National Institute for Health and Care Excellence

Draft

Multiple sclerosis in adults: management (update)

[H] Evidence review for non-pharmacological management of memory and cognition

NICE guideline <number>

Evidence reviews underpinning recommendations 1.5.37 to 1.5.40 and research recommendations in the NICE guideline

December 2021

Draft for Consultation

These evidence reviews were developed by National Guideline Centre, hosted by Royal College of Physicians

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN:

Contents

1 Non-pharmacological management of memory and cognition

1.1 Review question

For adults with MS, including people receiving palliative care, what is the clinical and cost effectiveness of interventions for memory and cognitive problems?

1.1.1 Introduction

Cognitive changes and problems are a common symptom of multiple sclerosis (MS), both at early and later stages of the disease. People can experience a wide range of difficulties including attention to task, memory, efficiency and speed of processing information, multi-tasking and divided attention as well as executive function skills such as planning, judging and decision making. The nature of cognitive impairment is affected by lesion site and number of lesions as well as the process of disease progression. There is a complex interplay between cognition and other MS symptoms including anxiety, depression, fatigue, pain and sleep. This can significantly impact on many aspects of daily life such as looking after self, managing medications, running household chores, ability to work and maintain employment, interpersonal relationships and social roles and participate fully in leisure and social activities affecting the overall quality of life. As such, it is an important component of assessment and care for people with MS.

As a result of the role of cognition in MS care there is a rapidly growing body of research considering different cognitive rehabilitation approaches and programmes for people with MS. Neuropsychological rehabilitation with a diverse range of strategies and techniques tailored to individual need and circumstance including computerised training delivery, is a promising intervention to support areas of cognition affected by MS However, there is no current national standard and significant regional variations on what level of intensity or how neuropsychological and cognitive rehabilitation is offered (for example, individual, group, computerised) and by who this is provided. Neuropsychological rehabilitation also needs to be guided by clear assessment and formulation processes which also are varied across the country and can be limited by access to specialist clinical psychologists and neuropsychologists. However, this requires further evidence review to guide standard practice for supporting cognitive rehabilitation in people with MS and establish the cost effectiveness.

Pharmacological agents are not currently used to treat memory and cognitive problems in the MS population and this review therefore focuses on non-pharmacological interventions.

1.1.2 Summary of the protocol

For full details see the review protocol in Appendix A.

Table 1: PICO characteristics of review question

Population	Inclusion: Adults (≥18 years) with MS, including people receiving palliative care. Exclusion: Children and young people (≤18 years).
Interventions	Multi-domain cognitive/neuropsychological rehabilitation • Brain Training Apps such as luminosity

- Neuropsychological intervention for example neuropsychological Compensatory Training (NCT)
- Computer aided ‘Cognifit Personal Coach’ for cognition
- MS-Rehab computerised tool
- Psychoeducation
- Insight and awareness (typically termed as 'metacognitve training or metacognitive strategies')

Speed of information processing

- Time Pressure Management Training (TPM)

Attention and Working Memory

- CogMed Working Memory Training
- Attention Process Training (APT)
- Computer aided RehaCom module ‘Divided Attention’ for attention

Memory

- External compensatory strategies
- Errorless Learning Techniques
- Personal assistant apps
- Computer aided RehaCom module ‘memory and Attention’
- Computer aided (VILAT-G 1.0) training for memory
- Story memory technique (SMT)
- Computer aided memory retraining programme (SCRP)

Executive Function

- Goal Management Training (GMT)
- Problem Solving Training
- Computer aided RehaCom module ‘Plan a Day’ for organization and planning
- Interventions for apathy

Cognition

- Social Cognition Training
- Cognitive rehabilitation programmes
- Psychotherapy/counselling relating to cognitive impairment

Interventions aimed at improving language

- Retraining type approaches
- Compensatory type approaches (for example, use of communication aids)

Interventions aimed improving perception

- Psychoeducation
- Retraining type approaches (repeated practice on identifying specific objects/patterns)
- Compensatory type approaches (for example, labelling objects)

Combinations may be included as most rehabilitation programmes with a clinician (rather than computerised focus) will be multi-factorial as they will take into account the whole presentation rather than just focus on one part.

