0.8)[35]	-	MD 0.2 lower (0.65 lower to 0.25 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
fast walking speed (m/s) change from baseline to 10 weeks (Better indicated by higher values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0.05(0.17)[36]	0.01(0.19)[35]	-	MD 0.04 higher (0.04 lower to 0.12 higher)	VERY LOW	CRITICAL
2 min walking distance (m) change from baseline to 10 weeks (Better indicated by higher values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	2.8(14.4)[36]	0.7(13.40)[35]	-	MD 2.1 higher (4.37 lower to 8.57 higher)	VERY LOW	CRITICAL
fast walking speed (m/s) change from baseline to 22 weeks (Better indicated by higher values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-0.02(0.19)[36]	0.01(0.18)[35]	-	MD 0.03 lower (0.12 lower to 0.06 higher)	VERY LOW	CRITICAL
2 min walking distance (m) change from baseline to 22 weeks (Better indicated by higher values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance training	Control	Relative (95% CI)	Absolute		
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-1.6(15.6)[36]	1.6(9)[35]	-	MD 3.2 lower (9.1 lower to 2.7 higher)	VERY LOW	CRITICAL
10 min walking distance (m) change from baseline to 12 weeks (Better indicated by higher values)												
Tarikci 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-4.73(9.1)[51]	1.45(9.1)[48]	-	MD 6.18 lower (9.75 to 2.61 lower)	VERY LOW	CRITICAL
ADVERSE EVENTS												
AE - stiffness MSIS-88 overall change from baseline to 10 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-3.6(7.6)[36]	-0.5(6)[35]	-	MD 3.1 lower (6.28 lower to 0.08 higher)	VERY LOW	CRITICAL
AE - muscle spasm MSIS-88 overall change from baseline to 10 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-2(6.2)[36]	0.5(6)[35]	-	MD 2.5 lower (5.34 lower to 0.34 higher)	VERY LOW	CRITICAL
AE - stiffness MSIS-88 overall change from baseline to 22 weeks (Better indicated by lower values)												
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-0.5(7)[36]	-0.7(7.7)[35]	-	MD 0.2 higher (3.23 lower to 3.63 higher)	LOW	CRITICAL
AE - muscle spasm MSIS-88 overall change from baseline to 22 weeks (Better indicated by lower values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Resistance training**	**Control**	**Relative (95% CI)**	**Absolute**		
Dodd 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	1.1(8.2)[36]	-1.1(7.5)[35]	-	MD 2.2 higher (1.45 lower to 5.85 higher)	VERY LOW	CRITICAL
CARER PERCEPTIONS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.
[C] Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.

Table 106: Clinical evidence profile: aerobic training versus control

Quality assessment	Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	aerobic	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
MFIS total change from baseline 8 weeks (Better indicated by lower values)												
Kargarfard 2012	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-9.8(10.1)[10]	15.3(8)[11]	-	MD 25.1 lower (32.94 to 17.26 lower)	LOW	CRITICAL
MFIS (phys) change from baseline 8 weeks (Better indicated by lower values)												
Kargarfard 2012	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-5.2(5.4)[10]	8.8(4.6)[11]	-	MD 14 lower (18.31 to 9.69 lower)	LOW	CRITICAL
MFIS (psychosocial) change from baseline 8 weeks (Better indicated by lower values)												
Kargarfard 2012	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-2.7(7)[10]	5.9(8.3)[11]	-	MD 8.6 lower (15.15 to 2.05 lower)	VERY LOW	CRITICAL
MFIS (cog) change from baseline 8 weeks (Better indicated by lower values)												
Kargarfard 2012	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-1.9(1.9)[10]	0.5(2)[11]	-	MD 2.4 lower (4.07 to 0.73 lower)	VERY LOW	CRITICAL
MFIS at 6 weeks (Better indicated by lower values)												
Hebert 2011	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	44.3(16.40[13]	52.1(17.1)[13]	-	MD 7.8 lower (20.68 lower to 5.08 higher)	LOW	CRITICAL
MFIS at 10 weeks (Better indicated by lower values)												
Hebert 2011	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	44.7(16.3)[13]	52.6(17.4)[13]	-	MD 7.9 lower (20.86 lower to 5.06 higher)	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	aerobic	Control	Relative (95% CI)	Absolute		
Proportion with improvement in MFIS at 3 weeks (better indicated by higher proportion)												
Dettmers 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	6/9 (66.7%)	9/10 (90%)	RR 0.74 (0.45 to 1.23)	234 fewer per 1000 (from 495 fewer to 207 more)	VERY LOW	CRITICAL
Proportion with improvement in MFIS (motor) at 3 weeks(better indicated by higher proportion)												
Dettmers 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	8/9 (88.9%)	9/10 (90%)	RR 0.99 (0.72 to 1.35)	9 fewer per 1000 (from 252 fewer to 315 more)	VERY LOW	CRITICAL
FSS change from baseline to 12 weeks (Better indicated by lower values)												
Geddes 2009 Ahmadhi 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	-0.24(0.72)[8] -1.56(0.98)[10]	-0.17(0.49)[4] 0.06(0.74)[10]	-	MD 0.84 lower (2.36 lower to 0.68 higher)	VERY LOW	CRITICAL
Proportion with improvement in HAQUAMS (motor) at 3 weeks(better indicated by higher proportion)												
Dettmers 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	5/9 (55.6%)	7/10 (70%)	RR 0.79 (0.39 to 1.62)	147 fewer per 1000 (from 427 fewer to 434 more)	VERY LOW	CRITICAL
Fatigue Severity Scale 4-12 weeks (Better indicated by lower values)												
Geddes 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	-0.24(0.74)[8] 4.4(1.9)[13]	-0.17(0.49)[4] 5(1.9)[13]	-	MD 0.17 lower (0.79 lower to	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	aerobic	Control	Relative (95% CI)	Absolute		
Mostert 2002 Van den berg 2006							-4.5(7.7)[8]	-4.4(7.8)[8]		0.46 higher)		
FUNCTIONAL OUTCOMES												
Dynamic gait index – change from baseline to 2 weeks (Better indicated by higher values)												
Gervasoni 2014	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	2.16(2.175)[15]	2.07(2.175)[15]	-	MD 0.09 higher (1.47 lower to 1.65 higher)	VERY LOW	CRITICAL
6MWT 6 weeks (Better indicated by higher values)												
Hebert 2011	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	1112(391)[13]	1072(375)[13]	-	MD 40 higher (254.5 lower to 334.5 higher)	VERY LOW	CRITICAL
6MWT 10-12 weeks (Better indicated by higher values)												
Geddes 2009 Hebert 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	65.7(24.4)[8] 1054(449)[13]	46.75(37.25)[4] 1105(284)[13]	-	MD 17.59 higher (22.24 lower to 57.42 higher)	VERY LOW	CRITICAL
Increase in walking distance (m) from baseline to 3 weeks (Better indicated by higher values)												
Dettmers 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	650(474)[15]	97(70)[15]	-	MD 553 higher (310.53 to 795.47 higher)	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**aerobic**	**Control**	**Relative (95% CI)**	**Absolute**		
10-metre timed walk (s) (Better indicated by lower values)												
Van den Berg 2006	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-3.1(2.5)[8]	-0.6(1.4)[8]	-	MD 2.5 lower (4.49 to 0.51 lower)	VERY LOW	CRITICAL
Guys neurological disability scale (Better indicated by lower values)												
Van den Berg 2006	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	4.1(8.6)[8]	4.3(9.5)[8]	-	MD 0.2 lower (9.08 lower to 8.68 higher)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Table 107: Clinical evidence profile: Aerobic training versus neurorehabilitation

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic exercise	versus neurological rehabilitation	Relative (95% CI)	Absolute		
FUNCTIONAL OUTCOMES												
walking distance (m) in 6 minutes (Better indicated by higher values)												
Rampello 2007	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision	none	332(108)[11]	308(110)[11]	-	MD 24 higher (67.1 lower to 115.1 higher)	VERY LOW	CRITICAL
FATIGUE												
See narrative review												
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Table 108: Clinical evidence profile: Mixed aerobic and resistance training versus control

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic and resistance training	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
Change in MFIS (total) from baseline to 12 weeks (Better indicated by lower values)												
Garret 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-7.5(14.29)[63]	-1.1(11.8)[49]	-	MD 6.4 lower (11.24 to 1.56 lower)	VERY LOW	CRITICAL
Change in MFIS (physical subscale) from baseline to 12 weeks (Better indicated by lower values)												
Garret 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-3.9(6.8)[63]	-1.1(11.8)[49]	-	MD 4.3 lower (6.42 to 2.18 lower)	VERY LOW	CRITICAL
Change in MFIS (cognitive subscale) from baseline to 12 weeks (Better indicated by lower values)												
Garret 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision	none	-2.1(4.17)[63]	-0.51(4.18)[49]	-	MD 1.59 lower (3.15 lower to 0.03 lower)	VERY LOW	CRITICAL
Change in fatigue severity scale from baseline to 5 weeks (Better indicated by lower values)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-10.75(7.27)[12]	3(4.1)[12]	-	MD 13.75 lower (from 18.48 lower to 9.02 lower)	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic and resistance training	Control	Relative (95% CI)	Absolute		
Fatigue severity scale at 12 weeks (Better indicated by lower values)												
Learmouth 2012	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	5(1.8)[15]	6.2(0.7)[10]	-	MD 1.2 lower (2.21 to 0.19 lower)	VERY LOW	CRITICAL
Change in Knee extensor Fatigue Index from baseline to 26 weeks - FEMALES (Better indicated by lower values)												
Surakka 2004	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-3.3(13.7)[30]	4.3(13.9)[31]	-	MD 7.6 lower (14.53 to 0.67 lower)	VERY LOW	CRITICAL
Change in Knee flexor Fatigue Index from baseline to 26 weeks - FEMALES (Better indicated by lower values)												
Surakka 2004	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-1.9(9.9)[30]	5.3(10)[31]	-	MD 7.2 lower (12.19 to 2.21 lower)	VERY LOW	CRITICAL
Change in Knee extensor Fatigue Index from baseline to 26 weeks - MALES (Better indicated by lower values)												
Surakka 2004	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	2(14.8)[30]	1.8(13.6)[31]	-	MD 0.2 higher (6.95 lower to 7.35 higher)	VERY LOW	CRITICAL
Change in Knee flexor Fatigue Index from baseline to 26 weeks - MALES (Better indicated by lower values)												
Surakka 2004	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	2.8(13.6)[30]	2.4(12.8)[31]	-	MD 0.4 higher (6.23 lower to	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic and resistance training	Control	Relative (95% CI)	Absolute		
										7.03 higher)		
FUNCTIONAL OUTCOMES												
Timed 25-Foot Walk Test (Better indicated by lower values)												
Learmouth 2012	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	14.9(13.6)[15]	13.1(8.6)[10]	-	MD 1.8 higher (6.91 lower to 10.51 higher)	VERY LOW	CRITICAL
Timed up and go (Better indicated by lower values)												
Learmouth 2012 Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	18.4(14.9)[15] -0.99(1.03)[12]	16.2(11)[10] 0.95(1.26)[12]	-	MD 1.91 lower (2.82 lower to 0.99 lower)	LOW	CRITICAL
QUALITY OF LIFE												
Leeds MS quality of life (Better indicated by lower values)												
Learmouth 2012	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	10.9(3.9)[15]	12.4(3.1)[10]	-	MD 1.5 lower (4.25 lower to 1.25 higher)	VERY LOW	CRITICAL
CARER PERCEPTIONS												
No studies found covering this outcome												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic and resistance training	Control	Relative (95% CI)	Absolute		
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.
[C] Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.

Table 109: Clinical evidence profile: Mixed aerobic and resistance training + CBT versus usual care

Quality assessment	Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic/res and CBT	Usual care	Relative (95% CI)	Absolute		
QUALITY OF LIFE												
EQ-5D at 3 months (Better indicated by higher values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0.744(0.204)[54]	0.684(0.263)[53]	-	MD 0.06 higher (0.03 lower to 0.15 higher)	LOW	CRITICAL
EQ-5D at 9 months (Better indicated by higher values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	0.739(0.249)[49]	0.734(0.252)[50]	-	MD 0.01 higher (0.09 lower to 0.1 higher)	MOD	CRITICAL
MSQoL-54 at 3 months (Better indicated by higher values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	68.1(20.3)[54]	60.6(19.2)[53]	-	MD 7.5 higher (0.01 to 14.99 higher)	LOW	CRITICAL
MSQoL-54 at 9 months (Better indicated by higher values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	65.9(20.1)[49]	60.4(21.1)[50]	-	MD 5.5 higher (2.62 lower to 13.62 higher)	LOW	CRITICAL
FATIGUE												
MFIS total at 3 months (Better indicated by lower values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	35.8(18.2)[54]	43.2(17.3)[53]	-	MD 7.4 lower (14.13 to 0.67 lower)	LOW	CRITICAL
MFIS phys at 3 months (Better indicated by lower values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	17.9(18.2)[54]	43.2(17.3)[53]	-	MD 3.3 lower (6.56 to 0.04 lower)	LOW	CRITICAL
MFIS cog at 3 months (Better indicated by lower values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	14.9(9.6)[54]	17.7(8.2)[53]	-	MD 2.8 lower (6.18 lower to 0.58 higher)	LOW	CRITICAL
MFIS psych at 3 months (Better indicated by lower values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic/res and CBT	Usual care	Relative (95% CI)	Absolute		
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	2.9(2.2)[49]	4.2(2.1)[50]	-	MD 1.3 lower (2.11 to 0.49 lower)	LOW	CRITICAL
MFIS total at 9 months (Better indicated by lower values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	39.6(16.6)[49]	41.3(18.8)[50]	-	MD 1.7 lower (8.68 lower to 5.28 higher)	MOD	CRITICAL
MFIS phys at 9 months (Better indicated by lower values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	20.1(7.8)[49]	20.7(8.5)[50]	-	MD 0.6 lower (3.81 lower to 2.61 higher)	MOD	CRITICAL
MFIS cog at 9 months (Better indicated by lower values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	16(8.8)[49]	16.7(9.6)[50]	-	MD 0.7 lower (4.33 lower to 2.93 higher)	MOD	CRITICAL
MFIS psych at 9 months (Better indicated by lower values)												
Carter 2014	RCT	serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	3.5(1.9)[49]	4(2.4)[50]	-	MD 0.5 lower (1.35 lower to 0.35 higher)	LOW	CRITICAL
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic/res and CBT	Usual care	Relative (95% CI)	Absolute		
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 110: Clinical evidence profile: Supervised resistance and balance versus control

Quality assessment	Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
Fatigue Severity Scale - change from baseline to 8 weeks (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-9.5(2.8)[14]	-5.2(5.3)[9]	-	MD 4.3 lower (8.06 to 0.54 lower)	VERY LOW	CRITICAL
QUALITY OF LIFE OUTCOMES												
SF-36 - Physical functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	21.2(14.4)[14]	7.7(7.4[9]	-	MD 13.5 higher (4.54 to 22.46 higher)	LOW	CRITICAL
SF-36 - Role-physical functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	34(30.1)[14]	5(44.7)[9]	-	MD 29 higher (4.19 lower to 62.19 higher)	VERY LOW	CRITICAL
SF-36 - Bodily pain (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	8.8(5.8)[14]	2(2.1)[9]	-	MD 6.8 higher (3.47 to 10.13 higher)	LOW	CRITICAL
SF-36 - General health (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	4.3(8.4)[14]	3.2(11.7)[9]	-	MD 1.1 higher (7.72 lower to 9.92 higher)	VERY LOW	CRITICAL
SF-36 - Vitality (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	9(19.3)[14]	11(20.4)[9]	-	MD 2 lower (18.73 lower to 14.73 higher)	VERY LOW	CRITICAL
SF-36 - Social functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	3.4(23.1)[14]	5(16.7)[9]	-	MD 1.6 lower (17.89 lower to 14.69 higher)	VERY LOW	CRITICAL
SF-36 - Role-emotional functioning (Better indicated by higher values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance	Control	Relative (95% CI)	Absolute		
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	24.2(49.6)[14]	19.9(50.5)[9]	-	MD 4.3 higher (37.69 lower to 46.29 higher)	VERY LOW	CRITICAL
SF-36 - Mental health (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	7.2(13.4)[14]	7(6.7)[9]	-	MD 0.2 higher (8.07 lower to 8.47 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
10 m walking test s (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	-1.9(1.2)[14]	0.1(0.8)[9]	-	MD 2 lower (2.82 to 1.18 lower)	LOW	CRITICAL
Timed up and go test (s) (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-1.3(1.2)[14]	-0.2(0.8)[9]	-	MD 1.1 lower (1.92 to 0.28 lower)	VERY LOW	CRITICAL
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.
[C] Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.

Table 111: Clinical evidence profile: Home based resistance and balance versus control

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home based balance and resistance	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
Fatigue Severity Scale - change from baseline to 8 weeks (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-0.4(2.1)[10]	-5.2(5.3)[9]	-	MD 4.8 higher (1.1 to 8.5 higher)	VERY LOW	CRITICAL
QUALITY OF LIFE OUTCOMES												
SF-36 - Physical functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	12.1(6)[10]	7.7(7.4)[9]	-	MD 4.4 higher (1.7 lower to 10.5 higher)	VERY LOW	CRITICAL
SF-36 - Role-physical functioning (Better indicated by higher values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Home based balance and resistance**	**Control**	**Relative (95% CI)**	**Absolute**		
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-5(20.9)[10]	5(44.7)[9]	-	MD 10 lower (41.95 lower to 21.95 higher)	VERY LOW	CRITICAL
SF-36 - Bodily pain (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	2(2.1)[10]	2(2.1)[9]	-	MD 0 higher (1.89 lower to 1.89 higher)	VERY LOW	CRITICAL
SF-36 - General health (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	2.4(11.5)[10]	3.2(11.7)[9]	-	MD 0.8 lower (11.25 lower to 9.65 higher)	VERY LOW	CRITICAL
SF-36 - Vitality (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	12(22.5)[10]	11(20.4)[9]	-	MD 1 higher (18.29 lower to 20.29 higher)	VERY LOW	CRITICAL
SF-36 - Social functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	10(13.6)[10]	5(16.7)[9]	-	MD 5 higher (8.79 lower to 18.79 higher)	VERY LOW	CRITICAL
SF-36 - Role-emotional functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-6.7(27.8)[10]	19.9(50.5)[9]	-	MD 26.6 lower (63.82 lower to 10.62 higher)	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home based balance and resistance	Control	Relative (95% CI)	Absolute		
SF-36 - Mental health (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	3(6.7)[10]	7(6.7)[9]	-	MD 4 lower (10.03 lower to 2.03 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
10 m walking test s (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	-0.08(0.7)[10]	0.1(0.8)[9]	-	MD 0.18 lower (0.86 lower to 0.5 higher)	VERY LOW	CRITICAL
Timed up and go test secs (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	0.2(0.5)[10]	-0.2(0.8)[9]	-	MD 0.4 higher (0.21 lower to 1.01 higher)	VERY LOW	CRITICAL
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Table 112: Clinical evidence profile: Supervised resistance and balance versus home based resistance and balance

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance and balance	Home based resistance and balance	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
Fatigue Severity Scale- change from baseline to 8 weeks (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-9.5(2.8)[14]	-0.4(2.1)[10]	-	MD 9.1 lower (11.06 to 7.14 lower)	LOW	CRITICAL
QUALITY OF LIFE OUTCOMES												
SF-36 - Physical functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	21.2(14.4)[14]	12.1(6)[10]	-	MD 9.1 higher (0.69 to 17.51 higher)	VERY LOW	CRITICAL
SF-36 - Role-physical functioning (Better indicated by higher values)												
Cakit	RCT	Very	No serious	No serious	No serious	none	34(30.1)[14]	-5(20.9)[10]	-	MD 39 higher (18.59 to	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance and balance	Home based resistance and balance	Relative (95% CI)	Absolute		
2010		serious[A]	inconsistency	indirectness	imprecision[B]					59.41 higher)		
SF-36 - Bodily pain (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	8.8(5.8)[14]	-2(2.1) [10]		MD 6.8 higher (3.49 to 10.11 higher)	LOW	CRITICAL
SF-36 - General health (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	4.3(8.4)[14]	2.4(11.5)[10]	-	MD 1.9 higher (6.48 lower to 10.28 higher)	VERY LOW	CRITICAL
SF-36 - Vitality (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	9(19.3)[14]	12(22.5)[10]	-	MD 3 lower (20.22 lower to 14.22 higher)	VERY LOW	CRITICAL
SF-36 - Social functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	3.4(23.1)[14]	10(13.6)[10]	-	MD 6.6 lower (21.35 lower to 8.15 higher)	VERY LOW	CRITICAL
SF-36 - Role-emotional functioning (Better indicated by higher values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	24.2(49.6)[14]	-6.7(27.8)[10]	-	MD 30.9 higher (0.28 lower to 62.08 higher)	VERY LOW	CRITICAL
SF-36 - Mental health (Better indicated by higher values)												
Cakit	RCT	Very	No serious	No serious	Very serious	none	7.2(13.4)[14]	3(6.7)[10]	-	MD 4.2 higher (3.96	VERY	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance and balance	Home based resistance and balance	Relative (95% CI)	Absolute		
2010		serious[A]	inconsistency	indirectness	imprecision[B]					lower to 12.36 higher)	LOW	
FUNCTIONAL OUTCOMES												
10 m walking test s (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-1.9(1.2)[14]	-0.08(0.7)[10]	-	MD 1.82 lower (2.58 to 1.06 lower)	LOW	CRITICAL
Timed up and go test secs (Better indicated by lower values)												
Cakit 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-1.3(1.2)[14]	0.2(0.5)[10]	-	MD 1.5 lower (2.2 to 0.8 lower)	LOW	CRITICAL
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.
[C] Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.

Table 113: Clinical evidence profile: Vestibular rehabilitation versus control

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vestibular rehabilitation	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
MFIS total score at 6 weeks (Better indicated by lower values)												
Hebert 2011	RCT	Serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	29.5(15.8)[12]	52.1(17.1)[13]	-	MD 22.6 lower (35.5 to 9.7 lower)	MOD	CRITICAL
MFIS total score at 10 weeks (Better indicated by lower values)												
Hebert 2011	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	30.3(20.8)[12]	52.6(17.4)[13]	-	MD 22.3 lower (37.4 to 7.2 lower)	LOW	CRITICAL
FUNCTIONAL OUTCOMES												
Change from baseline 6MWT at 6 weeks (Better indicated by higher values)												
Hebert 2011	RCT	Serious[A]	No serious inconsistency	Very Serious indirectness	Serious imprecision[B]	none	85.1(159.5)[12]	1072(375)[13]	-	MD 62.7 higher (81.1 to 206.5 higher)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vestibular rehabilitation	Control	Relative (95% CI)	Absolute		
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.
[C] Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.

Table 114: Clinical evidence profile: Yoga versus control

Quality assessment	Mean (sd) [n] (in study order)	Effect	Quality	Importance

							OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV					
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Yoga	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
% change in FFS from baseline to 8 weeks (Better indicated by lower values)												
Ahmadi 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-38.7(40.5)[11]	1.4(2.3)[10]	-	MD 40.1 lower (64.07 to 16.13 lower)	LOW	CRITICAL
Change in MFIS (total) from baseline to 12 weeks (Better indicated by lower values)												
Garrett 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-5.8(23.0)[63]	-1.1(11.8)[49]	-	MD 4.7 lower (11.28 lower to 1.88 higher)	VERY LOW	CRITICAL
Change in MFIS (physical subscale) from baseline to 12 weeks (Better indicated by lower values)												
Garrett 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-2.1(6.4)[67]	0.4(4.7)[49]	-	MD 2.5 lower (4.51 to 0.49 lower)	VERY LOW	CRITICAL
Change in MFIS (cognitive subscale) from baseline to 12 weeks (Better indicated by lower values)												
Garrett 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-0.96(3.6)[67]	-0.5(4.2)[49]	-	MD 0.45 lower (1.9 lower to 1 higher)	LOW	CRITICAL
QUALITY OF LIFE OUTCOMES												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Yoga	Control	Relative (95% CI)	Absolute		
MSQOL physical change from baseline to 8 weeks (Better indicated by higher values)												
Ahmadi 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	6.8(8.1)[11]	-0.6(6.9)[10]	-	MD 7.35 higher (0.92 to 13.78 higher)	VERY LOW	CRITICAL
MSQOL mental change from baseline to 8 weeks (Better indicated by higher values)												
Ahmadi 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	18.2(13.2)[11]	5.0(41.5)[10]	-	MD 13.14 higher (13.74 lower to 40.02 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
% change in 10m timed walk from baseline to 8 weeks (Better indicated by lower values)												
Ahmadi 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-7.4(14.9)[11]	3.38(6.6)[10]	-	MD 10.78 lower (20.49 to 1.07 lower)	VERY LOW	CRITICAL
% change in 2 min walk distance from baseline to 8 weeks (Better indicated by higher values)												
Ahmadi 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	9.96(7.2)[11]	-2.9(5.8)[10]	-	MD 12.85 higher (7.28 to 18.42	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Yoga	Control	Relative (95% CI)	Absolute		
										higher)		
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Table 115: Clinical evidence profile: Yoga versus aerobic exercise

Quality assessment	Mean (sd) [n] (in study order)	Effect	Quality	Importance

							OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV					
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Yoga**	**Aerobic exercise**	**Relative (95% CI)**	**Absolute**		
FATIGUE OUTCOMES												
FFS change from baseline to 8 weeks (Better indicated by lower values)												
Ahmadi 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	-1.54(1.54)[11]	-1.56(0.98)[10]	-	MD 0.02 higher (1.07 lower to 1.11 higher)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No studies found covering this outcome												
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.
[C] Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.

Table 116: Clinical evidence profile: Mixed aerobic/resistance versus yoga

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic/resistance	Yoga	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
MFIS total change from baseline to 12 weeks (Better indicated by lower values)												
Garrett 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-7.5(14.29)[63]	-5.8(23)[63]	-	MD 1.7 lower (8.39 lower to 4.99 higher)	LOW	CRITICAL
MFIS physical change from baseline to 12 weeks (Better indicated by lower values)												
Garrett 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-3.9(6.8)[63]	-2.1(6.4)[63]	-	MD 1.8 lower (4.09 lower to 0.49 higher)	VERY LOW	CRITICAL
MFIS cognitive change from baseline to 12 weeks (Better indicated by lower values)												
Garrett 2013A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-2.1(4.2)[63]	-0.96(3.6)[63]	-	MD 1.14 lower (2.5 lower to 0.22	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mixed aerobic/resistance**	**Yoga**	**Relative (95% CI)**	**Absolute**		
										higher)		
MFIS total at 24 weeks (Better indicated by lower values)												
Garrett 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	32.9(14.6)[41]	33.9(19.2)[36]	-	MD 1 lower (8.7 lower to 6.7 higher)	LOW	CRITICAL
QUALITY OF LIFE												
No studies found covering this outcome												
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.
[C] Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.

Table 117: Clinical evidence profile: Resistance training + standard exercise versus standard exercise

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance training + standard exercise	Standard exercise	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
Change in FSS from baseline to 12 weeks (Better indicated by lower values)												
Hayes 2011A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-0.94(0.98)[10]	-1.38(0.87)[10]	-	MD 0.44 higher (0.37 lower to 1.25 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
Change in TUG from baseline to 12 weeks (Better indicated by lower values)												
Hayes 2011A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	0.2(2.7)[10]	0.69(5.8)[10]	-	MD 0.49 lower (4.44 lower to 3.46 higher)	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance training + standard exercise	Standard exercise	Relative (95% CI)	Absolute		
Change in 6MWT from baseline to 12 weeks (Better indicated by lower values)												
Hayes 2011A	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	0.03(0.17)[10]	0.04(0.133)[10]	-	MD 5 higher (64.11 lower to 74.11 higher)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Table 118: Clinical evidence profile: massage versus mixed aerobic/resistance

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**massage**	**Mixed aerobic/res**	**Relative (95% CI)**	**Absolute**		
FATIGUE												
FSS at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-8.08(7.58)[12]	-10.75(7.27)[12]	-	MD: 2.67 higher lower (from 3.27 lower to 8.61 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).

[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 119: Clinical evidence profile: massage/exercise versus usual care

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Massage/exercise**	**Usual care**	**Relative (95% CI)**	**Absolute**		
FATIGUE												
FSS at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No imprecision[B]	none	-9.41(10.63)[12]	3(4.11)[12]	-	MD: 12.41 lower (from 18.86 lower to 6.28 lower)	LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).

[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 120: Clinical evidence profile: massage/exercise versus mixed aerobic/res

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Massage/exercise	Mixed aerobic/res	Relative (95% CI)	Absolute		
FATIGUE												
FSS at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision	none	-9.41(10.63)[12]	-10.75(7.27)[12]	-	MD: 1.34 higher (from 5.95 lower to 8.63 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 121: Clinical evidence profile: massage versus usual care

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	massage	Usual care	Relative (95% CI)	Absolute		
FATIGUE												
FSS at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	-8.08(7.58)[12]	3(4.11)[12]	-	MD: 11.08 lower (from 15.96 to 6.2 lower)	LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).

[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 122: Clinical evidence profile: Wii balance versus control

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Wii balance	control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
See narrative review												
FUNCTIONAL OUTCOMES												
Berg balance scale at 4 weeks (higher better)												
Brichetto 2013	RCT	serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	54.6(2.2)[18]	49.7(3.9)[18]	-	MD 4.9 higher (2.83 higher to 6.97 higher)	MOD	IMPORTANT
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of blinding of participants and health care professionals. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

Table 123: Clinical evidence profile: Electromagnetic field therapy versus placebo device

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electromagnetic field therapy	Placebo device	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
MFIS total at 12 weeks (Better indicated by lower values)												
Piatkowski 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	26.8(12.1)[19]	36.7(13.2)[18]	-	MD 9.9 lower (18.11 to 1.69 lower)	VERY LOW	CRITICAL
MFIS (phys)at 12 weeks (Better indicated by lower values)												
Piatkowski 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	14.1(5.8)[19]	17.7(6.5)[18]	-	MD 3.6 lower (7.58 lower to 0.38 higher)	VERY LOW	CRITICAL
MFIS (cog) at 12 weeks (Better indicated by lower values)												
Piatkowski 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	10.4(6.8)[19]	15.8(6.4)[18]	-	MD 5.4 lower (9.65 to 1.15 lower)	VERY LOW	CRITICAL
MFIS (psychosocial) at 12 weeks (Better indicated by lower values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electromagnetic field therapy	Placebo device	Relative (95% CI)	Absolute		
Piatkowski 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	2(1.6)[19]	3.1(1.6)[18]	-	MD 1.1 lower (2.13 lower to 0.07 lower)	VERY LOW	CRITICAL
FSS at 12 weeks (Better indicated by lower values)												
Piatkowski 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	3.5(1.3)[19]	4.7(1.6)[18]	-	MD 1.2 lower (2.14 to 0.26 lower)	VERY LOW	CRITICAL
change from baseline in self-reported fatigue score (lower better) (Better indicated by lower values)												
Richards 1997	RCT	serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	-0.87(1.55)[15]	-0.23(0.697)[15]	-	MD 0.64 lower (1.5 lower to 0.22 higher)	MOD	CRITICAL
Improvement in fatigue score (higher better)												
Lappin 2003	RCT	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	none	0.50(0.65)[117]	0.36(0.65)[117]	-	MD: 0.14 higher (0.01 higher to 0.27 higher)	HIGH	CRITICAL
FUNCTIONAL OUTCOMES												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electromagnetic field therapy	Placebo device	Relative (95% CI)	Absolute		
MSFC total at 12 weeks (Better indicated by higher values)												
Piatkowski 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	-0.3(1.8)[19]	0(0.8)[18]	-	MD 0.3 lower (1.19 lower to 0.59 higher)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Table 124: Clinical evidence profile: Cognitive behavioural therapy (CBT) versus control

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CBT	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
Fatigue score at 8-10 weeks (Better indicated by lower values)												
Van Kessel 2008 Moss-Morris 2012	RCT	Serious[A]	Serious inconsistency[C]	No serious indirectness	Serious imprecision[B]	none	7.9(4.34)[35] 12.4(6.8)[23]	11.6(5.3)[37] 19.6(5.2)[17]	-	Random MD 5.09 lower (8.47 to 1.72 lower)	VERY LOW	CRITICAL
Fatigue score at 5 months (Better indicated by lower values)												
Van Kessel 2008	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	9(5.3)[35]	11.1(4.6)[37]	-	MD 2.12 lower (4.41 lower to 0.17 higher)	LOW	CRITICAL
Fatigue score at 8 months (Better indicated by lower values)												
Van Kessel 2008	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	10.4(6.4)[35]	12.5(5.2)[37]	-	MD 2.12 lower (4.82 lower to 0.58 higher)	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CBT	Control	Relative (95% CI)	Absolute		
Fatigue related impairment (Work and Social Adjustment Scale) 8 weeks (Better indicated by lower values)												
Van Kessel 2008	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	16.1(9.98)[35]	19.7(37)[37]	-	MD 3.58 lower (8.13 lower to 0.97 higher)	LOW	CRITICAL
Fatigue related impairment (Work and Social Adjustment Scale) 5 months (Better indicated by lower values)												
Van Kessel 2008	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	13.4(8.3)[35]	19.2(37)[37]	-	MD 5.86 lower (9.99 to 1.73 lower)	LOW	CRITICAL
Fatigue related impairment (Work and Social Adjustment Scale) 8 months (Better indicated by lower values)												
Van Kessel 2008	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	14.97(9.9)[35]	10.5(37)[37]	-	MD 5.19 lower (9.9 to 0.48 lower)	LOW	CRITICAL
MFIS score at 10 weeks (Better indicated by lower values)												
Moss-Morris 2012	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	9(3.8)[23]	12.9(3.9)[17]	-	MD 3.88 lower (6.28	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CBT	Control	Relative (95% CI)	Absolute		
										to 1.48 lower)		
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Table 125: Clinical evidence profile: Fatigue management programme/energy conservation versus control

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fatigue management programme/energy conservation	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
MFIS total at 5/6 weeks (Better indicated by lower values)												
Hugos 2010 Garcia 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	39.8(13.4)[15] 59.6(23.1)[13]	44.4(12.97)[15] 63.3(26)[10]	-	MD 4.45 lower (13.02 lower to 4.12 higher)	LOW	CRITICAL
Proportion with clinically relevant improvements (10 or more) in MFIS (higher better)												
Kos 2007	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	4/24 (16.7%)	7/16 (43.8%)	RR 0.38 (0.13 to 1.09)	271 fewer per 1000 (from 381 fewer to 39 more)	VERY LOW	CRITICAL
								43.8%		272 fewer per 1000 (from 381 fewer to 39 more)		
MFIS cognitive at 5/6 weeks (Better indicated by lower values)												
Mathiowet	RCT	Very	No serious	No serious	No serious	none	MD(SE): -2.55(1.19)		-	MD 2.80 lower (4.22	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fatigue management programme/energy conservation	Control	Relative (95% CI)	Absolute		
z 2010 Finlayson 2011 Garcia 2013		serious[A]	inconsistency	indirectness	imprecision[B]		MD(SE): -3.12(0.954) MD(SE): -1.2(3.106)			lower to 1.39 lower)		
MFIS physical at 5/6 weeks (Better indicated by lower values)												
Mathiowetz 2010 Finlayson 2011 Garcia 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE): -3.71(1.19) MD(SE): -2.53(1.02) MD(SE): -2.6(2.6)		-	MD 3 lower (4.45 lower to 1.54 lower)	LOW	CRITICAL
MFIS psychosocial at 5/6 weeks (Better indicated by lower values)												
Mathiowetz 2010 Finlayson 2011 Garcia 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE): -6.1(2.10) MD(SE): -6.01(1.93) MD(SE): -0.2(4.74)		-	MD 5.57 lower (8.24 lower to 2.9 lower)	LOW	CRITICAL
MFIS total at 4.25 months (Better indicated by lower values)												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fatigue management programme/energy conservation	Control	Relative (95% CI)	Absolute		
Garcia 2013	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	58.7(30.3)[13]	79.4(24.5)[10]	-	MD 20.7 lower (43.1 lower to 1.7 higher)	LOW	CRITICAL
MFIS (phys)at 4.25 months (Better indicated by lower values)												
Garcia 2013	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	20.2(7.8)[13]	23.6(7.7)[10]	-	MD 3.4 lower (9.78 lower to 2.98 higher)	LOW	CRITICAL
MFIS (cog) at 4.25 months (Better indicated by lower values)												
Garcia 2013	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	14.6(6.4)[13]	21.1(6.8)[10]	-	MD 6.5 lower (11.97 to 1.03 lower)	LOW	CRITICAL
MFIS (psychsocial)at 4.25 months (Better indicated by lower values)												
Garcia 2013	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	28(13.5)[13]	24.7(11.3)[10]	-	MD 6.7 lower (16.84 lower to 3.44 higher)	LOW	CRITICAL
FSS at 4.25 months (Better indicated by lower values)												
Garcia 2013	RCT	Serious[A]	No serious	No serious	Very serious	none	5.21(1.3)[13]	4.9(1.3)[10]	-	MD 0.31 higher	VERY	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fatigue management programme/energy conservation	Control	Relative (95% CI)	Absolute		
			inconsistency	indirectness	imprecision[B]					(0.76 lower to 1.38 higher)	LOW	
FSS at 5/6 weeks (Better indicated by lower values)												
Hugos 2010 Finlayson 2011 Garcia 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE): 1.73(4.13) MD(SE): -0.18(0.153) MD(SE): 0.08(0.466)		-	MD 0.15 lower (0.44 lower to 0.13 higher)	LOW	CRITICAL
Global fatigue severity at 10 weeks (Better indicated by lower values)												
Thomas 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE) -0.03(0.158)		-	MD 0.03 lower (0.34 lower to 0.28 higher)	LOW	CRITICAL
Global fatigue severity at 5.5 months (Better indicated by lower values)												
Thomas 2013	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	MD(SE) -0.36(0.143)		-	MD 0.36 lower (0.64 lower to 0.08 lower)	VERY LOW	CRITICAL
Fatigue self-efficacy scale at 10 weeks (Better indicated by higher values)												
Thomas	RCT	Serious[A]	No serious	No serious	Serious	none	MD(SE) 9(2.55)		-	MD 9 higher (4	VERY	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fatigue management programme/energy conservation	Control	Relative (95% CI)	Absolute		
2013			inconsistency	indirectness	imprecision[B]					higher to 14 higher)	LOW	
Fatigue self-efficacy scale at 5.5 months (Better indicated by higher values)												
Thomas 2013	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	MD(SE) 6(3.06)		-	MD 6 higher (0.00 lower to 12 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
MSSE at 6 weeks (Better indicated by higher values)												
Hugos 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	1391(237)[15]	1285(237)[15]	-	MD 106 higher (63.62 lower to 275.62 higher)	VERY LOW	CRITICAL
MSIS-29 at 10 weeks (Better indicated by lower values)												
Thomas 2013	RCT	Serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE) 1.44(1.94)		-	MD 1.44 higher (2.37 lower to 5.25 higher)	LOW	CRITICAL
MSIS-29 at 5.5 months (Better indicated by lower values)												
Thomas	RCT	Serious[A]	No serious	No serious	No serious	none	MD(SE) -1.56(2.5)		-	MD 1.56 lower (6.46	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fatigue management programme/energy conservation	Control	Relative (95% CI)	Absolute		
201			inconsistency	indirectness	imprecision[B]					lower to 3.34 higher)		
QUALITY OF LIFE												
SF-36 vitality at 6 weeks (Better indicated by higher values)												
Mathiowetz 2010 Finlayson 2011	RCT	Serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	MD(SE): 11.64(3.12) MD(SE):6.68(4.47)		-	MD 10.01 higher (5 higher to 15.03 higher)	LOW	CRITICAL
SF-36 role emotional at 6 weeks (Better indicated by higher values)												
Mathiowetz 2010 Finlayson 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE): 13.23(10.16) MD(SE):8.69(6.31)		-	MD 9.95 higher (0.55 lower to 20.46 higher)	LOW	CRITICAL
SF-36 mental health at 6 weeks (Better indicated by higher values)												
Mathiowetz 2010 Finlayson 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	MD(SE): 6.12(3.10) MD(SE):5.32(2.1)		-	MD 5.57 higher (2.16 higher to 8.98 higher)	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fatigue management programme/energy conservation	Control	Relative (95% CI)	Absolute		
SF-36 social function at 6 weeks (Better indicated by higher values)												
Mathiowetz 2010 Finlayson 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	MD(SE): 6.06(4.34) MD(SE):7.54(3.97)		-	MD 6.86 higher (1.13 higher to 12.6 higher)	VERY LOW	CRITICAL
SF-36 general health at 6 weeks (Better indicated by higher values)												
Mathiowetz 2010 Finlayson 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE): 0.81(3.15) MD(SE):3.37(2.34)		-	MD 2.46 higher (1.22 lower to 6.14 higher)	LOW	CRITICAL
SF-36 role physical at 6 weeks (Better indicated by higher values)												
Mathiowetz 2010 Finlayson 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	MD(SE): 15.18(7.3) MD(SE):18.06(4.76)		-	MD 17.2 higher (9.39 higher to 25.02 higher)	VERY LOW	CRITICAL
SF-36 physical function at 6 weeks (Better indicated by higher values)												
Mathiowetz	RCT	Very	No serious	No serious	No serious	none	MD(SE): 1.75(3.11)		-	MD 1.36 higher	LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fatigue management programme/energy conservation	Control	Relative (95% CI)	Absolute		
z 2010 Finlayson 2011		serious[A]	inconsistency	indirectness	imprecision[B]		MD(SE):1.2(1.95)			(1.88 lower to 4.59 higher)		
SF-36 bodily pain at 6 weeks (Better indicated by higher values)												
Mathiowetz 2010 Finlayson 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE): 2.69(4.58) MD(SE):5.02(3.08)		-	MD 4.29 higher (0.71 lower to 9.3 higher)	LOW	CRITICAL
SF-36 self-efficacy at 6 weeks (Better indicated by higher values)												
Finlayson 2011	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	MD(SE): 0.14(0.25) MD(SE):5.02(3.08)		-	MD 0.14 higher (0.35 lower to 0.63 higher)	LOW	CRITICAL
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Table 126: Clinical evidence profile: Mindfulness training versus control

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mindfulness training	Control	Relative (95% CI)	Absolute		
FATIGUE OUTCOMES												
MFIS change from baseline to 8 weeks (Better indicated by lower values)												
Grossman 2010	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-6.2(9.7)[75]	-0.36(9.7)[75]	-	MD 5.83 lower (8.94 to 2.72 lower)	VERY LOW	CRITICAL
MFIS change from baseline to 6 months (Better indicated by lower values)												
Grossma	RCT	Very	No serious	No serious	Serious	none	-	0.09(12.45)[75]	-	MD 6.03	VERY LOW	CRITICAL

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mindfulness training	Control	Relative (95% CI)	Absolute		
n 2010		serious [A]	inconsistency	indirectness	imprecision [B]		5.94(12.8)[75]			lower (10.08 to 1.98 lower)		
HAQUAMS change from baseline to 8 weeks (Better indicated by lower values)												
Grossman 2010	RCT	Very serious [A]	No serious inconsistency	No serious indirectness	Serious imprecision [B]	none	-0.18(0.39)[75]	0.09(0.43)[75]	-	MD 0.27 lower (0.4 to 0.14 lower)	VERY LOW	CRITICAL
HAQUAMS change from baseline to 6 months (Better indicated by lower values)												
Grossman 2010	RCT	Very serious [A]	No serious inconsistency	No serious indirectness	Serious imprecision [B]	none	-0.13(0.53)[75]	0.05(0.52)[75]	-	MD 0.18 lower (0.35 to 0.01 lower)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No studies found covering this outcome												
FUNCTIONAL OUTCOMES												

Quality assessment							Mean (sd) [n] (in study order) OR Proportion with event (%) OR Overall MD (SE) if analysed using GIV		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mindfulness training	Control	Relative (95% CI)	Absolute		
No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment (serious inconsistency) if the I^2 value was 50-74 and by two increments (very serious inconsistency) if the I^2 value was >75.*

Narrative review for outcomes not appropriate for meta-analysis

Aerobic exercise versus control

McCullagh 2008 reported their results as medians and interquartile ranges (IQR), and so these could not be analysed in Revman. The results, which showed a clear advantage to aerobic exercise in reducing fatigue and improving function, are shown in Table 127 below.

Table 127: Results from McCullagh 2008 for aerobic exercise versus control

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

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

Aerobic training versus neurorehabilitation

Rampello 2007 reported their results for fatigue and quality of life as medians and ranges, and so these could not be analysed in Revman. The results, which showed no difference between aerobic exercise and neurorehabilitation in reducing fatigue and quality of life, are shown in Table 128 below.

Table 128: Results from Rampello 2007 for aerobic exercise versus control

	Aerobic training N=11 [median (range)]	Neurological rehab N=11 [median (range)]	p
MFIS total median range	29 (4-56)	26 (3-67)	0.86
MFIS physical median range	14 (4-23)	13 (3-26)	0.89
MFIS cognitive median range	8 (0-36)	10 (0-40)	0.71
MFIS psychosocial median range	3 (0-7)	2 (0-6)	0.92
MSQOL-54 Overall quality of life median range	28 (10-82)	36 (20-82)	
MSQOL-54 physical median range	59 (44-81)	57 (41-81)	
MSQOL-54 mental health median	66 (24-90)	66 (32-87)	

range			

Motivational interviewing versus control

Bombardier 2008 reported their results as medians and interquartile ranges (IQR), and so these could not be analysed in Revman. The results, which showed a clear advantage to motivational interviewing in reducing fatigue and mental quality of life, but a possible disadvantage in terms of physical quality of life and no clear effect in improving function, are shown in Table 129 below.

Table 129: Results from Bombardier 2008 for aerobic exercise versus control

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

Wii balance versus resistance training

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

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

Resistance training versus Yoga

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

Table 130: Results from Velikonja 2010 for resistance training versus yoga

Variable	Climbing (n=10)			Yoga (n=10)		
	baseline	10 weeks	p	baseline	10 weeks	p
MFIS total	40(36.5-53)	27(21.5-45.5)	0.015	32(22-42)	23(20.5-36)	0.057
MFIS cog	17(8.5-21.5)	8(6-19.5)	0.024	12(4.5-14.3)	7(3.8-12.5)	0.282
MFIS ps	3(1.5-6)	3(1-5.5)	0.334	4(1-4.5)	3(0.8-4)	0.234
MFISphys	25(21.5-28.5)	19(9-26.5)	0.021	17.5(14.3-24.5)	18(9.8-19)	0.064

Spasticity MSA	10(8.5-18.3)	12.5(10-17.3)	0.574	9.3(3.5-18.4)	8.8(5.5-17.1)	0.673
EDSSpyr	4(3-4)	3(2.5-4)	0.046	2.5(2-4)	2(2-3.3)	0.317

Individualised rehabilitation versus group wellness intervention

Plow 2009 did not provide data for between group analyses except effect sizes. However, the paper reported that the modified fatigue impact scale and SF36-36 did not differ significantly between groups at post-test.

10.3.4 Economic evidence

Published literature

Two economic evaluations were identified with a relevant comparison and have been included in this review.[245,250] These studies are summarised in the economic evidence profiles below (Table 131 and Table 132) and the economic evidence tables in Appendix H.

See also the economic article selection flow chart in Appendix E.

Table 131: Economic evidence profile: Group based fatigue management programme (FACETS) and current local practice versus current local practice

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Thomas 2013[245] (UK)	Directly applicable	Minor limitations (a)	Within-trial analysis (RCT) of adults with clinical definite MS diagnosis (FSS total score >4; ambulant) receiving either current local practice or group based management programme (FACETS) and current local practice. Analysis of individual level data for health outcomes, EQ-5D and resource use, with unit costs applied. Follow-up: 5.5 months (4 months after final session)	£488 (b)	-0.02 QALYs (c)	Current local practice dominates FACETS (£ per QALY)	A probabilistic sensitivity analysis was undertaken to analyse the impact of the uncertainty in the level of staff input for FACETS programme delivery on costs. The mean cost of the intervention was £453 with 95% of estimates in the range of £331 to £585 per participant.

(a) No probabilistic sensitivity analysis for ICER and short follow-up
(b) 2010 GBP. Costs incorporated are: FACETS programme including training, equipment, session facilitators (two Band 7 therapists), venue hire, refreshments, printing, administrative support. Cost for NHS and social care (over a 3 month period) assessed at 4 months follow up for both interventions.
(c) QALYs derived from EQ-5D (from patients, tariff used not stated) with maximum QALY equalling 0.46, assuming full health over 24 weeks.
Abbreviations: EQ-5D = Euroqol 5 dimensions (scale: 0.0 [death] to 1.0 [full health]; FACETS: Fatigue Applying Cognitive behavioural and Energy effectiveness Techniques to lifeStyle; FSS = Fatigue Severity Scale; ICER = incremental cost-effectiveness ratio; NR = not reported; QALYs = quality-adjusted life years; RCT = randomised control trial.

Table 132: Economic evidence profile: Aerobic and resistance exercise and CBT programme (EXIMS) and current local practice versus current local practice

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Tosh 2014[250] (UK)	Directly applicable	Minor limitations (a)	Within-trial analysis (RCT) of adults with clinically definite MS diagnosis; EDSS score 1.0–6.5; able to walk a 10-metre distance and physically able to participate in exercise three times per week receiving either current local	£466 (b)	0.046 QALYs (c)	£10,137 per QALY gained	Probability cost effective (£20,000 threshold): 75% Scenario analyses conducted: • Scenario 1 (EDSS score): <4.0 = dominated; ≥4.0 = £5,092 per QALY gained

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
			practice or a programme incorporating aerobic and resistance exercise and CBT (EXIMS) and current local practice. Analysis of individual level data for health outcomes, EQ-5D and resource use, with unit costs applied. Follow-up: 9 months (6 months after final session)				• Scenario 2 (GLTEQ score): >14 = £9,558 per QALY; <14 = £11,470 per QALY gained • Scenario 3 (private provision of intervention): £11,938 per QALY gained • Scenario 4 (SF-6D utility score): £19,783 per QALY gained

(a) Short follow-up
(b) 2011 GBP. Costs incorporated are: EXIMS programme including staff, equipment, and overheads. Costs for NHS and social care services over 9 month period (intervention start to end of follow-up) assessed for both interventions.
(c) QALYs derived from EQ-5D (from patients, tariff used not stated).
Abbreviations: CBT = cognitive behavioural therapy; EDSS = Expanded Disability Status Scale; EQ-5D = Euroqol 5 dimensions (scale: 0.0 [death] to 1.0 [full health]; EXIMS = EXercise Intervention for people with MS; GLTEQ = Godin Leisure Time Exercise Questionnaire; QALYs = quality-adjusted life years; RCT = randomised control trial; SF-6D = Short form 6 dimension.