Comparisons	• Interventions will be compared to each other, placebo/sham, or usual care. • Waiting list control • Supportive therapy (dedicated time with a supportive clinician)

Outcomes	All outcomes are considered equally important for decision making and therefore have all been rated as critical.
	• Objective Measures
	  ○ Cognitive functions, such as memory, attention, executive functions, processing speed, for example, symbol digit modality test (SMDT)
	• Subjective Measures
	  ○ 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, activities.(for example Canadian Occupational Performance measure, Cognitive failure questionnaire, perceived deficits questionnaire
	  ○ Self-efficacy/self-management (MS self-efficacy scale)
	• Functional Measures
	  ○ Medication management/ adherence to medication
	  ○ Mood
	  ○ Fatigue (MS fatigue scale includes cognition (perhaps include this- if score reported separately?)
	  ○ Activities of daily living (ADL).
	• Vocational Measures
	  ○ Employment
	  ○ Training
	  ○ Social engagement
	  ○ Relationship satisfaction/ Impact on carers.
	• Engagement Measures
	  ○ Completion/adherence rates
	  ○ Acceptability
	  ○ Satisfaction
	Validated measures will be prioritised. If no evidence is available, non-validated may be considered.
	Follow-up:
	• 3-6 months (minimum of 3 months but can include 1-3 months and downgrade)
	• >6 months – 1 year (data from >1 year follow up may be included but will be downgraded)

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

	Published NMAs and IPDs will be considered for inclusion. Cross over trials will be excluded as many interventions are around learning where it would not be possible to do a cross-over trial as the information cannot be 'unlearned'

1.1.3 Methods and process

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

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

1.1.4 Effectiveness evidence

1.1.4.1 Included studies

Sixty-three studies (from seventy-one papers) were included in the review; these are summarised in Table 2 below. Evidence from these studies is summarised in the clinical evidence summary below (Tables 3-42).

All included studies were randomised controlled trials (RCTs), as crossover trials were not accepted for this review given many cognitive interventions will involve learning where information cannot be unlearned.

Population

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

- The study used a threshold for memory or cognitive impairments as an inclusion criterion in the study (e.g., only those with score below a threshold indicating impairment on a particular cognitive test)
- The study did not use a threshold for memory or cognitive impairment for inclusion, but it was clear from the paper that memory and/or cognition was the primary outcome or one of the primary outcomes
- The study did not use a threshold for memory or cognitive impairment for inclusion and it was listed as a secondary outcome, but it was clear from the paper that memory and/or cognition was one of the focuses of the paper
- The study did not focus on any particular MS symptom and memory and/or cognition was emphasised as an important outcome

Most studies had cognitive impairment as one of the inclusion criteria, the definition of which varied between studies but was often based on baseline scores on one or more cognitive tests with or without patient-reported subjective cognitive complaints.

Of studies that reported the proportion of participants with different types of MS, the majority reported relapsing-remitting MS as the most common type of MS among participants, though four studies reported progressive MS to be the most common type of MS within the study. Expanded Disability Status Scale (EDSS) scores (range 0 – 10) most commonly fell into the <6.0 score (indicating more severe disability) category across studies.

Severity of cognitive impairment was unclear for many studies, though in many cases severe impairment appeared to be an exclusion criterion. Severe mood disorders/psychiatric conditions were often excluded from studies. In terms of disease-modifying treatment status, this was unclear for many studies – some reported that the majority were using these treatments while for others it was not mentioned.

<u>Interventions and comparisons covered by the evidence</u>

There was a wide range of interventions and comparisons covered by the included studies, limiting the amount of pooling that was possible across interventions. Interventions ranged from general cognitive rehabilitation programmes targeting cognition in general (more than one domain, for example attention, memory and executive function) and using multiple techniques (for example practical exercises, lectures and group discussions) to very specific

programmes focusing on a single cognitive domain, for example specific techniques for improving memory.

In terms of the frequency of the intervention, how they were delivered and who they were delivered by, this also varied widely across studies. For frequency of sessions, some interventions involved up to five sessions weekly over a few weeks and were more intensive than other studies where, for example, the sessions were weekly and extended over a longer period of time. Many interventions that were solely computerised were performed remotely by participants with technical assistance, and support was often available at all times. Interventions that involved training of specific techniques, such as specific techniques for memory or psychological techniques such as mindfulness and mental visual imagery, were usually performed in person with the support of healthcare professionals, such as therapists or neuropsychologists. These techniques were also usually spread over a longer time period than more intensive, home-based, computerised training programmes. Most interventions lasted between one and four months.

Controls also varied across studies, with some being more active and controlling for contact with professionals (for example, meeting and performing and discussing non-cognitive exercises) while others involve no contact (for example, waitlist control).

Interventions listed in the protocol were not an exhaustive list, but of specific programmes or modules mentioned in the protocol, no evidence was identified for the MS-Rehab computerised tool.

In some cases, interventions touched on areas listed in the protocol but contained other elements and were not solely focused on that skill or domain. For example, time pressure management training, interventions for apathy and social cognitive training were part of some interventions but not the sole focus of interventions.

Evidence was identified for the following interventions and comparisons:

<u>Multi-domain cognitive/neuropsychological rehabilitation</u>

- **General cognitive rehabilitation – multi-component** (different types of strategies combined, e.g., computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) – n=11 studies
    - vs. **control** (for example, non-cognitive exercises, waitlist control, no training/intervention or usual care + freely available games) in n=9 studies
    - vs. **psychoeducation and information-sharing only** (group sessions with interpersonal training, discussion of life experiences and sharing of scientific information about MS) in n=1 study
    - vs. **non-specific cognitive rehabilitation** programme (including information about the disease, management, relaxation, physical activity coaching and global cognitive stimulation) in n=1 study

- **General cognitive rehabilitation + outpatient rehabilitation – multi-component and tailored to individual deficits** (different types of strategies combined, e.g., computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) – n=1 study
    - vs. **control** (no treatment) in n=1 study

- **General cognitive rehabilitation - Goal Attainment Scaling goals for coping with cognitive challenges with cognitive rehabilitation, tailored to individual + usual rehabilitation** – n= 1 study

  - vs. **control – usual multidisciplinary rehabilitation only** (including physical activity, lectures with information about MS, opportunity to consult with clinical psychologist and attend lectures on cognitive and psychological aspects of MS, and the offer of neuropsychological assessment with feedback) in n=1 study

- **Multi-domain skills training (e.g., computer or pen/pencil tasks) without additional strategies (e.g., computer training for multiple domains such as attention, memory and information processing speed** – does not include other techniques such as teaching learning techniques or psychoeducation) – n=7 studies
  - vs. **control** (for example, no training/intervention, waitlist control, usual care only or usual care + sham computer exercises such as puzzles and reading magazines) in n=7 studies

- **Multi-domain skills training tailored to individual (e.g., computer or pen/pencil tasks) without additional strategies (e.g., computer training for multiple domains such as attention, memory and information processing speed** – does not include other techniques such as teaching learning techniques or psychoeducation) – n=2 studies
  - vs. **control** (for example, conversation about patient's disease perception, family and working life with aim of not exercising cognitive ability or no training) in n=2 studies

- **Brain training apps/games** – n=7 studies
  - vs. **control** (for example, ordinary computer games without features of adaptive cognitive remediation programmes, waitlist control or no training) in n=7 studies

- **Mental visual imagery** – n=1 study
  - vs. control (sham verbal intervention involving constructing discussions about texts from websites with neuropsychologist guidance) in n=1 study

- **Mindfulness** – n=3 studies (one study reports multiple comparisons)
  - vs. **control** (no intervention/waitlist control) in n=2 studies (waitlist control group also compared with general cognitive rehabilitation group below)
  - vs. **general cognitive rehabilitation (different types of strategies combined, e.g., computer training for skills + teaching other strategies)** in n=1 study (also compared with a waitlist control group above)
  - vs. **outpatient visits with counselling** in n=1 study

<u>Focus on information processing speed</u>

- **Cognitive rehabilitation focused on processing speed + occupational therapy** – n=1 study
  - vs. occupational therapy alone (consisting of various physical exercises) in n=1 study

- **Cognitive rehabilitation software focused on processing speed** – n=2 studies
  - vs. control (no training or active control involving similar time on task and engagement) in n=2 studies

<u>Focus on information processing speed and working memory</u>

- **Visual n-back programs focused on working memory and processing speed** – n=1 study
    - vs. control (involving sham training with the same tasks but without increasing difficulty) in n=1 study

Focus on attention/working memory

- **Computer-aided RehaCom training for attention** (includes 'Divided Attention' or other tasks said to be focused on attention, with or without memory-specific modules) – n=1 study
    - vs. active control (visuomotor coordination task using in-house software) in n=1 study

- **Computer aided training for attention/working memory** – n=2 studies
    - vs. **control** (for example, standard medical care only or control task of watching natural history DVDs) in n=2 studies