New cost-effectiveness analysis

One RCT identified in the clinical review (Cakit 2010[34]) which evaluated the effects on mobility and fatigue of resistance and balance training in different settings (supervised or home based training) compared to no intervention (control), reported SF-36 scores at baseline and after 8 weeks (end of intervention). Based on this study the NCGC were able to undertake a simple cost-utility analysis via mapping of SF-36 data to EQ-5D. Methods were consistent with the NICE reference case unless otherwise stated. The summary of the results can be found in Table 133 below and the details of the analysis can be found in the following paragraphs.

Table 133: Economic evidence profile: Home based resistance and balance vs. control (comparison 1) and supervised resistance and balance vs. home based resistance and balance (comparison 2)

Study	Applicability	Limitations	Other comments	Incremental cost per year	Incremental effects (QALY)	ICER	Uncertainty
NCGC analysis	Directly applicable	Potentially serious limitations (a)	Population: people with multiple sclerosis. Time horizon: one year. Comparators: 1) Control 2) Home-based resistance and balance 3) Supervised resistance and balance Based on an RCT included in the clinical review[34]	2-1: £52 3-2: £398 (b)	2-1: 0.011 QALY 3-2: 0.052 QALY (c)	2 vs. 1: £7,152 per QALY 3 vs. 2: £7,619 per QALY	Sensitivity analysis was conducted with a shorter time horizon of 8 weeks. Assuming the improvement in quality of life is not maintained beyond the 8 week intervention duration, the ICER increased to £31,633 per QALY and £49,526 per QALY for comparison 1 and 2 respectively.

(a) *Analysis based on a single RCT[34]; utilities were estimated through a mapping function which is associated with limitations. Cost of a cycling machine and downstream costs were excluded from the analysis.*

(b) *Cost of staff time only.*

(c) *Difference in QALY calculated as the incremental change in EQ-5D score between baseline and follow-up using an algorithm that mapped SF-36 scores to EQ-5D scores. The improvement in EQ-5D was assumed to be maintained, beyond the 8 week intervention period, over 1 year.*

Methods

This analysis was based on a study by Cakit (2010)[34] which included people who had clinically or laboritorially definite relapsing-remitting or secondary progressive MS with an EDSS≤6 and who were able to stand independently > 3 secs. There were three comparators which were control (no intervention), home based resistance and balance and supervised resistance and balance. This study reported SF-36 data that could be mapped to EQ-5D allowing quality-adjusted life years (QALYs) to be estimated and cost-effectiveness to be explored.

This simple deterministic cost-utility analysis took an NHS perspective. Due to the limited follow-up time of the clinical data, a one year time horizon was considered. While it is possible that benefits may persist longer than one year it was not considered reliable to extrapolate any further as the trial only reported relevant results at 8 weeks. Methods were consistent with the NICE reference case unless otherwise stated. Costs and QALYs were not discounted due to the short time horizon.

Effectiveness was expressed as quality adjusted life years (QALYs); this was estimated through the mapping of changes in SF-36 scores obtained from Cakit (2010)[34] to EQ-5D values using an algorithm by Ara and Brazier (2008).[8]

QALYs

Preferably, direct EQ-5D data measuring treatment effect on health-related quality of life would be used to estimate QALYs but this was not available. Ara and Brazier (2008)[8] provides us with a mapping function to estimate EQ-5D scores from SF-36 scores. Regression model 4 was used as this is the recommended model when comparing incremental differences between study arms or changes over time.

Mapped EQ-5D scores for the three interventions are reported in Table 134.

Table 134: Mapped EQ-5D from SF-36 scores

Intervention	SF-36 at baseline and 8 weeks (a)	EQ-5D score at baseline (b)	EQ-5D score at 8 weeks (b)	Estimated change in EQ-5D at 8 weeks
Intervention 1: control (n=9)	See Table 110	0.6818	0.7612	0.0793
Intervention 2: Home based resistance and balance (n=10)	See Table 112	0.4369	0.5269	0.0900
Intervention 3: Supervised resistance and balance (n=14)	See Table 112	0.6127	0.7549	0.1423

(a) *From Cakit et al. (2010)*[34]
(b) *Calculated by using regression model 4 by Ara and Brazier (2008)*[8]

In the base case, QALY gains for each intervention were estimated assuming the effectiveness throughout the year is similar to the effectiveness observed at 8 weeks (that is the difference in EQ-5D between interventions is constant and the effectiveness of the interventions is sustained throughout a year even after the intervention is discontinued), therefore the QALY gain corresponds to the improvement in EQ-5D value.

A sensitivity analysis was conducted with a shorter time horizon of 8 weeks; this assumed the mean change in EQ5D over the 8 week trial duration is maintained over trial duration only (that is, the difference in EQ-5D between the interventions is lost at 8 weeks when the intervention is discontinued). In this case QALYs are calculated by multiplying the EQ-5D scores by the number of life-years (8 weeks / 52 weeks = 0.15 life years).

Cost

Costs of each intervention were estimated based on published unit costs and within trial resource use. Costs and key assumptions made for the costing are summarised in Table 135. The cost of a cycling machine was not included; however when the cost of the machine is spread over the lifetime of the equipment and the amount of usage, the cost per patient per session is expected to be low. Downstream costs were not incorporated as it is unclear what these would be. Feasibly there could be savings in terms of reduced healthcare visits related to fatigue and mobility issues but there is no clinical evidence to support this.

Table 135: Cakit 2010 intervention costs

Intervention	Resource use estimate based on Cakit 2010	Unit cost of staff time (£ per hour)	Total cost of intervention
Intervention 1: Control	No resource use	n/a	£0
Intervention 2: Home based resistance and balance	4 phone calls from research staff in the RCT. We assume each phone call lasts 15 minutes and conducted by community physiotherapist (band 7)	£52	£52
Intervention 3: Supervised resistance and balance	In the RCT, 16 group sessions observed by physiatrist, each session consisting of: • 90 minutes of cycling repetitions • 25 minutes balance • 10 minutes • warm-up/stretching Total session duration was 125 minutes in the study. Groups described as small, therefore assumed to be 4 people per group. We excluded the cost of cycling machine. We assume the supervision is conducted by hospital physiotherapist (band 7).	£54	£450

Source: PSSRU 2013[47]
Abbreviations: n/a = not applicable; RCT = randomised control trial.

Model validation

The model was developed in consultation with the GDG; inputs and results were presented to and discussed with the GDG for clinical validation and interpretation.

The model was systematically checked by the health economist undertaking the analysis; this included checking that results were plausible given inputs. The model was peer reviewed by a second experienced health economist from the NCGC; this included systematic checking of the model calculations.

Computations and estimation of cost effectiveness

The model was constructed in Microsoft Excel 2010 and allowed for the calculation of the incremental cost effectiveness ration (ICER). The ICER is calculated by dividing the difference in costs associated with two alternatives by the difference in QALYs. The decision rule then applied is that if the ICER falls below a given cost per QALY threshold the result is considered to be cost effective. If both costs are lower and QALYs are higher the option is said to dominate and an ICER is not calculated.

$$ICER = \frac{Costs(B) - Costs(A)}{QALYs(B) - QALYs(A)}$$

Where: Costs(A) = total costs for option A; QALYs(A) = total QALYs for option A

Cost-effective if:

- ICER < Threshold

When there are more than two comparators, as in this analysis, options must be ranked in order of increasing cost then options ruled out by dominance or extended dominance before calculating ICERs excluding these options. An option is said to be dominated, and ruled out, if another intervention is less costly and more effective. An option is said to be extendedly dominated if a combination of two other options would prove to be less costly and more effective.

Interpreting results

NICE's report 'Social value judgements: principles for the development of NICE guidance'[163] sets out the principles that GDGs should consider when judging whether an intervention offers good value for money. In general, an intervention was considered to be cost effective if either of the following criteria applied (given that the estimate was considered plausible):

- The intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- The intervention costs less than £20,000 per quality-adjusted life-year (QALY) gained compared with the next best strategy.

Results

Base case

The results of the base case analysis with a one-year time horizon (assuming persistence of effect beyond the trial follow-up) is reported in Table 136.

Table 136: Results of incremental deterministic analysis – 1 year time horizon

Intervention	Total costs (mean per patient)	Incremental costs versus previous intervention	Total QALYs (mean per patient)	Incremental QALYs versus previous intervention	ICER versus previous intervention (£ per QALY gained)
Intervention 1: Control	£0	n/a	0.079	n/a	n/a
Intervention 2: Home based resistance and balance	£52	£52	0.090	0.011	£7,152 per QALY
Intervention 3: Supervised resistance and balance	£450	£398	0.142	0.052	£7,619 per QALY

Abbreviations: ICER = incremental cost effectiveness ratio; n/a = not applicable.

Sensitivity analysis

The results of the sensitivity analysis, with an 8-week time horizon (in line with the trial data), are reported in Table 137.

Table 137: Results of incremental deterministic analysis – 8 week time horizon

Intervention	Total costs (mean per patient)	Incremental costs versus previous intervention	Total QALYs (mean per patient)	Incremental QALYs versus previous intervention	ICER versus previous intervention (£ per QALY gained)
Intervention 1: Control	£0	n/a	0.012	n/a	n/a
Intervention 2: Home based resistance and balance	£52	£52	0.014	0.002	£31,633 per QALY
Intervention 3: Supervised resistance and balance	£450	£398	0.022	0.008	£49,526 per QALY

Abbreviations: ICER = incremental cost effectiveness ratio; n/a = not applicable.

Discussion

With the one-year time horizon and assuming the improvement in quality of life is maintained after the intervention is completed, 'supervised resistance and balance' was found to be the most effective option (highest QALYs) and the most cost effective of the three options considered. In the sensitivity analysis however that took an 8-week time horizon assuming that the improvement in quality of life is not maintained beyond the 8-week intervention duration, neither the supervised nor the home-based interventions are cost effective because the QALY gain is not sufficient to justify the additional cost of the interventions. This shows that the conclusion is sensitive to the assumption regarding persistence of treatment effect beyond the trial follow-up.

This analysis has some limitations: it is based on a single RCT with a limited number of participants and all the limitations of the clinical data also apply to this economic analysis. This analysis does not include all intervention costs, for example the cost of the cycling machine. However, when the cost of the machine is spread over the lifetime of the equipment and the amount of usage, the cost per patient per session is expected to be low. Downstream costs have not been included in the analysis as they were unclear from the clinical evidence. Feasibly there could be savings in terms of reduced healthcare visits related to fatigue and mobility issues.

In addition, the model is based EQ-5D estimates mapped from the generic health-related quality of life instrument SF-36. The regression model selected to map the SF-36 score to EQ-5D score (model 4) does not utilise the score from the physical role domain or the vitality (energy/fatigue) dimensions. However, the authors state that these dimensions add little to either the goodness of fit or the accuracy of the scores generated by the models.[8] Furthermore, the regression models by Ara and Brazier (2008)[8] have not been validated in people with MS specifically.

Unit costs

Relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

10.3.5 Evidence statements

10.3.5.1 Clinical

Resistance training versus control

Very low quality evidence from 1 RCT comprising 71 participants showed that resistance training was clinically effective compared to control in terms of MFIS (total) at 10 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 71 participants showed that resistance training was clinically effective compared to control in terms of MFIS (physical) at 10 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 35 participants showed that resistance training was clinically effective compared to control in terms of MFI-20 general fatigue at 12 weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 35 participants showed that resistance training was clinically effective compared to control in terms of MFI-20 physical fatigue scale at 12 weeks, with no imprecision.

Low quality evidence from 1 RCT comprising 35 participants showed that resistance training was clinically effective compared to control in terms of the SF-36 physical quality of life scale at 12 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 99 participants showed that resistance training was clinically harmful compared to control in terms of 10m walking distance at 12 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 71 participants showed that resistance training was clinically effective compared to control in terms of stiffness (MSIS-88 sub-scale) at 10 weeks, with no imprecision.

Very low quality evidence from 3 RCTs comprising 205 participants showed that there were no clinically important differences between resistance training and control treatment in the MFIS cognitive and psychosocial sub-scales at 10 weeks, MFIS total, physical, cognitive and psychosocial sub-scales at 22 weeks, and FSS and MUSIQOL at 12 weeks. There were also no clinically important group differences in quality of life as measured by SF-36 mental at 12 weeks and WHOQol at 10 and 22 weeks. Functional outcomes of fast walking speed and 2 minute walking distance at 10 weeks and 22 weeks, and 10 minute walking distance at 12 weeks were also clinically similar across groups. A similar lack of clinically important group differences was observed for muscle spasm at 12 and 22 weeks, and stiffness at 22 weeks. These inconclusive outcomes ranged from precise to seriously imprecise.

Aerobic training versus control

Low quality evidence from 1 RCT comprising 21 participants showed that aerobic training was clinically effective compared to control in terms of MFIS total at 8 weeks, with no imprecision.

Low quality evidence from 1 RCT comprising 21 participants showed that aerobic training was clinically effective compared to control in terms of MFIS physical at 8 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 21 participants showed that aerobic training was clinically effective compared to control in terms of MFIS psychosocial at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 21 participants showed that aerobic training was clinically effective compared to control in terms of MFIS cognitive at 8 weeks, with serious imprecision.

Low quality evidence from 2 RCTs comprising 32 participants showed that aerobic training was clinically effective compared to control in terms of FSS at 8-12 weeks, with very serious imprecision.

Very low quality evidence from 1 RCT comprising 15 participants showed that aerobic training was clinically harmful compared to control in terms of the number with improvements in MFIS (motor) at 3 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 12 participants showed that aerobic training was clinically harmful compared to control in terms of the number with improvements in HAQUAMS (motor) at 3 weeks, with very serious imprecision.

Low quality evidence from 1 RCT comprising 30 participants showed that aerobic training was clinically effective compared to control in terms of walking distance at 3 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 16 participants showed that aerobic training was clinically effective compared to control in terms of 10m timed walk, with serious imprecision.

Low to very low quality evidence from 6 RCTs comprising 125 participants showed that there were no clinically important differences between aerobic training and control treatment in the MFIS total at 6 and 10 weeks, MFIS motor at 3 weeks and FSS at 4-12 weeks. Functional outcomes of dynamic gait index at 2 weeks, 6 min walk test at 6-12 weeks and Guys Neurological Disability (GND) scale were also clinically similar across groups. These inconclusive outcomes ranged from precise to very seriously imprecise.

Mixed aerobic/resistance training versus control

Very low quality evidence from 1 RCT comprising 112 participants showed that mixed aerobic and resistance training was clinically effective compared to control in terms of MFIS (total) at 12 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 112 participants showed that mixed aerobic and resistance training was clinically effective compared to control in terms of MFIS (physical) at 12 weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 24 participants showed that mixed aerobic and resistance training was clinically effective compared to control in terms of FSS at 5 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 25 participants showed that mixed aerobic and resistance training was clinically effective compared to control in terms of FSS at 12 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 71 participants showed that mixed aerobic and resistance training was clinically effective compared to control in terms of Knee extensor fatigue index in women at 26 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 71 participants showed that mixed aerobic and resistance training was clinically effective compared to control in terms of Knee flexor fatigue index in women at 26 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 71 participants showed that mixed aerobic and resistance training was clinically effective compared to control in terms of Leeds MS quality of life, with serious imprecision.

Low to very low quality evidence from 4 RCTs comprising 210 participants showed that that there was no difference in clinical effectiveness between mixed aerobic and resistance training and control treatment in the MFIS cognitive at 12 weeks, and the knee flexor and extensor fatigue index at 26 weeks in men. Functional outcomes of timed 25 foot walk and timed get up and go were also clinically similar across groups. These inconclusive outcomes ranged from precise to very seriously imprecise.

Aerobic training versus neurorehabilitation

Very low quality evidence from 1 RCT comprising 11 participants showed that there was no difference in clinical effectiveness between aerobic training and neurorehabilitation in terms of MFIS total, physical, cognitive and psychosocial subscales, MSQOL-54 overall, physical and mental subscales, and 6 minute walking distance. The 6 minute walk outcome was very seriously imprecise. Precision of the fatigue and quality of life subscales was not estimable due to the narrative nature of results.

Massage versus usual care

Very low quality evidence from 1 RCT comprising 24 participants showed that massage was clinically effective compared to usual care in terms of fatigue at 5 weeks, with no imprecision.

Massage with aerobic/resistance exercise versus usual care

Very low quality evidence from 1 RCT comprising 24 participants showed that massage combined with aerobic/resistance exercise was clinically effective compared to usual care in terms of fatigue at 5 weeks, with no imprecision.

Massage versus aerobic/resistance exercise alone

Very low quality evidence from 1 RCT comprising 24 participants showed that there was no difference in clinical effectiveness between massage and aerobic/resistance exercise in terms of fatigue at 5 weeks, with serious imprecision.

Massage with aerobic/resistance exercise versus aerobic/resistance exercise alone

Very low quality evidence from 1 RCT comprising 24 participants showed that there was no difference in clinical effectiveness between massage combined with aerobic/resistance exercise and aerobic/resistance exercise alone in terms of fatigue at 5 weeks, with very serious imprecision.

Yoga versus aerobic training

Very low quality evidence from 1 RCT comprising 21 participants showed that there was no difference in clinical effectiveness between Yoga and aerobic training in terms of fatigue at 8 weeks, with very serious imprecision.

Wii balance board versus control

Moderate quality evidence from 1 RCT comprising 36 participants showed that in comparison to control, Wii balance board exercises had no clear effects in terms of fatigue.

Moderate quality evidence from 1 RCT comprising 36 participants showed that Wii balance board exercises were clinically effective compared to control in terms of balance, with no imprecision.

Mixed aerobic/resistance training plus CBT versus control

Low quality evidence from 1 RCT comprising 107 participants showed that mixed aerobic and resistance exercise was clinically effective compared to control in terms of the psychological domain of the MFIS scale, with serious imprecision.

Low to moderate quality evidence from 1 RCT comprising 107 participants showed that there was no difference in clinical effectiveness between mixed aerobic and resistance exercise coupled with CBT and control treatment in terms of quality of life (EQ-5D or MSQoL-54) at 3 or 9 months, or most indices of fatigue (MFIS total, physical, cognitive at 3 months and MFIS total, physical, cognitive and psychosocial at 9 months), with a range of precision from no imprecision to serious imprecision.

Supervised resistance and balance training versus control

Very low quality evidence from 1 RCT comprising 23 participants showed that supervised resistance and balance training was clinically effective compared to control in terms of effects on fatigue at 8 weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 23 participants showed that supervised resistance and balance training was clinically effective compared to control in terms of the physical domain of SF-36 at 8 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 23 participants showed that supervised resistance and balance training was clinically effective compared to control in terms of the role-physical functioning domain of SF-36 at 8 weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 23 participants showed that supervised resistance and balance training was clinically effective compared to control in terms of the bodily pain domain of SF-36 at 8 weeks, with no imprecision.

Low quality evidence from 1 RCT comprising 23 participants showed that supervised resistance and balance training was clinically effective compared to control in terms of 10 m walking time at 8 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 23 participants showed that supervised resistance and balance training was seriously imprecise but clinically effective compared to control in terms of Timed up and Go test at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 23 participants showed that there was no difference in clinical effectiveness between supervised resistance and balance exercise and control treatment in terms of the general health, vitality, social functioning, role emotional and mental health domains of SF-36 at 8 weeks , with very serious imprecision.

Home based resistance and balance training versus control

Very low quality evidence from 1 RCT comprising 19 participants showed that home resistance and balance training was clinically harmful compared to control in terms of fatigue at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 19 participants showed that home resistance and balance training was clinically harmful compared to control in terms of the role-emotional functioning domain of SF-36 at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 19 participants showed that home resistance and balance training was clinically harmful compared to control in terms of the mental health domain of SF-36 at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 19 participants showed that home resistance and balance training was clinically harmful compared to control in terms of on Timed up and Go test at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 19 participants showed that there was no difference in clinical effectiveness between home resistance and balance exercise and control treatment in terms of the 10 metre walking test, nor the physical functioning, role-physical, bodily pain, general health, vitality, or social functioning domains of SF-36 at 8 weeks, with very serious imprecision.

Supervised resistance and balance training versus home based resistance and balance training

Low quality evidence from 1 RCT comprising 24 participants showed that supervised resistance and balance training was clinically effective compared to home based resistance and balance training in terms of fatigue at 8 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 24 participants showed that supervised resistance and balance training had seriously imprecise but clinically effective compared to home based resistance and balance training in terms of the physical domain of SF-36 at 8 weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 24 participants showed that supervised resistance and balance training was clinically effective compared to home based resistance and balance training in terms of the role-physical functioning domain of SF-36 at 8 weeks, with no imprecision.

Low quality evidence from 1 RCT comprising 24 participants showed that supervised resistance and balance training was clinically effective compared to home based resistance and balance training in terms of the bodily pain of SF-36 at 8 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 24 participants showed that supervised resistance and balance training was clinically effective compared to home based resistance and balance training in terms of the role-emotional domain of SF-36 at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 24 participants showed that supervised resistance and balance training was clinically effective compared to home based resistance and balance training in terms of the mental health domain of SF-36 at 8 weeks, with very serious imprecision.

Low quality evidence from 1 RCT comprising 24 participants showed that supervised resistance and balance training was clinically effective compared to home based resistance and balance training in terms of 10 m walking time at 8 weeks, with no imprecision.

Low quality evidence from 1 RCT comprising 24 participants showed that supervised resistance and balance training was clinically effective compared to home based resistance and balance training in terms of the Timed up and Go test at 8 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 24 participants showed that there was no difference in clinical effectiveness between supervised resistance and balance exercise and home based resistance and balance training in terms of the general health, vitality, and social functioning domains of SF-36 at 8 weeks, with serious to very serious imprecision.

Vestibular rehabilitation versus control

Moderate quality evidence from 1 RCT comprising 25 participants showed that vestibular rehabilitation training was clinically effective compared to control in terms of MFIS total score at 6 weeks, with no imprecision.

Low quality evidence from 1 RCT comprising 25 participants showed that vestibular rehabilitation training was clinically effective compared to control in terms of MFIS total score at 10 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 25 participants showed that vestibular rehabilitation training was clinically effective compared to control in terms of 6 minute walk test change from baseline to 6 weeks, very with serious imprecision.

Yoga versus control

Low quality evidence from 1 RCT comprising 21 participants showed that yoga was clinically effective compared to control in terms of Fatigue at 8 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 21 participants showed that Yoga was clinically effective compared to control in terms of MFIS physical score at 12 weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 21 participants showed that yoga was clinically effective compared to control in terms of 2 minute walk distance at 8 weeks, with no imprecision.

Very low quality evidence from 1 RCT comprising 21 participants showed that Yoga was clinically effective compared to control in terms of MSQoL physical score at 12 weeks, with serious imprecision.

Low to very low quality evidence from 1 RCT comprising 21 participants showed that there was no difference in clinical effectiveness between Yoga and control treatment in terms of MFIS total and MFIS cognitive at 12 weeks, nor MSQoL mental score at 8 weeks, with serious imprecision.

Mixed aerobic/resistance versus yoga

Low to very low quality evidence from 2 RCTs comprising 203 participants showed that that there was no difference in clinical effectiveness between mixed aerobic/resistance exercise and yoga in terms of fatigue at 12 weeks (MFIS total, physical, cognitive and psychosocial scores) and 24 weeks (MFIS total score) with precision ranging from no serious imprecision to serious imprecision.

Resistance training and standard exercise versus standard exercise

Very low quality evidence from 1 RCT comprising 20 participants showed that resistance training and standard exercise was clinically harmful compared to control in terms of fatigue at 12 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 20 participants showed that there was no difference in clinical effectiveness between resistance training and standard exercise and standard exercise alone in terms of Timed up and Go and 6 minute walk test at 12 weeks, with very serious imprecision.

Electromagnetic field therapy versus placebo device

Very low quality evidence from 1 RCT comprising 37 participants showed that electromagnetic field therapy was clinically effective compared to control in terms of MFIS total at 12 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 37 participants showed that electromagnetic field therapy was clinically effective compared to control in terms of MFIS physical at 12 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 37 participants showed that electromagnetic field therapy was clinically effective compared to control in terms of MFIS cognitive at 12 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 37 participants showed that electromagnetic field therapy was clinically effective compared to control in terms of MFIS psychosocial at 12 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 37 participants showed that electromagnetic field therapy was clinically effective compared to control in terms of FSS at 12 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 37 participants showed that there was no difference in clinical effectiveness between electromagnetic field therapy and placebo in terms of self-reported fatigue, clinician graded fatigue or MSFC total at 12 weeks, with precision ranging from no serious imprecision to very serious imprecision.

Cognitive behavioural therapy versus control

Very low quality evidence from 2 RCTs comprising 112 participants showed that CBT was clinically effective compared to control in terms of fatigue score at 8-10 weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 72 participants showed that CBT was clinically effective compared to control in terms of fatigue related impairment (Work and Social adjustment scale) at 5 months, with serious imprecision.

Low quality evidence from 1 RCT comprising 72 participants showed that CBT was clinically effective compared to control in terms of fatigue related impairment (Work and social adjustment scale) at 8 months, with serious imprecision.

Low quality evidence from 1 RCT comprising 40 participants showed that CBT was clinically effective compared to control in terms of MFIS total score at 10 weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 72 participants showed that there was no difference in clinical effectiveness between CBT and control treatment in terms of fatigue scores at 5 and 8 months, and fatigue related impairment (Work and social adjustment scale) at 8 weeks, with serious imprecision.

Fatigue management /energy conservation versus control

Very low quality evidence from 1 RCT comprising 40 participants showed that fatigue management /energy conservation was clinically harmful compared to control in terms of the number of people with clinically relevant improvements in MFIS score, with serious imprecision.

Low quality evidence from 1 RCT comprising 23 participants showed that fatigue management /energy conservation was clinically effective compared to control in terms of MFIS total score at 4.25 months, with serious imprecision.

Low quality evidence from 1 RCT comprising 23 participants showed that fatigue management /energy conservation was clinically effective compared to control in terms of MFIS cognitive score at 4.25 months, with serious imprecision.

Low quality evidence from 1 RCT comprising 23 participants showed that fatigue management /energy conservation was clinically effective compared to control in terms of MFIS psychosocial score at 4.25 months, with serious imprecision.

Low quality evidence from 3 RCTs comprising 373 participants showed that fatigue management /energy conservation was clinically effective compared to control in terms of MFIS physical score at 5-6 weeks, with no imprecision.

Low quality evidence from 1 RCT comprising 146 participants showed that fatigue management /energy conservation was clinically effective compared to control in terms of Fatigue self-efficacy scale at 10 weeks, with serious imprecision.

Low to very low quality evidence from 5 RCT comprising 549 participants showed that there was no difference in clinical effectiveness between fatigue management /energy conservation and control in terms of MFIS total at 5/6 weeks, MFIS physical at 4.25 months, FSS at 4.25 months, FSS at 5/6 weeks, MFIS cognitive at 5/6 weeks, MFIS psychosocial at 5/6 weeks, global fatigue severity at 10 weeks, global fatigue severity at 5.5 months, fatigue self-efficacy scale at 5.5 months, all domains of the SF-36 at 6 months, the MSSE at 6 weeks, and the MSIS-29 at 10 weeks and 5.5 months. Precision varied between no serious imprecision and serious imprecision.

Mindfulness training versus control

Very low quality evidence from 1 RCT comprising 150 participants showed that mindfulness training was clinically effective compared to control in terms of MFIS total score at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 150 participants showed that mindfulness training was clinically effective compared to control in terms of MFIS total score at 6 months, with serious imprecision.

Very low quality evidence from 1 RCT comprising 150 participants showed that mindfulness training was clinically effective compared to control in terms of HAQUAMS score at 8 weeks, with serious imprecision.

Very low quality evidence from 1 RCT comprising 150 participants showed that there was no difference in clinical effectiveness between mindfulness training and control in terms of HAQUAMS score at 6 months, with serious imprecision.

10.3.5.2 Economic

One cost-utility analysis found that in adults with clinical definite MS diagnosis, current local practice was dominant (less costly and more effective) compared to group based fatigue management programme (FACETS) and current local practice for treating fatigue. This analysis was assessed as directly applicable with minor limitations.

One cost-utility analysis found that in adults with MS, aerobic and resistance exercise in combination with CBT and usual care was cost effective compared to usual care for treating fatigue (ICER: £10,137 per QALY gained). This analysis was assessed as directly applicable with minor limitations.

One original cost-utility analysis found that in adults with MS, with a one year time horizon supervised resistance and balance training was more effective and the most cost-effective option (ICER: £7,619 per QALY) compared to control and home based resistance and balance training for treating fatigue and mobility. This analysis was assessed as directly applicable and with potential serious limitations.

10.3.6 Recommendations and link to evidence

Recommendations	**49. Assess and offer treatment to people with MS who have fatigue for anxiety, depression, difficulty in sleeping, and any potential medical problems such as anaemia or thyroid disease.** **50. Explain that MS-related fatigue may be precipitated by heat, overexertion and stress or may be related to the time of day.** **51. Consider mindfulness-based training, cognitive behavioural therapy or fatigue management for treating MS-related fatigue.** **52. Advise people that aerobic, balance and stretching exercises including yoga may be helpful in treating MS-related fatigue.** **53. Consider supervised exercise programmes involving moderate progressive resistance training and aerobic exercise to treat people with MS who have mobility problems and/or fatigue.** **54. Consider a comprehensive programme of aerobic and moderate progressive resistance activity combined with cognitive behavioural techniques for fatigue in people with MS with moderately impaired mobility (an EDSS[cc] score of greater than or equal to 4).** **55. Consider vestibular rehabilitation for people with MS who have fatigue or mobility problems associated with limited standing balance.** **56. Encourage people with MS to keep exercising after treatment programmes end for longer term benefits (see Behaviour change: individual approaches NICE public health guideline 49).** **57. Help the person with MS continue to exercise, for example by referring them to exercise referral schemes.** **58. If more than one of the interventions recommended for mobility or fatigue are suitable, offer treatment based on which the person prefers and whether they can continue the activity when the treatment programme ends.**
Relative values of different outcomes	As with pharmacological treatments for fatigue, the GDG noted the subjective nature of fatigue outcome measures. Most non-pharmacological studies of fatigue used the Fatigue Severity Scale or the Modified Fatigue Impact Scale. Quality of life outcomes were also considered when available, and timed walking distances were used as functional measures of fatigue. Most studies examined a programme or course of therapy/treatment/activity. Where possible, the GDG valued long-term sustained improvements in outcomes after the course had ended. For example, cognitive behavioural therapy and

cc Expanded Disability Status Scale.

	mindfulness training showed benefits in fatigue for over three months after the programme ended. Resistance training for fatigue was of benefit if measured at 12 weeks (the end of therapy) but not of benefit when measured at 22 weeks, which may be explained by a reduction in self-directed exercise over the follow up period. No studies assessed return to normal activities as an outcome, but the GDG thought this would be a useful measure in future studies.
Trade off between clinical benefits and harms	Clinical benefit was considered to be present if there was improvement in scales of fatigue, or in overall functioning. There appeared to be clinically beneficial reductions in fatigue from moderately intensive resistance training, aerobic training and balance training, as well as yoga, electromagnetic field therapy and vestibular rehabilitation in people with balance deficits. Cognitive behavioural therapy, mindfulness based training, and fatigue management/energy conservation were also beneficial. Unsupervised resistance training at home appeared to worsen fatigue and intellectual functioning, although this may be a result of poorer compliance. The GDG agreed that unsupervised exercise programmes did carry a risk of injury and worsening of function. Very high intensity resistance training was also shown to cause a harm, in comparison to standard resistance training. Other therapies had minimal known risks or these were not measured. The GDG did not prioritise different outcomes but listed all therapies with evidence of benefit in one or more relevant outcomes.
Economic considerations	There are costs associated with assessing and treating people with fatigue for anxiety, depression, difficulty in sleeping, and any potential medical problems such as anaemia or thyroid disease. The GDG considered identifying and treating the underlying cause of fatigue justified the cost. One cost-utility analysis was identified which found that in adults with clinical definite MS diagnosis, current local practice was dominant (less costly and more effective) compared to a group based fatigue management programme (FACETS) and current local practice for treating fatigue. In this study, the group based fatigue management programme and current local practice resulted in a decrease in QALYs compared to current local practice. The authors suggest that a longer term follow-up may be required for improvements as a result of changes in attitudes and lifestyle (central to the FACETS programme) to impact on quality of life. The GDG agreed with the authors and noted that this study, as well as the other studies included in the clinical review showed improvements in scales of fatigue. Furthermore, two studies identified in the clinical review (Mathiowetz 2005 and Finlayson 2011) showed improvements in SF-36 subscales for the group based fatigue management group compared to control. The cost of the FACETS programme was estimated to be £453 per participant. Therefore, based on this cost and the evidence of clinically meaningful improvements in fatigue, the GDG felt that these programmes are likely to be cost effective. No economic evidence was identified for mindfulness or CBT. The GDG considered the unit costs of group-based mindfulness interventions (£357 per user) and individual CBT interventions (£726 per user). The GDG felt the benefits in terms of improvements in scales of fatigue and overall functioning justified the cost of the intervention. Furthermore, the GDG discussed that in current practice; CBT may be conducted as a group and therefore would be less costly per user. No economic evidence was identified for yoga. The cost of the time spent by healthcare professionals in providing advice to people with MS on yoga is likely to be minimal. The clinical evidence showed beneficial effects of yoga on fatigue and therefore the provision of advice on yoga is likely to be cost effective. A simple cost-utility analysis was undertaken by the NCGC based on the results of an RCT by Cakit (2010)[34,34] evaluating the effects of supervised and

	unsupervised progressive resistance and balance training compared to no intervention on mobility and fatigue. The cost of each intervention was estimated based on published unit costs and within trial resource use. Quality of life values were estimated by mapping SF-36 scores to EQ-5D values using an algorithm by Ara and Brazier (2008).[8,8] Two time horizons were considered, 8 weeks to reflect the duration of the intervention and one year which assumed that the effectiveness of the intervention was maintained after it is completed. With a one year time horizon, supervised training was the most cost effective option. With the 8 week time horizon neither supervised nor unsupervised training were cost-effective compared to control. The GDG agreed that supervised programmes were preferable to unsupervised ones. They also discussed the importance of selecting activities that can people can continue following the end of a supervised treatment programme. A cost-utility analysis found that in adults with MS, aerobic and resistance exercise in combination with CBT and usual care was cost effective compared to usual care for treating fatigue. A scenario analysis found that compared to usual care, the intervention was cost effective in people with more severely impaired mobility (EDSS >4) but was dominated (more costly and less effective) in people with moderately impaired mobility (EDSS<4). No economic evidence was identified for vestibular rehabilitation. The GDG considered that for people with fatigue or mobility problems associated with sensory deficits, such an intervention, which would be conducted by a physiotherapist or occupational therapist, is likely to be cost-effective.
Quality of evidence	The evidence was mostly very low to low quality. Furthermore, for most of the individual outcomes of a therapy, there were only one or two studies. The population was noted to be limited to relapsing remitting MS with an EDSS less than seven in most studies, and therefore may be less applicable to other patients with MS. The economic evidence for group based fatigue management programme (FACETS) and current local practice compared to current local practice was assessed as directly applicable with minor limitations. The economic evidence for supervised versus home based resistance and balance training versus control was assessed as directly applicable with potential serious limitations.
Other considerations	Fatigue was acknowledged as a prominent symptom in MS and seems to be different to physiological fatigue. However there is no accepted definition of fatigue and studies do not define fatigue or differentiate between types of fatigue. The GDG looked at the programme of therapy itself and not the type of staff or healthcare professionals used. It is assumed that any of our recommended therapies would be delivered by a person or persons competent in that field. Mood was recognised as an important component of fatigue. Stress, depression and sleep disturbance may contribute to increased fatigue and should be considered when managing people with MS and fatigue. Other medical disorders such as hypothyroidism should be considered too. Interventions to reduce heat sensitive fatigue were considered by the GDG. No high quality studies have been carried out into the management of heat sensitive fatigue for people with MS. The GDG agreed that this was an important area and that further research was required.

10.4 Non-pharmacological management of mobility

10.4.1 Introduction

Reduced mobility is a common manifestation of the gradual decline in function that may occur in MS. Causes include muscle weakness, spasticity, disordered balance, co-ordination problems and visual deficits. Although some of these causes may be amenable to pharmacological treatment, non-pharmacological methods may be particularly useful in addressing causes related to motor control.

10.4.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological programmes (including self-management programmes) for mobility?

For full details see review protocol in Appendix C.

Table 138: PICO characteristics of review question

Population	• Adults with MS only
Intervention/s	Any non-pharmacological management programme, including self-management programmes , for example: • Multidisciplinary rehabilitation/programmes • Self-management programmes • Treatment programmes for various symptoms • FACETS prog, energy conservation programs, mindfulness (Grossman Paul), exercise (John Saxton), Getting To Grips (MS Society), stretching, standing, splinting, gym prescription, diet, yoga, tai chai, pilates, relaxation, lycra garments
Comparison/s	Usual treatment or placebo
Outcomes	• mobility [symptoms or measures (ie FSS)] Also, any of the following outcomes, provided the treatment has been directed at **impaired mobility**: • Quality of life • Function (i.e. EDSS, ambulation measures, MSIS, Guys scale etc) • carer perceptions • Incidence of adverse events
	Systematic reviews, RCTs. Include cross-over studies.

10.4.3 Clinical evidence

Summary of included studies

25 RCTs were found on non-pharmacological interventions for mobility [2]; [23]; [24]; [34]; [41]; [50]; [53]; [54]; [73,74]; [70]; [89]; [93]; [96]; [120]; [125]; [158]; [182]; [186]; [192]; [198]; [233]; [242]; [252]; [262]. The Cochrane review[195,195] on exercise therapy for multiple sclerosis was checked for relevant included papers (the review presented the outcomes narratively).

Table 139: Summary of studies included in the review

Study	Intervention/comparison	Population characteristics	N randomised /analysed
Dettmers 2009[53]	Aerobic versus control	Mean age intervention 45.8 control 39.7, Modified Fatigue Impact Scale 36.8/41.8; EDSS <4.5; mostly female; mostly RR	30/30, but depended on outcome
Harvey 1999[89]		Mean age intervention 49 control 43, ; time since disease duration 5-10	20/15
Hebert 2011[97]		Mean age intervention 43 control 50, MFIS ≥ 45; ambulant >100m with/without aids;	26/26
Mostert 2002[158]		Mean age intervention 45 control 44, EDSS 1-6.5; mostly relapsing progressive	37/26
Van den Berg 2006[252]		Able to walk 10m in <60 secs;	19/16
Rampello 2007[192]	Aerobic versus neurorehabiliation	Aged 20-55, EDSS<7;	11/11
Bjarnadottir 2007[23]	Aerobic + resistance versus control	Age<50 years MS; EDSS < 4, ability to ride a stationary bicycle. Mostly female; all RR	23/19
Garrett 2013A[73,74]		Aged c50; mostly RR;	151/112
Hayes 2011[93]		18-65; ambulatory with/without assistive devices	22/19
Romberg 2005[198]		Aged between 30 and 55 yrs, clinically and/or laboratory-defined MS and an EDSS score of 1.0 to 5.5 (inclusive); aged 43`	95/95
Learmouth 2012[120]		Age c50, EDSS 5-6.5; MMSE >24; mostly female; years since onset 13.4/12.6	32/25
Garrett 2013A[73,74]	Aerobic + resistance versus yoga	Aged c50; mostly RR;	151/126
DeBolt 2004[50]	Resistance training versus control	Age c50, Ability to walk (with or without assistive devices) at least 20 m without rest. Mostly female	37/36
Dodd 2011[54]		Age c50, AI score of 2-4; 41/71 MFIS > 38	76/71
Harvey 1999[89]		Aged 43-49; time since disease duration 5-10	20/15
Tarakci 2013[242]		Age c40, EDSS 2-6.5; FSS 39.3/39.9; mostly RR	110/99
Cakit 2010[34];	Supervised resistance/balance versus control	Age 35-43, EDSS≤6; able to stand independently > 3 secs;	30/23

Study	Intervention/comparison	Population characteristics	N randomised /analysed
	home based resistance/balance versus control	EDSS≤6; able to stand independently > 3 secs;	30/19
	Supervised versus home based resistance/balance	EDSS≤6; able to stand independently > 3 secs;	30/24
Plow 2014[182]	Home resistance + pamphlets versus control	Age c48, Ability to walk 25 feet with or without cane	30/30
Solari 1999[233]	Hospital stretching + aerobic versus home stretching + aerobic	Mean age intervention 63 control 48, EDSS between 3.0 and 6.5, 48-63% women; RR 22%	50/50
Fuller 1996[70]	Inpatient physiotherapy versus control	Mean age 46, recent deterioration in their ability to walk or transfer to and from a wheelchair	45/45
Wiles 2001[262]	Outpatient physiotherapy versus control	Mean age 47, Able to walk 5 m with or without a mechanical aid.; median EDSS 6-6.5; symptom duration 12	42/40
	Home physiotherapy versus control		42/40
Lord 1998[125]	Task orientated versus facilitated physiotherapy	Mean age 54-62; able to walk 10 m with or without supervision;; disease duration 14-18	23/20
Prosperini 2013[186];	Balance versus control	18-50 years; RR or SP; EDSS ≤5.5; ability to walk without resting for >100m; disease duration 9-12	36/34
Claerbout 2012[41]	Whole body vibration + physiotherapy versus physiotherapy	Mean age 39-48; EDSS 3-7; disease duration 10-12 yrs	55/47
Hayes 2011[93]	High resistance + standard exercise versus standard exercise	Aged 18-65; EDSS 5.2; 11/19 women	20/19
Ahmadi 2010[2]	Yoga versus control	Mean age 32-36; EDSS 1-4; DMDs allowed; disease duration 5 years	21/21
Garret 2013A[73]		Aged c50; mostly RR;	148/112
Garret 2013A[73]	Yoga versus mixed resistance/aerobic	Aged c50; mostly RR;	157/126
Garret 2013 [74]			157/79
Hebert 2011[97]	Vestibular rehab versus control	Mean age 43-50; MFIS ≥ 45; ambulant >100m with/without aids;	25/25
	Vestibular rehab versus aerobic		
Bombardier 2008[24]	Motivational interviewing versus control	Mean age 45-47; EDSS<6; able to walk 90m without assistance; all types of MS	130/130

Table 140: Clinical evidence profile: Aerobic versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic versus control mean (SD) [n]	Control mean (SD)[n]	Relative (95% CI)	Absolute		
2 min walk (m) (follow-up 7 weeks; Better indicated by higher values)												
Van den Berg 2006	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	10.8 (6.7)[8]	5.8 (7.8) [8]	-	MD 5 higher (2.13 lower to 12.13 higher)	VERY LOW	CRITICAL
6 Minute Walk Test (feet) (6 wks) (follow-up 6 weeks; Better indicated by higher values)												
Herbert 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	1112.1 (391.3) [13]	1,071.6 (375) [13]	-	MD 40.5 higher (254.12 lower to 335.12 higher)	VERY LOW	CRITICAL
6 Minute Walk Test (feet) (10 wks) (follow-up 7 weeks; Better indicated by higher values)												
Herbert 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	1053.9 (448.7) (13)	1,100.5 (284) [13]	-	MD 46.6 lower (335.26 lower to 242.06 higher)	VERY LOW	CRITICAL
10 m timed walk (secs) (follow-up 7 weeks; Better indicated by lower values)												
Van den Berg 2006	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-3.1 (2.5) [8]	-0.6 (1.4) [8]	-	MD 2.5 lower (4.49 to 0.51 lower)	VERY LOW	CRITICAL
Increase in walking distance from baseline (m) (follow-up 3 weeks; Better indicated by higher values)												
Dettmers 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	650 (474) [15]	97 (70) [15]	-	MD 553 higher (310.53 to 795.47 higher)	VERY LOW	CRITICAL
Increase in walking time from baseline (min) (follow-up 3 weeks; Better indicated by higher values)												
Dettmers	randomised	very	no serious	no serious	very	none	11.3 (6)	1.3 (1)	-	MD 10 higher		CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic versus control mean (SD) [n]	Control mean (SD)[n]	Relative (95% CI)	Absolute		
2009	ed trials	serious[a]	inconsistency	indirectness	serious[b]		[15]	[15]		(6.92 to 13.08 higher)	VERY LOW	L
Fatigue Severity Scale (follow-up 4-7 weeks; Better indicated by lower values)												
Mostert 2002 Van den Berg 2006	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	4.4 (1.9) [13] -4.5 (7.7) [8]	5 (1.9) [13] -4.4 (7.8) [8]	-	MD 0.58 lower (2.02 lower to 0.85 higher)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (6 wks) (follow-up 6 weeks; Better indicated by lower values)												
Herbert 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	44.3 (16.4) [13]	52.1 (17.1) [13]	-	MD 7.8 lower (20.68 lower to 5.08 higher)	LOW	CRITICAL
Multiple Sclerosis Impact Scale (10 wks) (follow-up 10 weeks; Better indicated by lower values)												
Herbert 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	44.7 (16.3 [13]	52.6 (17.4) [13]	-	MD 7.9 lower (20.86 lower to 5.06 higher)	VERY LOW	CRITICAL
Proportion improvement in Multiple Sclerosis Impact Scale change from baseline (follow-up 3 weeks)												
Dettmers 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	6/9 (66.7%)	90%	RR 0.74 (0.45 to 1.23)	234 fewer per 1000 (from 495 fewer to 207 more)	VERY LOW	CRITICAL
Proportion improvement in Multiple Sclerosis Impact Scale (motor) from baseline (follow-up 3 weeks)												
Dettmers 2009	randomised trials	very serio	no serious inconsistency	no serious indirectness	very serious[b]	none	8/9 (88.9%)	90%	RR 0.99 (0.72 to	9 fewer per 1000 (from 252 fewer	VERY	CRITICAL

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic versus control mean (SD) [n]	Control mean (SD)[n]	Relative (95% CI)	Absolute	Quality	Importance
		us[a]							1.35)	to 315 more)	LOW	
Proportion Improvement in HAQUAMS (motor) from baseline (follow-up 3 weeks)												
Dettmers 2009	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	5/9 (55.6%)	70%	RR 0.79 (0.39 to 1.62)	147 fewer per 1000 (from4273 fewer to 434 more)	VERY LOW	CRITICAL
Guys Neurological Disability Scale (follow-up 7 weeks; Better indicated by higher values)												
Van den Berg 2006	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	4.1 (8.6) [8]	4.3 (9.5)[8]	-	MD 0.2 lower (9.08 lower to 8.68 higher)	VERY LOW	CRITICAL
Work activity (follow-up 4 weeks; Better indicated by higher values)												
Mostert 2002	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	2.6 (0.6) [13]	2.7 (0.9) [13]	-	MD 0.1 lower (0.69 lower to 0.49 higher)	VERY LOW	CRITICAL
Sport activity (follow-up 4 weeks; Better indicated by higher values)												
Mostert 2002	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	2 (0.4) [13]	1.7 (0.4) [13]	-	MD 0.3 higher (0.01 lower to 0.61 higher)	VERY LOW	CRITICAL
Leisure activity (follow-up 4 weeks; Better indicated by higher values)												
Mostert 2002	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	2.5 (0.8)[13]	1.7 (0.4)[13]	-	MD 0.1 higher (0.52 lower to 0.72 higher)	VERY LOW	CRITICAL
Quality of life												
No evidence for this outcome												

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic versus control mean (SD) [n]	Control mean (SD)[n]	Relative (95% CI)	Absolute	Quality	Importance
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 141: Clinical evidence profile: Aerobic versus neurorehabilitation

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic mean (SD) [n]	Neurorehabilitationl mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
Walking distance (m) (follow-up 8 weeks; Better indicated by higher values)												
Rampello 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	332 (108)[11]	308 (110)[11]	-	MD 24 higher (67.1 lower to 115.1 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic mean (SD) [n]	Neurorehabilitationl mean (SD) [n]	Relative (95% CI)	Absolute		
Walking speed (m/min) (follow-up 8 weeks; Better indicated by lower values)												
Rampello 2007	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	55 (18[11]	51 (18)[11]	-	MD 4 higher (11.04 lower to 19.04 higher)	VERY LOW	CRITICAL
Quality of life												
No evidence for this outcome												
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 142: Aerobic + resistance versus control

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic + resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
MSIS-29 vs (physical component) mean change (follow-up 12 weeks; Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-6.9 (15.49)[63]	0.3 (14.97)[49]	-	MD 7.2 lower (12.87 to 1.53 lower)	VERY LOW	CRITICAL
Leeds MS quality of life (follow-up 12 weeks; Better indicated by lower values)												
Learmonth 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.9(3.9)[15]	12.4 (3.1)[10]	-	MD 1.5 lower (4.25 lower to 1.25 higher)	VERY LOW	CRITICAL
MS Functional Composite mean change (follow-up 6 months; Better indicated by higher values)												
Romberg 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.11 (0.35)[47]	-0.13(0.4648	-	MD 0.24 higher (0.08 to 0.41 higher)	VERY LOW	CRITICAL
MSQOL-54 Mental component (follow-up 6 months; Better indicated by higher values)												
Romberg 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	71.2 (20.6)[47]	70.4(21.3)[48]	-	MD 0.8 higher (7.63 lower to 9.23 higher)	LOW	CRITICAL
MSQOL-54 Physical component (follow-up 6 months; Better indicated by higher values)												
Romberg 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	63 (17.8)[47]	63.3 (16/6[48]	-	MD 0.3 lower (7.22 lower to 6.62 higher)	LOW	CRITICAL
Timed 25-Foot Walk Test (s) (follow-up 12 weeks; Better indicated by lower values)												
Learmonth 2012 Romberg 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	No serious imprecision	none	14.9 (13.6) [15] 0.19 (0.49)[47]	13.1 (8.6 [10] -0.12 (0.49)[48]	-	MD 0.30 (0.11 to 0.50 higher)	LOW	CRITICAL
6 minute walk test (m) (follow-up 24 weeks; Better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic + resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Learmonth 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	262.2 (127.4) [15]	215.8 (175.7 [10]	-	MD 46.4 higher (80.15 lower to 172.95 higher)	VERY LOW	CRITICAL
Timed up and go (s) (follow-up 24 weeks; Better indicated by lower values)												
Learmonth 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	18.4 (14.95)[15]	16.22(11) [10]	-	MD 2.18 higher (8 lower to 12.36 higher)	VERY LOW	CRITICAL
PhoneFITT (follow-up 24 weeks; Better indicated by higher values)												
Learmonth 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	78.2 (35.5)[15]	54.6 (16.7)[10]	-	MD 23.6 higher (2.87 to 44.33 higher)	VERY LOW	CRITICAL
Paced Auditory Serial Additions Test change score (follow-up 6 months; Better indicated by higher values)												
Romberg 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.092 (0.71)[47]	-0.16(0.71)[48]	-	MD 0.25 higher (0.03 lower to 0.54 higher)	VERY LOW	CRITICAL
Nine Hole Peg Test change score (follow-up 6 months; Better indicated by higher values)												
Romberg 2005	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.07 (0.37)[47]	-0.11)[48]	-	MD 0.18 higher (0.03 to 0.33 higher)	VERY LOW	CRITICAL
Berg Balance Scale (follow-up 24 weeks; Better indicated by higher values)												
Learmonth 2012	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	46.7 (10.6)[15]	40.9 (15.2)[10]	-	MD 5.8 higher (5.04 lower to 16.64 higher)	VERY LOW	CRITICAL
Modified Fatigue Impact Scale (total score) (follow-up 12 weeks; Better indicated by lower values)												
Garrett	randomis	very	no serious	no serious	serious[b]	none	-7.5 (14.29[63]	-1.1	-	MD 6.4 lower		CRITICA

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic + resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
2013	ed trials	serious[a]	inconsistency	indirectness				(11.83)[49]		(11.24 to 1.56 lower)	VERY LOW	L
Modified Fatigue Impact Scale (physical) (follow-up 12 weeks; Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-3.9 (6.75)[63]	-1.1 (11.83)[49]	-	MD 4.3 lower (6.42 to 2.18 lower)	VERY LOW	CRITICAL
Modified Fatigue Impact Scale (cognitive) change score (follow-up 12 weeks; Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-2.1 (4.17)[63]	-0.51 (4.18)[49]	-	MD 1.59 lower (3.15 to 0.03 lower)	VERY LOW	CRITICAL
Fatigue severity scale (follow-up 24 weeks; Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	5 (1.8)[15]	6.2 (0.7)[10]	-	MD 1.2 lower (2.21 to 0.19 lower)	VERY LOW	CRITICAL
Activities balance confidence (follow-up 24 weeks; Better indicated by lower values)												
Learmonth 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	79.8 (28.3)[15]	60.9 (35.6)[10]	-	MD 18.9 higher (7.41 lower to 45.21 higher)	VERY LOW	CRITICAL
Hospital anxiety and disability scale (follow-up 6 months; Better indicated by lower values)												
Learmonth 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	11.7 (5.9)[15]	13.8(6.6[10]	-	MD 2.1 lower (7.16 lower to 2.96 higher)	VERY LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												
Adverse events												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic + resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 143: Clinical evidence profile: Aerobic + resistance versus yoga

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic + resistance mean (SD) [n]	Yoga mean (SD) [n]	Relative (95% CI)	Absolute		
Multiple Sclerosis Impact Scale-29 v2 change score (physical) (12 wks) (Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-6.9(15.49)[63]	-4 (13.90[63]	-	MD 2.9 lower (8.04 lower to 2.24 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Impact Scale-29 v2 (physical) (24 wks) (Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	27.7 (16.2)[63]	34 (21.8)[37]	-	MD 6.3 lower (14.38 lower to 1.78 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Impact Scale-29 v2 (psychological) (24 wks) (Better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Aerobic + resistance mean (SD) [n]**	**Yoga mean (SD) [n]**	**Relative (95% CI)**	**Absolute**		
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	23.4 (14.8[41]	30.1 (20.9)[37]	-	MD 6.7 lower (14.82 lower to 1.42 higher)	VERY LOW	CRITICAL
6 min walking test m(24 wks) (Better indicated by higher values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	313.9 (104.9)[34]	281.7(112.5)[37]	-	MD 32.2 higher (18.37 lower to 82.77 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (total) change score (12 wks) (Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-7.5 (14.29)[63]	-5.8 (23.02[63]	-	MD 1.7 lower (8.39 lower to 4.99 higher)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (total) (24 wks) (Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32.9 (`4.6)[41]	33.9 (19.2)[36]	-	MD 1 lower (8.7 lower to 6.7 higher)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (cognitive) change score 12 wks (Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-2.1 (4.17)[63]	-0.96 (3.57)[63]	-	MD 1.14 lower (2.5 lower to 0.22 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (physical) 12 wks change score (Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-3.9 (6.75)[63]	-2.1 (6.35)[63]	-	MD 1.8 lower (4.09 lower to 0.49 higher)	VERY LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic + resistance mean (SD) [n]	Yoga mean (SD) [n]	Relative (95% CI)	Absolute		
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 144: Resistance versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
MusiQoL (follow-up 12 weeks; Better indicated by lower values)												
Tarakci 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	1.98 (5)[51]	-0.4 (5) [48]	-	MD 2.38 higher (0.41 to 4.35 higher)	VERY LOW	CRITICAL
WHOQOL-BREF QoL change from baseline (follow-up 10 weeks; Better indicated by higher values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.4 (0.9) [36]	0.1 (0.8)[35]	-	MD 0.3 higher (0.1 lower to 0.7 higher)	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
WHOQOL-BREF QoL change from baseline (follow-up 22 weeks; Better indicated by higher values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-0.1(1.1)[36]	0.1(0.8)[35]	-	MD 0.2 lower (0.65 lower to 0.25 higher)	LOW	CRITICAL
WHOQOL-BREF health change from baseline ((follow-up 10 weeks; Better indicated by higher values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.3 (1.2)[36]	-0.1 (1)[35]	-	MD 0.4 higher (0.11 lower to 0.91 higher)	LOW	CRITICAL
WHOQOL-BREF health change from baseline (follow-up 22 weeks; Better indicated by higher values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0.1 (1.1)[36]	0.1(1)[35]	-	MD 0 higher (0.49 lower to 0.49 higher)	MODERATE	CRITICAL
WHOQOL-BREF physical health change from baseline (follow-up 10 weeks; Better indicated by higher values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	1.8 (3.4)[36]	0.3 (2.8)[35]	-	MD 1.5 higher (0.05 to 2.95 higher)	LOW	CRITICAL
WHOQOL-BREF physical health change from baseline (follow-up 22 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.3 (3.3)[36]	0.9 (3.2)[35]	-	MD 0.6 lower (2.11 lower to 0.91 higher)	LOW	CRITICAL
10 m walking test (s) (follow-up 12 weeks; Better indicated by lower values)												
Tarakci 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	4.73(9.05)[51]	1.45 (9.06)[48]	-	MD 6.18 lower (9.75 to 2.61 lower)	VERY LOW	CRITICAL
Fast walking speed (m/s) change from baseline (follow-up 10 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0.05 (0.17)[36]	0.01(0.19)[35]	-	MD 0.04 higher (0.04 lower to 0.12 higher)	MODERATE	CRITICAL
Fast walking speed (m/s) change from baseline (follow-up 22 weeks; Better indicated by higher values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-0.02 (0.19)[36]	0.01 (0.18)[35]	-	MD 0.03 lower (0.12 lower to 0.06 higher)	MODERATE	CRITICAL
2 min walk distance (m) change from baseline (follow-up 10 weeks; Better indicated by higher values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	2.8 (14.4)[36]	0.7 (13.4[35]	-	MD 2.1 higher (4.37 lower to 8.57 higher)	LOW	CRITICAL
2 min walk distance (m) change from baseline (follow-up 22 weeks; Better indicated by higher values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-1.6 (15.6)[36]	1.6(9)[35]	-	MD 3.2 lower (9.1 lower to 2.7 higher)	VERY LOW	CRITICAL
Power (W/kg) (follow-up 10 weeks; Better indicated by higher values)												
De Bolt 2004	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	3.95 (1.23)[19]	3.68(1.22)[17]	-	MD 0.27 higher (0.53 lower to 1.07 higher)	VERY LOW	CRITICAL
Power (W) (follow-up 10 weeks; Better indicated by higher values)												
De Bolt 2004	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	282(65.44)[19]	290.04 (110.23)[17]	-	MD 8.04 lower (68.14 lower to 52.06 higher)	VERY LOW	CRITICAL
Balance AP sway (cm/s) (follow-up 10 weeks; Better indicated by lower values)												
De	randomis	very	no serious	no serious	serious[b]	none	0.382	0.412	-	MD 0.03 lower	VERY	CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Bolt 2004	ed trials	serious[a]	inconsistency	indirectness			(0.212)[19]	(0.256)[17]		(0.18 lower to 0.12 higher)	LOW	L
Balance ML sway (cm/s) (follow-up 10 weeks; Better indicated by lower values)												
De Bolt 2004	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0.212 (0.13)[19]	0.235 (0.129)[17]	-	MD 0.02 lower (0.11 lower to 0.06 higher)	VERY LOW	CRITICAL
Balance velocity sway (cm/s) (follow-up 10 weeks; Better indicated by lower values)												
De Bolt 2004	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	1.727 (0.778)[19]	1.748 (0.49)[17]	-	MD 0.02 lower (0.44 lower to 0.4 higher)	VERY LOW	CRITICAL
Up and Go (s) (follow-up 10 weeks; Better indicated by lower values)												
De Bolt 2004	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	9.15 (2.26)[19]	11.08 (5.21)[17]	-	MD 1.93 lower (4.61 lower to 0.75 higher)	VERY LOW	CRITICAL
Fatigue Severity Scale (follow-up 12 weeks; Better indicated by lower values)												
Tarakci 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-8.26 (16.9)[51]	3.29 (16.9)[48]	-	MD 11.55 lower (18.21 to 4.89 lower)	VERY LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (total) change from baseline (follow-up 10 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-10.2 (11.2)[36]	-3 (14.1)[35]	-	MD 7.2 lower (13.13 to 1.27 lower)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (total) change from baseline (follow-up 22 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-2.9 (12.8)[36]	-4.8 (12.4) [35]	-	MD 1.9 higher (3.96 lower to	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
										7.76 higher)		
Multiple Sclerosis Fatigue Impact Scale (physical) change from baseline (follow-up 10 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-2.6 (6.8)[36]	-2.1 (5.4)[35]	-	MD 0.5 lower (3.35 lower to 2.35 higher)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (physical) change from baseline (follow-up 22 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious2	none	-5.9 (5.9)[36]	-1.8 (6.8)[35]	-	MD 4.1 lower (7.06 to 1.14 lower)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (cognitive) change from baseline (follow-up 10 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-3.2 (5.9)[36]	-1.7 (6.9)[35]	-	MD 1.5 lower (4.49 lower to 1.49 higher)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (cognitive) change from baseline (follow-up 22 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-0.2 (7) [36]	-2.1 (6.3)[35]	-	MD 1.9 higher (1.2 lower to 5 higher)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (psychosocial) change from baseline (follow-up 10 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	-1.1(1.6)[36]	-0.4(2.4)[35]	-	MD 0.7 lower (1.65 lower to 0.25 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (psychosocial) change from baseline (follow-up 22 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-0.1(2)[36]	-0.5 (2.2)[35]	-	MD 0.4 higher (0.58 lower to 1.38 higher)	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Resistance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
AEs stiffness MSIS-88 change from baseline (follow-up 10 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-3.6 (7.6)[36]	-0.5 (6)[35]	-	MD 3.1 lower (6.28 lower to 0.08 higher)	LOW	CRITICAL
AEs stiffness MSIS-88 change from baseline (follow-up 22 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-0.5 (7)[36]	-0.7 (7.7)[35]	-	MD 0.2 higher (3.23 lower to 3.63 higher)	MODERATE	CRITICAL
AEs muscle spasm MSIS-88 change from baseline (follow-up 10 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-2 (6.2)[36]	0.5 (6)[35]	-	MD 2.5 lower (5.34 lower to 0.34 higher)	LOW	CRITICAL
AEs muscle spasm MSIS-88 change from baseline (follow-up 22 weeks; Better indicated by lower values)												
Dodd 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	1.1 (8.2)[36]	-1.1(7.5[35]	-	MD 2.2 higher (1.45 lower to 5.85 higher)	LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 145: Clinical evidence profile: Supervised resistance + balance versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
SF-36 - Physical functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	21.2 (14.4)[14]	7.7 (7.4)[9]	-	MD 13.5 higher (4.54 to 22.46 higher)	LOW	CRITICAL
SF-36 - Role-physical functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	34 (30.1) [14]	5 (44.7) [9]	-	MD 29 higher (4.19 lower to 62.19 higher)	VERY LOW	CRITICAL
SF-36 - Bodily pain change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	8.8 (5.8)[14]	2 (2.1)[9]	-	MD 6.8 higher (3.47 to 10.13 higher)	LOW	CRITICAL
SF-36 - General health change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	4.3 (8.4)[14]	3.2 (11.7) [9]	-	MD 1.1 higher (7.72 lower to 9.92 higher)	VERY LOW	CRITICAL
SF-36 – Vitality change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	9 (19.3 [14]	11 (20.4)[9]	-	MD 2 lower (18.73 lower to 14.73 higher)	VERY LOW	CRITICAL
SF-36 - Social functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	23.1 (23.1)[14]	5 (16.7)[9]	-	MD 1.6 lower (17.89 lower to 14.69 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
SF-36 - Role-emotional functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	24.2 (49.6)[14]	19.9 (50.5)[9]	-	MD 4.3 higher (37.69 lower to 46.29 higher)	VERY LOW	CRITICAL
SF-36 - Mental health change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	7.2 (13.4)[14]	7 (6.7)[9]	-	MD 0.2 higher (8.07 lower to 8.47 higher)	VERY LOW	CRITICAL
10 m walking test s change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-1.9 (1.2)[14]	0.1 (0.8)[9]	-	MD 2 lower (2.82 to 1.18 lower)	VERY LOW	CRITICAL
Duration of exercise (mins) change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	8.4 (3.8) [14]	3.3 (5.3) [9]	-	MD 5.1 higher (1.11 to 9.09 higher)	VERY LOW	CRITICAL
Tolerated maxi wkload on bicycle change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	123.6 (18)[14]	22 (13.03) [9]	-	MD 101.6 higher (88.9 to 114.3 higher)	LOW	CRITICAL
Timed up and go test (s) change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-1.3 (1.2)[14]	-0.2 (0.8)[9]	-	MD 1.1 lower (1.92 to 0.28 lower)	VERY LOW	CRITICAL
Dynamic Gait Index change score (follow-up 8 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	2.7 (0.5)[14]	0.4 (0.4)[9]	-	MD 2.3 higher (1.93 to 2.67 higher)	LOW	CRITICAL
Functional reach change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	7.3 (2.4)[14]	-1 (2.04)[9]	-	MD 8.3 higher (6.47 to 10.13 higher)	LOW	CRITICAL
Fatigue Severity Scale - change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-9.5 (2.8)[14]	-5.2 (5.3)[9]	-	MD 4.3 lower (8.06 to 0.54 lower)	VERY LOW	CRITICAL
Falls Efficacy Scale change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-11.3 (7.8[14]	-2.6 (3.1)[9]	-	MD 8.7 lower (13.26 to 4.14 lower)	VERY LOW	CRITICAL
Beck Depression Inventory change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-5.5 (5.3)[14]	-1.6 (6)[9]	-	MD 3.9 lower (8.7 lower to 0.9 higher)	LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 146: Clinical evidence profile: Supervised resistance + balance versus home resistance + balance

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance mean (SD) [n]	home resistance + balance mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
SF-36 - Physical functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	21.2 (14.4)[14]	12.1 (6)[10]	-	MD 9.1 higher (0.69 to 17.51 higher)	VERY LOW	CRITICAL
SF-36 - Role-physical functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	34 (30.1)[14]	-5 (20.9)[10]	-	MD 39 higher (18.59 to 59.41 higher)	LOW	CRITICAL
SF-36 - Bodily pain change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	8.8 (5.8)[14]	14 (2)[10]	-	MD 6.8 higher (3.49 to 10.11 higher)	LOW	CRITICAL
SF-36 - General health change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	4.3 (8,4)[14]	2.4 (11.4)[10]	-	MD 1.9 higher (6.48 lower to 10.28 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance mean (SD) [n]	home resistance + balance mean (SD) [n]	Relative (95% CI)	Absolute		
SF-36 - Vitality change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	9 (19.3)[14]	12 (22.5)[10]	-	MD 3 lower (20.22 lower to 14.22 higher)	VERY LOW	CRITICAL
SF-36 - Social functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cait 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	3.4 (23.1)[14]	10 (13.6)[10]	-	MD 6.6 lower (21.35 lower to 8.15 higher)	VERY LOW	CRITICAL
SF-36 - Role-emotional functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	24.2 (49.6)[14]	-6.7 (27.8)[10]	-	MD 30.9 higher (0.28 lower to 62.08 higher)	VERY LOW	CRITICAL
SF-36 - Mental health change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	7.2 (13.4)[14]	3 (6.7)[10]	-	MD 4.2 higher (3.96 lower to 12.36 higher)	VERY LOW	CRITICAL
10 m walking test (s) change score s(follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-1.9 (1.2)[14]	-0.08 (0.7)[10]	-	MD 1.82 lower (2.58 to 1.06 lower)	LOW	CRITICAL

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance mean (SD) [n]	home resistance + balance mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
Duration of exercise change score (mins) (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	8.4(3.8)[14]	1.8(0.5) [10]	-	MD 6.6 higher (4.59 to 8.61 higher)	LOW	CRITICAL
Tolerated maxi wkload on bicycle change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	123.6(18)[14]	36 (8.2)[10]	-	MD 87.6 higher (76.89 to 98.31 higher)	LOW	CRITICAL
Timed up and go test (s) change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-1.3 (1.2)[14]	0.2 (0.5)[10]	-	MD 1.5 lower (2.2 to 0.8 lower)	LOW	CRITICAL
Dynamic Gait Index change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	2.7 (0.5)[14]	0.2 (0.4) [10]	-	MD 2.5 higher (2.14 to 2.86 higher)	LOW	CRITICAL
Functional reach change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	7.3 (2.4)[14]	0.2 (1.8)[10]	-	MD 7.1 higher (5.42 to 8.78 higher)	LOW	CRITICAL
Fatigue Severity Scale change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit	randomi	very	no serious	no serious	no serious	none	9.5(2.8)[14]	-0.4	-	MD 9.1 lower		CRITICA

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised resistance + balance mean (SD) [n]	home resistance + balance mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
2010	sed trials	serious[a]	inconsistency	indirectness	imprecision			(2.1)[10]		(11.06 to 7.14 lower)	LOW	L
Falls Efficacy Scale change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-11.3 (7.8)[14]	-2.1(1.3) [10]	-	MD 9.2 lower (13.36 to 5.04 lower)	LOW	CRITICAL
Beck Depression Inventory change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-5.5(5.3)[14]	1.6 (3.6)[10]	-	MD 7.1 lower (10.66 to 3.54 lower)	LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 147: Clinical evidence profile: Home based resistance and balance versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home based resistance and balance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
SF-36 - Physical functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	12.1 (6)[10]	7.7 (7.4)[9]	-	MD 4.4 higher (1.7 lower to 10.5 higher)	VERY LOW	CRITICAL
SF-36 - Role-physical functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	-5 (20.9)[10]	5 (44.7)[9]	-	MD 10 lower (41.95 lower to 21.95 higher)	VERY LOW	CRITICAL
SF-36 - Bodily pain change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	2 (2.1)[10]	2 (2.1)[9]	-	MD 0 higher (1.89 lower to 1.89 higher)	VERY LOW	CRITICAL
SF-36 - General health change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	2.4 (11.5)[10]	3.2(11.7)[9]	-	MD 0.8 lower (11.25 lower to 9.65 higher)	VERY LOW	CRITICAL
SF-36 - Vitality change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	12 (22.5)[10]	11 (20.4)[9]	-	MD 1 higher (18.29 lower to 20.29 higher)	VERY LOW	CRITICAL
SF-36 - Social functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	10 (13.6)[10]	5 (16.7)[9]	-	MD 5 higher (8.79 lower to 18.79 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home based resistance and balance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
SF-36 - Role-emotional functioning change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-6.7 (27.8)[10]	19.9 (50.5)[9]	-	MD 26.6 lower (63.82 lower to 10.62 higher)	VERY LOW	CRITICAL
SF-36 - Mental health change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	3 (6.7)[10]	7(6.7)[9]	-	MD 4 lower (10.03 lower to 2.03 higher)	VERY LOW	CRITICAL
10 m walking test (s) change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	-0.08(0.7)[10]	0.1 (0.8)[9]	-	MD 0.18 lower (0.86 lower to 0.5 higher)	VERY LOW	CRITICAL
Duration of exercise (mins) change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	1.8 (0.5)[10]	3.3 (5.3)[9]	-	MD 1.50 lower (4.98 lower to 1.98 higher)	VERY LOW	CRITICAL
Tolerated maxi wkload on bicycle change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	36(8.2)[10]	22 (13.03)[9]	-	MD 14 higher (4.09 to 23.91 higher)	VERY LOW	CRITICAL
Timed up and go test (s) change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.2 (0.5)[10]	-0.2 (0.8)[9]	-	MD 0.4 higher (0.21 lower to 1.01 higher)	VERY LOW	CRITICAL
Dynamic Gait Index change score (follow-up 8 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home based resistance and balance mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.2 (0.4)[10]	0.4 (0.4)[9]	-	MD 0.2 lower (0.56 lower to 0.16 higher)	VERY LOW	CRITICAL
Functional reach change score (follow-up 8 weeks; Better indicated by higher values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.2 (1.8)[10]	-1 (2.04)[9]	-	MD 1.2 higher (0.54 lower to 2.94 higher)	VERY LOW	CRITICAL
Fatigue Severity Scale change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-0.4 (2.1)[10]	-5.2 (5.3)[9]	-	MD 4.8 higher (1.1 to 8.5 higher)	VERY LOW	CRITICAL
Falls Efficacy Scale change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	-2.1 (1.3)[10]	-2.6 ~(3.1)[9]	-	MD 0.5 higher (1.68 lower to 2.68 higher)	VERY LOW	CRITICAL
Beck Depression Inventory change score (follow-up 8 weeks; Better indicated by lower values)												
Cakit 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	1.6 (3.6)[10]	-1.6 (6)[9]	-	MD 3.2 higher (1.31 lower to 7.71 higher)	VERY LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).

[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 148: Clinical evidence profile: Home resistance + pamphlets versus control

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home resistance + pamphlets mean (SD)[n]	Control mean (SD)[n]	Relative (95% CI)	Absolute	Quality	Importance
SF-12 (physical) (follow-up 24 weeks; Better indicated by higher values)												
Plow 2014	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	45.27 (9.47)[14]	41.86 (11.53)[16]	-	MD 3.41 higher (4.11 lower to 10.93 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Impact Scale (follow-up 24 weeks; Better indicated by lower values)												
Plow 2014	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	52.29 (23.51)[14]	65.38 (28.02)[16]	-	MD 13.09 lower (31.53 lower to 5.35 higher)	VERY LOW	CRITICAL
6 minute walk test m (follow-up 24 weeks; Better indicated by higher values)												
Plow 2014	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	406.42 (99.79)[14]	333.09 (115.77)[16]	-	MD 73.33 higher (3.81 lower to 150.47 higher)	VERY LOW	CRITICAL
Up and Go s (follow-up 24 weeks; Better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home resistance + pamphlets mean (SD)[n]	Control mean (SD)[n]	Relative (95% CI)	Absolute		
Plow 2014	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	8.07 (2.65)[14]	10.5 (5.23)[16]	-	MD 2.43 lower (5.34 lower to 0.48 higher)	VERY LOW	CRITICAL

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).

[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 149: Clinical evidence profile: Hospital stretching + aerobic versus home stretching + aerobic

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Hospital stretching + aerobic mean (SD) [n]	home stretching + aerobic mean (SD) [n]	Relative (95% CI)	Absolute		
SF-36 physical change score (follow-up 15 weeks; Better indicated by higher values)												
Solari	randomis	seriou	no serious	no serious	serious[b]	none	3.24 (6.49)[27]	0.26	-	MD 2.98 higher		CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Hospital stretching + aerobic mean (SD) [n]	home stretching + aerobic mean (SD) [n]	Relative (95% CI)	Absolute		
1995	ed trials	s[a]	inconsistency	indirectness				(7.9)[23]		(1.07 lower to 7.03 higher)	LOW	L
SF-36 mental change score (follow-up 15 weeks; Better indicated by higher values)												
Solari 1995	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	2.08 (9.7)[27]	-1.81 (7.75)[23]	-	MD 3.89 higher (0.95 lower to 8.73 higher)	LOW	CRITICAL
Mobility change score (follow-up 15 weeks; Better indicated by higher values)												
Solari 1995	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0 (0.87)[27]	-0.54 (1.22)[23]	-	MD 0.54 higher (0.06 lower to 1.14 higher)	LOW	CRITICAL
Self-care change score (follow-up 15 weeks; Better indicated by higher values)												
Solari 1995	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.32 (1.34)[27]	-1.18 (3.08)[23]	-	MD 1.5 higher (0.14 to 2.86 higher)	LOW	CRITICAL
Locomotion change score (follow-up 15 weeks; Better indicated by higher values)												
Solari 1995	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	0.28 (0.89)[27]	-0.41 (0.91)[23]	-	MD 0.69 higher (0.19 to 1.19 higher)	LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 150: Clinical evidence profile: Inpatient physiotherapy versus control

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Inpatient physiotherapy mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
Frenchay Activities Index (follow-up 9 weeks; Better indicated by higher values)												
Fuller 1986	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	27.8 (8.2)[23]	27.3 (6.5)[22]	-	MD 0.5 higher (3.81 lower to 4.81 higher)	VERY LOW	CRITICAL
Rivermead Mobility Index (follow-up 9 weeks; Better indicated by higher values)												
Fuller 1986	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	9.4 (3.2)[23]	11.1 (3.3)[22]	-	MD 1.7 lower (3.6 lower to 0.2 higher)	VERY LOW	CRITICAL
Barthel ADL (follow-up 9 weeks; Better indicated by lower values)												
Fuller 1986	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	17.4 (2.9)[23]	18.2 (1.7)[22]	-	MD 0.8 lower (2.18 lower to 0.58 higher)	VERY LOW	CRITICAL
Nottingham Extended ADL Index (follow-up 9 weeks; Better indicated by higher values)												
Fuller 1986	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	34.7 (8.7)[23]	36.7 (7.3)[22]	-	MD 2 lower (6.68 lower to 2.68 higher)	VERY LOW	CRITICAL
Nottingham ADL mobility (follow-up 9 weeks; Better indicated by higher values)												
Fuller 1986	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	8.2 (3.5)[23]	8.7 (2.8)[22	-	MD 0.5 lower (2.35 lower to	VERY	CRITICAL

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Inpatient physiotherapy mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
								]		1.35 higher)	LOW	
Nottingham ADL housework (follow-up 9 weeks; Better indicated by higher values)												
Fuller 1986	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	6.4 (3.5)[23]	8.7 (2.8)[22]	-	MD 0.3 lower (2.35 lower to 1.75 higher)	VERY LOW	CRITICAL
Five-metre walk or transfer (s)(follow-up 9 weeks; Better indicated by lower values)												
Fuller 1986	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	11.43 (5.59)[23]	11.01 (8.21)[22]	-	MD 0.42 higher (3.7 lower to 4.54 higher)	VERY LOW	CRITICAL
Quality of life												
No evidence for this outcome												
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).

[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 151: Clinical evidence profile: Outpatient physiotherapy versus control

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Outpatient physiotherapy mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
Rivermead mobility index (follow-up 8 weeks; Better indicated by higher values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.5 (3.5)[40]	9.1 (3.9)[40]	-	MD 1.4 higher (0.22 lower to 3.02 higher)	VERY LOW	CRITICAL
Assessor global mobility change score (post treatment) (follow-up 8 weeks; Better indicated by higher values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	62 (17)[40]	42 (11)[40]	-	MD 20 higher (13.73 to 26.27 higher)	LOW	CRITICAL
Assessor global mobility change score (follow-up) (follow-up 13 weeks; Better indicated by higher values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	44 (11)[40]	46 (11)[40]	-	MD 2 lower (6.82 lower to 2.82 higher)	VERY LOW	CRITICAL
HADS-anxiety (follow-up 8 weeks; Better indicated by lower values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	6.4 (4.4)[40]	8 (5.3)[40]	-	MD 1.6 lower (3.73 lower to 0.53 higher)	VERY LOW	CRITICAL
HADS-depression (follow-up 8 weeks; Better indicated by lower values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	6.5 (3.9)[40]	7.6 (4.7)[40]	-	MD 1.1 lower (2.99 lower to 0.79 higher)	VERY LOW	CRITICAL
Quality of life												
No evidence for this outcome												
Carer perceptions												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Outpatient physiotherapy mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 152: Clinical evidence profile: Home physiotherapy versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home physiotherapy mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Rivermead mobility index (follow-up 8 weeks; Better indicated by higher values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10.6 (2.9)[40]	9.1 (3.9)[40]	-	MD 1.5 higher (0.01 lower to 3.01 higher)	VERY LOW	CRITICAL
Assessor global mobility change score (post treatment) (follow-up 8 weeks; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Home physiotherapy mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	65 (17)[40]	42 (11)[40]	-	MD 23 higher (16.73 to 29.27 higher)	LOW	CRITICAL
Assessor global mobility change score (follow-up) (follow-up 13 weeks; Better indicated by higher values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious2	none	44 (14)[40]	42 (11)[40]	-	MD 2 lower (7.52 lower to 3.52 higher)	VERY LOW	CRITICAL
HADS-anxiety (follow-up 8 weeks; Better indicated by lower values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	6.6 (4.5)[40]	8 (5.3)[40]	-	MD 1.4 lower (3.55 lower to 0.75 higher)	VERY LOW	CRITICAL
HADS-depression (follow-up 8 weeks; Better indicated by lower values)												
Wiles 2001	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	5.9 (3.9)[40]	7.6 (4.7)[40]	-	MD 1.7 lower (3.59 lower to 0.19 higher)	VERY LOW	CRITICAL
Quality of life												
No evidence for this outcome												
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous* variables.

Table 153: Clinical evidence profile: Task orientated vs Facilitation

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Task orientated mean (SD) [n]	Facilitation mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
Walking time change score (s) (follow-up 6-8 weeks; Better indicated by lower values)												
Lord 1998	randomised trials	Very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	-9.3 (14.7)[10]	-6 (4.7[)10]	-	MD 3.3 lower (12.87 lower to 6.27 higher)	VERY LOW	CRITICAL
Stride length change score (cm) (follow-up 6-8 weeks; Better indicated by higher values)												
Lord 1998	randomised trials	Very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	17.6 (18.6)[10]	15.7 (19.4)[10]	-	MD 1.9 higher (14.76 lower to 18.56 higher)	VERY LOW	CRITICAL
Global Gait Score change score (follow-up 6-8 weeks; Better indicated by lower values)												
Lord 1998	randomised trials	Very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	6.5 (9.1)[10]	4.7 (3.8)[10]	-	MD 1.8 higher (4.31 lower to 7.91 higher)	VERY LOW	CRITICAL
Berg Balance Test change score (follow-up 6-8 weeks; Better indicated by higher values)												
Lord 1998	randomised trials	Very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	8.5 (7.6)[10]	7.2 (5.5)[10]	-	MD 1.3 higher (4.51 lower to 7.11 higher)	VERY LOW	CRITICAL
Rivermead Mobility Index change score (follow-up 6-8 weeks; Better indicated by higher values)												
Lord 1998	randomised trials	very seriou	no serious inconsistency	no serious indirectness	very serious[b]	none	1.2 (1.5)[10]	0.8 (0.7)[10]	-	MD 0.4 higher (0.63 lower to 1.43 higher)	VERY	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Task orientated mean (SD) [n]	Facilitation mean (SD) [n]	Relative (95% CI)	Absolute		
		s[a]									LOW	
Quality of life												
No evidence for this outcome												
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 154: Clinical evidence profile: Balance versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Balance versus control mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
MSIS-29 % change from baseline (follow-up 12 weeks; Better indicated by lower values)												
Prosperini 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-12 (27)[31]	2 (15)[30]	-	MD 14 lower (24.92 to 3.08 lower)	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Balance versus control mean (SD) [n]**	**Control mean (SD) [n]**	**Relative (95% CI)**	**Absolute**		
25-foot walking test % change from baseline (follow-up 12 weeks; Better indicated by lower values)												
Prosperini 2013	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-8 (18)[31]	-2 (14)[30]	-	MD 6 lower (14.08 lower to 2.08 higher)	LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 155: Clinical evidence profile: Whole body vibration versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Whole body vibration mean (SD) [n]**	**Control mean (SD) [n]**	**Relative (95% CI)**	**Absolute**		

Quality assessment							No of patients		Effect			
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Whole body vibration mean (SD) [n]**	**Control mean (SD) [n]**	**Relative (95% CI)**	**Absolute**	**Quality**	**Importance**
3 min walk test m change from baseline - Light (follow-up 3 weeks; Better indicated by higher values)												
Claerbout 2011	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	37.4 (34.3)[16]	20.4 (27.95[17]	-	MD 17 higher (4.42 lower to 38.42 higher)	VERY LOW	CRITICAL
3 min walk test (m) change from baseline - Full (follow-up 3 weeks; Better indicated by higher values)												
Claerbout 2011	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	45 (42.6)[14]	20.4 (27.95)[17]	-	MD 24.6 higher (1.37 lower to 50.57 higher)	VERY LOW	CRITICAL
Timed up and go test (s) change from baseline - Light (follow-up 3 weeks; Better indicated by higher values)												
Claerbout 2011	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-3.2 (4.7)[16]	0.8 (2.3)[17]	-	MD 4 lower (6.55 to 1.45 lower)	VERY LOW	CRITICAL
Timed up and go test (s) change from baseline - Full (follow-up 3 weeks; Better indicated by higher values)												
Claerbout 2011	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-0.8 (2.3)[14]	0.8 (2.3)[17]	-	MD 1.6 lower (3.23 lower to 0.03 higher)	VERY LOW	CRITICAL
Quality of life												
No evidence for this outcome												
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 156: Clinical evidence profile: Yoga versus control

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Yoga mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute	Quality	Importance
MSQoL physical change from baseline (follow-up 8 weeks; Better indicated by higher values)												
Ahmadi 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	6.75 (8.1)[11]	-0.6 (6.9)[10]	-	MD 7.35 higher (0.93 to 13.77 higher)	VERY LOW	CRITICAL
MSQoL mental change from baseline (follow-up 8 weeks; Better indicated by higher values)												
Ahmadi 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	18.18 (3.14)[11]	5.04 (41.52)[10]	-	MD 13.14 higher (12.66 lower to 38.94 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Impact Scale-29 v2 (follow-up 12 weeks; Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-4 (13.90)[63]	0.3 (14.97)[49]	-	MD 4.3 lower (9.72 lower to 1.12 higher)	VERY LOW	CRITICAL
2 min timed walk distance (m) % change from baseline (follow-up 8 weeks; Better indicated by higher values)												
Ahmadi 2010	randomised trials	Very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	9.96 (7.2)[11]	-2.89 (5.81)[10]	-	MD 12.85 higher (7.28 to 18.42 higher)	LOW	CRITICAL
10 m timed walk (m) % change from baseline (follow-up 8 weeks; Better indicated by lower values)												
Ahmadi 2010	randomised trials	Very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-7.4 (14.9)[11]	3.38 (6.6)[10]	-	MD 10.78 lower (20.49 to 1.07 lower)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Yoga mean (SD) [n]	Control mean (SD) [n]	Relative (95% CI)	Absolute		
Multiple Sclerosis Fatigue Impact Scale (total) (follow-up 12 weeks; Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-5.8 (23.02)[63]	-1.1 (11.83)[49]	-	MD 4.7 lower (11.28 lower to 1.88 higher)	VERY LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (physical) (follow-up 12 weeks; Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	-0.96 (3.57)[63]	-0.51 (4.18)[49]	-	MD 2.5 lower (4.55 to 0.45 lower)	VERY LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (cognitive) (follow-up 12 weeks; Better indicated by lower values)												
Garrett 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-0.96 (3.57)[63]	-0.51 (4.18)[49]	-	MD 0.45 lower (1.92 lower to 1.02 higher)	LOW	CRITICAL
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 157: Clinical evidence profile: Vestibular rehabilitation versus control

Quality assessment	No of patients	Effect	Quality	Importa

												nce
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Vestibular rehabilitation mean (SD) [n]**	**Control mean (SD) [n]**	**Relative (95% CI)**	**Absolute**		
6 Minute Walk Test feet (change from baseline to 6 wks) (Better indicated by higher values)												
Hebert 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	85.1 (159.5)[12]	22.4 (88.1)[13]	-	MD 62.7 higher (81.1 lower to 206.5 higher)	VERY LOW	CRITICAL
6 Minute Walk Test feet (10 wks) (Better indicated by higher values)												
Hebert 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	1,396.1 (330.5)[12]	1,100.5 (284)[13]	-	MD 295.6 higher (53.11 to 538.09 higher)	Low	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (6 wks) (Better indicated by lower values)												
Hebert 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	29.5 (15.8)[12]	52.1 (17.1)[13]	-	MD 22.6 lower (35.5 to 9.7 lower)	MODERATE	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (10 wks) (Better indicated by lower values)												
Hebert 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	30.3 (20.8)[12]	52.6 (17.4)[13]	-	MD 22.3 lower (37.4 to 7.2 lower)	MODERATE	CRITICAL
Quality of life												
No evidence for this outcome												
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).

[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 158: Clinical evidence profile: Vestibular rehabilitation versus aerobic

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Vestibular rehabilitation mean (SD) [n]**	**Aerobic mean (SD) [n]**	**Relative (95% CI)**	**Absolute**		
6 Minute Walking Test feet (6 wks) (Better indicated by higher values)												
Hebert 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	1,420.7 (283.6~)[12]	1,112.1 (391.3)[13]	-	MD 308.6 higher (42.16 to 575.04 higher)	LOW	CRITICAL
6 Minute Walking Test feet (10 wks) (Better indicated by higher values)												
Hebert 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	1,396.1 (330.5)[12]	1,053.9 (448.7)[13]	-	MD 342.2 higher (34.86 to 649.54 higher)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (6 wks) (Better indicated by lower values)												
Hebert 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	29.5 (15.8)[12]	44.3 (16.4)[13]	-	MD 14.8 lower (27.43 to 2.17 lower)	LOW	CRITICAL
Multiple Sclerosis Fatigue Impact Scale (10 wks) (Better indicated by lower values)												
Hebert 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	30.3 (20.8)[12]	44.7 (16.3)[13]	-	MD 14.4 lower (29.13 lower to 0.33 higher)	LOW	CRITICAL
Quality of life												
No evidence for this outcome												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vestibular rehabilitation mean (SD) [n]	Aerobic mean (SD) [n]	Relative (95% CI)	Absolute		
Carer perceptions												
No evidence for this outcome												
Adverse events												
No evidence for this outcome												

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Narrative review for outcomes not appropriate for meta-analysis

One study[89] comparing aerobic versus control reported the results in Table 159 below.

Table 159: Aerobic versus control

	No exercise N=5	Aerobic N=6	Resistance N=4
Walking speed m/s % change 10 m 50 m	 -7.4 -7.1	 +6.5 +2.5	 +9.6 +0.9
Pulse recovery time (secs)	-3.0	+1.9	+16.4
Chair transfer % change Time (seconds) No. of contacts	 -2.3 +30	 +4.8 +30.8	 +23.1 +42.0

P values for between group differences not reported

One study[192] comparing aerobic training versus neurorehabilitation reported the results in Table 160 below.

Table 160: Aerobic training versus neurorehabilitation

	Aerobic training N=11	Neurological rehab N=11
MFIS physical median range	14 (4-23)	13 (3-26)
MFIS cognitive median range	8 (0-36)	10 (0-40)
MFIS psychosocial median range	3 (0-7)	2 (0-6)
MSQOL-54 Overall quality of life median range	28 (10-82)	736 (20-82)
MSQOL-54 physical median range	59 (44-81)	57 (41-81)
MSQOL-54 mental health median range	66 (24-90)	66 (32-87)

MFIS Multiple Sclerosis Fatigue Impact Scale MSQOL Multiple Sclerosis Quality of Life

One study[92] comparing aerobic + resistance versus aerobic reported the results in Table 161 below

Table 161: Aerobic + resistance versus aerobic

	Resistance + aerobic (N=10)	Aerobic (N=9)
Timed Up and Go s	15.49	15.34
TMWSS 10-min walk self-selected pace m/s	0.87	0.87
TMWMP m/s 10-min walk maximum pace	1.05	1.19
6-Minute Walk Test m	409	280
BBS Berg Balance Scale /56 max	47	47
FSS Fatigue Severity Scale /10 max	5.1	4.5
Strength SUM	293.55	278.97

One study[23] comparing aerobic + resistance versus control reported the results in Table 162 below.

Table 162: Aerobic + resistance versus control

Outcome	Exercise (n=6)	Control (n=10)	MD(95% CIs)
EDSS change from baseline to 5 weeks	-0.2	-0.1	-0.07(-0.74 to 0.61)
SF36 physical function change from baseline to 5 weeks	-1.7	-0.5	-1.2(-16.1 to 13.8)
SF36 role physical change from baseline to 5 weeks	4.2	7.5	-3.3(-49.5 to 42.84)
SF36 bodily pain change from baseline to 5 weeks	15.5	-5.1	20.6(-8 to 49.2)
SF36 general health change from baseline to 5 weeks	5.8	-4	9.8(-5.7 to 25.4)
SF36 vitality change from baseline to 5 weeks	11.7	-7	18.7(0.08 to 37.25)
SF36 social function change from baseline to 5 weeks	20.3	2.8	17.5(-15.5 to 50.58)
SF36 role emotion change from baseline to 5 weeks	22.3	-0.1	22.43(-34.7 to 79.5)
SF36 mental health change from baseline to 5 weeks	-0.7	4.8	-5.47(-27.7 to 16.8)

P values not reported

One study[24] comparing motivational interviewing versus control reported the results in Table 163 below.

Table 163: Motivational interviewing versus control

Outcome	Motivational interviewing N=70	Control N=60	P
Health Promotion Lifestyle Profile HPLP total	0.2 (0.0 to 0.3)	0.0 (-0.2 to 0.2)	<.01
MS Fatigue Impact Scale	-1 (-9.5 to 0.5)	0 (-7 to 5)	0.02
SF-36 mental component	3.6 (0.3 to 8.0)	0.7 (-2.7 to 6.3)	0.02
SF-36 Physical component	-0.3 (-3.4 to 2.1)	1.0 (-2.8 to 5.1)	0.11
TMT-A s	0.0 (-6.0 to 2.0)	-2.0 (-8.5 to 0.5)	0.15
TMT-B s	-3.5 (-23.0 to 2.0)	-2.0 (-14.5 to 9.0)	0.14
Bicycle ergometer time s	0 (-45 to 23)	0 (-34 to 31)	0.62
Self-selected walking speed	-0.4 (-2.0 to 0.5)	0.0 (-1.7 to 1.0)	0.28

10.4.4 Economic evidence

Published literature

One economic evaluation was identified with two relevant comparisons and has been included in this review.[262] This study is summarised in the economic evidence profile below and the economic evidence table in Appendix H.

See also the economic article selection flow chart in Appendix E.

Table 164: Economic evidence profile: outpatient rehabilitation versus no therapy (comparison 1) and home rehabilitation versus no therapy (comparison 2)

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Wiles 2001[262] (UK) (a)	Partially applicable (b)	Very serious limitations (c)	Within trial analysis (RCT). Follow-up = 8 weeks.	1) £11 2) £25 (d)	Rivermead mobility index, MD 1) 1.4 2) 1.5 Balance time, MD 1) 4.82 2) 5.49 Walk A, MD 1) -14 2) -14 Nine hole peg test, MD 1) -18 2) -13 Assessor global mobility change score, MD 1) 19.8 2) 22.4 VAS patient mobility, MD 1) 25.2 2) 24.2	n/a	NR

(a) Study also includes comparison of outpatient rehabilitation versus home rehabilitation. This comparison was reviewed in the rehabilitation setting question and is available in the economic evidence table in Appendix H.
(b) Costs consequence analysis.
(c) Source of unit costs unclear. No sensitivity analysis conducted.
(d) Cost components considered: employment cost of physiotherapist and mileage.
Abbreviations: MD = mean difference; n/a = not applicable; NR = not reported; RCT = randomised control trial.

New cost-effectiveness analysis

One RCT identified in the clinical review (Cakit 2010)[34] which evaluated the effects on mobility and fatigue of resistance and balance training in different setting (supervised or home based training) compared to no intervention, reported SF-36 scores at baseline and after 8 weeks (end of intervention). Based on this study, the NCGC undertook a simple cost-utility analysis. This analysis is reported in full in section 10.3.4.

Unit costs

Relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

10.4.5 Evidence statements

10.4.5.1 Clinical

Aerobic versus control

Very low quality evidence from one RCT comprising sixteen participants showed that aerobic was clinically effective compared to control in terms of 2 min walk, with very serious imprecision

Very low quality evidence from one RCT comprising sixteen participants showed that aerobic was clinically effective compared to control in terms of 10m timed walk, with very serious imprecision

Very low quality evidence from one RCT comprising thirty participants showed that aerobic was clinically effective compared to control in terms of increase in walking distance from baseline , with very serious imprecision

Very low quality evidence from one RCT comprising thirty participants showed that aerobic was clinically effective compared to control in terms of increase in walking time from baseline , with very serious imprecision

Very low quality evidence from one RCT comprising fifteen participants showed that aerobic was clinically harmful compared to control in terms of proportion improvement in Multiple Sclerosis Impact Scale change from baseline, with very serious imprecision

Very low quality evidence from one RCT comprising twelve participants showed that aerobic was clinically harmful compared to control in terms of HAQUAMS from baseline, with very serious imprecision

Low to very low quality evidence from one or two RCTs (per outcome) comprising between sixteen to forty two participants showed there was no difference in clinical effectiveness between aerobic and control in terms of the outcomes below , with serious to very serious imprecision:

- 6 min walk test
- Fatigue Severity Scale
- Multiple Sclerosis Fatigue Impact Scale
- Proportion improvement in Multiple Sclerosis Impact Scale
- Guys Neurologic al Disability Scale
- Work activity
- Sport activity
- Leisure activity

Aerobic versus neurorehabilitation

Very low quality evidence from one RCT comprising twenty two participants showed that there was no difference in clinically effectiveness between aerobic and neurorehabilitation in terms of walking distance or walking speed, with very serious imprecision

Aerobic + resistance versus control

Very low quality evidence from one RCT comprising one hundred and twelve participants showed that aerobic + resistance was clinically effective compared to control in terms of Modified Fatigue Impact Scale (total score, physical, cognitive), with serious imprecision

Very low quality evidence from one RCT comprising twenty five participants showed that aerobic + resistance was clinically effective compared to control in terms of Fatigue Severity Scale, with serious imprecision

Very low quality evidence from one RCT comprising twenty five participants showed that aerobic + resistance was clinically effective compared to control in terms of Activities, balance, confidence, with serious imprecision

Low to very low quality evidence from one or two RCT s (per outcome) comprising between twenty five to two hundred and seven participants showed there was no difference in clinical effectiveness between aerobic + resistance and control in terms of the outcomes below , with no serious imprecision, serious or very serious imprecision:

- Multiple Sclerosis Impact Scale (physical component)
- Leeds MS quality of life
- MS Functional Composite
- MSQOL-54 (mental, physical)
- 6 min walk
- Timed walk test
- Timed 25 foot walk test
- Timed up and go
- Paced auditory serial additions
- Nine hole peg test
- Berg balance test
- Hospital anxiety and disability scale

Aerobic + resistance versus yoga

Very low quality evidence from one RCT comprising between seventy seven to one hundred and twenty six participants showed there was no difference in clinical effectiveness between aerobic + resistance and yoga in terms of the outcomes below , with no serious imprecision or serious or imprecision:

- Multiple Sclerosis Impact Scale (physical, psychological)
- 6 min walking test
- Multiple Sclerosis Impact Scale (total, cognitive, physical)

Resistance versus control

Low quality evidence from one RCT comprising seventy one participants showed that resistance was more clinically effective compared to control in terms of WHOQOL-BREF physical health (10 wks) change, with serious imprecision

Very low quality evidence from one RCT comprising seventy one participants showed that resistance was more clinically effective compared to control in terms of 10 m walking test, with serious imprecision

Very low quality evidence from one RCT comprising ninety nine participants showed that resistance was more clinically effective compared to control in terms of the Fatigue Severity Scale, with serious imprecision

Low quality evidence from one RCT comprising seventy one participants showed that resistance was more clinically effective compared to control in terms of the Multiple Sclerosis Fatigue Impact Scale (total), with serious imprecision

Low quality evidence from one RCT comprising seventy one participants showed that resistance was more clinically effective compared to control in terms of the Multiple Sclerosis Fatigue Impact Scale (physical), with serious imprecision

Low quality evidence from one RCT comprising seventy one participants showed that resistance was more clinically effective compared to control in terms of stiffness Multiple Sclerosis Impact Scale-88, with serious imprecision

Moderate to very low quality evidence from one RCT (per outcome) containing between thirty six and ninety nine participants showed there was no difference in clinical effectiveness between resistance and control in terms of the outcomes below , with no serious imprecision, serious or imprecision:

- MusiQoL
- WHOQOL (QoL, health, physical (22 wks)
- Fast walking speed
- 2 min walk distance
- Power W/Kg
- Power W
- Balance ML sway
- Balance velocity sway
- Balance AP
- Up and Go
- Multiple Sclerosis Fatigue Impact Scale (total (22 wks), physical (10 wks), cognitive, psychosocial)
- AEs (stiffness, spasm)

Supervised resistance + balance versus control

Low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of SF-36 physical functioning, with no serious imprecision

Very low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of SF-36 role-physical functioning, with serious imprecision

Low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of SF-36 bodily pain, with no serious imprecision

Very low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of 10 m walking test, with serious imprecision

Very low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of duration of exercise, with serious imprecision

Low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of tolerated maxi workload on bicycle, with no serious imprecision

Very low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of timed up and go test, with serious imprecision

Low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of dynamic gait index, with no serious imprecision

Low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of functional reach, with no serious imprecision

Very low quality evidence from one RCT containing twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of Fatigue Severity Scale, with serious imprecision

Very low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of Falls efficacy scale, with serious imprecision

Low quality evidence from one RCT comprising twenty three participants showed that supervised resistance + balance was more clinically effective than control in terms of Beck Depression Inventory, with no serious imprecision

Very low quality evidence from one RCT (per outcome) comprising twenty three participants showed there was no difference in clinical effectiveness between supervised resistance + balance and control in terms of the outcomes below , with very serious imprecision:

- SF-36 (general health)
- SF-36 (vitality)
- SF-36 (social functioning)
- SF-36 (role-emotional)
- SF-36 (mental health)

Supervised resistance + balance versus home resistance + balance

Very low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of SF-36 (physical functioning), with serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of SF-36 (role-physical), with no serious imprecision

Very low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of SF-36 (bodily pain), with serious imprecision

Very low quality evidence from one RCT containing twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of SF-36 (social functioning), with very serious imprecision

Very low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of SF-36 (role-emotional functioning), with serious imprecision

Very low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of SF-36 (mental health), with very serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of 10 m walking test, with no serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of duration of exercise, with no serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of tolerated maxi wkload on bicycle, with no serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of timed up and go, with no serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of dynamic gait index, with no serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of functional reach, with no serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of Fatigue Severity Scale, with no serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of falls efficacy scale, with no serious imprecision

Low quality evidence from one RCT comprising twenty four participants showed that supervised resistance + balance was more clinically effective than home resistance + balance in terms of Beck Depression Inventory, with no serious imprecision

Very low quality evidence from one RCT (per outcome) comprising twenty four participants showed there was no difference in clinical effectiveness between supervised resistance + balance and home resistance + balance in terms of the outcomes below , with very serious imprecision:

- SF-36 (bodily pain)

- SF-36 (vitality)

Home resistance + balance versus control

Very low quality evidence from one RCT comprising nineteen participants showed that home resistance + balance was more clinically effective than control in terms of SF-36 (physical functioning), with serious imprecision

Very low quality evidence from one RCT comprising nineteen participants showed that home resistance + balance was more clinically effective than control in terms of tolerated maxi wkload on bicycle, with serious imprecision

Very low quality evidence from one RCT comprising nineteen participants showed that home resistance + balance was more clinically effective than control in terms of functional reach, with serious imprecision

Very low quality evidence from one RCT comprising nineteen participants showed that home resistance + balance was more clinically effective than control in terms of Fatigue Severity Scale, with serious imprecision

Very low quality evidence from one RCT comprising nineteen participants showed that home resistance + balance was more clinically effective than control in terms of Beck Depression Inventory, with serious imprecision

Very low quality evidence from one RCT (per outcome) comprising nineteen participants showed there was no difference in clinical effectiveness between home resistance + balance and control in terms of the outcomes below, with serious or very serious imprecision:

- SF-36 (role-physical)
- SF-36 (bodily pain)
- SF-36 (general health)
- SF-36 (vitality)
- SF-36 (social functioning)
- SF-36 (role emotional)
- SF-36 (mental health)
- 10 min walking test
- Duration of exercise
- Timed up and go
- Dynamic gait index
- Falls efficacy scale

Home resistance + pamphlets versus control

Very low quality evidence from one RCT comprising 30 participants showed home resistance + pamphlets was more clinically effective than control in terms of 6 minute walk test, with very serious imprecision

Very low quality evidence from one RCT (per outcome) comprising thirty participants showed there was no difference in clinical effectiveness between home resistance + pamphlets versus control in terms of the outcomes listed below, with serious imprecision:

- SF-12 (physical)
- Multiple Sclerosis Impact Scale
- Up and Go

Hospital stretching + aerobic versus home stretching + aerobic

Low quality evidence from one RCT comprising fifty participants showed hospital stretching + aerobic was more clinically effective than home stretching + aerobic in terms of SF-36 (mental change), with serious imprecision

Low quality evidence from one RCT comprising fifty participants showed hospital stretching + aerobic was more clinically effective than home stretching + aerobic in terms of locomotion, with serious imprecision

Low quality evidence from one RCT (per outcome) comprising fifty participants showed there was no difference in clinical effectiveness between hospital stretching + aerobic versus home stretching + aerobic in terms of the outcomes below, with serious imprecision:

- SF-36 (physical)
- Mobility
- Self-care

Inpatient physiotherapy versus control

Very low quality evidence from one RCT comprising forty five participants showed inpatient physiotherapy was more clinically effective than control in terms of Rivermead Mobility Index, with serious imprecision

Very low quality evidence from one RCT (per outcome) comprising forty five participants showed there was no difference in clinical effectiveness between inpatients physiotherapy versus control in terms of the outcomes below, with serious or very serious imprecision:

- Frenchay Activities Index
- Barthel ADL
- Nottingham Extended ADL (mobility, housework)
- Five metre walk or transfer

Outpatient physiotherapy versus control

Low quality evidence from one RCT comprising eighty participants showed outpatient physiotherapy was more clinically effective than control in terms of Assessor global mobility change score (post treatment), with no serious imprecision

Very low quality evidence from one RCT (per outcome) comprising eighty participants showed there was no difference in clinical effectiveness between outpatient physiotherapy versus control in terms of the outcomes below, with serious imprecision:

- Rivermead Mobility Index
- Assessor global mobility (follow-up)
- HADS-anxiety
- HADS- depression

Home physiotherapy versus control

Low quality evidence from one RCT comprising eighty participants showed home physiotherapy was more clinically effective than control in terms of Assessor global mobility change score (post treatment) , with no serious imprecision

Very low quality evidence from one RCT (per outcome) comprising eighty participants showed there was no difference in clinical effectiveness between home physiotherapy versus control in terms of the outcomes below, with serious imprecision:

- Rivermead Mobility Index
- Assessor global mobility (follow-up)
- HADS-anxiety
- HADS- depression

Task orientated versus facilitation

Very low quality evidence from one RCT comprising twenty participants showed task orientated was more clinically effective than facilitation in terms of walking time change score, with very serious imprecision

Very low quality evidence from one RCT (per outcome) comprising twenty participants showed there was no difference in clinical effectiveness between task orientated versus facilitation in terms of the outcomes below , with very serious imprecision:

- Stride length change score
- Global Gait Score change score
- Berg Balance Test change score
- Rivermead Mobility Index change score

Balance versus control

Low quality evidence from one RCT (per outcome) comprising sixty one participants showed there was no difference in clinical effectiveness balance versus control in terms of the outcomes below, with no serious imprecision:

- MSIS-29 % change from baseline
- 25-foot walking test % change from baseline

Whole body vibration versus control

Very low quality evidence from one RCT comprising thirty three participants showed whole body vibration (light) was more clinically effective than control in terms of 3 min walking test change score, with serious imprecision

Very low quality evidence from one RCT comprising thirty one participants showed whole body vibration (full) was more clinically effective than control in terms of 3 min walk test change score, with serious imprecision

Very low quality evidence from one RCT comprising thirty three participants showed whole body vibration (light) was more clinically effective than control in terms of timed up and go test change score, with serious imprecision

Very low quality evidence from one RCT comprising thirty one participants showed whole body vibration (full) was more clinically effective than control in terms of timed up and go test change score, with serious imprecision

Yoga versus control

Very low quality evidence from one RCT comprising twenty one participants showed yoga was more clinically effective than control in terms of MSQoL physical change from baseline, with serious imprecision

Low quality evidence from one RCT comprising twenty one participants showed yoga was more clinically effective than control in terms of 2 min timed walk distance % change from baseline, with no serious imprecision

Very low quality evidence from one RCT comprising one hundred and twelve participants showed yoga was more clinically effective than control in terms of Multiple Sclerosis Fatigue Impact Scale (physical), with serious imprecision

Very low quality evidence from one RCT (per outcome) comprising twenty one or one hundred and twelve participants showed there was no difference in clinical effectiveness yoga versus control in terms of the outcomes below, no serious imprecision or serious imprecision:

- MSQoL mental change from baseline
- Multiple Sclerosis Impact Scale-29 v2
- 10 m timed walk % change from baseline
- Multiple Sclerosis Impact Scale (total)
- Multiple Sclerosis Impact Scale (cognitive)

Vestibular rehabilitation versus control

Low quality evidence from one RCT comprising twenty five participants showed vestibular rehabilitation was more clinically effective than control in terms of 6 minute walk test (6, 10 wks), with serious imprecision

Moderate quality evidence from one RCT comprising twenty five participants showed vestibular rehabilitation was more clinically effective than control in terms of Multiple Sclerosis Fatigue Impact Scale (6, 10 wks), with no serious imprecision

Vestibular rehabilitation versus aerobic

Very low quality evidence from one RCT comprising twenty five participants showed vestibular rehabilitation was more clinically effective than aerobic in terms of 6 minute walk test (baseline to 6 wks), with very serious imprecision

Low quality evidence from one RCT comprising twenty five participants showed vestibular rehabilitation was more clinically effective than aerobic in terms of Multiple Sclerosis Fatigue Impact Scale (6, 10 wks), with serious imprecision.

10.4.5.2 Economic

One cost-consequence analysis found that outpatient and home rehabilitation were more costly and effective than no therapy for treating mobility (£11 and £25 more per patient, 1.4 and 1.5 mean difference improvement in the primary outcome, the Rivermead mobility index, per patient, respectively). This analysis was assessed as partially applicable with very serious limitations.

One original cost-utility analysis found that in adults with MS, with a one year time horizon supervised resistance and balance training was more effective and the most cost-effective option (ICER: £7,619 per QALY) compared to control and home based resistance and balance training for treating fatigue and mobility. This analysis was assessed as directly applicable and with potential serious limitations.

10.4.6 Recommendations and link to evidence

Note: some programmes used to treat mobility are also used in treatment of fatigue. Evidence for treatment of fatigue and the GDG considerations are in section 10.3.6.

Recommendations	**59. Ensure people with MS and mobility problems have access to an assessment to establish individual goals and discuss ways in which to achieve them. This would usually involve rehabilitation specialists and physiotherapists with expertise in MS.** **Consider supervised exercise programmes involving moderate progressive resistance training and aerobic exercise to treat people with MS who have mobility problems and/or fatigue.** **Consider vestibular rehabilitation for people with MS who have fatigue or mobility problems associated with limited standing balance.** **Encourage people with MS to keep exercising after treatment programmes end for longer term benefits (see Behaviour change: individual approaches NICE public health guideline 49).** **Help the person with MS continue to exercise, for example by referring them to exercise referral schemes.** **If more than one of the interventions recommended for mobility or fatigue are suitable, offer treatment based on which the person prefers and whether they can continue the activity when the treatment programme ends.**
Relative values of different outcomes	For mobility, a wide range of outcomes had been measured in included studies. These included fatigue scales, quality of life, timed walking distances, and balance. Mobility outcomes included validated measures such as the timed walk test. Most studies examined a programme or course of therapy/treatment/activity. Where possible, the GDG valued long-term sustained improvements in outcomes after the course had ended, however very few studies reported outcomes in the post-intervention phase. A number of the studies did not report on quality of life but it was noted that one study showed clinically important benefits on the SF-36 for supervised resistance plus balance compared to home resistance plus balance or no treatment. Furthermore the GDG felt that there was evidence of clinical effectiveness from a number of comparisons that included an aerobic exercise component. No studies assessed changes in the ability ro perform activities of daily living as an outcome, but the GDG thought this would be a useful measure in future

	studies.
Trade off between clinical benefits and harms	The GDG agreed that unsupervised exercise programmes did carry a risk of injury and worsening of function. Other therapies had minimal known risks or these were not measured. Clinical benefit was considered to be present if there was improvement in scales of fatigue or mobility, or in overall functioning. The GDG did not prioritise different outcomes but listed all therapies with evidence of benefit in one or more relevant outcomes.
Economic considerations	A simple cost-utility analysis was undertaken by the NCGC based on the results of an RCT by Cakit (2010)[34,34] evaluating the effects of supervised and unsupervised progressive resistance and balance training compared to no intervention on mobility and fatigue. The cost of each intervention was estimated based on published unit costs and within trial resource use. Quality of life values were estimated by mapping SF-36 scores to EQ-5D values using an algorithm by Ara and Brazier (2008).[8,8] Two time horizons were considered, 8 weeks to reflect the duration of the intervention and one year which assumed that the effectiveness of the intervention was maintained after it is completed. With a one year time horizon, supervised training was the most cost effective option. With the 8 week time horizon neither supervised nor unsupervised training were cost-effective compared to control. The GDG agreed that supervised programmes were preferable to unsupervised ones. They also discussed the importance of selecting activities that can people can continue following the end of a supervised treatment programme. No economic evidence was identified for vestibular rehabilitation. The GDG considered that for people with fatigue or mobility problems associated with sensory deficits, such an intervention, which would be conducted by a physiotherapist or occupational therapist, is likely to be cost-effective. One cost-consequence analysis was presented which found that outpatient and home rehabilitation were more costly and effective than no therapy for treating mobility.
Quality of evidence	The evidence was almost all of very low to low quality. Furthermore, for most of the individual outcomes of a therapy, there were only one or two studies. The population was noted to be limited to relapsing remitting MS with an EDSS less than seven in most studies, and therefore may be less applicable to other patients with MS. The economic evidence for outpatient and home rehabilitation compared to no rehabilitation was assessed as partially applicable with very serious limitations. The economic evidence for supervised versus home based resistance and balance training versus control was assessed as directly applicable with potential serious limitations.
Other considerations	The GDG looked at the programme of therapy itself and not the type of staff or healthcare professionals used. It is assumed that any of our recommended therapies would be delivered by a person or persons competent in that field.

10.5 Non-pharmacological management of pain

10.5.1 Introduction

NICE have developed a clinical guideline on the pharmacological management of neuropathic pain and this has included people with MS. The guideline scope included non-pharmacological management of pain in people with MS. People with MS should have access to pain management expertise and this review examined MS specific studies only.

10.5.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological programmes (including self-management programmes) for pain?

For full details see review protocol in Appendix C.

Table 165: PICO characteristics of review question

Population	Adults with MS only
Intervention/s	Any non-pharmacological management programme, including self-management programmes , for example: • Multidisciplinary rehabilitation/programmes • Self-management programmes • Treatment programmes for various symptoms • FACETS prog, energy conservation programs, mindfulness (Grossman Paul), exercise (John Saxton), Getting To Grips (MS Society), stretching, standing, splinting, gym prescription, diet, yoga, tai chai, pilates, relaxation, lycra garments
• Comparison/s	• Usual treatment or placebo • Two active interventions compared to each other
Outcomes	• pain [symptoms or measures (ie FSS)] Also, any of the following outcomes, provided the treatment has been directed at **pain**: • Quality of life • Function (i.e. EDSS, ambulation measures, MSIS, Guys scale etc) • carer perceptions • Incidence of adverse events
	Systematic reviews, RCTs. Include cross-over studies.

10.5.3 Clinical evidence

Six RCTs were found [4,36,101,133,154,259].

The non-pharmacological treatments for pain used were:

- TENS
- Hydrotherapy
- Reflexology
- Progressive muscle relaxation
- Self-hypnosis

- Cognitive restructuring
- Combination of hypnosis and cognitive restructuring
- Anodal transcranial direct current stimulation

Table 166: Summary of studies included in the review

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised/ analysed	Analysis
Al-Smadi 2003[4]	TENS 1 (4 Hz, 200µs) 45 minutes 3 times a week for 6 weeks TENS 2 (110 Hz, 200µs) 45 minutes 3 times a week for 6 weeks Placebo TENS 45 minutes 3 times a week for 6 weeks	Age 34-65 years; stable low back pain (present at least 3 months and had not responded to conventional treatments)	15/15	VAS for current low back pain, right and left leg pain; Leeds Multiple Sclerosis Quality of Life Questionnaire; Roland Morris Disability Questionnaire; SF-36; McGill Pain Questionnaire; blinded assessment at baseline, week 6 (end of treatment) and week 10 (4 week follow up)
Castro-Sanchez 2012[36]	Ai-Chi exercise in swimming pool Abdominal breathing and contraction-relaxation exercises in therapy room	Mean age 46 (9.97) for Ai-Chi group and 50 (12.31) for controls; gender: 26 female/10 male for Ai-Chi group and 24/13 for controls. EDSS 6.3 (0.8) vs. 5.9 (0.9). Years since diagnosis: 10.7 (9.1) vs. 11.9 (8.7). Type of MS: primary progressive 6 vs. 9; secondary progressive: 9 vs. 12; not known 21 vs. 16. Mean pain VAS 8.3 (1.2) vs. 7.8 (1.6). All differences non-significant.	73/71	Pain, disability, spasm, depression, fatigue, autonomy at baseline, 20 weeks (end of treatment), 4 and 10 weeks follow up
Hughes 2009[101]	Precision reflexology Sham reflexology	Mean age 50 (11.1) precision reflexology and 53 (11.0) sham. Gender: 30 females/ 5 males vs. 29 females/ 7 males. EDSS 5.8 (0.95) vs. 6.2 (0.8). Years since diagnosis 12.9 (8.9) vs. 12.2 (8.4). Type of MS: benign 0 vs. 1; relapsing-remitting 16 vs. 12; primary-progressive 4	71/67 at week 10; 67 at week 16; 66 at week 22	Pain VAS; McGill Pain Questionnaire; Roland Morris Disability Questionnaire; spasticity VAS; Multiple Sclerosis Impact Scale (MSIS)-29; Modified Fatigue Impact Scale (MFIS) Physical; Modified Fatigue Impact Scale (MFIS) Cognitive; Modified Fatigue Impact Scale (MFIS) Psychological; Fatigue

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised/ analysed	Analysis
		vs. 4; secondary progressive 6 vs. 13; not known 9 vs. 6. Level of pain (baseline VAS) 7.5 (1.3) vs. 7.9 (1.5). All not significantly different.		Severity Scale; Beck Depression Inventory; Barthel Index; blinded assessment at baseline, week 10 (end of treatment) and weeks 16 and 22 (follow up)
Masoudi 2013 [133]	Progressive muscle relaxation training No treatment	Age 18/35 20-30 yrs, 17/35 31-40 yrs. 23/35 female in active treatment group and 20/35 yrs 20-30 yrs, 15/35 31-40 yrs, 22/35 female	70/70	Pain VAS
Mori 2010[154]	Anodal transcranial direct current stimulation Sham transcranial direct current stimulation	Mean age 44.8 (27.5) years; 11 females/8 males. Mean 42.8 years (5 females, 5 males) in active treatment group and 46.3 years (6 females, 3 males) in sham group.	19/19	Pain VAS, anxiety VAS, Short Form McGill Questionnaire, Multiple Sclerosis Quality of Life-54, Beck Depression Inventory at baseline, at end of 5-day treatment week, and at weeks 2, 3 and 4 (1, 2 and 3 week follow ups)
Warke 2004[259]	TENS 1 (4 Hz, 200µs) 45 minutes twice a day and at any time when a painful episode occurred for 6 weeks TENS 2 (110 Hz, 200µs) 45 minutes twice a day and at any time when a painful episode occurred for 6 weeks Placebo TENS 45 minutes twice a day and at any time when a painful episode occurred for 6 weeks	Age range 37 to 71 years. Baseline VAS pain scores: 58.4 (SEM 8.00) for TENS 1; 64.00 (SEM 10.18) for TENS 2 and 51.00 (SEM 6.04) for placebo TENS	15	VAS for current low back pain, McGill Pain Questionnaire, Barthel Index, Rivermead Index, Roland Morris Disability Questionnaire, Leeds Multiple Sclerosis Quality of Life Questionnaire at week 1 (pre-treatment), week 6 (post-treatment), week 10 (4-week follow up) and week 32 (6 month follow up)

Table 167: Clinical evidence profile: Progressive muscle relaxation training

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Progressive muscle relaxation mean (SD) [N]	Control mean (SD) [N]	Relative (95% CI)	Absolute		
Pain VAS (follow-up 3 months; Better indicated by lower values)												
Masoudi 2013	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	3.97 (1.72) [35]	8.14 (0.94) [35]	-	MD 4.17 lower (4.82 to 3.52 lower)	LOW	CRITICAL

[a] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

Table 168: Clinical evidence profile: Anodal direct current stimulation versus sham

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anodal transcranial direct current stimulation mean (SD)[N]	Sham mean (SD) [N]	Relative (95% CI)	Absolute		
Pain - Week 1 (end of treatment) (follow-up end of treatment; Better indicated by lower values)												
Mori	randomis	very	no serious	no serious	very	none	45.5 (34.78)[10]	89.3	-	MD 43.8 lower	VERY	CRITICA

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Anodal transcranial direct current stimulation mean (SD)[N]	Sham mean (SD) [N]	Relative (95% CI)	Absolute		
2010	ed trials	serious[a]	inconsistency	indirectness	serious[b]			(25.8)[9]		(71.16 to 16.44 lower)	LOW	L
Pain - Week 2 (1 week follow up) (follow-up post treatment + 1 weeks; Better indicated by lower values)												
Mori 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	40.3 (31.94)[10]	85.2 (18.9)[9]	-	MD 44.9 lower (68.23 to 21.57 lower)	VERY LOW	CRITICAL
Pain - Week 3 (2 weeks follow up) (follow-up post treatment + 2 weeks; Better indicated by lower values)												
Mori 2010	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	40.4 (31.31)[10]	84.7 (26.1)[9]	-	MD 44.3 lower (70.13 to 18.47 lower)	VERY LOW	CRITICAL

[a] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*

[b] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variable*

Narrative review for outcomes not appropriate for meta-analysis

The study by Al-Smadi 2003[4] was a pilot study into the use of TENS (group 1: 4Hz, 200µs or group 2: 110Hz, 200µs) versus placebo TENS; each was applied by a researcher for 45 minutes 3 times a week for 6 weeks. There were only 5 patients in each group (underpowered). Not all baseline data were shown, but of those that were shown, there were significant differences between the groups at baseline but final scores (SEM) not change scores (SEM or SD) were reported. Also, the use of two intervention groups would mean double counting the control groups in RevMan. Therefore it would be misleading to use these data in meta-analysis. The authors report no significant differences between the groups on any outcome measure (VAS, right/leg pain, Leeds Multiple Sclerosis Quality of Life questionnaire, Roland Morris Disability Questionnaire, McGill Pain Questionnaire, SF-36 physical and mental) [LOW quality for methodological limitations]

The study by Warke 2004[259] was similar to the study by Al-Smadi 2003 (the two authors were working in the same department, both authors on both papers). This was also pilot study into the use of TENS (group 1: 4Hz, 200µs or group 2: 110Hz, 200µs) versus placebo TENS, but using self-applied TENS for 45 minutes twice a day and at any time when a painful episode occurred (rather than being treated by a researcher 3 times a week as in the Al-Smadi study). There were only 5 patients in each group (underpowered). Not all baseline data were shown, but of those that were shown, there were significant differences between the groups at baseline. The use of two intervention groups would mean double counting the control groups in RevMan. Therefore it would be misleading to use these data in meta-analysis. The authors report no significant differences between the groups on any outcome measure or within groups over time (McGill Pain Questionnaire pain rating and affective sub-scale, VAS, Barthel Index, Rivermead Mobility Index, Roland Morris Disability Questionnaire, Leeds Multiple Sclerosis Quality of Life Questionnaire, SF-36 physical and mental) [LOW quality for methodological limitations].

The study by Castro-Sanchez 2012[36] compared Ai-Chi exercise in a swimming pool to breathing and contraction-relaxation exercises in a therapy room. Outcomes were presented as medians and standard deviations rather than as means and standard deviations making the data unsuitable for RevMan. The authors report that the experimental group showed a significant and clinically relevant decrease in pain intensity versus baseline, with a reduction in VAS of 50% that was maintained for up to 10 weeks. Significant improvements were also observed in spasm, fatigue, disability and autonomy, while few changes were observed in the control group. [MODERATE quality for methodological limitations]

The study by Hughes 2009[101] compared precision reflexology with sham reflexology (standardised foot massage avoiding points representative of common areas of pain associated with MS). Outcomes were presented as median (IQR) rather than means and standard deviations making the data unsuitable for RevMan. The authors reported that a significant and clinically important decrease in pain intensity was observed in both groups compared with baseline; median VAS scores were reduced by 50% following treatment and were maintained for up to 12 weeks. Significant decreases were also observed for fatigue, depression, disability, spasm and quality of life. Precision reflexology was not superior to sham but the authors suggest that the improvement in symptoms might be due to a placebo effect or stimulation of reflex points in the feet using the non-specific massage. [HIGH quality for methodological limitations]

The study by Mori 2010[154] compared Anodal transcranial direct current stimulation with sham transcranial direct current stimulation. VAS pain intensity data were presented and are shown in Forest plots; the other outcome measures were only shown graphically. On the Short Form McGill Questionnaire and the Multiple Sclerosis Quality of Life-54, the authors reported that scores were reduced in the active group compared with the control group after the first week and this effect persisted until the last evaluation. There were no effects of treatment on the Beck Depression Inventory or VAS for anxiety [LOW quality for methodological limitations].

10.5.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

Relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

10.5.5 Evidence statements

10.5.5.1 Clinical

TENS versus placebo

Low quality evidence from two RCTs comprising 15 participants each showed there was no statistically significant differences on the VAS, right/leg pain, Leeds Multiple Sclerosis Quality of Life questionnaire, Roland Morris Disability Questionnaire, McGill Pain Questionnaire, SF-36 physical and mental or McGill Pain Questionnaire pain rating and affective sub-scale, VAS, Barthel Index, Rivermead Mobility Index, Roland Morris Disability Questionnaire, Leeds Multiple Sclerosis Quality of Life Questionnaire, SF-36 physical and mental.

Ai-chai versus breathing and contraction-relaxation exercises

Moderate quality evidence from 1 RCT comprising 71 participants reported a statistically and clinically significant relevant decrease in pain intensity versus baseline, with a reduction in VAS of 50%. Statistically significant improvements were reported for spasm, fatigue, disability and autonomy.

Reflexology versus placebo

High quality evidence from one RCT comprising 66/67 participants reported a statistically and clinically significant important decreases in pain intensity and VAS in both reflexology and sham groups. Both groups reported statistically significant decreases in fatigue, depression, disability, spasm and quality of life.

Progressive muscle relaxation training versus control

Low quality evidence RCT comprising 70 participants showed that progressive muscle relaxation training was of clinical benefit compared to control for pain, with no serious imprecision

Anodal versus sham

Very low quality evidence from 1 RCT comprising nineteen participants showed that anodal stimulation was of clinical benefit compared to sham for pain (week 1, 2, 3) with very serious imprecision

10.5.5.2 Economic

No relevant economic evaluations were identified.

10.5.6 Recommendations and link to evidence

Recommendations	**60. Treat neuropathic pain in people with MS according to Neuropathic pain – pharmacological management (NICE clinical guideline 173) and refer to pain services if appropriate.** **61. Be aware that musculoskeletal pain is common in people with MS and is usually secondary to problems with mobility and posture. Assess musculoskeletal pain, offer treatment to the person and refer them as appropriate.**
Relative values of different outcomes	In reviewing the evidence for the non-pharmacological treatment and management of pain the GDG noted that the outcomes were mainly subjective but did not include validated quality of life measures in any of the studies. The importance of measuring outcomes beyond the treatment phase was considered important but was not reported in the majority of studies.
Trade off between clinical benefits and harms	There was a low probability of adverse events with the interventions reported in the studies. A reduction in pain is likely to improve a person's quality of life, health and well-being. Although evidence was not formally reviewed for the pharmacological management of pain in MS as this was outside of the scope of this guideline, it is recognised that this management has to be balanced against the potential adverse effects of pharmacological treatments.
Economic considerations	No relevant economic evaluations were identified. The unit costs of individual based physiotherapy interventions (£234–416 for one intervention episode), TENS devices (£34–191 each), group-based mindfulness interventions (£357 per user) and individual CBT interventions (£594-1,188 per user) were presented. Given the lack of clear clinical evidence and the considerable cost to the NHS, the GDG felt further research was required into the use of non-pharmacological interventions for pain in people with MS. No economic evidence was reviewed on the pharmacological management of pain (see trade off section above). There are costs associated with referring people to pain services and treating neuropathic pain. The GDG considered however that this was standard practice for people with or without MS and they wanted to reinforce the importance of addressing these needs.
Quality of evidence	The studies for non-pharmacological management were at high risk of bias mainly due to lack of allocation concealment and blinding. Five out of the six studies were likely to be underpowered. Four of the studies were not appropriate for reporting using GRADE and were reported narratively. These studies reported the statistical significant of the estimations of effect and judgements on it was not possible to make judgements on clinical importance. For this reason a recommendation was not made for reflexology even despite the high quality evidence. Also, the pharmacological management of pain was outside the scope of this guideline and the GDG were unable to compare the effectiveness of non-pharmacological interventions with pharmacological interventions.
Other considerations	The GDG considered overall that people with MS should have access to pain management expertise and that individual patients will benefit from individual review by healthcare professionals to ascertain the cause of pain and to try

	treatments. They did not wish to make any recommendations related to the non-pharmacological management of pain associated with MS but they were aware that there is existing NICE guidance on the pharmacological management of neuropathic pain which is relevant to people with MS. The GDG agreed that people with MS who experience pain should be referred to pain services if symptoms persisted after first line treatment. A consensus recommendation was therefore made to reinforce current good practice in relation to existing NICE guidance and to ensure referral to pain services as appropriate

10.6 Non-pharmacological management of spasticity

10.6.1 Introduction

Spasticity describes both stiffness and muscle spasms and is a common problem for people with MS. Chapter 9.1 examines pharmacological options for management of MS. This chapter reports evidence for non-pharmacological programmes for treatment of spasticity.

10.6.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of non-pharmacological programmes (including self-management programmes) for spasticity?

For full details see review protocol in Appendix C.

Table 169: PICO characteristics of review question

Population	Adults with MS only
Intervention/s	Any non-pharmacological management programme, including self-management programmes , for example: • Multidisciplinary rehabilitation/programmes • Self-management programmes • Treatment programmes for various symptoms • FACETS prog, energy conservation programs, mindfulness (Grossman Paul), exercise (John Saxton), Getting To Grips (MS Society), stretching, standing, splinting, gym prescription, diet, yoga, tai chi, pilates, relaxation, lycra garments
Comparison/s	Usual treatment or placebo
Outcomes	• spasticity [symptoms or measures (ie Ashworth scale)] Also, any of the following outcomes, provided the treatment has been directed at **spasticity**: • Quality of life • Function (i.e. EDSS, ambulation measures, MSIS, Guys scale etc) • carer perceptions • Incidence of adverse events
Review strategy	Systematic reviews, RCTs. Include cross-over studies.

10.6.3 Clinical evidence

Summary of included studies

13 RCTs[12,78,119,124,149,155,167,168,194,218,226,242,256] were found, covering 16 different comparisons, as shown in Table 170.

Table 170: Summary of studies included in the review

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised/analysed

Study	Intervention/comparison	Mean MS characteristics where available (group-specific data designated by intervention / comparator)	N randomised/analysed
Schyns 2009[218]	Vibration therapy with exercise versus exercise alone	Mean age was c47. Range of duration from diagnosis was 10 months to 23 years	16/12
Tarakci 2013[242]	Resistance training versus control	EDSS 2-6.5; FSS 39.3/39.9; mostly RR	110/99
Velikonja 2010[256]	Sports climbing versus yoga	RR, PP or SP; 26-50 years, EDSS <7; $EDSS_{pyr}$ >2	10/10
Nielsen 1996[167]	Repetitive magnetic stimulation (thoracic) versus placebo	MAS 19.8/14.4; age 44; duration of MS: 12/13	38/35
Nilsagard 2006[168,168]	Cooling garment versus control	EDSS 4; mean age 52	48/43
Richards 1997[194]	Pulsed electromagnetic field versus placebo	EDSS 5.13/4.98	30/30
Lappin 2003 [119]		72% had duration MS ≥ 4yrs; 57% moderately disabled or worse	145/117
Baker 2007[12]	Therapeutic standing versus home exercise plan	SDSS>6; stable symptoms for 3 months	6/6
Mori 2011[155]	Transcranial magnetic stimulation versus placebo	MAS 2.1/2.4; EDSS 2-6	30/30
Siev-Ner 2003[226]	Reflexology versus control	MAS 5.1/3.3; duration MS: 11.9/13.4 years	71/53
Miller 2007A[149]	TENS 8 hrs/day versus TENS 1 hr/day	Stable MS for 3 months; increased tone in at least 1 LL	37/32
Livesley 1992[124]	Electrical Neuromuscular stimulation	37/40 had MS. 2 with spinal injuries and 1 CVA. MS duration 10 years	40/39
Gervasoni 2014[78]	Massage versus massage/exercise versus exercise versus usual care	EDSS 3.8; time since diagnosis 87-149 months (range between groups)	48/48

Table 171: Clinical evidence profile: vibration and exercise versus exercise only

Quality assessment							proportion of patients with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vibration therapy with exercise	exercise alone	Relative (95% CI)	Absolute		
SPASTICITY												
Proportion with improvement from baseline to 4 weeks in modified Ashworth score in QUADRICEPS												
Schyns 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	9/18 (50%)	5/21 (23.8%)	RR 2.1 (0.86 to 5.13)	262 more per 1000 (from 33 fewer to 983 more)	VERY LOW	CRITICAL
Proportion with improvement from baseline to 4 weeks in modified Ashworth score in HAMS												
Schyns 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	2/18 (11.1%)	4/21 (19%)	RR 0.58 (0.12 to 2.82)	80 fewer per 1000 (from 168 fewer to 347 more)	VERY LOW	CRITICAL
Proportion with improvement from baseline to 4 weeks in modified Ashworth score in HIP ADDUCTORS												
Schyns 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	5/18 (27.8%)	5/21 (23.8%)	RR 1.17 (0.4 to 3.39)	40 more per 1000 (from 143 fewer to 569 more)	VERY LOW	CRITICAL
Proportion with improvement from baseline to 4 weeks in modified Ashworth score in GASTROCNEMIUS												

Quality assessment							proportion of patients with event (%)		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vibration therapy with exercise	exercise alone	Relative (95% CI)	Absolute		
Schyns 2009	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	1/18 (5.6%)	3/21 (14.3%)	RR 0.39 (0.04 to 3.42)	87 fewer per 1000 (from 137 fewer to 346 more)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No papers found												
FUNCTION												
No papers found												
IMPACT ON CARERS												
No papers found												
ADVERSE EVENTS												
No papers found												

[A] *Outcomes were downgraded by two increments because the study had likely attrition bias and inadequate blinding.*
[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment if the I^2 value was 50-74 and by two increments if the I^2 value was >75.*

Table 172: Clinical evidence profile: resistance exercise versus control

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Resistance exercise**	**Control**	**Relative (95% CI)**	**Absolute**		
SPASTICITY												
Change from baseline to 12 weeks in R hip flexion modified Ashworth score (higher worse) (Better indicated by lower values)												
Tarakci 2013	RCT	Very serious[A]	No inconsistency	No indirectness	Very serious imprecision[B]	none	-0.67(1.17)[51]	0.13(1.17)[48]	-	MD 0.8 lower (1.26 to 0.34 lower)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No papers found												
FUNCTION												
No papers found												
IMPACT ON CARERS												
No papers found												
ADVERSE EVENTS												
No papers found												

[A] *Outcomes were downgraded by two increments because the study had likely selection and attrition bias and inadequate blinding.*
[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*
[C] *Outcomes were downgraded by one increment if the I^2 value was 50-74 and by two increments if the I^2 value was >75.*

Table 173: Clinical evidence profile: Mid thoracic magnetic stimulation versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Magnetic stimulation (at mid thorax)	versus placebo	Relative (95% CI)	Absolute		
SPASTICITY												
Modified Ashworth Score at 1 day (Better indicated by lower values)												
Nielsen 1996	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	16.3(6.2)[18]	13.2(7.8)[17]	-	MD 3.1 higher (1.59 lower to 7.79 higher)	VERY LOW	CRITICAL
Modified Ashworth Score at 8 days (Better indicated by lower values)												
Nielsen 1996	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	18.3(7.4)[18]	13.5(7.3)[17]	-	MD 4.8 higher (0.07 lower to 9.67 higher)	VERY LOW	CRITICAL
Modified Ashworth Score at 16 days (Better indicated by lower values)												
Nielsen 1996	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	19(9.4)[18]	13.2(9)[17]	-	MD 5.8 higher (0.3 lower to 11.9 higher)	VERY LOW	CRITICAL
spasticity self score (relative to fixed baseline score of 5) at 1 day (Better indicated by lower values)												
Nielsen 1996	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	6.1(1.6)[18]	6.5(1.8)[17]	-	MD 0.4 lower (1.53 lower to 0.73 higher)	VERY LOW	CRITICAL
spasticity self score (relative to fixed baseline score of 5) at 8 days (Better indicated by lower values)												
Nielsen 1996	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	4.8(1)[18]	5.8(1.6)[17]	-	MD 1 lower (1.89 to 0.11 lower)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Magnetic stimulation (at mid thorax)**	**versus placebo**	**Relative (95% CI)**	**Absolute**		
spasticity self-score (relative to fixed baseline score of 5) at 16 days (Better indicated by lower values)												
Nielsen 1996	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	4.6(0.8)[18]	5.2(1.9)[17]	-	MD 0.6 lower (1.58 lower to 0.38 higher)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No papers found												
FUNCTION												
No papers found												
IMPACT ON CARERS												
No papers found												
ADVERSE EVENTS												
No papers found												

[A] *Outcomes were downgraded by two increments because the study had likely selection and attrition bias and inadequate blinding.*

[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 174: Clinical evidence profile: reflexology versus control

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Reflexology**	**Control**	**Relative (95% CI)**	**Absolute**		
SPASTICITY												
MAS change from baseline to 11 weeks (Better indicated by lower values)												
Siev-Ner 2003	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	-2.09(3.01)[27]	0.2(1.72)[26]	-	MD 2.29 lower (3.6 to 0.98 lower)	LOW	CRITICAL
QUALITY OF LIFE												
No papers found												
FUNCTION												
No papers found												
IMPACT ON CARERS												
No papers found												
ADVERSE EVENTS												
No papers found												

[A] *Outcomes were downgraded by two increments because the study had likely selection bias and inadequate blinding.*

Table 175: Clinical evidence profile: pulsed electromagnetic stimulation versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulsed electromagnetic stim	placebo	Relative (95% CI)	Absolute		
SPASTICITY												
change from baseline in self-reported spasticity score at 8 weeks (lower better) (Better indicated by lower values)												
Richards 1997	RCT	serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision	none	-0.8(0.89)[15]	-0.17(1.47)[15]	-	MD 0.63 lower (1.5 lower to 0.24 higher)	MOD	CRITICAL
Improvement in spasm score at 4 weeks (higher better)												
Lappin 2003	RCT	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	none	0.24(0.79)[117]	0.11(0.88)[117]	-	MD: 0.13 higher (0.00 higher to 0.26 higher)	HIGH	CRITICAL
QUALITY OF LIFE												
No papers found												
FUNCTION												
No papers found												
IMPACT ON CARERS												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulsed electromagnetic stim	placebo	Relative (95% CI)	Absolute		
No papers found												
ADVERSE EVENTS												
No papers found												

[A] *Outcomes were downgraded by two increments because the study had likely selection and attrition bias and inadequate blinding.*

Table 176: Clinical evidence profile: Electrical neuromuscular stimulation versus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulsed electromagnetic stim	placebo	Relative (95% CI)	Absolute		
SPASTICITY												
Patient subjective improvement in spasticity at 6 weeks												
Livesley 1992	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	9/20 (45%)	4/19 (21.1%)	2.14(0.79 to 5.79)	241 more per 1000 (from 44 fewer to 376 more)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulsed electromagnetic stim	placebo	Relative (95% CI)	Absolute		
Patient subjective improvement in spasticity at 3 months												
Livesley 1992	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	2/20(10%)	3/19 (15.8%)	0.63 (0.12 to 3.38)	58 fewer per 1000 (from 139 fewer to 376 more)	VERY LOW	CRITICAL
QUALITY OF LIFE												
No papers found												
FUNCTION												
Rivermead gross function at 6 weeks												
Livesley 1992	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Likely to be very serious	none	Median (IQR): 9(6-10.5)	Median (IQR): 11(5-11)	NS difference reported	-	VERY LOW	CRITICAL
Rivermead leg function at 6 weeks												
Livesley 1992	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Likely to be very serious	none	Median (IQR): 8(4.5-10)	Median (IQR): 9(4-10)	NS difference reported	-	VERY LOW	CRITICAL
IMPACT ON CARERS												
No papers found												
ADVERSE EVENTS												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulsed electromagnetic stim	placebo	Relative (95% CI)	Absolute		
No papers found												

[A] *Outcomes were downgraded by two increments because the study had likely selection and attrition bias and inadequate blinding.*
[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID <u>or</u> the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 177: Clinical evidence profile: massage versus usual care

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	massage	Usual care	Relative (95% CI)	Absolute		
SPASTICITY												
Modified Ashworth Scale at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-0.54(0.55)[12]	0.33(0.46)[12]	-	MD: 0.87 lower (from 1.28 lower to 0.46 lower)	LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												

No studies found covering this outcome												
CARER PERCEPTIONS												
No studies found covering this outcome												
ADVERSE EVENTS												
No studies found covering this outcome												

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 178: Clinical evidence profile: massage versus aerobic/res exercise

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	massage	Mixed aerobic/res	Relative (95% CI)	Absolute		
SPASTICITY												
Modified Ashworth Scale at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Very serious imprecision[B]	none	-0.54(0.55)[12]	-0.47(0.66)[12]	-	MD: 0.07 lower (from 0.56 lower to 0.42 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												

No studies found covering this outcome
ADVERSE EVENTS
No studies found covering this outcome

[A] *Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).*
[B] *Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Table 179: Clinical evidence profile: massage versus massage plus aerobic/res exercise

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	massage	massage plus aerobic/res exercise	Relative (95% CI)	Absolute		
SPASTICITY												
Modified Ashworth Scale at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-0.54(0.55)[12]	-0.14(0.77)[12]	-	MD: 0.4 lower (from 0.94 lower to 0.14 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												
No studies found covering this outcome												
CARER PERCEPTIONS												

No studies found covering this outcome
ADVERSE EVENTS
No studies found covering this outcome

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 180: Clinical evidence profile: aerobic/res exercise versus control

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic/res	Control	Relative (95% CI)	Absolute		
SPASTICITY												
Modified Ashworth Scale at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	No serious imprecision[B]	none	-0.47(0.66)[12]	0.33(0.46)[12]	-	MD: 0.8 lower (from 1.26 lower to 0.34 higher)	LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												
No studies found covering this outcome												

CARER PERCEPTIONS	
No studies found covering this outcome	
ADVERSE EVENTS	
No studies found covering this outcome	

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Table 181: Clinical evidence profile: aerobic/res exercise versus massage plus aerobic/res ex

Quality assessment							Mean (sd) [n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed aerobic/res	Massage + mixed aerobic/res	Relative (95% CI)	Absolute		
SPASTICITY												
Modified Ashworth Scale at 5 weeks (lower better)												
Negahban 2013	RCT	Very serious[A]	No serious inconsistency	No serious indirectness	Serious imprecision[B]	none	-0.47(0.66)[12]	-0.14(0.77)[12]	-	MD: -0.33 lower (from 0.9 lower to 0.24 higher)	VERY LOW	CRITICAL
FUNCTIONAL OUTCOMES												
No studies found covering this outcome												
QUALITY OF LIFE												

No studies found covering this outcome
CARER PERCEPTIONS
No studies found covering this outcome
ADVERSE EVENTS
No studies found covering this outcome

[A] Outcomes were downgraded for lack of allocation concealment, lack of blinding, attrition bias or other major risks of bias. One major risk of bias in an outcome led to a downgrade by one increment (serious risk of bias) and two or more major risks of bias led to a downgrade by two increments (very serious risk of bias).
[B] Outcomes were downgraded by one increment (serious imprecision) if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments (very serious imprecision) if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Narrative review for outcomes not appropriate for meta-analysis

Vibration and exercise versus exercise only

The Eight MS spasticity scale (MSSS-88) component results were not presented in a manner allowing analysis in review manager in Schyns et al. 2009[218]. The article reported a non-significant difference for 6 of the components (ADL, social functioning, stiffness, gait, body movement and emotional health). For MSSS-88 spasm a greater improvement in score was reported with vibration (p=0.02). For MSSS-88 pain, within-group results were reported but they were impossible to interpret as the direction of effect was very ambiguously described.

Functional variables were analysed in a paired analysis, but the reporting of results was again unsuitable for review manager. The table below summarises these functional results

Table 182: Functional variables

	Paired analysis
10m walk test	Vibration caused median improvement of 1 second, no vibration median improvement of 0.5 seconds. NS difference: P=0.561
TUG	Vibration caused median improvement of 1.25 seconds, no vibration median improvement of 1.5 seconds. NS difference: P=0.720
MSIS physical	Vibration caused median improvement of 1 point, no vibration median improvement of 4 points. NS difference: P=0.760
MSIS psychological	Vibration caused median improvement of 2 points, no vibration median improvement of 0 seconds. NS difference: P=0.634

Sports climbing versus Yoga

Velikonja et al. 2010[256] carried out a non-parametric analysis, and results are in the table below, in medians and an undefined range. Overall, yoga appeared to lead to greater improvements in spasticity than sports climbing, but no between group statistical analysis was performed, so precision of this difference is not available.

Table 183: Sports Climbing versus Yoga

Variable	Climbing (n=10)			Yoga (n=10)		
	baseline	10 weeks	p	baseline	10 weeks	p
Spasticity MSA	10(8.5-18.3)	12.5(10-17.3)	0.574	9.3(3.5-18.4)	8.8(5.5-17.1)	0.673

TENS 8 hours per day versus TENS 1 hour per day

In this cross-over study, Miller 2007A[149] made no within-subject comparisons of the interventions, and merely reported the significance of the pre-post changes in each treatment separately. A narrative summary of their results is given below, and this suggests that the 8 hour treatment was more effective.

Table 184: TENS 1 or 8 hours per day

	8 versus 1 hour TENs
Global spasticity scale	Larger reduction* in GSS for the 8 hour than 60 minute treatment but no between-treatment data or variances provided (except in low resolution graph)
Penn spasm scale	Larger reduction** in PSS for the 8 hour than 60 minute treatment but no between-treatment data or variances provided (except in low resolution graph)
VAS (10 point) of effects on muscle spasm and pain	Larger reduction** in VAS for the 8 hour than 60 minute treatment but no between-treatment data or variances provided (except in low resolution graph)

*= significant ($p<0.05$) pre-post improvements in both treatments

**= significant ($p<0.05$) pre-post improvements in 8 hour treatment but not in 1 hour treatment.

<u>Transcranial magnetic stimulation versus placebo</u>

Mori et al. 2011[155] performed no between-group analyses. Post test data were given for all variables for the TMS and exercise group but only for the MAS for TMS alone. No comparable 2 week data were provided for the sham and exercise group as only 2 month data were provided for that group (but not the other 2 groups). Post-test values are given in the table below.

Table 185: Transcranial magnetic stimulation versus placebo

	TMS + exercise	TMS alone	Exercise + placebo
Modified Ashworth scale	1.3(0.4)	1.6(0.8)	Not available at comparable 2 week follow up
MSSS-88	53.2(10.9)	Data not given	Not available at comparable 2 week follow up
FSS	31.6(4.6)	Data not given	Not available at comparable 2 week follow up
Barthel index	95(1.85)	Data not given	Not available at comparable 2 week follow up
MSQoL Phys	64.8(2.7)	Data not given	Not available at comparable 2 week follow up
MSQol mental	Data not given	Data not given	Not available at comparable 2 week follow up

However, it appeared from low resolution graphs that the biggest improvements from baseline in MAS, MSSS-88, FSS, Barthel index, MSQoL physical were in the combined TMS and exercise group. Within (pre/post) group analyses showed that significant improvements were only seen in the TMS/exercise group, except for MAS, where there were significant improvements also seen in the TMS alone group.

Standing therapy versus home exercise programme

Baker 2007[12] showed no difference between 30 minutes of standing and a home exercise plan in reducing spasticity. Median (IQR) values at 3 weeks are given in the table below. The paired data was analysed using the appropriate Wilcoxon signed ranks test.

Table 186: Standing therapy versus home exercise

	Exercise	Standing	p
R hip flexion Ashworth	1.5(1)	1(2.25)	1
L hip flexion Ashworth	2(2)	2(2.5)	0.56
R hip Abduction Ashworth	2(1.5)	2(1)	0.56
L hip Abduction Ashworth	2(1)	2(1.5)	0.56
R knee flex Ashworth	1.5(1.2)	1.5(2.25)	0.47
L knee flexion Ashworth	3(2)	1(0.5)	0.45
R ankle DF Ashworth	2(1.25)	1.5(1.25)	0.56
L ankle DF Ashworth	1.5(2)	1(0.5)	0.33
Penn Spasm frequency R	2.5(2.5)	1(3.2)	1
Penn Spasm frequency L	2(2.2)	2(3.2)	0.317

Cooling versus control

In a cross-over study, Nilsagard 2006[168] presented results as median (IQR). The analysis appears to have used a parallel group, rather than paired approach. There was no clear difference in spasticity reduction between groups.

Table 187: Cooling versus control

	Cooling Median (IQR) n=43	Control Median (IQR) N=43	p
Change in modified Ashworth score from baseline	-0.5(-1.25 to 0.5)	0(-0.5 to 0.62)	0.296

10.6.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

Relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

10.6.5 Evidence statements

10.6.5.1 Clinical

Vibration and exercise versus exercise only

Very low quality evidence from one RCT comprising 12 participants showed that vibration and exercise was clinically effective compared to exercise only in terms of the proportion of people with improvements in quadriceps spasticity at 4 weeks, with very serious imprecision.

Very low quality evidence from one RCT comprising 12 participants showed that there was no difference in clinical effectiveness between vibration and exercise and exercise alone in terms of hamstring, adductor or gastrocnemius spasticity at 4 weeks, with very serious imprecision.

Resistance exercise versus control

Very low quality evidence from one RCT comprising 99 participants showed that resistance exercise was clinically effective compared to control in terms of hip flexor spasticity, with serious imprecision.

Mid thoracic magnetic stimulation versus placebo

Very low quality evidence from one RCT comprising 35 participants showed that mid-thoracic magnetic stimulation was clinically effective compared to placebo in terms of a subjective self-rating of spasticity at 8 days, with serious imprecision.

Very low quality evidence from one RCT comprising 35 participants showed that there was no difference in clinical effectiveness between mid-thoracic magnetic stimulation and placebo in terms of objective measures of spasticity at 1, 8 or 16 days, and subjective measures at 1 or 16 days, with serious imprecision

Reflexology versus control

Low quality evidence from one RCT comprising 51 participants showed that reflexology was clinically effective compared to placebo in terms of an objective rating of spasticity change from baseline to 11 weeks, with no imprecision.

Pulsed electromagnetic versus placebo

Moderate quality evidence from one RCT comprising 30 participants showed that there was no difference in clinical effectiveness between pulsed electromagnetic stimulation and placebo in terms of a subjective self-rating of spasticity at 8 weeks, with no imprecision.

High quality evidence from one RCT comprising 234 participants showed pulsed electromagnetic stimulation was clinically effective compared to placebo in terms of spasm score at 4 weeks, with no imprecision.

Electrical neuromuscular stimulation versus placebo

Very low quality evidence from one RCT comprising 39 participants showed that electrical neuromuscular stimulation was clinically effective compared to placebo in terms of a subjective self-rating of spasticity at 6 weeks, with serious imprecision.

Very low quality evidence from one RCT comprising 39 participants showed that there was no difference in clinical effectiveness between electrical neuromuscular stimulation and spasm in terms of subjective measures of spasticity at 3 months, and functional measures at 6 weeks, with very serious imprecision.

Massage versus usual care

Low quality evidence from one RCT comprising 24 participants showed that massage was clinically effective compared to usual care in terms of an objective rating of spasticity at 5 weeks, with no imprecision.

Massage versus aerobic/resistance exercise

Very low quality evidence from one RCT comprising 24 participants showed that there was no difference in clinical effectiveness massage and aerobic/resistance exercise in terms of an objective rating of spasticity at 5 weeks, with very serious imprecision.

Massage versus massage + aerobic/resistance exercise

Very low quality evidence from one RCT comprising 24 participants showed that massage was clinically effective compared to massage and aerobic/resistance exercise in terms of an objective rating of spasticity at 5 weeks, with serious imprecision.

Aerobic/resistance exercise versus usual care

Low quality evidence from one RCT comprising 24 participants showed that aerobic/resistance exercise was clinically effective compared to usual care in terms of an objective rating of spasticity at 5 weeks, with no imprecision.

Aerobic/resistance exercise versus massage + aerobic/resistance exercise

Low quality evidence from one RCT comprising 24 participants showed that there was no difference in clinical effectiveness between aerobic/resistance exercise and massage + aerobic/resistance exercise in terms of an objective rating of spasticity at 5 weeks, with serious imprecision.

10.6.5.2 Economic

No relevant economic evaluations were identified.

10.6.6 Recommendations and link to evidence

Recommendations	
Relative values of different outcomes	A measure of spasticity was the most critical outcome as this was the most directly relevant outcome to this question. Quality of life, function, carer perceptions and adverse events were also regarded as critical.
Trade off between clinical benefits and harms	There were clinically important benefits on spasticity observed for Vibration and exercise, resistance exercise, resistance/aerobic exercise, mid thoracic magnetic stimulation, pulsed electromagnetic field therapy, electrical muscle stimulation, reflexology and massage, and none of these interventions had any reported adverse effects.
Economic considerations	No relevant economic evaluations were identified. Unit costs for individual based physiotherapy interventions (£234–416 for one intervention episode), a standard 6-month electrical neuromuscular stimulation treatment package using the Microstim 2 system (£840 for one initial assessment and five treatment sessions) and TENS devices (£34–191 each) were presented. With the lack of clear clinical evidence and the considerable cost to the NHS of these interventions, the GDG agreed that there was insufficient evidence to make a recommendation.
Quality of evidence	The quality of evidence was generally low or very low, largely due to imprecise estimates and methodological flaws such as a lack of allocation concealment or blinding, and attrition bias. The exception was the evidence for pulsed electromagnetic stimulation, which was moderate to high quality.
Other considerations	Despite the clinically significant findings for several treatments, the quality of evidence was not considered by the GDG to be adequate to make a recommendation for a specific technique or programme for people with spasticity. The GDG recognised that individual people with MS may benefit from advice from physiotherapists and occupational therapists and other specialists for advice on management, posture as outlined in recommendations and LETR in section 11.1.6.

10.7 Setting of rehabilitation

10.7.1 Introduction

People with MS regularly require rehabilitation. Access to rehabilitation expertise can be variable. While the type of rehabilitation that is appropriate for each patient will need to be individualised there is interest in whether the setting of rehabilitation is important in outcomes. Settings may influence the frequency and intensity of contact with healthcare professionals. However if outcomes were similar from different settings this would have implications for organisation of care for people with MS.

10.7.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of rehabilitation provided in different settings?

For full details see review protocol in Appendix C.

Table 188: PICO characteristics of review question

Population	Adults
Intervention/s	Any rehabilitation in the following settings: • Inpatient/residential • Outpatient/other including community • Home (with or without carer involvement)
Comparison/s	any of the above
Outcomes	• Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale. • Impact on carers. • Functional scales that quantify activity of daily living/levels of disability levels of disability, such as the Expanded Disability Status Scale (EDSS), the Multiple Sclerosis Functional Composite (MSFC), the Cambridge Multiple Sclerosis Basic Score (CAMBS), the Functional Assessment of Multiple Sclerosis (FAMS) or the National Fatigue Index (NFI). • Adverse effects of treatment.
Study design	Systematic reviews RCTs observational studies

10.7.3 Clinical evidence

Three studies were included in the review. One study compared outpatient to home rehabilitation[262], although the only form of rehabilitation compared was physiotherapy. Two studies compared inpatient rehabilitation to outpatient rehabilitation [68,251]. Ungaro et al. 2009[251] was originally found as a published abstract, but no full paper was available. The lead author was contacted and adequate details of the methodology and results were obtained to allow its inclusion.

The methodologies and populations of all 3 studies are summarised in Table 189. Evidence is summarised in the Grade table in section 1.2.1 and the narrative review section 10.7.3.1.1

Table 189: Summary of studies included in the review

Study	Population	Methods	Intervention	comparator
Wiles 2001[262]	Definite or probably MS with median EDSS of 6-6.5; "chronic MS".	RCT. Cross-over study. Assessor blinding, but no clinician or patient blinding. Low drop out.	Home physiotherapy: 2 sessions of 45 minutes each week on different days for 8 weeks, given at home by senior 1 physiotherapist. There were 2 PTs, each treating half the patients, and each treated the same patients at both sites. Treatments based on an individualised problem solving approach, focussing mainly on specific functional activities. The actual therapy given at home was based on the space and facilities available.	Outpatient physiotherapy: 2 sessions of 45 minutes each week on different days for 8 weeks, given in the physiotherapy department by senior physiotherapist. There were 2 PTs, each treating half the patients, and each treated the same patients at both sites. Treatments based on an individualised problem solving approach, focussing mainly on specific rehabilitation techniques.
Francabandera 1988[68]	Definite MS with EDSS 6-9; MS type not reported	RCT. No allocation concealment or ITT. No blinding.	Inpatient rehabilitation. Patient admitted to a 30 bed neurological unit. Daily physical and occupational therapy on an individualised basis. Average of 2x45 min PT sessions and 1 OT session per day. Bladder management, speech therapy and social services were also provided as needed.	Outpatient rehabilitation under the supervision of a neurologist. Received PT, OT, bladder management, speech therapy and social services. Equipment needs were also evaluated and appliances ordered as needed. Treatments were administered through community-based visiting nurse services or public health nurse services. Nurses were actively involved in all aspects of home rehab. The frequency and duration of visits were not described.
Ungaro 2009[251]	MS with EDSS 3.5-6.5; 9/21 Relapsing remitting, 12/21 secondary progressive.	Prospective cohort study. Allocation by geographical area. No blinding. No drop-outs.	Intensive inpatient rehabilitation programme in a neurorehabilitation department in Northern Italy. 2 hours per day for 5 days a week for 3 weeks. Each patient had a tailored rehabilitation program consisting of passive interventions (muscle stretching) and active interventions (like strength training, and balance/gait training).	Similar programme in an outpatient clinic in Southern Italy, but given 3 times a week for 1 hour per day for 6 months.

Home versus outpatient rehabilitation

Table 190: Clinical evidence summary

Quality assessment							Effect*		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	n	Absolute		
Quality of life - 0										
Rivermead mobility index (Better indicated by higher values)										
Wiles 2001	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	MD 0.1 higher (0.65 lower to 0.87 higher)	MODERATE	CRITICAL
Balance time (s) (Better indicated by higher values										
Wiles 2001	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	MD 0.68 higher (2.64 lower to 3.99 higher)	MODERATE	CRITICAL
Walk A (Better indicated by lower values)										
Wiles 2001	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	MD 0 higher (9 lower to 8 higher)	MODERATE	CRITICAL
Nine hole peg test (s) (Better indicated by lower values)										
Wiles 2001	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	MD 5 higher (9 lower to 19 higher)	MODERATE	CRITICAL
Assessor global mobility change score (Better indicated by higher values)										
Wiles 2001	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	40	MD 2.6 higher (3.2 lower to 8.4 higher)	LOW	IMPORTANT
VAS patient mobility (0-100) (Better indicated by higher values)										
Wiles 2001	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	MD 1 lower (7.8 lower to 5.8 higher)	MODERATE	CRITICAL

[A] *Downgraded for risk of bias because of a lack of patient or clinician blinding. Although this was beyond the control of the researchers, the lack of blinding may have led to some bias.*

[B] *Downgraded for imprecision because the upper confidence interval crossed the positive MID (half the baseline control group sd), indicating that the overall estimate was consistent with both no clinical effect and clinical benefit*

10.7.3.1 Outpatient Rehabilitation versus inpatient rehabilitation

10.7.3.1.1 Narrative review for outcomes not appropriate for GRADE

Francabandera 1988[68]

In this RCT study, large baseline differences for Incapacity Status Scale (ISS) and hours of weekly home assistance precluded a direct group comparison of the mean (sd) values at 3 months. The authors adjusted for baseline difference, but reported only the adjusted mean values, the ANOVA F ratio and categorical p values (less than or more than 0.05) Table 191). Compared to the outpatient group, the inpatient group had a lower (better) adjusted ISS at 3 months ($p<0.05$), but a higher (worse) numbers of hours of home assistance required per week ($P>0.05$). These data were not appropriate for GRADE as variance measures for each group could not be accurately estimated from the adjusted ANOVA statistics.

The quality rating of outcomes from this study were deemed low, due to a lack of allocation concealment, a lack of blinding, and no intention to treat analysis.

Table 191: Adjusted means for ISS and hours of required home assistance (Francabandera 1988[68]).

Outcome	Inpatient	Outpatient	Statistical result
ISS (lower better) at 3 months adjusted for baseline level	24.3	27.2	ANOVA result, adjusting for baseline ISS: $F(1,70)=4.3$; $p<0.05$ in favour of inpatient group
Hours of home assistance needed weekly (lower better) with adjustment for baseline difference	76.9	73.1	ANOVA result, adjusting for baseline hours of home assistance: $F(1,70)=0.17$; $p>0.05$

Ungaro et al. 2009[251]

This study was originally only available in abstract form, but on request the lead author provided more information on the methodology and results. The details below comprise all the information received.

In this non-randomised study, an intensive inpatient rehabilitation regimen and a less intensive but more prolonged outpatient rehabilitation regimen were compared. Although the comparison between the different interventions appears to be confounded by intensity and duration, these are intrinsic features of the two interventions being compared, and therefore arguably enable a pragmatic comparison.

There were two follow-up points, but these were at different times for each group. These reflected the different durations of each treatment, which were 3 weeks for the inpatient regimen and 6 months for the outpatient regimen. The first follow-up point was 3 weeks for the inpatient group and 3 months for the outpatient group. The second follow-up time was 3 months for the inpatient group and 6 months for the outpatient group. The inpatient group therefore had the first follow-up at the cessation of treatment and the second follow-up 10 weeks after the end of therapy. In contrast, the outpatient group had the first follow-up half way through their rehabilitation and the second follow-up at the cessation of their rehabilitation. Thus there is no coherence across groups for follow-up times in either absolute or relative terms. This was presumably to evaluate if the brief inpatient programme had any prolonged effects after its cessation, compared to a more continuous outpatient programme.

At the first follow-up point the inpatient group showed better improvements from baseline than the outpatient group for all outcomes, although there was uncertainty about the true direction of effect for the 10 metre walk time. At the second follow-up point, there were no clear differences. Although there was a tendency for the outpatient group to show better improvements from baseline than the inpatient group for dominant hand 9 HPT and Barthel Index, there was uncertainty about the true direction of effect. Between-group outcomes were presented in low-resolution graphs, and therefore the figures in the table below (Table 192) are estimates, although the directions of effect and p values are accurate. The results show that intensive inpatient rehabilitation is better than less intensive outpatient rehabilitation whilst both are on-going. However the carry-over effects of inpatient therapy are relatively low after its cessation and at final follow-up there were no differences between groups.

These data were not entered into GRADE because the data were based on visual estimation from a low-resolution graph. In addition, there was no way of estimating variance for the results at second follow-up, as accurate p values were not given.

Allocation to groups was by geographical location, and there was no blinding, but this study had no drop outs. Overall the quality rating was very low.

Table 192: Results from Ungaro et al. 2009

Outcome	Inpatient change from baseline	Outpatient change from baseline	Group difference (inpatient – outpatient)	p value
EDSS (lower better) at first follow-up	-1	+0.25	-1.25	p<0.0001
9-HPT non-dominant hand (lower better) at first follow-up	-9	-1.5	-7.5	p<0.02
9-HPT dominant hand (lower better) at first follow-up	-4	+1	-5	P<0.02
BI (higher better) at first follow-up	+4	-1	+5	<0.01
10m walk time (lower better) at first follow-up	-2	+1.5	-3.5	p=0.09
EDSS (lower better) at second follow-up	0	0	0	NS
9HPT non dominant hand (lower better) at second follow-up	+1	+1	0	NS
9HPT dominant hand (lower better) at second follow-up	+6	-6	+12	NS
BI (higher better) at second follow-up	+1	+1.5	-0.5	NS
10m walk time	+0.5	+0.5	0	NS

Outcome	Inpatient change from baseline	Outpatient change from baseline	Group difference (inpatient – outpatient)	p value
(lower better) at second follow-up				

10.7.4 Economic evidence

Published literature

One study was included with the relevant comparison.[262] This study is summarised in the economic evidence profile below (Table 193). See the study evidence tables in Appendix H.

One economic evaluations relating to this review question was identified but was excluded due to limited applicability.[185] This is summarised in Appendix H, with reasons for exclusion given.

See also the economic article selection flow chart in Appendix E.

Table 193: Economic evidence profile: Home versus outpatient rehabilitation

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Wiles 2001 [262] (UK)	Partially applicable (a)	Very serious limitations (b)	Within trial analysis (RCT). Follow-up = 8 weeks.	£14 (c)	Rivermead mobility index, MD 0.1 Balance time, MD 0.68 Walk A, MD 0 Nine hole peg test, MD 5 Assessor global mobility change score, MD 2.6 VAS patient mobility, MD -1	n/a	NR

(a) Costs consequence analysis.
(b) Source of unit costs unclear. No sensitivity analysis conducted.
(c) Cost components considered: employment cost of physiotherapist and mileage.
Abbreviations: MD = mean difference; n/a = not applicable; NR = not reported; RCT = randomised control trial.

10.7.5 Evidence statements

10.7.5.1 Clinical

Outpatients versus inpatients

No evidence statements were produced because of the nature of the evidence reported.

Home versus outpatients

Moderate quality evidence from one cross-over study comprising 40 participants showed that home rehabilitation and outpatient rehabilitation did not differ in terms of Rivermead mobility index, balance time, anbulation, upper limb function (9 hole peg test), global mobility score or VAS patient mobility.

10.7.5.2 Economic

One cost-consequence analysis found that home rehabilitation was more costly and effective than outpatient rehabilitation for treating mobility (£14 more per patient, 0.1 mean difference improvement in Rivermead mobility index per patient). This analysis was assessed as partially applicable with very serious limitations.

10.7.6 Recommendations and link to evidence

Recommendations	
Relative values of different outcomes	Quality of life was the most important outcome, as optimising quality of life is the main aim of all treatment. Outcomes describing motor function were the second most important, as functional improvement is a key aim of rehabilitation. Impact on carers was next in importance. Adverse events from treatment were regarded as of lowest importance as these are not expected to be serious after rehabilitation in any setting.
Trade off between clinical benefits and harms	No important harms were expected from rehabilitation in any setting, and so the small reported clinical benefits for inpatients compared to outpatients shown in two studies can be considered without the need to take into account any adverse effects. These relative benefits for inpatient therapy were for functional measures such as ISS, EDSS, Barthel index and upper limb function. These advantages were only observed when assessment was undertaken during or immediately after the treatment period. After a 9 week period after the cessation of inpatient rehabilitation the advantage for inpatient rehabilitation over outpatient rehabilitation disappeared. However it should be noted that outpatient therapy was still on-going at this point, and that outcomes in the two forms of rehabilitation were still comparable, indicating that the effects of inpatient therapy are reasonably persistent. No differences in clinical benefits or harms were found between home rehabilitation and outpatient rehabilitation.
Economic considerations	One cost–consequence analysis, also presented as part of the clinical evidence, was identified which found that home rehabilitation was more costly and effective than outpatient rehabilitation for treating mobility. The study had very serious limitations and when the GDG considered it in conjunction with the clinical evidence, they felt it was not strong enough evidence to specify which type of rehabilitation is most cost-effective. Therefore, the GDG recommended that the setting for rehabilitation should be decided by

	professionals and patients decisions and needs. The GDG were aware that costs to the NHS may differ from costs to patients and to those providing social care. Rehabilitation outside hospital, in community units and at home is likely to be less costly to the NHS than rehabilitation in a hospital. Non-hospital settings might have more costs for people with MS, their carers and local authorities.
Quality of evidence	The evidence considered varied in quality. The evidence considering inpatient versus outpatient rehabilitation was of low to very low quality. One study was an RCT, which was graded low. The RCT was limited by a lack of allocation concealment, blinding or ITT, as well as poor description of the intensity and duration of the different rehabilitation interventions. The other study was a cohort study, which was graded as very low. It was confounded by geographical location, with groups allocated according to area of residence. Despite fair comparability for known confounders at baseline, the lack of randomisation meant a high probability of residual confounding from unmeasured confounders. The RCT evidence considering home rehabilitation and outpatient rehabilitation was classed as moderate, downgraded due to a lack of patient or health care professional blinding. Though completely unavoidable, this may have led to bias. The economic evidence was assessed as partially applicable with very serious limitations.
Other considerations	The question about setting of rehabilitation is important as there is concern that appropriate facilities are not available for people with MS for rehabilitation. The GDG were aware that a focussed review looking at studies of similar interventions delivered in different settings would not answer all questions in this area but were interested to know what evidence there was. The GDG considered that the choice of appropriate setting of rehabilitation for a person with MS is complex and that multiple issues need to be considered. The needs of people with MS and type of appropriate rehabilitation will vary. The factors influencing appropriate setting of rehabilitation include availability of care support for people at home, geographical location, goals of individuals, and the type of rehabilitation required. The GDG considered that while attending for out-patient rehabilitation has costs for patients and may not be possible for people with very severe disease; attendance at out-patient rehabilitation is both a marker of patient motivation but may also be a source of motivation for the person with MS. The GDG also considered that apart from specific rehabilitation goals, people derive general support from involvement in rehabilitation. Ideally it would be useful to know which type of rehabilitation benefits which people most. This decision is currently the task of people involved in the care of people with MS who have expertise in this area. Further research would be helpful in this area but because the best setting of rehabilitation will always be influenced by multiple factors clinical judgement will always be important. The GDG decided not to make a specific recommendation for setting of rehabilitation. The GDG considered it important that a range of options are available. They considered that most people would prefer community or home options, but that this is not always possible and the availability of inpatient facilities is important.

11 Comprehensive review

11.1 Introduction

Multiple sclerosis is a chronic condition. Many people diagnosed with MS with live with the condition for many decades. Some will be in close contact with medical professionals because of the nature of their condition, other less frequently. The nature of MS is that it can cause problems in many different systems. Some of these may not be recognised by the person with MS as associated with MS. Regular review of people with chronic diseases occurs in other areas such as diabetes where structured reviews ensure that all important areas are reviewed.

11.2 Review question: Does the use of structured assessment(s) compared with non-structured assessment(s) improve patient and carer outcomes for people with MS? What is the optimal timing of a structured assessment? What should be the frequency of a structured assessment?

A summary of the review protocol is outlined in Table 194 below. For full details see review protocol in Appendix C.

Table 194: PICO characteristics of review question

Population	• Adults with MS • Mixed any % of adults with MS • Chronic neurological disability exclude dementia
Intervention/s	• Structured assessments of mood, cognition and daily activities, as recommended by the NSF-LTC • Guys neurological disability scale • UK neurological disability scale (Sharrack and Hughes) • MSIS-29
Comparison/s	• Non structured/standard assessments
Outcomes	Critical: • Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale. • Impact on carers. • Functional scales that quantify level of disability, • Cognitive functions • Psychological symptoms assessed by validated and disease-specific scales, questionnaire or similar instruments. • Hospitalisations • Outpatient appointments • Adverse effects of treatment. • Relapse rates • Access to services
Study design	• Systematic reviews • RCTs

11.3 Clinical evidence

One study was included in the review.[123]. Evidence from this study are summarised in the clinical GRADE evidence profile below (Table 196). See also the study selection flow chart in Appendix D, forest plots in Appendix I, study evidence tables in Appendix G and exclusion list in Appendix J.

We searched for randomised trials comparing the effectiveness of structured assessments with no-structured or standardised assessment. One RCT[123] was identified comparing screening patients for cognitive impairments compared with performing a full cognitive assessment. There were no papers reporting on the timing or frequency of the assessment.

Summary of included studies

Table 195: Summary of studies included in the review

Study	Intervention/comparison	Population	Outcomes	Comments
Lincoln 2002[123]	**Screening** Shortened version of the National adult reading test (NART) as a measure of premorbid intelligence and the brief repeatable battery (BRB-N) to evaluate verbal reasoning, visual memory, attention and speed of information processing. In addition the ambulation index was administered as a measure of physical mobility, the Guy's neurological disability scale (GNDS) as a measure of effects of multiple sclerosis and mood was assessed on the general health questionnaire-28 (GHQ-28). The occupational, educational history and the disease duration and course were established **Assessment** Patients received detailed cognitive assessment taking about 3 hrs. Patients were assessed on measures of memory, attention and executive functioning. Further assessments were selected on the basis of patients' performance. The assessments were selected according to the nature of	Patients with either clinically definite, clinically probable, or laboratory supported multiple sclerosis. Secondary progressive (SP) 33, relapsing remitting (RR) 35, primary progressive (PP) 6, unknown 5	Patients outcomes: General Health Questionnaire (GHQ) SF-36 Overall quality of life Satisfaction with quality of life Extended activities of daily living Everyday memory questionnaire (EMQ) Dysexecutive syndrome questionnaire (DEX) Memory aids questionnaire Carers outcomes: GHQ EMQ DEX	Third intervention arm not reported here.

Study	Intervention/comparison	Population	Outcomes	Comments
	the patients' problems so that they were representative of cognitive assessments used in clinical practice. An assistant psychologist under the supervision of a chartered clinical psychologist conducted the assessments. Formal psychological reports were sent to the patients' general practitioners and hospital staff involved in the patients' care. The information obtained was summarised for patients and when the patients agreed, their relatives.			

Table 196: Clinical evidence profile: Control versus structured assessment

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Structured assessments	Control	Relative (95% CI)	Median IQR p value	Quality	Importance
General Health Questionnaire (follow-up 4 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	77	72	-	Control 21.0 13-34 Assessment 21.0 13-31 p=0.73[c]	MODERATE	CRITICAL
General Health Questionnaire (follow-up 8 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	71	77	-	Control 18.0 13-35 Assessment 18.5 13-35 p=0.59 [c]	MODERATE	CRITICAL
SF-36 physical component score (follow-up 4 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	72	77	-	Control 25.6 21-45 Assessment 27.1 20-47 p=0.55 [c]	MODERATE	CRITICAL
SF-36 physical component score (follow-up 8 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	71	77	-	Control 30.0 25-38 Assessment 32.1 25-42 p=0.55 [c]	MODERATE	CRITICAL
SF-36 mental health component (follow-up 4 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	72	77	-	Control 44.7 36-55 Assessment 44.7 35-57 p=0.55 [c]	MODERATE	CRITICAL
Sf-36 mental health component (follow-up 8 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	71	77	-	Control 47.3 36-57 Assessment 49.3	MODERATE	CRITICAL

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Structured assessments	Control	Relative (95% CI)	Median IQR p value	Quality	Importance
										33-58 p=0.76 [c]		
Overall Quality of Life (follow-up 4 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	72	77	-	Control 7.0 5-8 Assessment 6.0 5-7 p=0.15 [c]	MODERATE	CRITICAL
Overall Quality of Life (follow-up 8 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	71	72	-	Control 6.5 5-8 Assessment 6.0 4-7 p=0.04 in favour of control [c]	MODERATE	CRITICAL
Satisfaction with quality of life (follow-up 4 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	72	77	-	Control 4.0 4-5 Assessment 4.0 4-5 p=0.32 [c]	MODERATE	CRITICAL
Satisfaction with quality of life (follow-up 8 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	71	77	-	Control 5.0 4-8 Assessment 4.0 3-5 p=0.04 in favour of control [c]	MODERATE	CRITICAL
Extended activities of daily living index (follow-up 4 months; Better indicated by higher values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	72	77	-	Control 48.0 37-60 Assessment 43.0 37-60 p=0.23 [c]	MODERATE	CRITICAL
Extended activities daily living index (follow-up 8 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	v[b]	none	71	77	-	Control 47.5 37-59 Assessment 44.5	MODERATE	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Structured assessments	Control	Relative (95% CI)	Median IQR p value		
										26-61 p=0.21 [c]		
Everyday memory questionnaire (follow-up 4 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	72	77	-	Control 16.5 7-42 Assessment 18.5 5-31 p=0.69 [c]	MODERATE	CRITICAL
Everyday memory questionnaire (follow-up 8 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	71	77	-	Control 14.0 7-37 Assessment 15.0 5-31 p=0.76 [c]	MODERATE	CRITICAL
Dysexecutive syndrome questionnaire (follow-up 4 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	72	77	-	Control 17.0 9-32 Assessment 16.0 7-31 p=0.77 [c]	MODERATE	CRITICAL
Dysexecutive syndrome questionnaire (follow-up 8 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	no serious imprecision [b]	none	71	72	-	Control 16.5 9-32 Assessment 18.0 7-31 p=0.98 [c]	MODERATE	CRITICAL
Memory aids questionnaire (follow-up 4 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	72	77	-	Control 10.0 7-15 Assessment 11.0 7-14 p=0.92 [c]	MODERATE	CRITICAL
Memory aids questionnaire (Better indicated by lower valuesDK)[123]												
Lincoln 2002	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	imprecision not assessed [b]	none	71	77	-	Control 10.0 7-14 Assessment 9.0 6-15 p=0.80 [c]	MODERATE	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Structured assessments	Control	Relative (95% CI)	Median IQR p value		
Carer General Health Questionnaire (follow-up 4 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed[b]	none	Not reported	Not reported	-	Control 22.0 14-31 Assessment 24.0 16-35 p=0.35[c]	MODERATE	CRITICAL
Carer General Health Questionnaire (follow-up 8 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed[b]	none	Not reported	Not reported	-	Control 18.0 13-30 Assessment 18.5 13-32 p=0.59[c]	MODERATE	CRITICAL
Carer everyday memory questionnaire (follow-up 4 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed[b]	none	Not reported	Not reported	-	Control 14.0 3-35 Assessment 11.5 4-28 p=0.90[c]	MODERATE	CRITICAL
Carers everyday memory questionnaire (follow-up 8 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed[b]	none	Not reported	Not reported	-	Control 10.0 3-31 Assessment 10.0 3-25 p=0.88[c]	MODERATE	CRITICAL
Carer dysexecutive syndrome questionnaire (follow-up 4 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed[b]	none	Not reported	Not reported	-	Control 17.0 9-33 Assessment 11.5 7-31 p=0.80[c]	MODERATE	CRITICAL
Carer dysexecutive syndrome questionnaire (follow-up 8 months; Better indicated by lower values)[123]												
Lincoln 2002	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	imprecision not assessed[b]	none	Not reported	Not reported	-	Control 10.0 9-32 Assessment 10.0 7-28 p=0.72[c]	MODERATE	CRITICAL

[a] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation*

concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis.
[b] *Median (IQR), imprecision could not be assessed*
[c] *p values relate to three arms, intervention arm not reported here. Significant results are for control versus assessment comparison*

11.4 Economic evidence

Published literature

No relevant economic evaluations comparing structured assessment(s) with non-structured assessment(s) were identified.

See also the economic article selection flow chart in Appendix E.

11.5 Evidence statements

11.5.1 Clinical

Moderate quality evidence from one RCT comprising 148 participants showed a clinically important harm for assessment compared to control for overall quality of life and satisfaction with quality of life (8 months).

11.5.2 Economic

No relevant economic evaluations were identified.

11.6 Recommendations and link to evidence

Recommendations	**62. Determine how often the person with MS will need to be seen based on:** o their needs, and those of their family and carers **and** o the frequency of visits needed for different types of treatment (such as review of disease-modifiying therapies, rehabilitation and symptom management). **63. Ensure all people with MS have a comprehensive review of all aspects of their care once a year.** **64. Ensure the comprehensive review is carried out by healthcare professionals with expertise in MS and its complications. Involve different healthcare professionals with expertise in specific areas of the review if needed.** **65. Tailor the comprehensive review to the needs of the person with MS assessing:** MS symptoms o mobility and balance including falls o need for mobility aids including wheelchair assessment o use of arms and hands o muscle spasms and stiffness

	o tremor bladder (see Urinary incontinence in neurological disease NICE clinical guideline 148), bowel (see Faecal incontinence NICE clinical guideline 49) and sexual function o sensory symptoms and pain o speech and swallowing (see Nutrition support in adults NICE clinical guideline 32) o vision o cognitive symptoms o fatigue o depression (see Depression in adults with chronic physical health problems NICE clinical guideline 91) and anxiety (see Generalised anxiety disorder and panic disorder NICE clinical guideline 113) o sleep o respiratory function. MS disease course o relapses in last year. General health o weight o smoking, alcohol and recreational drugs o exercise o access to routine health screening and contraception o care of other chronic conditions. Social activity and participation o family and social circumstances o driving and access to transport o employment o access to daily activities and leisure. Care and carers o personal care needs o social care needs o access to adaptations and equipment at home. **66. Refer any issues identified during the review of the person with MS to members of the MS multidisciplinary team and other appropriate teams so that they can be managed.** **67. Ensure people with MS are offered a medication review in line with Medicines adherence (NICE clinical guideline 76).** **68. Ensure people with MS have their bone health regularly assessed and reviewed in line with Osteoporosis: assessing the risk of fragility fracture (NICE clinical guideline 146).** **69. Check people with MS and severely reduced mobility at every contact for areas at risk of pressure ulcers (see the Pressure**

	ulcers NICE clinical guideline 179). **70. Discuss the care provided by carers and care workers as part of the person's care plan. Ensure carers know about their right to access a local authority carer's assessment and how to apply for one.** **71. Refer people with MS to palliative care services for symptom control and for end of life care when appropriate.**
Relative values of different outcomes	A range of outcomes were thought to be relevant to measuring the success of structured assessments. Quality of life (generic scales such as EQ-5D or the MS specific scales such as Leeds MS Quality of Life scale) and the impact on carers were important outcomes. Scales of function, disability and neuropsychology are also relevant. Other outcomes included were severity of multiple sclerosis, the number of hospitalisations, number of outpatient appointments, frequency of access to or use of services and number of relapses.
Trade-off between clinical benefits and harms	The assessments could ensure that relevant topics and systems are discussed with patients. The GDG thought that they work best as a framework rather than as an exhaustive list. A number of potential harms were also identified. A structured assessment could make a consultation constrained and rigid and not responsive to the patient's needs. The issues important may also differ according to type and severity of MS. Immobile patients need to be assessed for pressure sores and end-stage patients for palliative care needs, but these are not usually relevant to a newly diagnosed patient. Related to this, a structured list may worry patients that they are expected to develop all of the problems in the list. There was also concern that despite a large number of different scales in multiple sclerosis, very few were validated.
Economic considerations	No economic evidence on structured assessment was found. The GDG agreed that it would be good practice to provide guidance on topics that may need review via a structured assessment for people with MS and reported that any referrals as a result of this assessment would be for evidence based interventions or treatments. The GDG considered that they were informing the content of the consultation and were not recommending additional consultations.
Quality of evidence	Only one randomised trial of structured assessments was identified. This was despite extending the search criteria to look for the use of structural assessments in any chronic neurological condition. The study looked only at a structured 3-hour cognitive assessment in people with multiple sclerosis (predominantly relapsing-remitting and secondary progressive) and the results were communicated to health professionals involved in the patients care. This was compared with a less detailed screening assessment and the results were not shared with any health professionals. The study included 24 outcomes. No outcomes were better with a structured cognitive assessment. For two outcomes, quality of life after 8 months and satisfaction with quality of life at 8 months, screening was associated with a clinically important benefit compared to structured assessment (moderate quality). The GDG considered that this may be because patients were not offered suitable follow-up or treatment for any problems identified.
Other considerations	Stakeholders had identified appropriate assessments as an important topic. The GDG considered that regular review of individual symptoms and their management is required and this shoud be decided on an individual basis. Many people receiving DMTs and having other symptom management will be seen frequently.

	The GDG considered that outside the management of particular symptoms and problems people with MS would benefit from having a comprehensive review of all aspects of their care at least annually. People with one chronic medical problem often do not receive attention to other medical conditions and there are aspects of care such as bone health and skin care that may be neglected if there is a focus on specific disease processes only. Some issues associated with MS such as cognitive issues and urinary smptoms may not be recognised by the MS patient as being related to MS. Some aspects of this care are particularly important such as attention to weight when people are less mobile, and the encouragement to take part in exercise for its general health benefits as well as beneficial effects on muscle strength. The GDG were aware that people may use of recreational drugs such as cannabis for relief of MS symptoms and healthcare professionals may need to be aware of this. They included MS symptoms that may not be easily discussed such as sexual function and cognitive symptoms to alert practitioners to enquire about these. The GDG structured their list under headings of MS symptoms, general health, social activity and participation, and care and carers. The GDG considered that their list of topics could be used as an aide-memoire and did not wish to see a rigid list used which did not allow the professional to be responsive to individual patients and their needs. They considered that many professionals who are in contact with people with MS would cover the topics listed in a general review. The GDG considered that templates are designed for use for chronic disease conditions such as diabetes and MS care could be approached in the same way. The GDG used their collective experience to draw up a list of relevant topics for a general structured assessment in a patient with multiple sclerosis. A number of lists used in rehabilitation and by patient groups were also examined. It is likely that the emphasis in the review may change over time and that different healthcare professionals may either carry out the review or need to be involved in care of the patient. While healthcare professionals will not be able to carry out a social care assessment they need to be alert to social care needs and refer people for a social care assessment. The impact on carers and carers' needs is an important topic even if not relevant to all patients. In all these areas the GDG acknowledged that healthcare professionals need to act on their findings at review. People with MS may also benefit from access to the expertise of palliative care services for symptoms control. Mechanisms will need to be in palce to ensure communication and referral as a result of the review is appropriately carried out. In other disease areas these mechanisms have been supported by QoF, Quality Standards or enchanced services contracts.

12 Treating acute relapse of MS with steroids

12.1 Introduction

Steroids are administered to patients experiencing acute relapses of relapsing-remitting MS with the aim of reducing inflammation and speeding up recovery. However, even short courses of steroids are associated with adverse effects and these need to be balanced against the potential benefits. Steroids can be delivered orally or intravenously and there has been uncertainty as to whether intravenous therapy is more clinically or cost effective.

12.2 Review question: What is the clinical evidence of pharmacological management of acute relapse with steroids compared to placebo? If steroids are more effective than placebo, is there a difference in efficacy between IV and oral steroids?there a difference in efficacy and cost-effectiveness between steroids given at inpatients, outpatients (include day case), community or home?

For full details see review protocol in Appendix C. See Table 197 for a summary of the PICO characteristics.

Table 197: PICO characteristics of review question

Population	• Adults with MS
Intervention	• Methylprednisolone, prednisolone, dexamethasone • Methylprednisolone given intravenously • Steroids given in an inpatient setting
Comparison [comparisons only relate to the intervention labelled with the corresponding letter – for example, comparison (a) only relates to intervention (a)]	• Placebo • Oral adrenocorticotrophic hormone or methylprednisolone • IV steroids given in outpatients OR community OR home
Outcomes	• Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale • Relapse outcomes • Relapse severity rating (change in EDSS >2) • Relapse duration • Duration of relapse-free period post treatment • Functional scales that quantify level of disability, such as the Expanded Disability Status Scale (EDSS), the Multiple Sclerosis Functional Composite (MSFC), the Cambridge Multiple Sclerosis Basic Score (CAMBS), the Functional Assessment of Multiple Sclerosis (FAMS), the National Fatigue Index (NFI), the MS walking scale (MSWS-12). • Cognitive functions and physical symptoms. • Psychological symptoms assessed by validated and disease-specific scales,

	questionnaire or similar instruments. • Adverse events • Any adverse effects of treatment, • Adverse events leading to withdrawal • Steroids vs placebo: abdominal pain (including gastritis and dyspepsia), mood disturbance, avascular necrosis, sepsis • IV versus oral: abdominal pain, rash, mood disturbance, sepsis • Setting: anaphylactic reaction, abdominal pain, mood disturbance, sepsis
Exclusion	• Children younger than 18 years
The review strategy	• Systematic reviews • RCTs • Include cross-over trials • Include dosing studies

12.3 Clinical evidence

a) Steroids versus placebo
Six RCTs were identified[56; 60; 150; 148; 199; 219]. In addition one dosing study was identified [173]. A full report of the study by Rose was excluded as it was not possible to extract the data[199].

b) Intravenous (IV) versus oral steroids
A Cochrane systematic review was included[33]

c) Home versus outpatient setting
one RCT was identified[38].

A summary of study characteristics for all studies is presented in Table 198 below. Evidence from these are summarised in the clinical GRADE evidence profile (Table 199) and the narrative review section. See also the study selection flow chart in Appendix D, forest plots in Appendix I, study evidence tables in Appendix G and exclusion list in Appendix J.

Table 198: Summary of studies included in the review

Study	Population	Intervention	Comparison
Steroids versus placebo			
Durelli 1986[57]	N=23 Patients with clinically definite MS of relapsing-remitting form. All patients had had at least two bouts in the preceding 3 yrs and were in an exacerbation for less than 8 wks and more than 10 days without evidence of spontaneous improvement	Steroids N=13 Parenteral methylprednisolone (MP). 15 mg/kg/d IV days 1-3, 10 mg/kg/d IV days 4-6, 5 mg/kg/d IV days 7-9, 2.5 mg/kg/d IV days 10-12, 1 mg/kg/d IV days 13-15 Followed by Oral prednisone	Placebo N=10 0-15 days placebo Followed by oral prednisone as for intervention

Study	Population	Intervention	Comparison
		100 mg/d slowly tapered over 120 days. Antacids and potassium chloride were given to all patients	
Filipovic 1997[60]	Relapsing-remitting (RR) form of the MS and were in acute exacerbation, or had secondary progressive (SP) form of disease with either progressive worsening of neurologic disability during the last six months or acute superimposed exacerbation Exclusion criteria: received anticholinergic or antidepressive medication at the time of investigation, had a history of corticosteroid or other immunosuppressive medication in the last 6 mths, had CNS diseases, had hearing impairment Inclusion criteria: Mini Mental State score of 27 or higher	Steroids N=19 1000 mg methylprednisolone in a single dose per day for 5 days	Placebo N=21 Saline for 5 days
Milligan 1987[150]	N=22 (patients with chronic progressive disease not reported) Relapse was defined as the occurrence of one or more new, or a worsening of existing symptoms of less than 8 wks but more than 24 hrs which had not improved spontaneously at the time of entry into the trial.	Steroids IV methylprednisolone 500 mgs once daily over 5 days	Placebo
Sellebjerg 1998[219]	N=51 N=1 drop-outs in the steroid gp Inclusion criteria: aged 18 to 59 yrs with an attack of clinically definite, laboratory-supported definite or probably MS with a duration of no more than 4 weeks. All patients had relapsing-remitting MS. An	Steroid N=26 Total dose of oral methylprednisolone 3676 mg (500 mg once a day for 5 days followed by a tapering period during which 400, 300, 200, 100, 64, 48, 32, 16, 8 mg were administered on the 10 following	Placebo

Study	Population	Intervention	Comparison
	attack was defined as occurrence of new symptoms or recurrence of previously existing symptoms in the absence of systemic infection and with a duration of more than 24 hrs	days)	
Oliveri 1998[173]	Patients with clinically definite relapsing-remitting MS. Referred to the MS centre as a result of a clinical relapse with a loss of at least 1.0 point in their EDSS score. Inclusion criteria: Only patients who were seen within 2 wks of onset of the relapse, with at least one enhancing lesion on MRI and not in an improving phase. Relapse was defined as either the onset of new symptoms and signs, or deterioration in the existing symptoms and signs of at least 24 hrs in duration without concomitant fever.	High dose N=14 2 g/d iv methylprednisolone for five days	Low dose N=15 0.5 g/d iv methylprednisolone for five days
IV versus oral			
Burton 2012[33] Included studies: Alam 1993[5] Barnes 1997[16,17,221] plus unpublished data Martinelli 2008[131,132] plus unpublished data Ramo-tello 2011[190] plus unpublished data	Alam – patients with clinically definite MS with a relapse ≤ 4 weeks in duration Barnes – patients with clinically definite MS presenting with relapse ≤ 4 weeks in duration severe enough to merit steroid treatment Martinelli – patients with clinically definite MS presenting with relapse ≤ 2 weeks in duration that were moderate to severe in intensity with ≥ 1 gadolinium enhancing lesion on MRI done at the time of relapse Ramo-tello – patients presenting with a MS relapse deemed to require	Alam – 500 mg/d of intravenous methylprednisolone for three days. Placebo oral agent Barnes – 1 g/d intravenous methylprednisolone for three days. Placebo oral agent Martinelli – 1 g/d intravenous methylprednisolone for 5 days Ramo-tello 1000 mg intravenous methylprednisolone with placebo oral agent for three days	Alam – 500 mg/day oral methylprednisolone for three days. Placebo iv agent Barnes – oral medication was 48g/d x 7 days, then 24 mg/d x 7 days, then 12 mg/d for 7 days. Placebo iv agent Martinelli 2x500 mg oral methylprednisolone BID for 5 days Ramo-tello 1250 mg of oral methylprednisolone with placebo iv agent for 3 days

Study	Population	Intervention	Comparison
	steroid therapy		
Home versus outpatients			
Chataway 2006[38]	N=138 Patients older than 18 yrs, had clinically definite MS and had a sustained definite relapse of more than 24 hrs but less than 4 wks in duration. Patients were excluded if their relapse was minor, such that the clinician would not prescribe steroids; their relapse was severe enough to require hospitalisation; there was evidence of intercurrent infection; they had a history of adverse side-effects after previous steroid use; or they had previously participated in this trial. In the initial phase of the trial, patients were also excluded if they had never received intravenous steroids before because of worries about safety issues for those treated at home. However, this requirement was dropped after n=22 patients	Outpatients steroids N=69 1g methylprednisolone over 1 hr, daily, for 3 days Dedicated suite	Home steroids N=69 Patients left the relapse clinic with a 3-day supply of intravenous methylprednisolone. Arrangements made for delivery team to visit patient during next 3 days. Delivery team consisted of generally trained nurses who were experienced in at-home chemotherapy treatment and who had received an educational programme on MS. Treatment was provided by the specialist multiple sclerosis nursing team at the hospital

GRADE evidence

Table 199: Clinical evidence profile: Steroids versus placebo

Quality assessment							No of patients		Effect			
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Steroids versus placebo**	**Control**	**Relative (95% CI)**	**Absolute Mean/sd/p value**	**Quality**	**Importance**
Health-related quality of life												
0	No evidence available					-	-	-	-	-		CRITICAL
Positive response to treatment (follow-up 3 weeks)												
Miller 1961	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	11/22 (50%)	4/18 (22.2%)	RR 2.25 (0.86 to 5.88)	278 more per 1000 (from 31 fewer to 1000 more)	VERY LOW	CRITICAL
Subjective improvement (follow-up 15-56 days)												
Durelli 1986 Sellebjerg 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	35/39 (89.7%)	60%	RR 1.48 (1.11 to 1.98)	288 more per 1000 (from 66 more to 588 more)	VERY LOW	CRITICAL
Minimal/no disability (unclear follow-up)												
Rose 1968	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	10/70 (14.3%)	5/65 (7.7%)	RR 1.86 (0.67 to 5.15)	66 more per 1000 (from 25 fewer to 319 more)	VERY LOW	CRITICAL
Restricted to bed/wheelchair (unclear follow-up)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Steroids versus placebo**	**Control**	**Relative (95% CI)**	**Absolute Mean/sd/p value**		
Rose 1968	randomised trials	very serious[a]	no serious risk of bias	no serious indirectness	serious[b]	none	9/70 (12.9%)	17/65 (26.2%)	RR 0.49 (0.24 to 1.02)	133 fewer per 1000 (from 199 fewer to 5 more)	VERY LOW	CRITICAL
Scripps Neurological Rating Scale (follow up 1 week)												
Sellebjerg 1998	randomised trials	very serious[a]	no serious risk of bias	no serious indirectness	no serious imprecision[c]	none	26 Median change (IQR) 5 (2-8)	25 1 (-1 to 3)	P=0.006		LOW	CRITICAL
Scripps Neurological Rating Scale (follow up 8 weeks)												
Sellebjerg 1998	randomised trials	very serious[a]	no serious risk of bias	no serious indirectness	no serious imprecision[c]	none	26 Median change (IQR) 11 (3-15)	25 0 (-5 to 6)	P=0.0007		LOW	CRITICAL
Clinical improvement (EDSS score change of 1.0) (one week to 15 days)												
Durelli 1986 Milligan 1987	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	20/26 (76.9%)	25.6%	RR 2.98 (1.39 to 6.38)	507 more per 1000 (from 100 more to 1000 more)	LOW	CRITICAL
Clinical improvement (EDSS score change of 1.0) (four weeks)												
Milligan 1987	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	10/13 (76.9%)	2/8 (25%)	RR 3.08 (0.89 to 10.6)	520 more per 1000 (from 28 fewer to 1000 more)	VERY LOW	CRITICAL
Improvement in EDSS score (follow-up 5 days; Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Steroids versus placebo**	**Control**	**Relative (95% CI)**	**Absolute Mean/sd/p value**		
Filipovic 1997	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision[c]	none	19	21	-	Steroid mean 1.0 (SD 0.5) Placebo 0 (0)	LOW	CRITICAL
Improvement in EDSS (1 week; Better indicated by higher values)												
Sellebjerg 1998	randomised trials	very serious[a]	no serious risk of bias	no serious indirectness	no serious imprecision[c]	none	26 Median change (IQR) 0.5 (0 to 1.0)	25 0 (0 to 0)		P=0.02	LOW	CRITICAL
Improvement in EDSS (8 weeks; Better indicated by higher values)												
Sellebjerg 1998	randomised trials	very serious[a]	no serious risk of bias	no serious indirectness	no serious imprecision[c]	none	26 Median change (IQR) 1.0 (-0.5 to 1.5)	25 0 (-0.5 to 1.0)		P=0.01	LOW	CRITICAL
Relapse durations (days) (Better indicated by lower values)												
Durelli 1986	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	13	10	-	MD 12.7 lower (19.98 to 5.42 lower)	LOW	CRITICAL
Cognitive functions												
0	No evidence available					none	-	-	-	-		IMPORTANT
Psychological symptoms												
0	No evidence available					none	-	-	-	-		IMPORTANT

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Steroids versus placebo	Control	Relative (95% CI)	Absolute Mean/sd/p value		
Gastrointestinal symptoms (follow-up 8 weeks)												
Sellebjerg 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	10/26 (38.5%)	2/25 (8%)	RR 4.81 (1.17 to 19.8)	305 more per 1000 (from 14 more to 1000 more)	VERY LOW	IMPORTANT
Dysphoria (follow-up 8 weeks)												
Sellebjerg 1998	randomised trials	very serious[a]	no serious inconsistency	no serious indirectness	very serious[b]	none	6/26 (23.1%)	2/25 (8%)	RR 2.88 (0.64 to 12.97)	150 more per 1000 (from 29 fewer to 958 more)	VERY LOW	IMPORTANT

[a] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis.
[b] Outcomes were downgraded by one increment if either the lower MID or the upper MID were crossed by one or both of the 95% confidence intervals. Outcomes were downgraded by two increments if both MIDs were simultaneously crossed. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation (either side of the null line) for continuous variables.
[c] Imprecision could not be assessed because the MD and 95%CI could not be calculated

Table 200: Clinical evidence profile: Oral versus iv steroids

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Oral	Iv steroids	Relative (95% CI)	Absolute		
Proportion of patients with improvement on EDSS after steroid treatment at 4 weeks												
Alam	randomis	no	no serious	no serious	very	none	54/	83.5%	OR 0.6	83 fewer per 1000	LOW	CRITICAL

Quality assessment							No of patients		Effect			
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Oral**	**Iv steroids**	**Relative (95% CI)**	**Absolute**	**Quality**	**Importance**
1993 Barnes 1997 Martinelli 2009 Ramo 2012 unpublished data	ed trials	serious risk of bias	inconsistency	indirectness	serious[a]		99 (54.5%)		(0.28 to 1.26)	(from 249 fewer to 29 more)		
Change in Ambulation Index at week 1 after treatment with oral vs. intravenous steroids (Better indicated by lower values)												
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	42	38	-	MD 0 higher (0.39 lower to 0.39 higher)	HIGH	CRITICAL
Change in Ambulation Index at week 4 after treatment with oral vs. intravenous steroids (Better indicated by lower values)												
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[a]	none	42	38	-	MD 0.4 higher (0.11 lower to 0.91 higher)	MODERATE	CRITICAL
Relapse rate 6 months after treatment with oral vs. intravenous steroids (Better indicated by lower values)												
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[a]	none	42	38	-	MD 0.21 higher (0.06 lower to 0.48 higher)	MODERATE	CRITICAL
Relapse rate at one year after treatment with oral vs. intravenous steroids (Better indicated by lower values)												
Barnes 1997	randomised trials	no serious	no serious inconsistency	no serious indirectness	serious[a]	none	42	38	-	MD 0.34 higher (0.13 lower to 0.81	MODERATE	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Oral	Iv steroids	Relative (95% CI)	Absolute		
		risk of bias								higher)		
Relapse rate at years 1-2 after treatment with oral vs. intravenous steroids (Better indicated by lower values)												
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[a]	none	42	38	-	MD 0.21 higher (0.16 lower to 0.58 higher)	MODERATE	CRITICAL
Relapse rate at two years after treatment with oral vs. intravenous steroids (Better indicated by lower values)												
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[a]	none	42	38	-	MD 0.28 higher (0.08 lower to 0.64 higher)	MODERATE	CRITICAL
Proportion relapse free at 2 years after treatment with oral vs. intravenous steroids												
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[a]	none	10/42 (23.8%)	11/38 (28.9%)	OR 0.77 (0.28 to 2.08)	51 fewer per 1000 (from 187 fewer to 169 more)	LOW	
Mean number of days to next relapse after treatment with oral vs. intravenous steroids (Better indicated by lower values)[16,17];[221] plus unpublished data												
	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[a]	none	42	38	-	MD 47 lower (150.53 lower to 56.53 higher)	MODERATE	CRITICAL
Mean change in EDSS at first relapse within 2 year period after treatment with oral vs. intravenous steroids (Better indicated by lower values)												
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	42	38	-	MD 0.03 higher (0.47 lower to 0.53 higher)	HIGH	CRITICAL
Proportion hospitalized at week 1 after treatment with oral vs. intravenous steroids												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Oral	Iv steroids	Relative (95% CI)	Absolute		
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[a]	none	11/42 (26.2%)	10/38 (26.3%)	OR 0.99 (0.37 to 2.69)	2 fewer per 1000 (from 146 fewer to 227 more)	MODERATE	CRITICAL
Proportion hospitalized at week 4 after treatment with oral vs. intravenous steroids												
Barnes 1997	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[a]	none	2/42 (4.8%)	1/38 (2.6%)	OR 1.85 (0.16 to 21.26)	21 more per 1000 (from 22 fewer to 339 more)	LOW	CRITICAL
Proportion with rash												
Martineeli 2009 Ramo 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[a]	none	7/45 (15.6%)	9/44 (20.5%)	OR 0.72 (0.23 to 2.26)	48 fewer per 1000 (from 149 fewer to 163 more)	LOW	IMPORTANT
Proportion with mood disturbance												
Ramo 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[a]	none	13/24 (54.2%)	7/24 (29.2%)	OR 2.87 (0.87 to 9.45)	250 more per 1000 (from 28 fewer to 504 more)	MODERATE	IMPORTANT

[a] *Outcomes were downgraded by one increment if either the lower MID or the upper MID were crossed by one or both of the 95% confidence intervals. Outcomes were downgraded by two increments if both MIDs were simultaneously crossed. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation (either side of the null line) for continuous variables.*

Ramo-Tello[191] presented results for a) a parametric analysis using the per protocol population; b) a parametric analysis using the ITT population; and c) a non-parametric analysis, comparing EDSS final scores (mean, mean and median, respectively, each with 95% CI) scores at weeks 1, 4 and 28, noting no significant differences between oral and IV steroid groups. The authors therefore accepted the hypothesis of non-inferiority of oral versus IV steroids. As

EDSS is an ordinal scale, the parametric analysis was inappropriate and so GRADE tables have not been produced for this outcome. An improvement in the EDSS score was observed in both treatment groups at 1 and 4 weeks vs. baseline (p<0.001 for both groups at both time points).

Table 201: Clinical evidence profile: outpatients versus home

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Outpatients	Home	Relative (95% CI)	Absolute	Quality	Importance
MSIS-29 physical impact (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	Serious[a]	no serious inconsistency	no serious indirectness	Serious[b]	none	60	62	-	MD 4.5 lower (12.28 lower to 3.28 higher)	LOW	CRITICAL
MSIS-29 psychological impact (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	62	-	MD 0.2 higher (8.45 lower to 8.85 higher)	MODERATE	CRITICAL
MSWS-12 walking score (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	60	62	-	MD 2.6 lower (11.2 lower to 6 higher)	LOW	CRITICAL
SF-36 role emotional (follow-up 6 weeks; Better indicated by lower values)												
Cha	rando	serious	no serious	no serious	serious[b]	none	60	62	-	MD 9.3	LOW	CRITI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Outpatients	Home	Relative (95% CI)	Absolute		
taway 2011	mised trials	[a]	inconsistency	indirectness						higher (7.69 lower to 26.29 higher)		CAL
SF-36 role physical (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	60	62	-	MD 9.7 higher (2.67 lower to 22.07 higher)	LOW	CRITICAL
SF-36 pain (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	62	-	MD 3.4 higher (5.19 lower to 11.99 higher)	MODERATE	CRITICAL
SF-36 energy and vitality (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	62	-	MD 1 higher (5.31 lower to 7.31 higher)	MODERATE	CRITICAL
SF-36 general health perceptions (follow-up 6 weeks; Better indicated by lower values)												
Chataway	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	60	62	-	MD 3.7 lower (9.77 lower to	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Outpatients	Home	Relative (95% CI)	Absolute		
2011										2.37 higher)		
SF-36 social functioning (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	serious[b]	none	60	62	-	MD 5.2 lower (15.26 lower to 4.86 higher)	LOW	CRITICAL
SF-36 physical functioning (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	62	-	MD 2.4 higher (4.68 lower to 9.48 higher)	MODERATE	CRITICAL
SF-36 mental health (follow-up 6 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	62	-	MD 1.6 higher (4.72 lower to 7.92 higher)	MODERATE	CRITICAL
MSRMS access to care (follow-up 1 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious[a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	66 Median 11.1 (IQR 4.2 to 27.8)	68 11.1 (5.6 to 22.2)	-	Difference in means -1.0 (-7.7 to 5.8)	MODERATE	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Outpatients**	**Home**	**Relative (95% CI)**	**Absolute**		
										P=0.868		
MSRMS coordination of care (follow-up 1 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	66 Median 12.1 (IQR 3.0 to 18.6)	68 4.5 (3.0 to 11.4)	-	Difference in means 3.8 (0.65 to 7.1) p=0.024	MODERATE	CRITICAL
MSRMS Information (follow-up 1 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	66 Median 28.6 (IQR 9.5 to 45.2)	68 28.6 (14.3 to 47.6)	-	Difference in means -2.9 (-10.2 to 4.4) p=0.367	MODERATE	CRITICAL
MSRMS Interpersonal care (follow-up 1 weeks; Better indicated by lower values)												
Chataway 2011	randomised trials	serious [a]	no serious inconsistency	no serious indirectness	no serious imprecision	none	66 Median 5.6 (IQR 0.0 to 13.0)	68 7.4 (1.9 to 16.7)	-	Difference in means -1.9 (-5.3 to 1.5) p=0.130	MODERATE	CRITICAL

[a] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis.*
[b] *Outcomes were downgraded by one increment if either the lower MID or the upper MID were crossed by one or both of the 95% confidence intervals. Outcomes were downgraded by two increments if both MIDs were simultaneously crossed. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation (either side of the null line) for continuous variables.*

Narrative review

Using a parametric analysis, Olivieri [173] compared EDSS mean scores at days 7, 15, 30 and 60, noting no significant differences between high and low dose IV steroid groups. As EDSS is an ordinal scale, the parametric analysis was inappropriate and so GRADE tables have not been produced for this outcome.

12.4 Economic evidence

Published literature

No relevant economic evaluations comparing steroids with placebo or comparing intravenous with oral steroids were identified.

One study was included comparing outpatient with home administration of steroids.[38] This is summarised in the economic evidence profile below (Table 20). See also the economic article selection flow chart in Appendix E and study evidence tables in Appendix H.

One study that met the inclusion criteria was selectively excluded[197] as it was a cost analysis conducted with the assumption that IV methyl prednisolone has the same efficacy in all settings; in addition, a more recent and better quality UK economic paper was available[38]. This is summarised in Appendix K, with reasons for exclusion given.

Table 202: Economic evidence profile: Outpatient versus home administration of intravenous steroids

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Chataway 2006 (UK NHS)[38]	Partially applicable (a)	Minor limitations	Outpatient administration of 1g methylprednisolone over 1 hour, daily, for 3 days in a dedicated suite. Home administration of intravenous methylprednisolone. Patients left hospital with a 3-day supply. Delivery team consisted of generally trained nurses who were experienced in at-home chemotherapy treatment and who had received an educational programme on MS.	£145 (b)	Similar efficacy of both interventions (c)	Home administration is cheaper with possibly no difference in health outcomes	Univariate sensitivity analysis was conducted which showed that if the charge for health at home increased by 51% or more or if NHS salaries were reduced by 50%, it would become cheaper to treat patients in the hospital. If direct non-medical costs (transport, childcare costs) were included, outpatient treatment would cost £205 more than home treatment.

(d) Health outcomes not presented as QALYS.
(e) 2006 UK pounds. Costs included direct medical costs (salaries, equipment, drugs, hospital overheads and investigations).
(f) See clinical evidence review- section 12.3 and evidence Appendix G.

While in the study by Chataway 2006 the cost of home administration was estimated as a fixed price of £354 for three days, the GDG expressed some concerns that this is an underestimate and that some trusts would not be able to deliver the treatment at home for the cost described in the study.

We calculated the cost of home administration based on the resources required to deliver this service in practice. Each treatment course would equate to the cost of three and half hours of home visiting by a community nurse (band 6) per hour of home visiting. Form the PSSRU, the cost per hour of home visiting is £61 including travel;[46] therefore the total cost of home administration of iv steroids would be £61 * 3.5 = £213.

Unit costs

The unit costs for iv and oral methyprednisolone are provided in Appendix M.

12.5 Evidence statements

12.5.1 Clinical

Steroids versus placebo

Very low quality evidence from 1 RCT comprising 40 participants showed that steroids were clinically effective compared to placebo in terms of positive response to treatment, with serious imprecision.

Very low quality evidence from 1 RCT comprising 56 participants showed that steroids were clinically effective compared to placebo in terms of subjective improvement, with serious imprecision.

Very low quality evidence from 1 RCT comprising 135 participants showed there was no difference in clinical effectiveness between steroids and placebo in terms of minimal/no disability, with very serious imprecision.

Very low quality evidence from 1 RCT comprising 135 participants showed that steroids were clinically effective compared to placebo in terms of proportion restricted to bed/wheelchair, with serious imprecision.

Low quality evidence from 1 RCT comprising 51 participants showed that steroids were clinically effective compared to placebo in terms of clinical improvement (EDSS score change of 1) at one week to 15 days, with no serious imprecision.

Very low quality evidence from 1 RCT comprising 21 participants showed that steroids were clinically effective compared to placebo in terms on clinical improvement (EDSS score change of 1) at four weeks, with serious imprecision.

Low quality evidence from 1 RCT comprising 23 participants showed that steroids were clinically effective compared to placebo in terms of relapse duration, with no imprecision.

Very low quality evidence from 1 RCT comprising 51 participants showed that steroids were clinically harmful compared to placebo in terms of gastrointestinal symptoms, with serious imprecision.

Very low quality evidence from 1 RCT comprising 51 participants showed that steroids were clinically harmful compared to placebo in terms of dysphoria, with very serious imprecision.

High dose versus low dose

Very low quality evidence from 1 RCT comprising 29 participants showed that there was no difference in clinical effectiveness between high dose and low steroids in terms of EDSS score at day 7, 15, 30 and 60, with very serious imprecision.

Oral versus IV steroids

Low quality evidence from 4 RCTs comprising 200 participants showed that there was no difference in clinical effectiveness between iv and oral steroids in terms of proportion of patients with improvement in EDSS, with very serious imprecision.

Low quality evidence from 1 RCT comprising on 48 participants showed that there was no difference in clinical effectiveness between iv and oral steroids in terms of numbers of relapses, with very serious imprecision.

Low to high quality evidence from 1 RCT comprising 80 participants showed that there was no difference in clinical effectiveness between iv and oral steroids in terms of change in ambulation index, relapse rate, proportion relapse free, mean change in EDSS, proportion hospitalise, proportion with rash.

Moderate quality evidence from 1 RCT comprising 80 participants showed that iv steroids were clinically effective compared to oral steroids in terms of mean number of days to next relapse after treatment, with serious imprecision.

Moderate quality evidence from 1 RCT comprising 80 participants showed that iv steroids were clinically less likely to cause mood disturbance compared to oral steroids, with serious imprecision.

Outpatients versus home

Low to moderate quality evidence from 1 RCT comprising 122 participants showed that there was no difference in clinical effectiveness between outpatient and home steroids in terms of MSIS-29 physical, MSIS-29 psychological, MSIS walking, SF -36 emotional, SF-36 role physical, SF-36 energy, SF-36 social, SF-36 physical, SF-36 mental or EDSS, with no to serious imprecision.

Moderate quality evidence from 1 RCT comprising 122 participants showed that home steroids were clinically effective compared to outpatient steroids in terms of SF-36 pain, with no serious imprecision.

12.5.2 Economic

One cost–consequence analysis found that that home administration was less costly and similarly effective than outpatient administration of intravenous steroids for treating acute relapse in people with MS. This analysis was assessed as partially applicable with minor limitations.

12.6 Recommendations and link to evidence

Recommendations	<u>Treating a relapse</u> **72. Develop local guidance and pathways for timely treatment of relapses of MS. Ensure follow-up is included in the guidance and pathway.** **73. Non-specialists should discuss a person's diagnosis of relapse**

	and whether to offer steroids with a healthcare professional with expertise in MS because not all relapses need treating with steroids. **74. Offer treatment for relapse of MS with oral methylprednisolone 0.5 g daily for 5 days.** **75. Consider intravenous methylprednisolone 1 g daily for 3-5 days as an alternative for people with MS:** o in whom oral steroids have failed or not been tolerated **or** o who need admitting to hospital for a severe relapse or monitoring of medical or psychological conditions such as diabetes or depression. **76. Do not prescribe steroids at lower doses than methylprednisolone 0.5 g daily for 5 days to treat an acute relapse of MS.** **77. Do not give people with MS a supply of steroids to self-administer at home for future relapses.**
Relative values of different outcomes	Quality of life was the most important outcome, but was not addressed in the steroids versus placebo or oral steroids versus IV steroids studies. There were also few patient reported outcome measures. Relapse duration was regarded as an important outcome, as an important aim of steroid therapy is to quicken recovery from relapses. Improved functional recovery (ie, EDSS, Scripps neurological rating scale) was also regarded as important as another aim of steroid therapy is to optimise full functional recovery.
Trade off between clinical benefits and harms	There were few relevant clinical adverse effects reported for IV or oral steroids. Oral steroids may present a significant risk of gastrointestinal symptoms, but this potential harm was not evaluated for IV steroids. There were no other clear harms reported.
Economic considerations	One cost–consequence analysis comparing outpatient with home administration of IV steroids was identified. The results found that home administration costs less than hospital administration with no difference in health outcomes. In addition, the unit cost of oral and IV methylprednisolone were presented to the GDG. The costs were similar for oral and IV methylprednisolone, £60 and £52 respectively, but when home administration of IV steroids was included, the cost difference between IV and oral (as a minimum) was £294 per patient. The GDG agreed that this is a likely underestimated administration cost, although there is no national figure to compare this to as the costing will depend on local arrangements. The GDG considered these costs and agreed that in many cases; in particular where no set-up for monitoring is available, offering an IV steroids service is unlikely to be cost-effective. However, the GDG agreed that it is reasonable to recommend that IV should be strongly considered in people with severe relapses or who have already taken oral steroids with no success or who need admitting for monitoring of medical or psychological conditions.
Quality of evidence	Outcomes relating to the steroids versus placebo question were graded as low

	or very low, largely because of risk of bias arising from a lack of allocation concealment and inadequate blinding. Despite much of the steroid versus placebo evidence being underpowered, clear effects were still seen in favour of steroids, many of these effects being clinically important. One small study analysed high and low dose (2g vs 0.5g) methylprednisolone and found no difference in outcome. The economic evidence was assessed as partially applicable with minor limitations
Other considerations	The GDG noted that some evidence for steroid use comes from older trials that had used ACTH but that ACTH is no longer used as a treatment option. The GDG considered that steroids are the common accepted treatment for relapse and that delivery is dependent on service organisation. The GDG noted that steroids appeared to reduce relapse duration by almost 13 days. This is corroborated by studies looking at neurological improvement 1-8 weeks afterwards (subjectively, using EDSS, or using Neurological Rating Scale). The available studies used different doses of steroids and there was no clear evidence on the most effective dose of steroids to use. Trial doses were typically between 500 mg and 1 g/day for iv steroids and 500 mg/day for oral steroids. Specialist opinion was that the standard regimen is 1g intravenous methylprednisolone for 3 days or 500mg oral methylprednisolone for 5 days, regardless of patient weight. The bio-availability of oral steroids is considered to be 70-80% that of intravenous doses. The GDG were aware that in Europe higher doses of steroids are used more routinely. The GDG considered that there was some evidence in favour of IV steroids. The GDG were aware of at least two ongoing trials of IV versus oral steroids – OMEGA trial (doses are oral 1400 mg/qd/day and iv 1000 mg/qd/day) and Copousep (oral and iv 1 g/day). On current evidence however the GDG considered that oral steroids were appropriate unless the patient was having a severe relapses or when oral steroids have failed or not been tolerated. Patients who are being admitted because of monitoring needs could also be given intravenous steroids. The GDG were also aware that there can be difficulty and therefore initial delay in obtaining and administering IV methylprednisolone. Since the GDG agreed to recommend oral methylprednisolone as the first option for treatment, this can be given to patients at home. This may be helpful for patients who are at some distance form specialist care. Relapse can be difficult to diagnose and steroids have side effects so the GDG considered there was no place for patients to be have steroid doses at home in case they needed them for a relapse. They also considered it important to stress that people suffering from MS relapse needed adequate steroid treatment and not lower doses associated with other conditions. The group thought it would be reasonable to consider primary care management of relapses, especially if hospital access is difficult because of, for example, geographical factors. However it is important that the specialist looking after the patient is made aware of the relapse because this will influence overall management of MS. It was also noted that pharmacies will probably not have methylprednisolone readily available, whereas hospitals will. The GDG agreed a recommendation that people with MS should not be given steroids to have at home to take in case of relapse. It is considered good practice to do this for conditions such as chronic obstructive pulmonary disease. However treatment of relapse of MS is very different. The doses of steroid required for treatment of relapse of MS are much higher with a consequent increase in risk of side effects; the diagnosis of relapse can be difficult and should be done in conjunction with a professional with expertise in MS; relapse can require referral to services such as social care and the

	occurrence of relapse should be considered in light of a person's treatment with disease modifying drugs.

Recommendations	**<u>Recognising a relapse</u>** **78. Diagnose a relapse of MS if the person:** o develops new symptoms **or** o has worsening of existing symptoms **and these last for more than 24 hours in the absence of infection or any other cause after a stable period of at least 1 month.** **79. Before diagnosing a relapse of MS:** o rule out infection – particularly urinary tract and respiratory infections **and** o discriminate between the relapse and fluctuations in disease or progression. **80. Assess and offer treatment for relapses of MS, that affect the person's ability to perform their usual tasks, as early as possible and within 14 days of onset of symptoms.** **81. Do not routinely diagnose a relapse of MS if symptoms are present for more than 3 months.**
Relative values of different outcomes	The outcome of recognition of a relapse can allow steroid use with the associated reduced severity and duration. It can also affect the choice of long term disease modifying treatment.
Trade off between clinical benefits and harms	It was thought that the benefits of recognition of a relapse and excluding infection outweighed any risks.
Economic considerations	Considering specific characteristics for the diagnosis of a relapse of multiple sclerosis does not have any economic implications.
Quality of evidence	The recommendations were informed by review of the McDonald criteria, the evidence review of use of steroids and GDG professional opinion.
Other considerations	The GDG developed these recommendations using their professional experience and informal consensus. The diagnosis of relapse is a clinical judgement which requires experience in management of people with MS. Fluctuations in symptoms can occur for reasons such as intercurrent infection or people may be developing progressive disease rather than suffering a relapse. The GDG considered it important to differentiate relapse from intercurrent infection such as urinary tract infection. They thought that testing for infection was important in all patients presenting with new neurological symptoms or signs. Urinary tract and respiratory infections were the commonest infections encountered in patients with MS. Diagnosis of relapse can be difficult and because of its implications for both acute and ongoing treatment a healthcare practitioner should always seek advice from a specialist in MS if he or she is not confident about recognising a relapse. In the majority of studies steroids were administered within 2-4 weeks of a

	relapse. The GDG recommended early use of steroids in a relapse and that if a patient presents late steroids should not routinely be given without neurologist review. Occasionally, people with PPMS may have superimposed relapses that can respond to steroids, but one must always consider other possibilities, more so than in RRMS.

Recommendations	<u>**Information about treating a relapse with steroids**</u> **82. Discuss the benefits and risks of steroids with the person with MS, taking into account the effect of the relapse on the person's ability to perform their usual tasks and their wellbeing.** **83. Explain the potential complications of high-dose steroids, for example temporary effects on mental health (such as depression, confusion and agitation) and worsening of blood glucose control in people with diabetes.** **84. Give the person with MS and their family members or carers (as appropriate) information that they can take away about side effects of high-dose steroids in a format that is appropriate for them.** **85. Ensure that the MS multidisciplinary team is told that the person is being treated for relapse, because relapse frequency may influence which disease-modifying therapies are chosen and whether they need to be changed.**
Relative values of different outcomes	
Trade off between clinical benefits and harms	No harms are likely to offset the benefits of information for the person with MS and their carers.
Economic considerations	The GDG considered that provision of information has minimal impact on time and resource use as it is routinely done in NHS settings. The GDG considered that written information would be readily available and that provision of such information should have negligible economic impact.
Quality of evidence	
Other considerations	The GDG used their experience and knowledge of the effects of steroids to inform the recommendations. The dose of steroids required to treat a relapse is high and people need information about the risks so that they can be alert to potential side effects. People need to be given information about the risks and benefits of steroid treatment and they should have access to written information about this. The occurrence of relapses may be an indication for review of disease modifying treatments and it is essential that pathways are in place to allow access to specialists who can diagnose relapse and to inform specialists providing longer term care about the occurrence of relapse so that treatment can be reviewed. A clear pathway for rapid access to MS services will help prevent delays to

	recognition of relapses. It will also help with continuity of care by ensuring that the specialist team looking after the patient are aware of the relapse.

Recommendations	**Medical, therapy and social care needs at time of relapse or exacerbation** **86. Identify whether the person having a relapse of MS or their family members or carers have social care needs and if so refer them to social services for assessment.** **87. Offer inpatient treatment to the person having a relapse of MS if their relapse is severe or if it is difficult to meet their medical and social care needs at home.** **88. Explain that a relapse of MS may have short-term effects on cognitive function.** **89. Identify whether the person with MS having a relapse or exacerbation needs additional symptom management or rehabilitation.**
Relative values of different outcomes	
Trade off between clinical benefits and harms	No harms are likely to offset the benefits of these strategies for the person with MS and their carers.
Economic considerations	There are costs associated with addressing other needs at the time of relapse, for example admission to hospital for those with severe relapse; however the GDG considered the benefit of addressing these needs justify the cost.
Quality of evidence	
Other considerations	The GDG used their experience and informal consensus to develop these recommendations. Relapse will be associated with change in symptoms and deterioration in function and this includes cognitive function. People may have social care requirements and referral to social care for assessment may be required. Early referral for rehabilitation may also be required. Severe relapses may necessitate admission to hospital for intensive therapy and intravenous methylprednisolone. The group noted that in occasional cases, patients may be admitted for treatments such as plasmapheresis for refractory relapses.

13 Other treatments

13.1 Vitamin D

13.1.1 Introduction

Low vitamin D levels have been associated with a number of conditions such as heart disease, diabetes, and multiple sclerosis. Many studies have shown that the prevalence of MS increases with distance from the equator, and this effect may be mediated by the lower levels of sunlight, and therefore lower serum vitamin D levels, in the more temperate zones. A recent study has shown a more direct association between lower serum vitamin D levels and faster and more severe progression of MS. It has therefore been suggested that low vitamin D levels may act as a trigger for MS or may affect the course of established MS.

There are no clear definitions of what vitamin D blood levels are optimal and insufficiency affects about 50% of adults in the UK at the end of the winter. The Department of Health advises that most people should be able to get all the vitamin D they need from their diet and moderate sun exposure. Supplementation is recommended for pregnant and breastfeeding women, people aged 65 and over, people with low sun exposure such as those who stay indoors a lot, or cover up when outside and children aged 6 months to 5 years.

13.1.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of pharmacological treatment with vitamin D?

For full details see review protocol in Appendix A.

Table 203: PICO characteristics of review question

Population	• Adults
Intervention/s	• Vitamin D
Comparison/s	• Usual treatment or placebo
Outcomes	• Quality of life • Relapse rates • Functional disability (i.e. Expanded Disability Status Scale [EDSS]) • Cognitive function (i.e. Scripps Neurological Rating Scale [SNRS]) • Incidence of adverse events
Study design	• Systematic reviews • RCTs • Include cross-over trials • Include dosing studies

13.1.3 Clinical evidence

Seven studies were included in the review. Five compared vitamin D supplementation to placebo or no treatment [32,110,156,225,231] and two compared a higher dose of vitamin D with a lower dose.[80,237] The methodologies and populations are summarised in Table 204. All studies contained patients with baseline serum vitamin D levels in the normal range, with the exception of Mosayebi 2011, where serum levels were low. Groups were all matched for serum vitamin D levels. Evidence is summarised in the clinical GRADE evidence profiles (Table 196 and Table 207). Evidence not appropriate for

GRADE because of its use of incomplete or non-parametric data was presented narratively in tables in sections 0 and 0.

Table 204: Summary of studies included in the review

Study	Population	Vitamin D dose and duration	Comparator
Burton 2010[32]	18-55 yrs; EDSS 0-6.5; continuation of DMDs allowed; serum vitamin D was 73 nmol/l in vitamin d group and 83 nmol/l in placebo group at baseline	**28,000 – 280,000IU vitamin D3 /week; Calcium also taken;** **1 year**	Nothing, but permitted to take up to 4000IU/day of vitamin D and supplemental calcium.
Shaygannejad 2012[225]	25-57yrs; EDSS<6; continuation of DMDs allowed; baseline Vitamin D serum levels not given but inclusion criterion was a level of >40 nmol/l at baseline.	**3.50 micrograms calcitriol/week (equivalent to 140 IU vitD3/week)** 1 year	Identical placebo
Soilu-Hanninen 2012[231]	18-55yrs; EDSS≤5; continuation of Interferon; serum vitamin D was 54 nmol/l in vitamin D group and 56 nmol/l in placebo group at baseline	Cholecalciferol containing **20,000 IU vitamin D3/week**; 1 year	Identical placebo
Kampman 2012[110]	18-50yrs; EDSS≤4.5; not stated if on DMDs; serum vitamin D was 55 nmol/l in vitamin d group and 57 nmol/l in placebo group at baseline	Cholecalciferol containing **20,000 IU vitamin D3/week** + 500mg calcium/day; 96 weeks	Identical placebo + 500mg calcium/day
Mosayebi 2011[156]	18-60yrs; EDSS 0-3.5; All received Interferon beta 1a	300000 IU vitamin D3/month via an intramuscular injection; 6 months	Identical placebo
Stein 2011[229]	Age >18;EDSS≤5; allowed to be on glatiramer acetate or interferon; serum vitamin D was 59 nmol/l in high dose group and 53.5 nmol/l in low dose group at baseline	1000IU vit D2 per day PLUS 2x6000IU vit D2 per day (i.e. **91,000 IU vitamin D2/week**)	1000IU vit D2 per day PLUS placebo
Golan 2013[80]	Age ≥18 years; interferon-β-treated patients with relapsing-remitting multiple sclerosis; 25-OH-D blood levels <75 nmol/L and EDSS score <7.	4,370 IU vitamin D3 per day (high dose)	800 IU vitamin D3 per day (low dose)

Vitamin D versus nothing or placebo

Table 205: Clinical evidence profile: Vitamin D versus control

Quality assessment							Meta-analysis data		Overall Effect		Quality	Importance
No of studies and reference	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vitamin D Event rate (%) OR Mean (sd) [n]	Control Event rate (%) OR Mean (sd) [n]	Relative (95% CI)	Absolute		
Quality of Life												
No evidence was available for this outcome												
Proportion of people with relapses by follow up time – 52-96 weeks												
3 Burton 2010 Shaygannejad 2012 Kampman 2012	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	Very serious[B]	none	18/85 (21%)	22/82 (26.8%) Median control event rate: 36%	RR 0.80 (0.46 to 1.36)	72 fewer per 1000 (from 194 fewer to 130 more)	VERY LOW	CRITICAL
Annualised Relapse Rate (at post-test, or change from baseline)												
3 Kampman 2012 Shaygannejad 2012 Soilu-Hanninen 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[B]	none	0.03(0.35)[35] 0.32(0.48)[25] 0.26(0.5)[29]	-0.07(0.35)[33] 0.4(0.58)[25] 0.28(0.6)[3	-	MD 0.04 higher (0.09 lower to 0.17 higher)	MODERATE	CRITICAL

Quality assessment							Meta-analysis data		Overall Effect		Quality	Importance
No of studies and reference	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Vitamin D Event rate (%) OR Mean (sd) [n]**	**Control Event rate (%) OR Mean (sd) [n]**	**Relative (95% CI)**	**Absolute**		
								0]				
Proportion of people with increased EDSS at end of trial												
1 Burton 2010	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	2/25 (8%)	9/24 (37.5%)	RR 0.21 (0.05 to 0.89)	296 fewer per 1000 (from 41 fewer to 356 fewer)	LOW	CRITICAL
Change in grip strength (higher better)												
1 Kampman 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[B]	none	-3.51(14.59)[35]	2.39(14.59)[33]	-	MD 1.12 lower (8.06 lower to 5.82 higher)	MODERATE	CRITICAL
Ambulation ability - 25 metre timed walk (lower better) - Change over 1 year follow up												
1 Soilu-Hanninen 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	-0.62(1)[31]	0.3(0.9)[30]	-	MD 0.92 lower (1.4 to 0.44 lower)	HIGH	CRITICAL
Ambulation ability - 25 metre timed walk (lower better) - Change over 96 weeks follow up												
1 Kampman 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[B]	none	0.08(0.7)[35]	0.11(0.71)[	-	MD 0.03 lower	MODER	CRITIC

Quality assessment							Meta-analysis data		Overall Effect		Quality	Importance
No of studies and reference	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vitamin D Event rate (%) OR Mean (sd) [n]	Control Event rate (%) OR Mean (sd) [n]	Relative (95% CI)	Absolute		
			y	s				33]		(0.36 lower to 0.30 higher)	ATE	AL
Ambulation ability - 10 metre timed walk (lower better)												
1 Soilu-Hanninen 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	-2.44(0.8)[31]	0.95(1.1)[30]	-	MD 3.39 lower (3.87 to 2.91 lower)	HIGH	CRITICAL
Change in Fatigue severity score (lower better)												
1 Kampman 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	0.28(1.36)[34]	0.27(1.33)[33]	-	MD 0.01 higher (0.63 lower to 0.65 higher)	HIGH	CRITICAL
Change in paced auditory serial addition test (higher better)												
1 Kampman 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[B]	none	4.11(6.49)[19]	1.48(6.50)[21]	-	MD 2.63 higher (1.43 lower to	MODERATE	CRITICAL

Quality assessment							Meta-analysis data		Overall Effect		Quality	Importance
No of studies and reference	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Vitamin D Event rate (%) OR Mean (sd) [n]**	**Control Event rate (%) OR Mean (sd) [n]**	**Relative (95% CI)**	**Absolute**		
										6.66 higher)		
Change in 9 hole peg test (lower better)												
1 Kampman 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[B]	none	0.16(2.56)[35]	-0.43(2.58)[33]	-	MD 0.59 higher (0.63 lower to 1.81 higher)	MODERATE	CRITICAL
Adverse event - constipation												
2 Burton 2010 Shaygannejad 2012	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	serious[B]	none	10/50 (20%)	4/49 (8.2%) Median control event rate: 8%	RR 2.31 (0.83 to 6.45)	105 more per 1000 (from 14 fewer to 436 more)	LOW	CRITICAL
Adverse event - dyspepsia												
1 Shaygannejad 2012	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	6/25 (24%)	2/25 (8%)	RR 3 (0.67 to 13.46)	160 more per 1000 (from 26 fewer to 997 more)	VERY LOW	CRITICAL

Quality assessment							Meta-analysis data		Overall Effect		Quality	Importance
No of studies and reference	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vitamin D Event rate (%) OR Mean (sd) [n]	Control Event rate (%) OR Mean (sd) [n]	Relative (95% CI)	Absolute		
Adverse event - fatigue												
1 Shaygannejad 2012	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	4/25 (16%)	5/25 (20%)	RR 0.8 (0.24 to 2.64)	40 fewer per 1000 (from 152 fewer to 328 more)	VERY LOW	CRITICAL
Adverse events - headache												
1 Shaygannejad 2012	randomised trials	serious[A]	no serious inconsistency	no serious indirectness	very serious[B]	none	2/25 (8%)	1/25 (4%)	RR 2 (0.19 to 20.67)	40 more per 1000 (from 32 fewer to 787 more)	VERY LOW	CRITICAL
Adverse event- diarrhoea												
1 Soilu-Hanninen 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[B]	none	5/34 (14.7%)	2/32 (6.3%)	RR 2.35 (0.49 to 11.28)	85 more per 1000 (from 32 fewer to 648 more)	LOW	CRITICAL
Adverse event- fever												

Quality assessment							Meta-analysis data		Overall Effect		Quality	Importance
No of studies and reference	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Vitamin D Event rate (%) OR Mean (sd) [n]**	**Control Event rate (%) OR Mean (sd) [n]**	**Relative (95% CI)**	**Absolute**		
1 Soilu-Hanninen 2012	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[B]	none	2/34 (5.9%)	5/32 (15.6%)	RR 0.38 (0.08 to 1.8)	97 fewer per 1000 (from 144 fewer to 125 more)	LOW	CRITICAL

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one. Methodological limitations comprised one of the following: unclear allocation concealment or the lack of blinding.
[B] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.33 for dichotomous outcomes with a negative effect (i.e. the greater the proportion with the outcome, the worse the clinical result), at 0.8 and 1.25 for dichotomous variables with a positive effect, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables. For continuous variables analysed with the standardised mean difference option, the MIDs were set half a standard deviation either side of the null line.

Narrative review

EDSS

EDSS data was not analysed as a continuous variable in review manager or GRADE because it is ordinal and not interval data. Five studies [32,110,156,225,231] assessed EDSS scores in the vitamin and placebo groups. Burton[32] measured the EDSS change from baseline to follow up, and reported a mean 23% improvement in the vitamin D group and a 30% worsening in the control group. This was non-significant on non-parametric testing. Burton[32] did analyse EDSS as a categorical variable too, and those results are included in GRADE.

Four studies [110,156,225,231] used parametric methods to compare changes from baseline across the two groups. Shaygannejad 2012[225] reported a change of 0 (0.38) in EDSS for the vitamin D group, and the placebo group showed a worsening of 0.24 (0.41) points, which significantly favoured vitamin D (p=0.03). Soilu-Hanninen[231] reported that the vitamin D group improved by 0.2(0.1) and the placebo group improving by 0.0290.1) points, which again favoured the vitamin group (p=0.00001). In contrast, Kampman 2012[110] showed no difference between groups, with the vitamin D group worsening by 0.16(0.71) and the placebo group worsening by 0.15 (0.71). Mosayebi 2011[156] did not present variance for change scores, but showed similar worsening of 0.21 and 0.17 points in the vitamin D and placebo groups respectively. Overall, it is difficult to assess the efficacy of vitamin D in improving EDSS from these studies as a result of the inappropriate statistical analysis of these results.

Annualised relapse rates

One study [32] also did not present data in a way that was amenable for meta-analysis. Annualised relapse rates (ARR) for the year before baseline and the year after baseline were presented as shown in Table 206 below:

Table 206: ARR data reported by Burton 2010.

	Vitamin D	Control
Baseline ARR [mean(sd)]	0.44(0.77)	0.54(0.72)
12 month ARR [mean(sd)]	0.26(0.62)	0.45(0.59)

Because of the large discrepancy at baseline, a simple comparison of the 12 month values would tend to over-estimate the treatment effect in favour of vitamin D. Variance of the change scores for each group were not reported. Hence the 12 month values were not included in the meta-analysis.

High dose versus low dose vitamin D

Table 207: Clinical evidence profile: High dose vitamin D versus low dose vitamin D

Quality assessment							Meta-analysis data		Overall Effect		Quality	Importance
No of studies and reference	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	High dose Vitamin D Event rate (%) OR Mean (sd) [n]	Low dose vitamin D Event rate (%) OR Mean (sd) [n]	Relative (95% CI)	Absolute		
Numbers with a relapse by exit (mostly 6 months) (lower better)												
1 Stein 2011[237]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	4/11 (36.4%)	0/12 (0%) Median control event rate: 0%	Peto OR: 11.26 (1.36 to 92.95)	360 more per 1000 (from 70 more to 660 more)	HIGH	CRITICAL
Annual relapse rate at 12 months (lower better)												
1 Golan 2013[80]	randomised trial	serious risk of bias[a]	no serious inconsistency	no serious indirectness	serious imprecision[b]	none	0.51 (0.34) [15]	0.34 (0.27) [15]	-	MD 0.17 higher (0.05 lower to 0.39 higher)	LOW	CRITICAL
Quality of life score (FAMS) at 12 months (lower better)												
1 Golan 2013[80]	randomised trial	serious risk of bias[a]	no serious inconsistency	no serious indirectness	very serious imprecision[b]	none	146.6 (45.5) [15]	142.7 (32.5) [15]	-	MD 3.9 higher (24.4 lower to 32.2 higher)	VERY LOW	CRITICAL

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one. Methodological limitations comprised one of the following: unclear allocation concealment or the lack of blinding.*
[B] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.33 for dichotomous outcomes with a negative effect (i.e. the greater the proportion with the outcome, the worse the clinical result), at 0.8 and 1.25 for dichotomous variables with a positive effect, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables. For continuous variables analysed with the standardised mean difference option, the MIDs were set half a standard deviation either side of the null line.*

Narrative review

EDSS

This outcome was not analysed as a continuous variable in review manager or GRADE because it is ordinal and not interval data. One study[237] presented EDSS data at baseline and 6 months follow up, as shown in Table 208 below:

Table 208: EDSS data reported by Mosayebi 2011.

	Vitamin D – high dose	Vitamin D – low dose
Baseline EDSS [median(IQR)]	2.5(2-4)	2(1-3)
6 month EDSS [median(IQR)]	3(2-4)	2(1-2)

Because of the large discrepancy at baseline, a simple comparison of the 6 month values would tend to over-estimate the treatment effect in favour of vitamin D. No variances were given for the change values of 0.5 worsening for the high dose group and no change in the low dose group.

Another study [80] used parametric methods to compare final scores at 6 months and at 12 months across the two groups. It reported a score at 6 months of 3.4 (2.3) in the high dose group and 3.6 (2.1) in the low dose group, and at 12 months of 3.3 (2.4) in the high dose group and 3.6 (2.3) in the low dose group. These were not significantly different from baseline in either group, or between groups.

13.1.4 Economic evidence

Published literature

No relevant economic evaluations comparing Vitamin D with control were identified.

See also the economic article selection flow chart in Appendix E.

13.1.5 Evidence statements

13.1.5.1 Clinical

Vitamin D versus control

Very low quality evidence from 3 studies comprising 167 participants showed that there was no difference in clinical effectiveness between vitamin D and placebo in terms of rate of relapse, with very serious imprecision.

Moderate quality evidence from 3 studies comprising 177 participants showed that there was no difference in clinical effectiveness between vitamin D and placebo in terms of a higher annualised relapse rate, with serious imprecision.

Low quality evidence from one study comprising 49 participants showed that vitamin D was clinically effective compared to placebo in terms of a lower proportion of people of worsened EDSS at 1 year, with serious imprecision.

Moderate quality evidence from one study comprising 67 participants showed that there was no difference in clinical effectiveness between vitamin D and placebo in terms of grip strength, with serious imprecision.

High quality evidence from one study comprising 61 participants showed that vitamin D was clinically effective compared to placebo in terms of the 25 metre walk at one year, with no imprecision.

Moderate quality evidence from one study comprising 67 participants showed that there was no difference in clinical effectiveness between vitamin D and placebo in terms of 25 metre walk at 96 weeks, with no imprecision.

High quality evidence from one study comprising 61 participants showed that vitamin D was clinically effective compared to placebo in terms of the 10 metre walk at one year, with no imprecision

High quality evidence from one study comprising 67 participants showed that there was no difference in clinical effectiveness between vitamin D and placebo in terms of fatigue severity score, with no imprecision.

Moderate quality evidence from one study comprising 61 participants showed that there was no difference in clinical effectiveness between vitamin D and placebo in terms of PASAT score, with no imprecision.

Moderate quality evidence from one study comprising 61 participants showed that there was no difference in clinical effectiveness between vitamin D and placebo in terms of the 9 hole peg test, with serious imprecision.

Low quality evidence from one study comprising 99 participants showed that vitamin D was clinically harmful compared to placebo in terms of constipation, with serious imprecision.

Very low quality evidence from one study comprising 50 participants showed that vitamin D was clinically harmful compared to placebo in terms of dyspepsia, with very serious imprecision.

Very low quality evidence from one study comprising 50 participants showed that there was no difference in clinical effectiveness between vitamin D and placebo in terms of fatigue or headache, with very serious imprecision.

Low quality evidence from one study comprising 66 participants showed that vitamin D was clinically harmful compared to placebo in terms of diarrhoea, with very serious imprecision.

Low quality evidence from one study comprising 66 participants showed that vitamin D was clinically harmful compared to placebo in terms of fever, with very serious imprecision.

High dose versus low dose vitamin D

High quality evidence from one study comprising 23 participants showed that high dose vitamin D was clinically harmful low dose in terms of rate of relapse at 6 months, with no imprecision.

Low quality evidence from one study comprising 30 participants showed that there was no difference in clinical effectiveness between vitamin D high dose and low dose in terms of annual relapse rate at 12 months, with serious imprecision.

Low quality evidence from one study comprising 30 participants showed that there was no difference in clinical effectiveness between vitamin D high dose and low dose in terms of quality of life score (FAMS) at 12 months, with very serious imprecision.

13.1.5.2 Economic

No relevant economic evaluations were identified.

13.1.6 Recommendations and link to evidence

Recommendations	**90. Do not offer vitamin D solely for the purpose of treating MS.**
Relative values of different outcomes	There were no direct measures of quality of life, which was considered the most critical outcome. Number of relapses (absolute and annualised rate) was also a critical outcome. EDSS and walking speed were regarded as critical measures of the progression of MS and functional impact on daily activities.
Trade-off between clinical benefits and harms	Relapse rates were not affected by vitamin D, when compared to placebo. However, two studies looking at high-dose and low-dose vitamin D found that relapse rates were significantly higher with high-dose vitamin D, suggesting a potential harm of higher doses. There were clinically important benefits for vitamin D compared to placebo in mobility but these were not sustained over time. In addition, vitamin D also led to a greater proportion of people with an improvement in EDSS compared to placebo. The studies also compared EDSS scores across groups as an interval variable in parametric analyses, again showing a relative benefit for vitamin D versus placebo; however this form of analysis is inappropriate for an ordinal variable and these results should be viewed with caution. The GDG noted that there were no serious adverse effects from vitamin D use, although there were clinically important gastrointestinal harms in terms of slightly increased rates of diarrhoea, dyspepsia, and constipation. Overall, the benefits observed for vitamin D were not felt to be large or consistent enough by the GDG to outweigh the harms.
Economic considerations	No cost effectiveness evidence was identified. The clinical data has proven inconclusive and prescribing vitamin D supplements would incur a cost to the NHS. Therefore, the GDG felt that their use was not cost-effective and recommended not to offer vitamin D for the management of MS.
Quality of evidence	The quality of the outcomes ranged from very low to high, but most studies were of high quality, with most having evidence of allocation and triple-blinding and management of missing data was good It was noted that there were only 7 studies meeting the criteria for inclusion. Five compared vitamin D to placebo, and two compared two different doses of vitamin D. Studies had used different doses of vitamin D and different preparations: ergocalciferol (D2) or cholecalciferol (D3), and with or without calcium supplementation. The GDG noted that vitamin D levels were normal in all participants at baseline in these studies, but it was not known if those assigned to placebo were taking

	any vitamin D outside of the study. Similarly, it was not known what disease-modifying treatments participants received and if this was different within or between studies. Most studies had a small sample size with less than 100 participants in every study, and as few as 23 in one study. The GDG thought that this would have led to a high chance of a type II error and failure to detect a beneficial effect of vitamin D.
Other considerations	The GDG discussed the association of MS prevalence with areas of latitude and the hypothesis that Vitamin D is associated with the pathogenesis of multiple sclerosis. Clinically this has been interpreted to mean that Vitamin D has a role to play in the management of MS. The GDG were concerned not to make any blanket recommendation on vitamin D use as there are other important reasons to use it. Particularly it is recommended in pregnancy and for bone health. The recommendation is only for use of vitamin D for management of MS. Further studies are needed to assess the benefit or harm of using vitamin D. Studies thus far have excluded people with primary progressive and secondary progressive MS, and these populations should also be investigated separately. The GDG was aware of one on-going study in Australia using different doses of vitamin D in people with MS.

13.2 Omega fatty acid compounds

13.2.1 Introduction

There have been suggestions that omega fatty acid compounds may be of benefit to people with MS. Omega fatty acid compounds include omega-3 and omega-6 fatty acids.

13.2.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of omega-3 fatty acids and omega-6 fatty acids?

For full details see review protocol in Appendix C.

Table 209: PICO characteristics of review question

Population	• Adults with MS
Intervention/s	• omega-3 fatty acids • omega-6 fatty acids
Comparison/s	• Usual treatment or placebo
Outcomes	• Quality of life • Functional disability (i.e. EDSS) • Pain • Incidence of adverse events • Relapse rates • Drop outs
Study design	• Systematic reviews, RCTs. Include cross-over studies.

13.2.3 Clinical evidence

Nine papers covering eleven RCTs were found. Five RCTs covered the *omega-6 versus placebo* comparison, and four covered the *omega-3 versus placebo* comparison. These RCTs are summarised in Table 37.

Evidence from both comparisons are summarised in the clinical GRADE evidence profiles below in sections 0 and 0. See also the study selection flow chart in Appendix D, forest plots in Appendix I, study evidence tables in Appendix G and exclusion list in Appendix J.

Some outcomes were not appropriate for meta-analysis as they lacked sufficient data, such as measures of variance. These have been reported in separate narrative sections in 0 and 0.

Summary of included studies

Table 210: Summary of randomised controlled trials included in the review

Study	Intervention details	Baseline characteristics where available (group-specific data designated by intervention / comparator)	n	Analysis
OMEGA-6 VERSUS PLACEBO				
Millar 1973[144]	17.2g linoleic acid daily for 2 years	EDSS 0-6; duration MS: 9.2/7.7 yrs; age 37.8/35.5	87	parallel
Paty 1983 / 1978 (essentially same study)[177,178]	17g linoleic acid daily for 30 months	EDSS 1-6; age 32 yrs	76	parallel
Bates 1977[20]	3.42g of linoleic acid + 360 mg linolenic acid daily for 2 years OR 11.5g of linoleic acid daily in the form of a spread (these groups were combined for the analysis)	Chronic progressive MS	152	parallel
Bates 1978[21]	2.92g of linoleic acid + 340 mg linolenic acid daily for 2 years OR 23g of linoleic acid daily in the form of a spread (these groups were combined for the analysis)	Acute remitting MS; duration MS 7/6 yrs; age 34/32 yrs	116	parallel
OMEGA-3 VERSUS PLACEBO				
Bates 1989[19]	1.71g of C20:5 and 1.14g of C22:6 per day for 2 years	Acute remitting MS; EDSS $\leq$ 6; duration MS 7/6 yrs; age 34/32 yrs	312	parallel
Weinstock-Guttman 2005[261]	EPA 1.98g and DHA 1.32 g / day for 1 year	RRMS; EDSS 1.9/2; MS duration 6.9/4.7 yrs; age 45/40	31	parallel
Torkidsen 2012[249]	EPA 1.35g and DHA 0.85 g / day for 6 months	RRMS; EDSS 1.9; MS duration 5/6 yrs; age 39/38 yrs	92	parallel
Ramirez-Ramirez 2013[189]	EPA 0.8g and DHA 1.6g / day for 1 year	EDSS: intervention: 2.1 (0.98); placebo 2.06 (0.84) Duration of disease: intervention: 7.14 (4.79); placebo 6.68 (5.69) years Age: intervention: 35.1 (7.6); placebo 34.7 (7.8) years	50	parallel

Omega-6 versus placebo

Table 211: Clinical evidence profile for omega-6 versus placebo

Quality assessment							Proportion with event OR Mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Linoleic acid(omega 6)	Control	Relative (95% CI)	Absolute		
Quality of life												
No studies reported on this outcome												
Global improvement over course of study												
Bates 1977 Millar 1973	RCT	Very serious[A]	None	None	Very serious[C]	none	12/112 (10.7%)	10/115 (8.7%) **8.5%**	RR 1.23 (0.56 to 2.74)	20 more per 1000 (from 37 fewer to 148 more)	VERY LOW	CRITICAL
Severity of relapses – score based on sensory, motor and visual criteria (higher worse)												
Millar 1973	RCT	Very serious[A]	None	None	serious[C]	none	17.9(28)[36]	34.6(28)[39]		MD: 16.7 lower (from 29.38 lower to 4.02 lower)	VERY LOW	CRITICAL
Number with 1 or more relapses												
Bates 1978 Millar 1973	RCT	Very serious[A]	None	None	None	none	76/94 (80.9%)	76/96 (79.2%) 78.8%	RR 1.02 (0.88 to 1.17)	16 more per 1000 (from 95 fewer to 134 more)	LOW	CRITICAL
Number with 3 or more relapses												
Bates 1978 Millar 1973	RCT	Very serious[A]	Very serious[B]	None	Very serious[C]	none	36/94 (38.3%)	39/96 (40.6%) 38.7%	Random RR 0.65 (0.16 to 2.63)	135 fewer per 1000 (from 325 fewer to 631	VERY LOW	CRITICAL

Quality assessment							Proportion with event OR Mean(sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Linoleic acid(omega 6)	Control	Relative (95% CI)	Absolute		
										more)		

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B]Outcomes were downgraded by one increment for serious inconsistency, as shown by the I^2value being between 50 and 74%. A double downgrade was applied for very serious inconsistency if I^2 was >75%. A random effects model was used for any inconsistent outcomes.
[C] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Narrative review for omega 6 versus placebo

Millar 1973 reported on worsening in disability status over 2 years. This outcome was unclearly derived from EDSS and functional ability. Linoleic acid had a worsening of 0.2 and the placebo group had a worsening of 0.5, but no variance measures were provided. The small difference was reported as non-significant.

Paty 1983, provided means for a variety of continuous outcomes, but no variance measures. Their results are supplied in Table 212.

Table 212: Results from Paty 1983[177,178]

	Linoleic acid	control
EDSS at 30 months	3.52	3.85
Change in Kurtze pyramidal function score	0.33	0.63
Change in Kurtze cerebellar function score	0.32	0.35
Change in Kurtze brain stem function score	0.52	0.53
Change in Kurtze sensory function score	0.36	0.22
Change in Kurtze bowel/bladder function score	0.10	0.15

	Linoleic acid	control
Change in Kurtze visual function score	0.03	0.16
Change in Kurtze mental function score	-0.02	-0.03
Number of relapses (refers to total number of relapses, not participants with relapses)	38	19
Mean relapse score (unclear definition)	16.9	23.1
Functional tests	No information, except that all were NS across groups	

Bates 1977 provided information on the relapses per patient year and the score per relapse in linoleic and placebo groups, but no variance measures or statistical measures were given. These are summarised in Table 213.

Table 213: Results from Bates 1977[20,20]

	Linoleic acid capsules (n=38)	Linoleic acid spread(n=38)	Placebo capsules (n=38)	Placebo spread (n=38)
Relapses per patient year	0.26	0.22	0.20	0.15
Score per relapse (higher worse)	11.1	12.8	20	10.3

Bates 1978 gathered data on clinical deterioration (as shown by the EDSS), duration of exacerbations and 'attack score' (measuring severity and duration of attacks. Unfortunately no raw data was provided.

Table 214: Results from Bates 1978[20,21]

Outcome	Findings
Clinical deterioration (as shown by EDSS)	"Significantly more patients had deteriorated than improved" in the linoleic acid capsule group. No data given. Also the number deteriorating in the linoleic acid capsule group was significantly greater than in the placebo capsules group (p<0.05); no data given.
Duration of exacerbations	The linoleic acid spread group "had attacks of significantly shorter duration" than those in the placebo spread group. No data given.
Attack score, measuring severity and duration of attacks (mean score per attack per patient)	The linoleic acid spread showed a significant benefit compared to the placebo spread group. No data given

Omega-3 versus placebo

Table 215: Clinical evidence profile omega 3 versus placebo

Quality assessment							Proportion with event OR Mean (sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Omega 3	Control	Relative (95% CI)	Absolute		
Quality of life												
No studies reported on this outcome												
EDSS worse												
Bates 1989 Torkilsden 2012	RCT	Very serious[A]	None	None	Serious[C]	none	82/191 (42.9%)	86/189 (45.5%) 32.7%	RR 0.96 (0.78 to 1.19)	13 fewer per 1000 (from 72 fewer to 62 more)	VERY LOW	CRITICAL
EDSS worse - 6 months												
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	6/46 (13%)	4/42 (9.5%) 9.5%	RR 1.37 (0.41 to 4.52)	35 more per 1000 (from 56 fewer to 334 more)	LOW	CRITICAL
EDSS worse - 2 years												
Bates 1989	RCT	Very serious[A]	None	None	Serious[C]	none	76/145 (52.4%)	82/147 (55.8%) 55.8%	RR 0.94 (0.76 to 1.16)	33 fewer per 1000 (from 134 fewer to 89 more)	VERY LOW	CRITICAL
SF36 (Phys) (Better indicated by higher values)												
Weinstock 2005	RCT	Very serious[A]	None	None	serious[C]	none	45.4(8.8)[13]	38.4(8.8)[14]	-	MD 7 higher (0.34 to 13.66 higher)	VERY LOW	CRITICAL
modified fatigue index scale at 6 months (Better indicated by lower values)												
Weinstock 2005	RCT	Very serious[A]	None	None	serious[C]	none	51.8(20.9)[15]	33.8(20.9)[16]	-	MD 18 higher (3.26 to 32.74 higher)	VERY LOW	CRITICAL
modified fatigue index scale at 12 months (Better indicated by lower values)												
Weinstock	RCT	Very	None	None	serious[C]	none	58.8(28.2)[15]	37.3(28.2)[16]	-	MD 21.5 higher (1.63 to 41.37	VERY	CRITICAL

Quality assessment							Proportion with event OR Mean (sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Omega 3	Control	Relative (95% CI)	Absolute		
2005		serious[A]								higher)	LOW	
Number with a relapse within 6 months												
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	10/46 (21.7%)	8/45 (17.8%) 17.8%	RR 1.22 (0.53 to 2.82)	39 more per 1000 (from 84 fewer to 324 more)	LOW	CRITICAL
Change in relapse rate at 1 yr compared to year prior to treatment (more negative better)												
Weinstock 2005	RCT	Very serious[A]	None	None	Very serious[C]	none	-0.79(1.1)[13]	-0.69(1.1)[14]	-	MD 0.1 lower (0.94 lower to 0.74 higher)	VERY LOW	CRITICAL
Relapse rate at 12 months												
Ramirez-Ramirez 2013	RCT	Serious[A]	None	None	Serious[C]	None	0.84 (0.9) [20]	1 (1) [19]	-	0.16 lower (0.76 lower to 0.44 higher)	LOW	CRITICAL
Overall adverse events												
Torkilsden 2012	RCT	None	None	None	serious[C]	none	34/46 (73.9%)	29/46 (63%) 63%	RR 1.17 (0.89 to 1.55)	107 more per 1000 (from 69 fewer to 346 more)	MOD	CRITICAL
hair loss												
Torkilsden 2012	RCT	None	None	None	serious[C]	none	3/46 (6.5%)	0/46 (0%) 0%	Peto OR 7.73 (0.78 to 76.2)	70 more per 1000 (from 20 lower to 150 more)	MOD	IMPORTANT
abdominal pain												
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	0/46 (0%)	3/46 (6.5%) 6.5%	RR 0.14 (0.01 to 2.69)	56 fewer per 1000 (from 64 fewer to 110 more)	LOW	IMPORTANT
cod liver oil gulp												

Quality assessment							Proportion with event OR Mean (sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Omega 3	Control	Relative (95% CI)	Absolute		
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	4/46 (8.7%)	1/46 (2.2%)	RR 4 (0.46 to 34.44)	66 more per 1000 (from 12 fewer to 736 more)	LOW	IMPORTANT
								2.2%				
Fatigue												
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	5/46 (10.9%)	4/46 (8.7%)	RR 1.25 (0.36 to 4.36)	22 more per 1000 (from 56 fewer to 292 more)	LOW	IMPORTANT
								8.7%				
Nausea												
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	3/46 (6.5%)	4/46 (8.7%)	RR 0.75 (0.18 to 3.17)	22 fewer per 1000 (from 71 fewer to 189 more)	LOW	IMPORTANT
								8.7%				
UTI												
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	4/46 (8.7%)	3/46 (6.5%)	RR 1.33 (0.32 to 5.63)	21 more per 1000 (from 44 fewer to 301 more)	LOW	IMPORTANT
								6.5%				
Arthralgia												
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	3/46 (6.5%)	3/46 (6.5%)	RR 1 (0.21 to 4.7)	0 fewer per 1000 (from 51 fewer to 240 more)	LOW	IMPORTANT
								6.5%				
LBP												
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	2/46 (4.3%)	3/46 (6.5%)	RR 0.67 (0.12 to 3.81)	21 fewer per 1000 (from 57 fewer to 183 more)	LOW	IMPORTANT
								6.5%				
Myalgia												
Torkilsden 2012	RCT	None	None	None	serious[C]	none	2/46 (4.3%)	10/46 (21.7%)	RR 0.2 (0.05 to 0.86)	174 fewer per 1000 (from 30 fewer to 206 fewer)	MOD	IMPORTANT
								21.7%				
Headache												

Quality assessment							Proportion with event OR Mean (sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Omega 3	Control	Relative (95% CI)	Absolute		
Torkilsden 2012	RCT	None	None	None	Very serious[C]	none	4/46 (8.7%)	4/46 (8.7%)	RR 1 (0.27 to 3.76)	0 fewer per 1000 (from 64 fewer to 240 more)	LOW	IMPORTANT
								8.7%				

[A] Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B]Outcomes were downgraded by one increment for serious inconsistency, as shown by the I^2 value being between 50 and 74%. A double downgrade was applied for very serious inconsistency if I^2 was >75%. A random effects model was used for any inconsistent outcomes.
[C] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.

Narrative review for omega 3 versus placebo

Weinstock-Guttman 2005[261] provided information on SF-36 and EDSS outcomes, but insufficient detail was provided. Global SF36 score was reported to stay the same in the omega 3 group, but to worsen in the placebo group, and this was reported to be non-significant. Group differences at follow up in EDSS were reported as a non-significant trend favouring omega 3.

Torkilsden 2012[249] reported on function, quality of life and fatigue, but insufficient details were given. Table 216 summarises their results for these outcomes.

Table 216: Results from Torkilsden 2012[248,249]

Outcome	Findings	P value
Change of multiple sclerosis functional composite scores	No group differences in changes detected. No data given, and it is unclear whether any changes were –ve or +ve.	0.53
Change of SF36 physical scores	No group differences in changes detected. No data given, and it is unclear whether any changes were –ve or +ve.	0.66
Change of SF36 mental scores	No group differences in changes detected. No data given, and it is unclear	0.53

Outcome	Findings	P value
	whether any changes were –ve or +ve.	
Change of FSS	No group differences in changes detected. No data given, and it is unclear whether any changes were –ve or +ve.	0.97

Ramirez-Ramirez 2013[189]reported EDSS scores at 6 and 12 months as parametric data: 6 months: omega group 2.1 (0.9) vs. placebo group 2.0 (0.8), p=0.73; 12 months: 2.2 (1.0) and 2.2 (0.8), p=0.66 respectively.

13.2.4 Economic evidence

Published literature

No relevant economic evaluations comparing omega-6 versus placebo or omega-3 versus placebo were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

13.2.5 Evidence statements

13.2.5.1 Clinical

Omega 6 versus placebo

Very low quality evidence from 2 RCTs comprising 227 participants showed that there was no clinical difference between omega 6 and placebo in terms of global improvement over the course of treatment, with very serious imprecision.

Very low quality evidence from 1 RCT comprising 75 participants showed that omega 6 was clinically effective compared to placebo in terms of severity of relapses, with serious imprecision.

Low quality evidence from 2 RCTs comprising 190 participants showed that there was no clinical difference between omega 6 and placebo in terms of the proportion of people with 1 or more relapses, with no imprecision.

Very low quality evidence from 3 RCTs comprising 190 participants showed that omega 6 was clinically effective compared to placebo in terms of the proportion of people with 3 or more relapses, with very serious imprecision.

Omega 3 versus placebo

Very low quality evidence from 2 RCTs comprising 380 participants showed that there was no clinical difference between omega 3 and placebo in terms of worsening of EDSS over the course of treatment, with serious imprecision.

Very low quality evidence from 1 RCT comprising 27 participants showed that omega 3 was clinically effective compared to placebo in terms of SF36 (physical), with serious imprecision.

Very low quality evidence from 1 RCT comprising 31 participants showed that omega 3 was clinically harmful compared to placebo in terms of modified fatigue index at 6 months, with serious imprecision.

Very low quality evidence from 1 RCT comprising 31 participants showed that omega 3 was clinically harmful compared to placebo in terms of modified fatigue index at 12 months, with serious imprecision.

Low quality evidence from 1 RCT comprising 91 participants showed that there was no clinical difference between omega 3 and placebo in terms of the proportion of people with a relapse within 6 months, with very serious imprecision.

Very low quality evidence from 1 RCT comprising 27 participants showed that there was no clinical difference between omega 3 and placebo in terms of the change in relapse rate at 1 year, with very serious imprecision.

Low quality evidence from 1 RCT comprising 39 participants showed that there was no clinical difference between omega 3 and placebo in terms of relapse rate, with serious imprecision.

Moderate quality evidence from 1 RCT comprising 92 participants showed that omega 3 was clinically harmful compared to placebo in terms of overall adverse events, with serious imprecision.

Moderate quality evidence from 1 RCT comprising 92 participants showed that omega 3 was clinically harmful compared to placebo in terms of hair loss, with serious imprecision.

Low quality evidence from 1 RCT comprising 92 participants showed that omega 3 was clinically effective compared to placebo in terms of abdominal pain, with very serious imprecision.

Low quality evidence from 1 RCT comprising 92 participants showed that omega 3 was clinically harmful compared to placebo in terms of cod liver oil gulp, fatigue, nausea, incidence of UTIs, arthralgia, low back pain and headache with very serious imprecision.

Moderate quality evidence from 1 RCT comprising 92 participants showed that omega 3 was clinically effective compared to placebo in terms of myalgia, with serious imprecision.

13.2.5.2 Economic

No relevant economic evaluations were identified.

13.2.6 Recommendations and link to evidence

Recommendations	**91. Do not offer omega-3 or omega-6 fatty acid compounds to treat MS. Explain that there is no evidence that they affect relapse frequency or progression of MS.**
Relative values of different outcomes	Quality of life was the most critical outcome. Important outcomes were functional disability (i.e. EDSS), pain, incidence of adverse events and relapse rates.
Trade off between clinical benefits and harms	The results of the omega-6 versus placebo studies showed no benefits in global improvement but a possible benefit in terms of decreasing the number and severity of relapses. The omega-3 versus placebo studies showed that omega-3 improves SF36 scores (physical component) at 6 months, but worsens fatigue and causes some adverse events. However, the vast majority of outcomes show that omega-3 has no clinically important effects.
Economic considerations	No relevant economic evaluation studies were found comparing omega-3 or omega-6 versus placebo. No unit costs were presented for omega-6 fatty acids as they are not available on prescription from the NHS. The cost of dietary sources of omega-6 fatty acids would fall on the patients. Prescribed capsules of omega-3 fatty acids cost around £534-£557 per year. Given that omega-3 and 6 fatty acids were judged not to be of clinical benefit to patients on the basis of current effectiveness evidence and prescribed capsules have a considerable cost to the NHS, their use was considered not to be cost effective.
Quality of evidence	Evidence was graded as LOW or VERY LOW for the 8 included studies. Most studies provided no evidence of allocation concealment and although most studies were described as 'blind' it was often unclear which parties were blinded.
Other considerations	The evidence available came from the 1970's when treatments available for MS were very different, in particular disease modifying drugs. The evidence therefore does not help understanding of how omega fatty acid compounds might be used in the present context. However no other evidence was available so the studies were included in the review. The GDG agreed that because omegas compounds show no appreciable effect with a relatively high cost, they should not be recommended. The GDG were aware that people may also buy these compounds over the counter and considered that people should be informed that there is no good quality evidence of benefit.

13.3 Acupuncture

13.3.1 Introduction

People with MS suffer from fatigue, spasticity and muscle pain. Acupuncture is commonly used in the treatment of musculoskeletal disorders and of pain and may be of benefit to MS patients. Acupuncture treatment is often associated with a holistic view of patients' needs which may be of value in management of a condition which may have effects on multiple organs.

13.3.2 Review question: For adults with MS, what is the clinical evidence and cost effectiveness of acupuncture?

For full details see review protocol in Appendix C.

Table 217: PICO characteristics of review question

Population	• Adults with MS
Intervention/s	• Acupuncture
Comparison/s	• Usual treatment or placebo
Outcomes	• Quality of life • Functional disability (i.e. EDSS) • Pain • Incidence of adverse events • Relapse rates • Drop outs
Study design	• Systematic reviews, RCTs. Include cross-over studies.

13.3.3 Clinical evidence

3 RCTs were found that covered the *acupuncture versus placebo* comparison [55;188;79]. These RCTs are summarised in Table 218.

Evidence from all three comparisons are summarised in the clinical GRADE evidence profiles below in section 0. See also the study selection flow chart in Appendix D, forest plots in Appendix I, study evidence tables in Appendix G and exclusion list in Appendix J.

Some outcomes were not appropriate for meta-analysis as they lacked sufficient data, such as measures of variance. These have been reported in a separate narrative section in 0.

Summary of included studies

Table 218: Summary of randomised controlled trials included in the review

Study	Intervention and comparator details	Baseline characteristics where available (group-specific data designated by intervention / comparator)	n	Analysis
ACCUPUNCTURE VERSUS PLACEBO				
Donnellan 2008[55]	Traditional Chinese acupuncture – 25 mins 2x per week for 5 consecutive weeks. Sites of acupuncture individualised. VERSUS Sham acupuncture (shallow and away from acupuncture points) – 25 mins 2x per week for 5 consecutive weeks.	Secondary progressive MS (SPMS); ambulant; aged 53/50	14	parallel
Quispe-Cabanillas 2012[188]	Electroacupuncture to Chinese acupuncture points - 30 mins 1x per week for 6 months. Electricity aimed to stimulate increased sensory input. VERSUS Sham electroacupuncture to Chinese acupuncture points - 30 mins 1x per week for 6 months (shallow, away from acupuncture points and with no current)	Relapsing remitting MS (RRMS); all on immunomodulatory drugs; EDSS: 2.3/3; duration MS: 8/9 yrs; age 36/40 yrs	31	parallel
Gibson 1999[79]	'Neural therapy' - injection of lignocaine to Chinese acupuncture points at ankle and head – two injection sessions for first week and another two in second week. VERSUS As above except that normal saline was injected instead of lignocaine in the first week, but this group had lignocaine injected in the second week (2 x 2 injection sessions as for the other group). Hence only week 1 assessment included.	mainly primary progressive MS (PPMS); EDSS: 5/4; duration MS: 9/12 yrs; age 42/40	21	parallel

Acupuncture versus placebo

Table 219: Clinical evidence profile for acupuncture versus placebo

Quality assessment							Proportion with event OR Mean (sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture	Control	Relative (95% CI)	Absolute		
Health related quality of life												
No studies reported on this outcome												
MSIS-29 (physical) change from baseline (Better indicated by lower values)												
Donnellan 2008	RCT	serious[A]	none	none	Very serious[C]	none	-14(15.5)[6]	-13.8(12.9)[7]	-	MD 0.6 lower (16.26 lower to 15.06 higher)	VERY LOW	CRITICAL
MSIS-29 (psychol) change from baseline (Better indicated by lower values)												
Donnellan 2008	RCT	serious[A]	none	none	serious[C]	none	-6(13.9)[6]	-23(21)[7]	-	MD 17 higher (2.12 lower to 36.12 higher)	LOW	CRITICAL
FSS change from baseline (Better indicated by lower values)												
Donnellan 2008	RCT	serious[A]	none	none	Very serious[C]	none	-0.5(1.1)[6]	-0.4(1)[7]	-	MD 0.1 lower (1.25 lower to 1.05 higher)	VERY LOW	CRITICAL
GHQ-12 change from baseline (Better indicated by lower values)												
Donnellan 2008	RCT	serious[A]	none	none	serious[C]	none	-3.3(4.3)[6]	-9.7(10.7)[7]	-	MD 6.4 higher (2.24 lower to 15.04 higher)	LOW	CRITICAL
AEs - muscle twitching/spasm												
Donnellan 2008	RCT	serious[A]	none	none	Very serious[C]	none	5/7 (71.4%)	4/7 (57.1%) 57.1%	RR 1.25 (0.56 to 2.77)	143 more per 1000 (from 251 fewer to 1000 more)	VERY LOW	IMPORTANT

Quality assessment							Proportion with event OR Mean (sd)[n]		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture	Control	Relative (95% CI)	Absolute		
AEs - temporary worsening of fatigue and weakness												
Donnellan 2008	RCT	serious[A]	none	none	Very serious[C]	none	2/7 (28.6%)	0/7 (0%) 0%	Peto OR 8.73 (0.49 to 156.28)	290 more per 1000 (from 80 less to 650 more)	VERY LOW	IMPORTANT
Bleeding [<5 secs]												
Donnellan 2008	RCT	serious[A]	none	none	NA	none	0/7 (0%)	0/7 (0%) 0%	not pooled	not pooled	NA	IMPORTANT
Bleeding [>10 secs]												
Donnellan 2008	RCT	serious[A]	none	none	serious[C]	none	7/7 (100%)	2/7 (28.6%) 28.6%	RR 3 (1.06 to 8.52)	572 more per 1000 (from 17 more to 1000 more)	LOW	IMPORTANT
number with improvements in at least one sub-scale of the Kutzke scale at 1 week												
Gibson 1999	RCT	serious[A]	none	none	serious[C]	none	8/11 (72.7%)	1/10 (10%) 10%	RR 7.27 (1.09 to 48.35)	627 more per 1000 (from 9 more to 1000 more)	LOW	CRITICAL
number with improvements of at least one point on EDSS at 1 week												
Gibson 1999	RCT	serious[A]	none	none	serious[C]	none	3/11 (27.3%)	0/10 (0%) 0%	Peto OR 8.34 (0.77 to 90.88)	270 more per 1000 (from 10 less to 560 more)	LOW	CRITICAL

[A] *Outcomes were downgraded by one increment if the weighted average number of serious methodological limitations across studies was one, and downgraded by two increments if the weighted average number of serious methodological limitation across studies were two or more. Methodological limitations comprised one or more of the following: unclear allocation*

concealment, the lack of blinding, or inadequate allowance for drop-outs in the analysis. Cross-over studies were not downgraded for selection bias, as the effects of such bias would only be expected to exert effects via an order effect, and so selection bias would be less serious a limitation than in a parallel trial.
[B]*Outcomes were downgraded by one increment for serious inconsistency, as shown by the I^2 value being between 50 and 74%. A double downgrade was applied for very serious inconsistency if I^2 was >75%. A random effects model was used for any inconsistent outcomes.*
[C] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID. Outcomes were downgraded by two increments if both MIDs were crossed by one or both of the 95% CIs. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous variables, and at 0.5 of the control group weighted mean standard deviation either side of the null line for continuous variables.*

Narrative review

Quispe-Cabanillas 2012[188] provided data on EDSS but there were flaws in their methods of analysis. They also provided data on FAMS, and pain in low resolution graphs so only their data summaries are possible to extract. Table 220 summarises their findings.

Table 220: Results from Quispe-Cabanillas 2012[188]

Outcome	Findings
EDSS 6 months	There was a reported trend (p=0.055) for comparison of post-test values, which were 2.2 in the electroacupuncture group (no variance given) and 3.3 in the sham group (no variance given). However taking into account baseline discrepancies there appeared to be no clear group difference in the change values (change was -0.1 in the acupuncture group and +0.3 in the sham group), although no variances were available for these change values. The paper also gives another p value from an ANOVA, adjusting for differing treatment durations: this shows a benefit for the electroacupuncture group. However there are no reports on differing durations of treatment elsewhere in the paper; furthermore, the baseline bias is still unaccounted for. Hence this result will not be used in this review.
FAMS	Results given in low resolution graphs. Although there were significant differences in favour of the electroacupuncture group at 3 (p=0.0026) and 6 months (p<0.001) these were confounded by a similar (if non-significant) trend at baseline.
Pain	Results given in low resolution graphs. There were significant differences in favour of the electroacupuncture group at 3 (p=0.0143) and 6 months (p<0.001). At baseline the groups were similar (p=0.42); in any case the sham group had lower pain at baseline, so this actually indicates the true effect size in favour of electroacupuncture was even greater than observed.

13.3.4 Economic evidence

No relevant economic evaluations comparing acupuncture versus placebo were identified.

See also the economic article selection flow chart in Appendix E.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided in Appendix M to aid consideration of cost effectiveness.

13.3.5 Evidence statements

13.3.5.1 Clinical

Acupuncture versus placebo

Very low quality evidence from one RCT comprising 13 participants showed that there was no clinical difference between acupuncture and placebo in terms of MSIS-29 (physical), with very serious imprecision.

Low quality evidence from one RCT comprising 13 participants showed that acupuncture was clinically harmful compared to placebo in terms of MSIS-29 (psychological), with serious imprecision.

Very low quality evidence from one RCT comprising 13 participants showed that there was no clinical difference between acupuncture and placebo in terms of changes in FSS, with very serious imprecision.

Low quality evidence from one RCT comprising 13 participants showed that acupuncture was clinically harmful compared to placebo in terms of GHQ-12 (psychological), with serious imprecision.

Very low quality evidence from one RCT comprising 14 participants showed that acupuncture was clinically harmful compared to placebo in terms of muscle twitching/spasm, with very serious imprecision.

Very low quality evidence from one RCT comprising 14 participants showed that acupuncture was clinically harmful compared to placebo in terms of temporary worsening of fatigue and weakness, with very serious imprecision.

Low quality evidence from one RCT comprising 14 participants showed that acupuncture was clinically harmful compared to placebo in terms of bleeding for >10 seconds, with serious imprecision.

Low quality evidence from one RCT comprising 14 participants showed that acupuncture was clinically effective compared to placebo in terms of improvements in at least one sub-scale of the Kutzke scale at 1 week, with serious imprecision.

Low quality evidence from one RCT comprising 14 participants showed that acupuncture was clinically effective compared to placebo in terms of improvements of at least one point on the EDSS scale at 1 week, with serious imprecision.

13.3.5.2 Economic

No relevant economic evaluations were identified.

13.3.6 Recommendations and link to evidence

Recommendations	
Relative values of different outcomes	Quality of life was considered the most important outcome. Other important outcomes were those measuring functional ability, pain, relapse rates and adverse events.
Trade off between clinical benefits and harms	The results of the studies showed some benefits from acupuncture for function (EDSS and Kutzke scale) and Pain. Harms were identified for acupuncture in some adverse events and MSIS-29.
Economic considerations	No relevant economic evaluation studies were found comparing acupuncture and placebo. The cost of treatment with acupuncture was estimated to be between £125-390 per course, based on the resource utilisation from two RCTs found in the clinical review. Given the lack of clear clinical evidence and the considerable cost to the NHS, the GDG felt further research was required into the use of acupuncture for pain and spasticity in people with MS.
Quality of evidence	Only 3 RCTs were found. All used placebo controls, with blinded assessors and patients. Two studies had unclear allocation concealment. The studies were all probably underpowered, with sample sizes from 14 to 31, which may partially explain the lack of clear effects.
Other considerations	The GDG did not have any other considerations.

14 Glossary

14.1 General terms

Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to conceal the randomised group allocation sequence from those involved in accepting people into the study. This is to prevent knowledge of allocation from influencing recruitment. For example, consider that the person recruiting people to the study knows the next participant is allocated to the intervention group, and also that the participant appears to have prognostic characteristics that could lead to a poor outcome. In such a situation the recruiter may decide to not recruit if he/she has a personal bias towards the intervention group. In order to avoid recruiters knowing the allocation sequence, the allocation process should therefore be administered by someone who is not responsible for recruiting participants.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive one particular intervention, for example placebo arm
Association	Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.
Baseline	The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs. This approach is very vulnerable to threats to internal validity.
Bias	Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The best way to do this is by sorting patients into study groups randomly. The purpose of 'blinding' or 'masking' is to protect against bias. A single-blinded study is one in which patients do not know which study group they are in (for example whether they are taking the experimental drug or a placebo). A double-blinded study is one in which neither patients nor the researchers/doctors know which study group the patients are in. A triple blind study is one in which neither the patients, clinicians or the people carrying out the statistical analysis know which treatment patients received.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	A study to find out the cause(s) of a disease or condition. This is done by

	comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may cause the condition. For example, a group of people with lung cancer might be compared with a group of people the same age that do not have lung cancer. The researcher could compare how long both groups had been exposed to tobacco smoke. Such studies are retrospective because they look back in time from the outcome to the possible causes of a disease or condition.
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.
Clinician	A healthcare professional who provides patient care. For example, a doctor, nurse or physiotherapist.
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and records what happens. See also observational study.
Comorbidity	A disease or condition that someone has in addition to the health problem being studied or treated.
Comparability	Similarity of the groups in characteristics likely to affect the study results (such as health status or age).
Concordance	This is a recent term whose meaning has changed. It was initially applied to the consultation process in which doctor and patient agree therapeutic decisions that incorporate their respective views, but now includes patient support in medicine taking as well as prescribing communication. Concordance reflects social values but does not address medicine-taking and may not lead to improved adherence.
Confidence interval (CI)	There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. The CI is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value. For example, a study may state that 'based on our sample findings, we are 95% certain that the 'true' population blood pressure is not higher than 150 and not lower than 110'. In such a case the 95% CI would be 110 to 150. A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment - often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for

	example, if a large number of patients have been studied).
Confounding factor	Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with. For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than exercise. Therefore age is a confounding factor.
Consensus methods	Techniques used to reach agreement on a particular issue. Consensus methods may be used to develop NICE guidance if there is not enough good quality research evidence to give a clear answer to a question. Formal consensus methods include Delphi and nominal group techniques.
Control group	A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any effects due to the treatment.
Cost–benefit analysis (CBA)	Cost-benefit analysis is one of the tools used to carry out an economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether the benefits exceed the costs.
Cost–consequences analysis (CCA)	Cost-consequence analysis is one of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with a suitable alternative. Unlike cost-benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is left to decision-makers to determine whether, overall, the treatment is worth carrying out.
Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost–utility analysis (CUA)	Cost-utility analysis is one of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
Credible interval (CrI)	The Bayesian equivalent of a confidence interval.
Decision analysis	An explicit quantitative approach to decision-making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.

Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be 'dominated' by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits - health effects - relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost-benefit analysis, cost-consequence analysis, cost-effectiveness analysis, cost-minimisation analysis and cost-utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extended dominance	If Option A is both more clinically effective than Option B and has a lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore more cost effective and should be preferred, other things remaining equal.
External validity	The extent to which the findings of a study can be generalised. For example, a study on children may have very limited external validity to an adult population.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.

Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
Health economics	Study or analysis of the cost of using and distributing healthcare resources.
Health-related quality of life (HRQoL)	A measure of the effects of an illness to see how it affects someone's day-to-day life.
Heterogeneity or Lack of homogeneity	The term is used in meta-analyses and systematic reviews to describe when the results of a test or treatment (or estimates of its effect) differ significantly in different studies. Such differences may occur as a result of differences in the populations studied the outcome measures used or because of different definitions of the variables involved. It is the opposite of homogeneity.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The extra cost linked to using one test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost-effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.
Incremental net benefit (INB)	The value (usually in monetary terms) of an intervention net of its cost compared with a comparator intervention. The INB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as: (£20,000 x QALYs gained) – Incremental cost.
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Internal validity	The extent to which changes in the dependent variable (usually a health outcome) is wholly caused by changes in the independent variable (usually the choice of treatment). If the possibility of other causes of changes in the dependent variable exist (such as the natural course of a disease) then internal validity is reduced. A control group is a powerful means to improve internal validity.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions

	could include action to help someone to be physically active or to eat a more healthy diet.
Intraoperative	The period of time during a surgical procedure.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by (1 minus specificity).
Long-term care	Residential care in a home that may include skilled nursing care and help with everyday activities. This includes nursing homes and residential homes.
Loss to follow-up	Loss to follow up refers to the loss of outcome data from patients who are unwilling or unable to attend outcome assessment sessions post intervention (see missing data).
Markov model	A method for estimating long-term costs and effects for recurrent or chronic conditions, based on health states and the probability of transition between them within a given time period (cycle).
Meta-analysis	A method often used in systematic reviews. Results from several studies of the same test or treatment are combined to estimate the overall effect of the treatment.
Missing data	Sometimes outcome data may be missing from some participants in a study. This may be due to action of researchers (for example excluding participants who do not comply with the intervention) or may be due to the choice of participants (unwilling to attend outcome assessment sessions after treatment). Missing data may lead to the study results being different to those that would have been gained had the data not been lost. This is especially likely if data is lost for reasons related to outcome. For example, people may drop out of treatment and refuse to attend follow up sessions if they responded poorly to a treatment. If there are systematic differences between groups in data lost for reasons related to outcome then there will be serious risk of bias.
Multivariate model	A statistical model for analysis of the relationship between 2 or more predictor (independent) variables and the outcome (dependent) variable.
Negative predictive value (NPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a negative test result who do not have the disease, and can be interpreted as the probability that a negative test result is correct. It is calculated as follows: true negative/(true negative + false negative)
Number needed to treat (NNT)	The average number of patients who need to be treated to get a positive outcome. For example, if the NNT is 4, then 4 patients would have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment. For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example,

	whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in one group with the probability of the same thing in another. An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group. Sometimes probability can be compared across more than 2 groups - in this case, one of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, relative risk, risk ratio.
Opportunity cost	The loss of other health care programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Perioperative	The period from admission through surgery until discharge, encompassing the pre-operative and post-operative periods.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had - over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
Positive predictive value	In screening or diagnostic tests: A measure of the usefulness of a screening

(PPV)	or diagnostic test. It is the proportion of those with a positive test result who have the disease, and can be interpreted as the probability that a positive test result is correct. It is calculated as follows: True positive / (true positive + false positive)
Postoperative	Pertaining to the period after patients leave the operating theatre, following surgery.
Post-test probability	In diagnostic tests: The proportion of patients with that particular test result who have the target disorder (post-test odds/[1 plus post-test odds]).
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Preoperative	The period before surgery commences.
Pre-test probability	In diagnostic tests: The proportion of people with the target disorder in the population at risk at a specific time point or time interval. Prevalence may depend on how a disorder is diagnosed.
Primary care	Healthcare delivered outside hospitals. It includes a range of services provided by GPs, nurses, health visitors, midwives and other healthcare professionals and allied health professionals such as dentists, pharmacists and opticians.
Primary outcome	The outcome of greatest importance, usually the one in a study that the power calculation is based on.
Product licence	An authorisation from the MHRA to market a medicinal product.
Prognosis	A probable course or outcome of a disease. Prognostic factors are patient or disease characteristics that influence the course. Good prognosis is associated with low rate of undesirable outcomes; poor prognosis is associated with a high rate of undesirable outcomes.
Prospective study	A research study in which the health or other characteristic of participants is monitored (or 'followed up') for a period of time, with events recorded as they happen. This contrasts with retrospective studies.
Publication bias	Publication bias occurs when researchers publish the results of studies showing that a treatment works well and don't publish those showing it did not have any effect. If this happens, analysis of the published results will not give an accurate idea of how well the treatment works. This type of bias can be assessed by a funnel plot.
Quality of life	See ‘Health-related quality of life’.
Quality-adjusted life year (QALY)	A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to 1 year of life in perfect health. QALYS are calculated by estimating the years of life remaining for a patient following a particular treatment or intervention and weighting each year with a quality of life score (on a scale of 0 to 1). It is often measured in terms of the person's ability to perform the activities of daily life, freedom from pain and mental disturbance.
Randomisation	Assigning participants in a research study to different groups without taking any similarities or differences between them into account. For example, it could involve using a random numbers table or a computer-generated random sequence. It means that each individual (or each group in the case of cluster randomisation) has the same chance of receiving each intervention.
Randomised controlled trial (RCT)	A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug or treatment. One group (the experimental group) receives the treatment being tested, the other (the comparison or control group) receives an alternative treatment, a dummy

	treatment (placebo) or no treatment at all. The groups are followed up to see how effective the experimental treatment was. Outcomes are measured at specific times and any difference in response between the groups is assessed. This method is used as the gold standard method of to reducing bias in comparisons of interventions.
RCT	See 'Randomised controlled trial'.
Receiver operated characteristic (ROC) curve	A graphical method of assessing the accuracy of a diagnostic test. Sensitivity is plotted on the y axis against the false positive rate (1 minus specificity) on the x axis at different diagnostic thresholds of the test. A perfect test will have a line adhering to the left and top sides of the graph and passing through the point at the top left hand corner corresponding to perfect sensitivity and specificity. A good test will be somewhere close to this ideal.
Reference standard	The test that is considered to be the best available method to establish the presence or absence of the outcome – this may not be the one that is routinely used in practice.
Relative risk (RR)	The ratio of the risk of disease or death among those exposed to certain conditions compared with the risk for those who are not exposed to the same conditions (for example, the risk of people who smoke getting lung cancer compared with the risk for people who do not smoke). If both groups face the same level of risk, the relative risk is 1. If the first group had a relative risk of 2, subjects in that group would be twice as likely to have the event happen. A relative risk of less than one means the outcome is less likely in the first group. Relative risk is sometimes referred to as risk ratio.
Reporting bias	See 'Publication bias'.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective study	A research study that focuses on the past and present. The study examines past exposure to suspected risk factors for the disease or condition. Unlike prospective studies, it does not cover events that occur after the study group is selected.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	Selection bias occurs if: a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn, or b) There are differences between groups of participants in a study in terms of how likely they are to get better.
Sensitivity	Sensitivity of a test for a disorder measures the probability of that test being positive if you really do have the disorder. It is thus a measure of how well a test detects the disorder it is testing for. If a diagnostic test for a disease has high sensitivity, it is likely to pick up most cases of the disease in people who have it (that is, give a 'true positive' result). If a test is too sensitive it may sometimes also give a positive result in people who don't have the disease (that is, give a 'false positive'). For example, if a test were developed to detect if a woman is 6 months pregnant, a very sensitive test would detect everyone who was 6 months pregnant, but would probably also include those who are 5 and 7 months pregnant. If the same test were more specific (sometimes referred to as having higher specificity), it would detect only those who are 6 months pregnant, and someone who was 5 months pregnant would get a negative result (a 'true

	negative'). But it would probably also miss some people who were 6 months pregnant (that is, give a 'false negative'). Breast screening is a 'real-life' example. The number of women who are recalled for a second breast screening test is relatively high because the test is very sensitive. If it were made more specific, people who don't have the disease would be less likely to be called back for a second test but more women who have the disease would be missed.
Sensitivity analysis	A means of representing uncertainty in the results of economic evaluations. Uncertainty may arise from missing data, imprecise estimates or methodological controversy. Sensitivity analysis also allows for exploring the generalisability of results to other settings. The analysis is repeated using different assumptions to examine the effect on the results. One-way simple sensitivity analysis (univariate analysis): each parameter is varied individually in order to isolate the consequences of each parameter on the results of the study. Multi-way simple sensitivity analysis (scenario analysis): 2 or more parameters are varied at the same time and the overall effect on the results is evaluated. Threshold sensitivity analysis: the critical value of parameters above or below which the conclusions of the study will change are identified. Probabilistic sensitivity analysis: probability distributions are assigned to the uncertain parameters and are incorporated into evaluation models based on decision analytical techniques (for example, Monte Carlo simulation).
Significance (statistical)	A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 ($p<0.05$).
Specificity	The proportion of true negatives that are correctly identified as such. For example in diagnostic testing the specificity is the proportion of non-cases correctly diagnosed as non-cases. See related term ‘Sensitivity’. In terms of literature searching a highly specific search is generally narrow and aimed at picking up the key papers in a field and avoiding a wide range of papers.
Stakeholder	An organisation with an interest in a topic that NICE is developing a clinical guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: • manufacturers of drugs or equipment • national patient and carer organisations • NHS organisations • organisations representing healthcare professionals.
Systematic review	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.
Time horizon	The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.
Treatment allocation	Assigning a participant to a particular arm of a trial.
Univariate	Analysis which separately explores each variable in a data set.
Utility	In health economics, a 'utility' is the measure of the preference or value that an individual or society places upon a particular health state. It is generally a number between 0 (representing death) and 1 (perfect health). The most widely used measure of benefit in cost-utility analysis is the quality-adjusted life year, but other measures include disability-adjusted life years (DALYs) and healthy year equivalents (HYEs).

14.2 MS related terms

Ataxia	A sign consisting of reduced voluntary co-ordination of muscle movements
Central Nervous System	The brain and spinal cord.
Clinically isolated syndrome	A person’s first neurological episode, caused by inflammation or demyelination of nerve tissue.
Cognition	Capacity to engage in mental activity essential for normal functioning
Disease modifying treatments / disease modifying drugs	Drugs designed to alter the course of MS. They are usually designed to reduce the frequency of relapses in relapsing remitting MS, and may therefore slow down functional deterioration.
Emotional lability	Involuntary laughing and crying related to a brain stem lesion
emotionalism	Involuntary laughing and crying related to a brain stem lesion
Exacerbation	In MS this refers to a relapse, a sudden onset of signs and symptoms due to a focal demyelinating lesion in the central nervous system, that usually resolves, partially or completely, within days to weeks.
L’hermittes sign	An electrical sensation that runs down the back and into the limbs. In many patients, it is elicited by bending the head forward.
McDonald criteria	These are diagnostic criteria for MS, involving consideration of attacks of neurological deterioration, objective clinical lesions and MRI findings.
Nystagmus	The involuntary horizontal and/or vertical movements of the eyes that may occur in response to disorders of balance.
Optic neuritis	A lesion of the optic nerve causing partial or complete loss of vision, blurring or selective dimming of colours in one eye. It is often accompanied by pain and may resolve in days or weeks.
Oscillopsia	The subjective sensation of horizontal and/or vertical movement of the visual field that is unexplained by movement of the observer or environment.
Primary progressive MS	A form of MS where, from the onset of the disease, the person experiences a continuous neurological deterioration without any periods of remission
Progressive MS	A form of MS where from the onset of the disease, or after an initial period of relapses and remissions, the person experiences a continuous neurological deterioration without any periods of remission
Pseudobulbar affect	Involuntary laughing and crying related to a brain stem lesion
Relapse	A sudden onset of signs and symptoms due to a focal demyelinating lesion in the central nervous system, that usually resolves, partially or completely, within days to weeks. These usually occur no more than twice a year.
Relapsing remitting MS	The most common form of MS, where relapses are followed by periods of partial or complete neurological recovery.
Secondary progressive MS	A form of MS where, after an initial period of relapses and remissions, the person experiences a continuous neurological deterioration without any periods of remission
Spasticity	Increased stiffness (tone) of muscles, often in response to a central nervous system lesion upsetting the normal input of excitatory and inhibitory activity into voluntary muscle cells.
Tremor	Rhythmic unsteadiness of muscle activity that may translate into trembling of extremities such as the hands or legs.

15 References

1 Ackerman KD, Heyman R, Rabin BS, Anderson BP, Houck PR, Frank E et al. Stressful life events precede exacerbations of multiple sclerosis. Psychosomatic Medicine. 2002; 64(6):916-920

2 Ahmadi, Nikbakh, Arastoo A, ., Habibi A-H. The Effects of a Yoga Intervention on Balance, Speed and Endurance of Walking, Fatigue and Quality of Life in People with Multiple Sclerosis. Journal of Human Kinetics. 2010; 23(-1):71-78

3 Ahmadi A, Arastoo AA, Nikbakht M, Zahednejad S, Rajabpour M. Comparison of the Effect of 8 weeks Aerobic and Yoga Training on Ambulatory Function, Fatigue and Mood Status in MS Patients. Iranian Red Crescent Medical Journal. 2013; 15(6):449-454

4 Al-Smadi J, Warke K, Wilson I, Cramp AFL, Noble G, Walsh DM et al. A pilot investigation of the hypoalgesic effects of transcutaneous electrical nerve stimulation upon low back pain in people with multiple sclerosis. Clinical Rehabilitation. 2003; 17(7):742-749

5 Alam SM, Kyriakides T, Lawden M, Newman PK. Methylprednisolone in multiple sclerosis: a comparison of oral with intravenous therapy at equivalent high dose. Journal of Neurology, Neurosurgery, and Psychiatry. 1993; 56(11):1219-1220

6 Amato MP, Goretti B, Viterbo RG, Portaccio E, Niccolai C, Hakiki B et al. Computer-assisted rehabilitation of attention in patients with multiple sclerosis: Results of a randomized, double-blind trial. Multiple Sclerosis. 2014; 20(1):91-98

7 Andreassen S, Wyller TB. Patients' experiences with self-referral to in-patient rehabilitation: A qualitative interview study. Disability and Rehabilitation. 2005; 27(21):1307-1313

8 Ara R, Brazier J. Deriving an algorithm to convert the eight mean SF-36 dimension scores into a mean EQ-5D preference-based score from published studies (where patient level data are not available). Value in Health. 2008; 11(7):1131-1143

9 Aragona M, Onesti E, Tomassini V, Conte A, Gupta S, Gilio F et al. Psychopathological and cognitive effects of therapeutic cannabinoids in multiple sclerosis: a double-blind, placebo controlled, crossover study. Clinical Neuropharmacology. 2009; 32(1):41-47

10 Armutlu K, Karabudak R, Nurlu G. Physiotherapy approaches in the treatment of ataxic multiple sclerosis: a pilot study. Neurorehabilitation and Neural Repair. 2001; 15(3):203-211

11 Averbuch-Heller L, Tusa RJ, Fuhry L, Rottach KG, Ganser GL, Heide W et al. A double-blind controlled study of gabapentin and baclofen as treatment for acquired nystagmus. Annals of Neurology. 1997; 41(6):818-825

12 Baker K, Cassidy E, Rone-Adams S. Therapeutic standing for people with multiple sclerosis: efficacy and feasibility. International Journal of Therapy & Rehabilitation. 2007; 14(3):104-109

13 **Baker LM. Sense making in multiple sclerosis: The information needs of people during an acute exacerbation. Qualitative Health Research. 1998; 8(1):106-120**

14 **Bamford CR, Sibley WA, Laguna JF. Swine influenza vaccination in patients with multiple sclerosis. Archives of Neurology. 1978; 35(4):242-243**

15 **Bandini F, Castello E, Mazzella L, Mancardi GL, Solaro C. Gabapentin but not vigabatrin is effective in the treatment of acquired nystagmus in multiple sclerosis: How valid is the GABAergic hypothesis? Journal of Neurology, Neurosurgery, and Psychiatry. 2001; 71(1):107-110**

16 **Barnes D, Hughes RA, Morris RW, Wade-Jones O, Brown P, Britton T et al. Randomised trial of oral and intravenous methylprednisolone in acute relapses of multiple sclerosis. Lancet. 1997; 349(9056):902-906**

17 **Barnes MP, Bateman DE, Cleland PG, Dick DJ, Walls TJ, Newman PK et al. Intravenous methylprednisolone for multiple sclerosis in relapse. Journal of Neurology, Neurosurgery and Psychiatry. 1985; 48(2):157-159**

18 **Barton JJ, Huaman AG, Sharpe JA. Muscarinic antagonists in the treatment of acquired pendular and downbeat nystagmus: a double-blind, randomized trial of three intravenous drugs. Annals of Neurology. 1994; 35(3):319-325**

19 **Bates D, Cartlidge NE, French JM, Jackson MJ, Nightingale S, Shaw DA et al. A double-blind controlled trial of long chain n-3 polyunsaturated fatty acids in the treatment of multiple sclerosis. Journal of Neurology, Neurosurgery, and Psychiatry. 1989; 52(1):18-22**

20 **Bates D, Fawcett PR, Shaw DA, Weightman D. Trial of polyunsaturated fatty acids in non-relapsing multiple sclerosis. BMJ. 1977; 2(6092):932-933**

21 **Bates D, Fawcett PR, Shaw DA, Weightman D. Polyunsaturated fatty acids in treatment of acute remitting multiple sclerosis. BMJ. 1978; 2(6149):1390-1391**

22 **Bever CTJ, Young D, Anderson PA, Krumholz A, Conway K, Leslie J et al. The effects of 4-aminopyridine in multiple sclerosis patients: results of a randomized, placebo-controlled, double-blind, concentration-controlled, crossover trial. Neurology. 1994; 44(6):1054-1059**

23 **Bjarnadottir OH, Konradsdottir AD, Reynisdottir K, Olafsson E. Multiple sclerosis and brief moderate exercise. A randomised study. Multiple Sclerosis. 2007; 13(6):776-782**

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

25 **Bowen C, MacLehose A, Beaumont J. Advanced multiple sclerosis and the psychosocial impact on families. Psychology & Health. 2011; 26(1):113-127**

26 **Bozek CB, Kastrukoff LF, Wright JM, Perry TL, Larsen TA. A controlled trial of isoniazid therapy for action tremor in multiple sclerosis. Journal of Neurology. 1987; 234(1):36-39**

27 **Brar SP, Smith MB, Nelson LM, Franklin GM, Cobble ND. Evaluation of treatment protocols on minimal to moderate spasticity in multiple sclerosis. Archives of Physical Medicine and Rehabilitation. 1991; 72(3):186-189**

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

29 **Brown RF, Tennant CC, Sharrock M, Hodgkinson S, Dunn SM, Pollard JD. Relationship between stress and relapse in multiple sclerosis: Part I. Important features. Multiple Sclerosis. 2006; 12(4):453-464**

30 **Brown RF, Tennant CC, Sharrock M, Hodgkinson S, Dunn SM, Pollard JD. Relationship between stress and relapse in multiple sclerosis: Part II. Direct and indirect relationships. Multiple Sclerosis. 2006; 12(4):465-475**

31 **Buljevac D, Hop WCJ, Reedeker W, Janssens ACJW, van der Meche FGA, van Doorn PA et al. Self reported stressful life events and exacerbations in multiple sclerosis: prospective study. BMJ. 2003; 327(7416):646**

32 **Burton JM, Kimball S, Vieth R, Bar-Or A, Dosch HM, Cheung R et al. A phase I/II dose-escalation trial of vitamin D3 and calcium in multiple sclerosis. Neurology. 2010; 74(23):1852-1859**

33 **Burton JM, O'Connor PW, Hohol M, Beyene J. Oral versus intravenous steroids for treatment of relapses in multiple sclerosis. Cochrane Database of Systematic Reviews (Online). 2012; 12:CD006921**

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

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

36 **Castro-Sanchez AM, Mataran-Penarrocha GA, Lara-Palomo I, Saavedra-Hernandez M, Arroyo-Morales M, Moreno-Lorenzo C. Hydrotherapy for the treatment of pain in people with multiple sclerosis: a randomized controlled trial. Evidence-Based Complementary and Alternative Medicine. 2012; 2012:473963**

37 **Cerasa A, Gioia MC, Valentino P, Nistico R, Chiriaco C, Pirritano D et al. Computer-assisted cognitive rehabilitation of attention deficits for multiple sclerosis: a randomized trial with FMRI correlates. Neurorehabilitation and Neural Repair. 2013; 27(4):284-295**

38 **Chataway J, Porter B, Riazi A, Heaney D, Watt H, Hobart J et al. Home versus outpatient administration of intravenous steroids for multiple-sclerosis relapses: a randomised controlled trial. Lancet Neurology. United Kingdom 2006; 5(7):565-571**

39 Chiaravalloti ND, DeLuca J, Moore NB, Ricker JH. Treating learning impairments improves memory performance in multiple sclerosis: a randomized clinical trial. Multiple Sclerosis. 2005; 11(1):58-68

40 Chiaravalloti ND, Moore NB, Nikelshpur OM, DeLuca J. An RCT to treat learning impairment in multiple sclerosis: The MEMREHAB trial. Neurology. 2013; 81(24):2066-2072

41 Claerbout M, Gebara B, Ilsbroukx S, Verschueren S, Peers K, Van Asch P et al. Effects of 3 weeks' whole body vibration training on muscle strength and functional mobility in hospitalized persons with multiple sclerosis. Multiple Sclerosis. 2012; 18(4):498-505

42 Cohen RA, Fisher M. Amantadine treatment of fatigue associated with multiple sclerosis. Archives of Neurology. 1989; 46(6):676-680

43 Collin C, Davies P, Mutiboko IK, Ratcliffe S. Randomized controlled trial of cannabis-based medicine in spasticity caused by multiple sclerosis. European Journal of Neurology. 2007; 14(3):290-296

44 Collin C, Ehler E, Waberzinek G, Alsindi Z, Davies P, Powell K et al. A double-blind, randomized, placebo-controlled, parallel-group study of Sativex, in subjects with symptoms of spasticity due to multiple sclerosis. Neurological Research. 2010; 32(5):451-459

45 Courts NF, Newton AN, McNeal LJ. Husbands and wives living with multiple sclerosis. Journal of Neuroscience Nursing. 2005; 37(1):20-27

46 Curtis L. Unit costs of health and social care. Canterbury: Personal Social Services Research Unit, University of Kent; 2012. Available from: http://www.pssru.ac.uk/archive/pdf/uc/uc2012/full-with-covers.pdf

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

48 Cutter NC, Scott DD, Johnson JC, Whiteneck G. Gabapentin effect on spasticity in multiple sclerosis: a placebo-controlled, randomized trial. Archives of Physical Medicine and Rehabilitation. 2000; 81(2):164-169

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

50 DeBolt LS, McCubbin JA. The effects of home-based resistance exercise on balance, power, and mobility in adults with multiple sclerosis. Archives of Physical Medicine and Rehabilitation. 2004; 85(2):290-297

51 Department of Health. NHS reference costs 2011-2012. 2012. [Last accessed: 24 March 2014]

52 Department of Health. NHS reference costs 2012-2013. 2013. [Last accessed: 24 March 2014]

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

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

55 Donnellan CP, Shanley J. Comparison of the effect of two types of acupuncture on quality of life in secondary progressive multiple sclerosis: a preliminary single-blind randomized controlled trial. Clinical Rehabilitation. 2008; 22(3):195-205

56 Durelli L, Baggio GF, Bergamasco B, Barile C, Cocito D, Delsedime M et al. Early immunosuppressive effect of parenteral methylprednisolone in the treatment of multiple sclerosis. Acta Neurologica. 1985; 7(3-4):338-344

57 Durelli L, Cocito D, Riccio A, Barile C, Bergamasco B, Baggio GF et al. High-dose intravenous methylprednisolone in the treatment of multiple sclerosis: clinical-immunologic correlations. Neurology. 1986; 36(2):238-243

58 Ehde DM, Kraft GH, Chwastiak L, Sullivan MD, Gibbons LE, Bombardier CH et al. Efficacy of paroxetine in treating major depressive disorder in persons with multiple sclerosis. General Hospital Psychiatry. 2008; 30(1):40-48

59 Eyssette M, Rohmer F, Serratrice G, Warter JM, Boisson D. Multi-centre, double-blind trial of a novel antispastic agent, tizanidine, in spasticity associated with multiple sclerosis. Current Medical Research and Opinion. 1988; 10(10):699-708

60 Filipovic SR, Drulovic J, Stojsavljevic N, Levic Z. The effects of high-dose intravenous methylprednisolone on event-related potentials in patients with multiple sclerosis. Journal of the Neurological Sciences. 1997; 152(2):147-153

61 Fink F, Rischkau E, Butt M, Klein J, Eling P, Hildebrandt H. Efficacy of an executive function intervention programme in MS: a placebo-controlled and pseudo-randomized trial. Multiple Sclerosis. 2010; 16(9):1148-1151

62 Finlayson M. Pilot study of an energy conservation education program delivered by telephone conference call to people with multiple sclerosis. Neurorehabilitation. 2005; 20(4):267-277

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

64 Flachenecker P. A new multiple sclerosis spasticity treatment option: Effect in everyday clinical practice and cost-effectiveness in Germany. Expert Review of Neurotherapeutics. 2013; 13(3 SUPPL. 1):15-19

65 Flavia M, Stampatori C, Zanotti D, Parrinello G, Capra R. Efficacy and specificity of intensive cognitive rehabilitation of attention and executive functions in multiple sclerosis. Journal of the Neurological Sciences. 2010; 288(1-2):101-105

66 Forbes A, While A, Dyson L, Grocott T, Griffiths P. Impact of clinical nurse specialists in multiple sclerosis--synthesis of the evidence. Journal of Advanced Nursing. 2003; 42(5):442-462

67 Forbes A, While A, Mathes L, Griffiths P. Evaluation of a MS specialist nurse programme. International Journal of Nursing Studies. 2006; 43(8):985-1000

68 Francabandera FL, Holland NJ, Wiesel-Levison P, Scheinberg LC. Multiple sclerosis rehabilitation: inpatient vs. outpatient. Rehabilitation Nursing : the Official Journal of the Association of Rehabilitation Nurses. 1988; 13(5):251-253

69 From A, Heltberg A. A double-blind trial with baclofen (Lioresal) and diazepam in spasticity due to multiple sclerosis. Acta Neurologica Scandinavica. 1975; 51(2):158-166

70 Fuller KJ, Dawson K, Wiles CM. Physiotherapy in chronic multiple sclerosis: A controlled trial. Clinical Rehabilitation. 1996; 10(3):195-204

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

72 Garcia Jalon EG, Lennon S, Hannan J, Murphy S, Lowe-Strong A. Energy conservation for people with MS-related fatigue: a pilot randomized controlled trial [corrected] [published erratum appears in PHYSIOTHER RES INT 2008 Dec;13(4):217]. Physiotherapy Research International. 2008; 13(3):139-140

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

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

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

76 Geisler MW, Sliwinski M, Coyle PK, Masur DM, Doscher C, Krupp LB. The effects of amantadine and pemoline on cognitive functioning in multiple sclerosis. Archives of Neurology. 1996; 53(2):185-188

77 Gelenberg AJ, Poskanzer DC. The effect of dantrolene sodium on spasticity in multiple sclerosis. Neurology. 1973; 23(12):1313-1315

78 Gervasoni E, Cattaneo D, Jonsdottir J. Effect of treadmill training on fatigue in multiple sclerosis: a pilot study. International Journal of Rehabilitation Research Internationale

Zeitschrift Fur Rehabilitationsforschung Revue Internationale De Recherches De Readaptation. 2014; 37(1):54-60

79 Gibson RG, Gibson SL. Neural therapy in the treatment of multiple sclerosis. Journal of Alternative and Complementary Medicine. 1999; 5(6):543-552

80 Golan D, Halhal B, Glass-Marmor L, Staun-Ram E, Rozenberg O, Lavi I et al. Vitamin D supplementation for patients with multiple sclerosis treated with interferon-beta: A randomized controlled trial assessing the effect on flu-like symptoms and immunomodulatory properties. BMC Neurology. 2013; 13(60)

81 Goodman AD, Brown TR, Cohen JA, Krupp LB, Schapiro R, Schwid SR et al. Dose comparison trial of sustained-release fampridine in multiple sclerosis. Neurology. 2008; 71(15):1134-1141

82 Goodman AD, Brown TR, Edwards KR, Krupp LB, Schapiro RT, Cohen R et al. A phase 3 trial of extended release oral dalfampridine in multiple sclerosis. Annals of Neurology. 2010; 68(4):494-502

83 Goodman AD, Brown TR, Krupp LB, Schapiro RT, Schwid SR, Cohen R et al. Sustained-release oral fampridine in multiple sclerosis: a randomised, double-blind, controlled trial. Lancet. 2009; 373(9665):732-738

84 Goodman AD, Cohen JA, Cross A, Vollmer T, Rizzo M, Cohen R et al. Fampridine-SR in multiple sclerosis: a randomized, double-blind, placebo-controlled, dose-ranging study. Multiple Sclerosis. 2007; 13(3):357-368

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

86 Gusev YI, Banach M, Simonow A, Skoromets A, Czlonkowska A, Shmidt T et al. Efficacy and safety of botulinum type a toxin in adductor spasticity due to multiple sclerosis. Journal of Musculoskeletal Pain. 2008; 16(3):175-188

87 Hader W, Duquette P, Auty A, Hashimoto S, Noseworthy J, Sawa G et al. A randomized controlled trial of amantadine in fatigue associated with multiple sclerosis. Canadian Journal of Neurological Sciences. 1987; 14(3):273-278

88 Hallett M, Lindsey JW, Adelstein BD, Riley PO. Controlled trial of isoniazid therapy for severe postural cerebellar tremor in multiple sclerosis. Neurology. 1985; 35(9):1374-1377

89 Harvey L, Smith AD, Jones R. The Effect of Weighted Leg Raises on Quadriceps Strength, EMG Parameters and Functional Activities in People with Multiple Sclerosis. Physiotherapy. 1999; 85(3):154-161

90 Hattori N, Hirayama T, Katayama Y. Cost-effectiveness analysis of intrathecal baclofen therapy in Japan. Neurologia Medico-Chirurgica. 2012; 52(7):482-487

91 Hawton A, Green C, Telford CJ, Wright DE, Zajicek JP. The use of multiple sclerosis condition-specific measures to inform health policy decision-making: mapping from the MSWS-12 to the EQ-5D. Mult Scler. 2012; 18(6):853-861

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

93 Hayes KC. Impact of extended-release dalfampridine on walking ability in patients with multiple sclerosis. Neuropsychiatric Disease and Treatment. 2011; 7:229-239

94 Haymarket Media Group. Monthly Index of Medical Specialities. 2013. Available from: http://www.mims.co.uk/home/ [Last accessed: 3 June 13 A.D.]

95 Healy BC, Ali EN, Guttmann CRG, Chitnis T, Glanz BI, Buckle G et al. Smoking and disease progression in multiple sclerosis. Archives of Neurology. 2009; 66(7):858-864

96 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. Annals of Neurology. 2012; 72:S34-S35

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

98 Hildebrandt H, Lanz M, Hahn HK, Hoffmann E, Schwarze B, Schwendemann G et al. Cognitive training in MS: effects and relation to brain atrophy. Restorative Neurology and Neuroscience. 2007; 25(1):33-43

99 Hoogstraten MC, Van Der Ploeg RJO, Burg W, Vreeling A, Van MS, Minderhoud JM. Tizanidine versus baclofen in the treatment of spasticity in multiple sclerosis patients. Acta Neurologica Scandinavica. 1988; 77(3):224-230

100 Hugenholtz H, Nelson RF, Dehoux E, Bickerton R. Intrathecal baclofen for intractable spinal spasticity--a double-blind cross-over comparison with placebo in 6 patients. Canadian Journal of Neurological Sciences. 1992; 19(2):188-195

101 Hughes CM, Smyth S, Lowe-Strong AS. Reflexology for the treatment of pain in people with multiple sclerosis: a double-blind randomised sham-controlled clinical trial. Multiple Sclerosis. 2009; 15(11):1329-1338

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

103 Hyman N, Barnes M, Bhakta B, Cozens A, Bakheit M, Kreczy K et al. Botulinum toxin (Dysport) treatment of upper leg adductor spasticity in multiple sclerosis: a prospective, randomised, double-blind, placebo controlled, dose ranging study. European Journal of Neurology. 1997; 4 Suppl 1:S82

104 Hyman N, Barnes M, Bhakta B, Cozens A, Bakheit M, Kreczy-Kleedorfer B et al. Botulinum toxin (Dysport) treatment of hip adductor spasticity in multiple sclerosis: a prospective, randomised, double blind, placebo controlled, dose ranging study. Journal of Neurology, Neurosurgery, and Psychiatry. 2000; 68(6):707-712

105 Jansen DEMC, Krol B, Groothoff JW, Post D. Evaluation of a transmural care model for multiple sclerosis patients. Journal of Neuroscience Nursing. 2006; 38(5):384-389

106 Johnson J. On receiving the diagnosis of multiple sclerosis: managing the transition. Multiple Sclerosis. 2003; 9(1):82-88

107 Johnson JGL and Smith P. Evaluation of MS Specialist Nurses: a review and development of the role: Part 1 - report of the National Survey of Mutliple Sclerosis (MS) Specialist Nurses in the United Kingdom. London. South Bank University/MS Resarch Trust, 2001. Available from: http://www.mstrust.org.uk/downloads/part1.pdf

108 Johnson JGL, Smith P, and Goldstone L. Evaluation of MS Specialist Nurses: a review and development of the role: Part 2: Case study of a new multiple sclerosis (MS) specialist nurse service in West Berkshire, England. London. South Bank University/MS Resarch Trust, 2001

109 Jonsson A, Korfitzen EM, Heltberg A, Ravnborg MH, Byskov-Ottosen E. Effects of neuropsychological treatment in patients with multiple sclerosis. Acta Neurologica Scandinavica. 1993; 88(6):394-400

110 Kampman MT, Steffensen LH, Mellgren SI, Jorgensen L. Effect of vitamin D3 supplementation on relapses, disease progression, and measures of function in persons with multiple sclerosis: exploratory outcomes from a double-blind randomised controlled trial. Multiple Sclerosis. 2012; 18(8):1144-1151

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

112 Keser I, Kirdi N, Meric A, Kurne AT, Karabudak R. Comparing routine neurorehabilitation program with trunk exercises based on bobath concept in multiple sclerosis: Pilot study. Journal of Rehabilitation Research and Development. 2013; 50(1):133-140

113 Kirker SGB, Young E, Warlow CP. An evaluation of a multiple sclerosis liaison nurse. Clinical Rehabilitation. 1995; 9(3):219-226

114 Koopman W. Needs assessment of persons with multiple sclerosis and significant others: using the literature review and focus groups for preliminary survey questionnaire development. Axon. 2003; 24(4):10-15

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

116 Kroll T, Neri MT. Experiences with care co-ordination among people with cerebral palsy, multiple sclerosis, or spinal cord injury. Disability and Rehabilitation. 2003; 25(19):1106-1114

117 Krupp LB, Coyle PK, Doscher C, Miller A, Cross AH, Jandorf L et al. Fatigue therapy in multiple sclerosis: results of a double-blind, randomized, parallel trial of amantadine, pemoline, and placebo. Neurology. 1995; 45(11):1956-1961

118 Lapierre Y, Bouchard S, Tansey C, Gendron D, Barkas WJ, Francis GS. Treatment of spasticity with tizanidine in multiple sclerosis. Canadian Journal of Neurological Sciences. 1987; 14(3 Suppl):513-517

119 Lappin MS, Lawrie FW, Richards TL, Kramer ED. Effects of a pulsed electromagnetic therapy on multiple sclerosis fatigue and quality of life: A double-blind, placebo controlled trial. Alternative Therapies in Health and Medicine. 2003; 9(4):38-48

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

121 Ledinek AH, Sajko MC, Rot U. Evaluating the effects of amantadin, modafinil and acetyl-L-carnitine on fatigue in multiple sclerosis--result of a pilot randomized, blind study. Clinical Neurology and Neurosurgery. 2013; 115 Suppl 1:S86-S89

122 Leigh RJ, Burnstine TH, Ruff RL, Kasmer RJ. Effect of anticholinergic agents upon acquired nystagmus: a double-blind study of trihexyphenidyl and tridihexethyl chloride. Neurology. 1991; 41(11):1737-1741

123 Lincoln NB, Dent A, Harding J, Weyman N, Nicholl C, Blumhardt LD et al. Evaluation of cognitive assessment and cognitive intervention for people with multiple sclerosis. Journal of Neurology, Neurosurgery, and Psychiatry. 2002; 72(1):93-98

124 Livesley E. Effects of Electrical Neuromuscular Stimulation on Functional Performance in Patients with Multiple Sclerosis. Physiotherapy. 1992; 78(12):914-917

125 Lord SE, Wade DT, Halligan PW. A comparison of two physiotherapy treatment approaches to improve walking in multiple sclerosis: a pilot randomized controlled study. Clinical Rehabilitation. 1998; 12(6):477-486

126 Loubser PG, Narayan RK, Sandin KJ, Donovan WH, Russell KD. Continuous infusion of intrathecal baclofen: long-term effects on spasticity in spinal cord injury. Paraplegia. 1991; 29(1):48-64

127 Lu L, Pearce H, Roome C, Shearer J, Lang IA, Stein K. Cost Effectiveness of Oromucosal Cannabis-Based Medicine (Sativex(R)) for Spasticity in Multiple Sclerosis. Pharmacoeconomics. 2012; 30(12):1157-1171

128 Macdonell C, Sorensen R, Pozzilli G, Nagels D, Laplaud B, de Jong A et al. Long-term prolongued-release fampridine treatment and health related quality of life outcomes: 12-month analysis of the ENABLE study. 2013; P665

129 Malcomson KS, Lowe-Strong AS, Dunwoody L. What can we learn from the personal insights of individuals living and coping with Multiple Sclerosis? Disability and Rehabilitation. 2008; 30(9):662-674

130 Mantynen A, Rosti-Otajarvi E, Koivisto K, Lilja A, Huhtala H, Hamalainen P. Neuropsychological rehabilitation does not improve cognitive performance but reduces perceived cognitive deficits in patients with multiple sclerosis: A randomised, controlled, multi-centre trial. Multiple Sclerosis. 2014; 20(1):99-107

131 Martinelli V, Pulizzi A, Annovazzi P, Rocca MA, Bucello S, Esposito F et al. A single blind, randomised MRI study comparing high-dose oral and intravenous methylprednisolone in treating MS relapses. Mutliple Sclerosis. 2007; 13(Suppl 2):S32-S33

132 Martinelli V, Rocca MA, Annovazzi P, Pulizzi A, Rodegher M, Boneschi FM et al. A short-term randomized MRI study of high-dose oral vs intravenous methylprednisolone in MS. Neurology. 2009; 73(22):1842-1848

133 Masoudi R, Sharifi FA, Mobasheri M, Moghadasi J. Evaluating the effectiveness of using a progressive muscle relaxation technique in reducing the pain of multiple sclerosis patients. Journal of Musculoskeletal Pain. 2013; 21(4):350-357

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

135 Mattioli F, Stampatori C, Scarpazza C, Parrinello G, Capra R. Persistence of the effects of attention and executive functions intensive rehabilitation in relapsing remitting multiple sclerosis. Multiple Sclerosis and Related Disorders. 2012; 1(4):168-173

136 Mattioli F, Stampatori C, Zanotti D, Parrinello G, Capra R. Efficacy and specificity of intensive cognitive rehabilitation of attention and executive functions in multiple sclerosis. Journal of the Neurological Sciences. 2010; 288(1-2):101-105

137 Matuska KM, Erickson B. Lifestyle balance: How it is described and experienced by women with multiple sclerosis. Journal of Occupational Science. 2008; 15(1):20-26

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

139 McKeown LP, Porter-Armstrong AP, Baxter GD. Caregivers of people with multiple sclerosis: experiences of support. Multiple Sclerosis. 2004; 10(2):219-230

140 Mendozzi L, Pugnetti L, Motta A, Barbieri E, Gambini A, Cazzullo CL. Computer-assisted memory retraining of patients with multiple sclerosis. Italian Journal of Neurological Sciences.: Springer-Verlag. 1998; 19(6):S431-S438

141 Meythaler JM, DeVivo MJ, Hadley M. Prospective study on the use of bolus intrathecal baclofen for spastic hypertonia due to acquired brain injury. Archives of Physical Medicine and Rehabilitation. 1996; 77(5):461-466

142 Meythaler JM, Guin-Renfroe S, Brunner RC, Hadley MN. Intrathecal baclofen for spastic hypertonia from stroke. Stroke; a Journal of Cerebral Circulation. 2001; 32(9):2099-2109

143 Middel B, Kuipers-Upmeijer H, Bouma J, Staal M, Oenema D, Postma T et al. Effect of intrathecal baclofen delivered by an implanted programmable pump on health related quality of life in patients with severe spasticity. Journal of Neurology, Neurosurgery, and Psychiatry. 1997; 63(2):204-209

144 Millar JH, Zilkha KJ, Langman MJ, Wright HP, Smith AD, Belin J et al. Double-blind trial of linoleate supplementation of the diet in multiple sclerosis. BMJ. 1973; 1(5856):765-768

145 Miller AE, Morgante LA, Buchwald LY, Nutile SM, Coyle PK, Krupp LB et al. A multicenter, randomized, double-blind, placebo-controlled trial of influenza immunization in multiple sclerosis. Neurology. 1997; 48(2):312-314

146 Miller CM. The lived experience of relapsing multiple sclerosis: a phenomenological study. Journal of Neuroscience Nursing. 1997; 29(5):294-304

147 Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee JC et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemedicine Journal and E-Health. 2011; 17(1):5-13

148 MILLER H, NEWELL DJ, RIDLEY A. Multiple sclerosis. Treatment of acute exacerbations with corticotrophin (A.C.T.H.). Lancet. 1961; 2(7212):1120-1122

149 Miller L, Mattison P, Paul L, Wood L. The effects of transcutaneous electrical nerve stimulation (TENS) on spasticity in multiple sclerosis. Multiple Sclerosis. 2007; 13(4):527-533

150 Milligan NM, Newcombe R, Compston DA. A double-blind controlled trial of high dose methylprednisolone in patients with multiple sclerosis: 1. Clinical effects. Journal of Neurology, Neurosurgery, and Psychiatry. 1987; 50(5):511-516

151 Mitsonis CI, Zervas IM, Mitropoulos PA, Dimopoulos NP, Soldatos CR, Potagas CM et al. The impact of stressful life events on risk of relapse in women with multiple sclerosis: A prospective study. European Psychiatry. 2008; 23(7):497-504

152 Mohr DC, Goodkin DE, Bacchetti P, Boudewyn AC, Huang L, Marrietta P et al. Psychological stress and the subsequent appearance of new brain MRI lesions in MS. Neurology. 2000; 55(1):55-61

153 Mokhtarian F, Shirazian D, Morgante L, Miller A, Grob D, Lichstein E. Influenza virus vaccination of patients with multiple sclerosis. Multiple Sclerosis (Houndmills, Basingstoke, England). 1997; 3(4):243-247

154 Mori F, Codeca C, Kusayanagi H, Monteleone F, Buttari F, Fiore S et al. Effects of anodal transcranial direct current stimulation on chronic neuropathic pain in patients with multiple sclerosis. Journal of Pain. 2010; 11(5):436-442

155 Mori F, Ljoka C, Magni E, Codeca C, Kusayanagi H, Monteleone F et al. Transcranial magnetic stimulation primes the effects of exercise therapy in multiple sclerosis. Journal of Neurology. 2011; 258(7):1281-1287

156 Mosayebi G, Ghazavi A, Ghasami K, Jand Y, Kokhaei P. Therapeutic effect of vitamin D3 in multiple sclerosis patients. Immunological Investigations. 2011; 40(6):627-639

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

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

159 Motl RW, McAuley E. Association between change in physical activity and short-term disability progression in multiple sclerosis. Journal of Rehabilitation Medicine. 2011; 43(4):305-310

160 Murray TJ. Amantadine therapy for fatigue in multiple sclerosis. Canadian Journal of Neurological Sciences Le Journal Canadien Des Sciences Neurologiques. 1985; 12(3):251-254

161 Myers LW, Ellison GW, Lucia M, Novom S, Holevoet M, Madden D et al. Swine influenza virus vaccination in patients with multiple sclerosis. Journal of Infectious Diseases. 1977; 136 Suppl:S546-S554

162 Mynors G, Perman S, and Morse M. Defining the value of MS Specialist Nurses. UK. Multiple Scelrosis Trust, 2012

163 National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. 2nd edition. London: National Institute for Health and Clinical Excellence; 2008. Available from: http://www.nice.org.uk/media/C18/30/SVJ2PUBLICATION2008.pdf

164 National Institute for Health and Clinical Excellence. The guidelines manual. London: National Institute for Health and Clinical Excellence; 2012. Available from: http://publications.nice.org.uk/the-guidelines-manual-pmg6

165 National Institute for Health and Clinical Excellence. Stroke rehabilitation: long term rehabilitation after stroke, 2013. Available from: http://guidance.nice.org.uk/CG162/Guidance/pdf/English

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

167 Nielsen JF, Sinkjaer T, Jakobsen J. Treatment of spasticity with repetitive magnetic stimulation; a double-blind placebo-controlled study. Multiple Sclerosis. 1996; 2(5):227-232

168 Nilsagard Y, Denison E, Gunnarsson LG. Evaluation of a single session with cooling garment for persons with multiple sclerosis--a randomized trial. Disability and Rehabilitation Assistive Technology. 2006; 1(4):225-233

169 North East Treatment Advisory Group. Frampridine (Fampyra) in multiple sclerosis. Newcastle UK. NHS Regionl Drug & Therapeutics Centre, 2012

170 Notcutt W. Results of a randomised withdrawal study of subjects with spasticity due to multiple sclerosis who were receiving long-term Sativex. Multiple Sclerosis. 2009; 15(9 Suppl):S258

171 Notcutt W, Langford R, Davies P, Ratcliffe S, Potts R. Multiple Sclerosis. 2012; 18(2):219-228

172 Novotna A, Mares J, Ratcliffe S, Novakova I, Vachova M, Zapletalova O et al. A randomized, double-blind, placebo-controlled, parallel-group, enriched-design study of nabiximols* (Sativex((R))), as add-on therapy, in subjects with refractory spasticity caused by multiple sclerosis. European Journal of Neurology : the Official Journal of the European Federation of Neurological Societies. 2011; 18(9):1122-1131

173 Oliveri RL, Valentino P, Russo C, Sibilia G, Aguglia U, Bono F et al. Randomized trial comparing two different high doses of methylprednisolone in MS: a clinical and MRI study. Neurology. 1998; 50(6):1833-1836

174 Ordia JI, Fischer E, Adamski E, Spatz EL. Chronic intrathecal delivery of baclofen by a programmable pump for the treatment of severe spasticity. Journal of Neurosurgery. 1996; 85(3):452-457

175 Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2013. Available from: http://www.oecd.org/std/ppp [Last accessed: 3 June 13 A.D.]

176 Orsnes GB, Sørensen PS, Larsen TK, Ravnborg M. Effect of baclofen on gait in spastic MS patients. Acta Neurologica Scandinavica. 2000; 101(4):244-248

177 Paty DW. Double-blind trial of linoleic acid in multiple sclerosis. Archives of Neurology. 1983; 40(11):693-694

178 Paty DW, Cousin HK, Read S, Adlakha K. Linoleic acid in multiple sclerosis: failure to show any therapeutic benefit. Acta Neurologica Scandinavica. 1978; 58(1):53-58

179 Piatkowski J, Kern S, Ziemssen T. Effect of BEMER magnetic field therapy on the level of fatigue in patients with multiple sclerosis-A randomised, double-blind controlled trial. Multiple Sclerosis. 2009; 15(9 Suppl. S):S255-S256

180 Piatkowski J, Kern S, Ziemssen T. Effect of BEMER magnetic field therapy on the level of fatigue in patients with multiple sclerosis: a randomized, double-blind controlled trial. Journal of Alternative and Complementary Medicine. 2009; 15(5):507-511

181 Pittas F, Ponsonby AL, van der Mei IAF, Taylor BV, Blizzard L, Groom P et al. Smoking is associated with progressive disease course and increased progression in clinical disability in a prospective cohort of people with multiple sclerosis. Journal of Neurology. 2009; 256(4):577-585

182 Plow M, Bethoux F, McDaniel C, McGlynn M, Marcus B. Randomized controlled pilot study of customized pamphlets to promote physical activity and symptom self-management in women with multiple sclerosis. Clinical Rehabilitation. 2014; 28(2):139-148

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

184 Potagas C, Mitsonis C, Watier L, Dellatolas G, Retziou A, Mitropoulos P et al. Influence of anxiety and reported stressful life events on relapses in multiple sclerosis: a prospective study. Multiple Sclerosis. 2008; 14(9):1262-1268

185 Pozzilli C, Brunetti M, Amicosante AM, Gasperini C, Ristori G, Palmisano L et al. Home based management in multiple sclerosis: results of a randomised controlled trial. Journal of Neurology, Neurosurgery, and Psychiatry. 2002; 73(3):250-255

186 Prosperini L, Fortuna D, Gianni C, Leonardi L, Marchetti MR, Pozzilli C. Home-based balance training using the Wii balance board: a randomized, crossover pilot study in multiple sclerosis. Neurorehabilitation and Neural Repair. 2013; 27(6):516-525

187 Pucci E, Branas P, D'Amico R, Giuliani G, Solari A, Taus C. Amantadine for fatigue in multiple sclerosis. Cochrane Database of Systematic Reviews. 2007;(1):CD002818

188 Quispe-Cabanillas JG, Damasceno A, von Glehn F, Brandao CO, Damasceno BP, Silveira WD et al. Impact of electroacupuncture on quality of life for patients with Relapsing-Remitting Multiple Sclerosis under treatment with immunomodulators: A randomized study. BMC Complementary and Alternative Medicine. 2012; 12:209

189 Ramirez-Ramirez V, Macias-Islas MA, Ortiz GG, Pacheco-Moises F, Torres-Sanchez ED, Sorto-Gomez TE et al. Efficacy of fish oil on serum of TNF alpha , IL-1 beta , and IL-6 oxidative stress markers in multiple sclerosis treated with interferon beta-1b. Oxidative Medicine and Cellular Longevity. 2013; 2013:709493

190 Ramo C, Grau L, Giner P, Ramio L, Brieva L, Saiz A et al. A multicentric, double blind randomized clinical and MRI study of high-dose oral vs intravenous methylprednisolone in acute relapses of multiple sclerosis. Neurology. 2012; 78(1 Meeting Abstract)

191 Ramo-Tello C, Grau-Lopez L, Tintore M, Rovira A, Ramio IT, Brieva L et al. A randomized clinical trial of oral versus intravenous methylprednisolone for relapse of MS. Multiple Sclerosis (Houndmills, Basingstoke, England). 2013;

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

193 Riazi A, Hobart JC, Lamping DL, Fitzpatrick R, Freeman JA, Jenkinson C et al. Using the SF-36 measure to compare the health impact of multiple sclerosis and Parkinson's disease with normal population health profiles. Journal of Neurology, Neurosurgery, and Psychiatry. 2003; 74(6):710-714

194 Richards TL, Lappin MS, Acosta-Urquidi J, Kraft GH, Heide AC, Lawrie FW et al. Double-blind study of pulsing magnetic field effects on multiple sclerosis. Journal of Alternative and Complementary Medicine. 1997; 3(1):21-29

195 Rietberg MB, Brooks D, Uitdehaag BMJ, Kwakkel G. Exercise therapy for multiple sclerosis. Cochrane Database of Systematic Reviews. 2009;(4)

196 Rinne U. Tizanidine treatment of spasticity in multiple sclerosis and chronic myelopathy. Current Therapeutic Research. 1980; 28:827-836

197 Robson LS, Bain C, Beck S, Guthrie S, Coyte PC, O'Connor P. Cost analysis of methylprednisolone treatment of multiple sclerosis patients. Canadian Journal of Neurological Sciences. 1998; 25:222-229:222-229

198 Romberg A, Virtanen A, Ruutiainen J. Long-term exercise improves functional impairment but not quality of life in multiple sclerosis. Journal of Neurology. 2005; 252(7):839-845

199 Rose AS, Kuzma JW, Kurtzke JF, Sibley WA, Tourtellotte WW. Cooperative study in the evaluation of therapy in multiple sclerosis; ACTH vs placebo in acute exacerbations. Preliminary report. Neurology. 1968; 18(6):Suppl-10

200 Rosenberg GA, Appenzeller O. Amantadine, fatigue, and multiple sclerosis. Archives of Neurology. 1988; 45(10):1104-1106

201 Rossini PM, Pasqualetti P, Pozzilli C, Grasso MG, Millefiorini E, Graceffa A et al. Fatigue in progressive multiple sclerosis: results of a randomized, double-blind, placebo-controlled, crossover trial of oral 4-aminopyridine. Multiple Sclerosis. 2001; 7(6):354-358

202 Rosti-Otajarvi E, Mantynen A, Koivisto K, Huhtala H, Hamalainen P. Neuropsychological rehabilitation has beneficial effects on perceived cognitive deficits in multiple sclerosis during nine-month follow-up. Journal of the Neurological Sciences. 2013; 334(1-2):154-160

203 Rosti-Otajarvi EM, Hamalainen PI. Neuropsychological rehabilitation for multiple sclerosis. Cochrane Database of Systematic Reviews. 2011; 2011(11

CD009131)

204 Roullet E, Verdier-Taillefer MH, Amarenco P, Gharbi G, Alperovitch A, Marteau R. Pregnancy and multiple sclerosis: a longitudinal study of 125 remittent patients. Journal of Neurology, Neurosurgery, and Psychiatry. 1993; 56(10):1062-1065

205 Roussan M, Terrence C, Fromm G. Baclofen versus diazepam for the treatment of spasticity and long-term follow-up of baclofen therapy. Pharmatherapeutica. 1985; 4(5):278-284

206 Runmarker B, Andersen O. Pregnancy is associated with a lower risk of onset and a better prognosis in multiple sclerosis. Brain : a Journal of Neurology. 1995; 118 (Pt 1):253-261

207 Rushton DN, Lloyd AC, Anderson PM. Cost-effectiveness comparison of tizanidine and baclofen in the management of spasticity. Pharmacoeconomics. 2002; 20(12):827-837

208 Sachais BA, Logue JN, Carey MS. Baclofen, a new antispastic drug. A controlled, multicenter trial in patients with multiple sclerosis. Archives of Neurology. 1977; 34(7):422-428

209 Sadovnick AD, Eisen K, Hashimoto SA, Farquhar R, Yee IM, Hooge J et al. Pregnancy and multiple sclerosis. A prospective study. Archives of Neurology. 1994; 51(11):1120-1124

210 Sailer M, Smid HGOM, Schoenfeld A, Hauser U, Heinze HJ. Effect of amantadine on cognitive processes in multiple sclerosis patients with fatigue: a double-blind placebo controlled cross-over study. European Journal of Neurology. 1996; 3 Suppl 4:61-62

211 Sampson FC, Hayward A, Evans G, Morton R, Collett B. Functional benefits and cost/benefit analysis of continuous intrathecal baclofen infusion for the management of severe spasticity. Journal of Neurosurgery. 2002; 96:1052-1057:1052-1057

212 Sampson FC, Hayward A, Evans G, Touch S, Morton R, Vloeburghs.M. et al. The effectiveness of intrathecal baclofen in the management of patients with severe spasticity. Sheffield. Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 2000

213 Sawa GM, Paty DW. The use of baclofen in treatment of spasticity in multiple sclerosis. Canadian Journal of Neurological Sciences. 1979; 6(3):351-354

214 Schiffer RB, Herndon RM, Rudick RA. Treatment of pathologic laughing and weeping with amitriptyline. New England Journal of Medicine. 1985; 312(23):1480-1482

215 Schmidt RT, Lee RH, Spehlmann R. Comparison of dantrolene sodium and diazepam in the treatment of spasticity. Journal of Neurology, Neurosurgery, and Psychiatry. 1976; 39(4):350-356

216 Schwartz CE, Foley FW, Rao SM, Bernardin LJ, Lee H, Genderson MW. Stress and course of disease in multiple sclerosis. Behavioral Medicine. 1999; 25(3):110-116

217 Schwid SR, Petrie MD, McDermott MP, Tierney DS, Mason DH, Goodman AD. Quantitative assessment of sustained-release 4-aminopyridine for symptomatic treatment of multiple sclerosis. Neurology. 1997; 48(4):817-821

218 Schyns F, Paul L, Finlay K, Ferguson C, Noble E. Vibration therapy in multiple sclerosis: a pilot study exploring its effects on tone, muscle force, sensation and functional performance. Clinical Rehabilitation. 2009; 23(9):771-781

219 Sellebjerg F, Frederiksen JL, Nielsen PM, Olesen J. Double-blind, randomized, placebo-controlled study of oral, high-dose methylprednisolone in attacks of MS. Neurology. 1998; 51(2):529-534

220 Shakespeare DT, Boggild M, Young C. Anti-spasticity agents for multiple sclerosis. Cochrane Database of Systematic Reviews. 2003;(4):CD001332

221 Sharrack B, Hughes RA, Morris RW, Soudain S, Wade-Jones O, Barnes D et al. The effect of oral and intravenous methylprednisolone treatment on subsequent relapse rate in multiple sclerosis. Journal of the Neurological Sciences. 2000; 173(1):73-77

222 Shatil E, Metzer A, Horvitz O, Miller A. Home-based personalized cognitive training in MS patients: a study of adherence and cognitive performance. Neurorehabilitation. 2010; 26(2):143-153

223 Shaygannejad V, Ashtari F. A randomized controlled crossover trial of aspirin and amantadine for fatigue in multiple sclerosis. Multiple Sclerosis. 2010; 16(2):274

224 Shaygannejad V, Ashtari F, Vaezi A. Effect of oral vitamin D (Calcitriol) therapy on clinical course of relapsing remitting MS. Multiple Sclerosis. 2010; 16(2):273-274

225 Shaygannejad V, Janghorbani M, Ashtari F, Dehghan H. Effects of adjunct low-dose vitamin d on relapsing-remitting multiple sclerosis progression: preliminary findings of a randomized placebo-controlled trial. Multiple Sclerosis International. 2012; 2012:452541

226 Siev-Ner I, Gamus D, Lerner-Geva L, Achiron A. Reflexology treatment relieves symptoms of multiple sclerosis: a randomized controlled study. Multiple Sclerosis. 2003; 9(4):356-361

227 Slof J, Gras A. Sativex(R) in multiple sclerosis spasticity: a cost-effectiveness model. Expert Review of Pharmacoeconomics and Outcomes Research. 2012; 12(4):439-441

228 Smith C, Birnbaum G, Carter JL, Greenstein J, Lublin FD. Tizanidine treatment of spasticity caused by multiple sclerosis: results of a double-blind, placebo-controlled trial. US Tizanidine Study Group. Neurology. 1994; 44(11 Suppl 9):S34-S42

229 Smolders J, Hupperts R, Barkhof F, Grimaldi LME, Holmoy T, Killestein J et al. Efficacy of vitamin D(3) as add-on therapy in patients with relapsing-remitting multiple sclerosis receiving subcutaneous interferon beta-1a: a Phase II, multicenter, double-blind, randomized, placebo-controlled trial. Journal of the Neurological Sciences. 2011; 311(1-2):44-49

230 Smolenski C, Muff S, Smolenski-Kautz S. A double-blind comparative trial of new muscle relaxant, tizanidine (DS 103-282), and baclofen in the treatment of chronic spasticity in multiple sclerosis. Current Medical Research and Opinion. 1981; 7(6):374-383

231 Soilu-Hanninen M, Aivo J, Lindstrom BM, Elovaara I, Sumelahti ML, Farkkila M et al. A randomised, double blind, placebo controlled trial with vitamin D3 as an add on treatment to interferon beta-1b in patients with multiple sclerosis. Journal of Neurology, Neurosurgery, and Psychiatry. 2012; 83(5):565-571

232 Solani A, Aquarone N, Pucci E, Martinelli V, Marrosu MG, Trojano M et al. Communicating the diagnosis of a multiple sclerosis - A qualitative study. Multiple Sclerosis. 2007; 13(6):763-769

233 Solari A, Filippini G, Gasco P, Colla L, Salmaggi A, La Mantia L et al. Physical rehabilitation has a positive effect on disability in multiple sclerosis patients. Neurology. 1999; 52(1):57-62

234 Solari A, Motta A, Mendozzi L, Pucci E, Forni M, Mancardi G et al. Computer-aided retraining of memory and attention in people with multiple sclerosis: a randomized, double-blind controlled trial. Journal of the Neurological Sciences. 2004; 222(1-2):99-104

235 Solari A, Uitdehaag B, Giuliani G, Pucci E, Taus C. Aminopyridines for symptomatic treatment in multiple sclerosis. Cochrane Database of Systematic Reviews. 2003;(2):CD001330

236 Starck M, Albrecht H, Pollmann W, Dieterich M, Straube A. Acquired pendular nystagmus in multiple sclerosis: an examiner-blind cross-over treatment study of memantine and gabapentin. Journal of Neurology. 2010; 257(3):322-327

237 Stein MS, Liu Y, Gray OM, Baker JE, Kolbe SC, Ditchfield MR et al. A randomized trial of high-dose vitamin D2 in relapsing-remitting multiple sclerosis. Neurology. 2011; 77(17):1611-1618

238 Stien R, Nordal HJ, Oftedal SI, Slettebo M. The treatment of spasticity in multiple sclerosis: a double-blind clinical trial of a new anti-spastic drug tizanidine compared with baclofen. Acta Neurologica Scandinavica. 1987; 75(3):190-194

239 **Stuifbergen AK, Blozis SA, Harrison TC, Becker HA. Exercise, functional limitations, and quality of life: A longitudinal study of persons with multiple sclerosis. Archives of Physical Medicine and Rehabilitation. 2006; 87(7):935-943**

240 **Stuifbergen AK, Becker H, Perez F, Morison J, Kullberg V, Todd A. A randomized controlled trial of a cognitive rehabilitation intervention for persons with multiple sclerosis. Clinical Rehabilitation. 2012; 26(10):882-893**

241 **Surakka J, Romberg A, Ruutiainen J, Aunola S, Virtanen A, Karppi SL et al. Effects of aerobic and strength exercise on motor fatigue in men and women with multiple sclerosis: a randomized controlled trial. Clinical Rehabilitation. 2004; 18(7):737-746**

242 **Tarakci E, Yeldan I, Huseyinsinoglu BE, Zenginler Y, Eraksoy M. Group exercise training for balance, functional status, spasticity, fatigue and quality of life in multiple sclerosis: a randomized controlled trial. Clinical Rehabilitation. 2013; 27(9):813-822**

243 **Tesar N, Bandion K, Baumhackl U. Efficacy of a neuropsychological training programme for patients with multiple sclerosis -- a randomised controlled trial. Wiener Klinische Wochenschrift. 2005; 117(21-22):747-754**

244 **The United Kingdom Tizanidine Trial Group. A double-blind, placebo-controlled trial of tizanidine in the treatment of spasticity caused by multiple sclerosis. Neurology. 1994; 44 Suppl 9(11):S70-S78**

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

246 **Thorne S, Con A, McGuinness L, McPherson G, Harris SR. Health care communication issues in multiple sclerosis: an interpretive description. Qualitative Health Research. 2004; 14(1):5-22**

247 **Tolosa ES, Soll RW, Loewenson RB. Letter: Treatment of spasticity in multiple sclerosis with dantrolene. JAMA. 1975; 233(10):1046**

248 **Torkildsen O, Bakke SJ, Beiske AG, Bjerve K, Bjornara B, Bjorna I et al. Omega-3 fatty acids treatment in relapsing-remitting multiple sclerosis. European Journal of Neurology. 2011; 18:49**

249 **Torkildsen O, Wergeland S, Bakke S, Beiske AG, Bjerve KS, Hovdal H et al. Omega-3 fatty acid treatment in multiple sclerosis (OFAMS Study): a randomized, double-blind, placebo-controlled trial. Archives of Neurology. 2012; 69(8):1044-1051**

250 **Tosh J, Dixon S, Carter A, Daley A, Petty J, Roalfe A et al. Cost effectiveness of a pragmatic exercise intervention (EXIMS) for people with multiple sclerosis: economic evaluation of a randomised controlled trial. Multiple Sclerosis (Houndmills, Basingstoke, England). 2014;**

251 **Ungaro D, Ciccone B, Judica E, Boneschi FM, Rossi P, Vivo P et al. Inpatient and outpatient rehabilitation in subjects with Multiple Sclerosis. A prospective and 6 months follow-up study. Journal of Neurology. 2009; 256:S31**

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

253 Van Der Walt A, Sung S, Spelman T, Marriott M, Kolbe S, Mitchell P et al. A double-blind, randomized, controlled study of botulinum toxin type A in MS-related tremor. Neurology. 2012; 79(1):92-99

254 van Diemen HA, Polman CH, van Dongen TM, van Loenen AC, Nauta JJ, Taphoorn MJ et al. The effect of 4-aminopyridine on clinical signs in multiple sclerosis: a randomized, placebo-controlled, double-blind, cross-over study. Annals of Neurology. 1992; 32(2):123-130

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

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

257 Vogt A, Kappos L, Calabrese P, Stocklin M, Gschwind L, Opwis K et al. Working memory training in patients with multiple sclerosis - comparison of two different training schedules. Restorative Neurology and Neuroscience. 2009; 27(3):225-235

258 Wade DT, Makela P, Robson P, House H, Bateman C. Do cannabis-based medicinal extracts have general or specific effects on symptoms in multiple sclerosis? A double-blind, randomized, placebo-controlled study on 160 patients. Multiple Sclerosis. 2004; 10(4):434-441

259 Warke K, Al SJ, Walsh DM, Lowe-Strong AS. Use of self-applied TENS for low back pain in people with multiple sclerosis... including commentary by Armutlu K. International Journal of Therapy and Rehabilitation. 2004; 11(6):275-280

260 Warner R, Thomas D, Martin R. Improving service delivery for relapse management in multiple sclerosis. British Journal of Nursing. 2005; 14(14):746-753

261 Weinstock-Guttman B, Baier M, Park Y, Feichter J, Lee-Kwen P, Gallagher E et al. Low fat dietary intervention with omega-3 fatty acid supplementation in multiple sclerosis patients. Prostaglandins, Leukotrienes, and Essential Fatty Acids. 2005; 73(5):397-404

262 Wiles CM, Newcombe RG, Fuller KJ, Shaw S, Furnival-Doran J, Pickersgill T P et al. Controlled randomised crossover trial of the effects of physiotherapy on mobility in chronic multiple sclerosis. Journal of Neurology, Neurosurgery and Psychiatry. United Kingdom 2001; 70(2):174-179

263 Wilson R. The Multiple Sclerosis Partnership Programme. International MS Journal. 1998; 5(1):30-34

264 Wollin JA, Yates PM, Kristjanson LJ. Supportive and palliative care needs identified by multiple sclerosis patients and their families. International Journal of Palliative Nursing. 2006; 12(1):20-26

265 Worthington J, Jones R, Crawford M, Forti A. Pregnancy and multiple sclerosis--a 3-year prospective study. Journal of Neurology. 1994; 241(4):228-233