- **High-intensity working memory training** (four times weekly for four weeks) – n=1 study (reporting multiple comparisons with all three groups compared to each other)
    - vs. **distributed working memory training** (two times weekly for eight weeks) in n=1 study
    - vs. **control** (no training) in n=1 study

- **Attention Process Training (focused, sustained, selective, alternating and divided attention, specific computer programme focused on attention) + multidisciplinary rehabilitation** – n=1 study
    - vs. multidisciplinary rehabilitation only (individualised, goal-oriented inpatient programme including physiotherapy sessions) in n=1 study

- **Reaction time tasks + usual rehabilitation programme** – n=1 study
    - vs. **active control** (software aiming to improve similar cognitive functions of selective attention, cognitive flexibility and working memory but with no time component) in n=1 study

Focus on memory (with or without attention components also included)

- **Computer-aided training for memory (with or without attention components also included)** – n=3 studies (one reporting multiple comparisons)
    - vs. **control** (no training) in n=2 studies (also compared with the non-specific cognitive retraining programme below in one study)
    - vs. **non-specific cognitive retraining programme** (visual tracking task and reaction-time tasks or training of motor skills, designed to train cognitive abilities other than memory) in n=2 studies (one study also compares this to the control group above)

- **Story Memory Technique** – n=5 studies
    - vs. control (for example, meeting with therapists for discussion of non-training tasks such as reading stories and answering questions or being exposed to same stories and target words but not being taught how to apply context and imagery to the material) in n=5 studies

- **External compensatory strategies (e.g., lists, diaries and visual mnemonics)** – n=2 studies (one study reporting multiple comparisons)

  - vs. **cognitive assessment with feedback only and no intervention** in n=1 study
  - vs. **restitution training** (internal ability to code, organise and retrieve information) in n=1 study (also compared to self-help group below)
  - vs. **self-help control group** (relaxation techniques and ways of coping taught) in n=1 study (also compared to the restitution training group above)

- **Group memory programme (various learning techniques, including internal and external aids)** – n=4 studies
  - vs. **control** (for example, usual care involving physiotherapy, occupational therapy and advice and information from nurses about cognition, or discussion of experiences and coping strategies without supporting cognitive rehabilitation) in n=4 studies

- **Behavioural intervention (Self-GEN trial) focused on teaching self-generation technique with metacognitive strategies** – n=1 study
  - vs. **control** (memory tasks but no self-generated learning and transfer instructions) in n=1 study

<u>Focus on executive function</u>

- **Executive function-specific training exercises** – n=2 studies
  - vs. **control** (no training or placebo tasks focusing on responding quickly to visual stimuli) in n=2 studies

- **Goal management programme** – n=1 study
  - vs. **control** (psychoeducation programme involving lectures on cognition in MS and discussion of experiences) in n=1 study

<u>Focus on improving language</u>

- **RehaCom verbal fluency training** – n=1 study
  - vs. **control** (no intervention) in n=1 study

1.1.4.2 Excluded studies

Four Cochrane reviews[18, 63, 69, 73] were identified and reviewed to assess relevance to this review.

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

- The review was not specific to memory or cognition[69, 73]
- The review used very broad intervention categories which were pooled together (for example any memory rehabilitation vs. control), whereas the protocol for this review breaks interventions down into more specific interventions (e.g., Story Memory Technique for memory)[18, 63]

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

See the excluded studies list in appendices.

1.1.5 Summary of studies included in the effectiveness evidence

Table 2: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Comments
Multi-domain cognitive/neuropsychological rehabilitation			
General cognitive rehabilitation – multi-component (different types of strategies combined, e.g., computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) vs. control			
Brissart 2020[5] France N=128	**ProCog-SEP extended cognitive rehabilitation programme:** psychoeducational advices and cognitive exercises which target verbal and non-verbal episodic memory, working memory, short-term memory, executive functions, and language vs. **Placebo programme:** non-cognitive exercises and discussion	MS diagnosis with EDSS score ≤6.0 Cognitive impairment: moderate (at least 2 cognitive functions of neuropsychological examination but not all)	
Jonsson 1993[34] Denmark N=40	**Cognitive training and neuro-psychotherapy:** cognitive training involved compensation, substitution and direct training of concentration and memory using exercises and learning compensatory strategies. Visuospatial and orientation difficulties trained using mosaic games and practical exercises. Neuro-psychotherapy also used to realise and accept present cognitive and behavioural functioning learning how best to use resources. **Control - non-specific mental stimulation:** met with therapist and discussed films, newspaper articles and played games with no relation to specific cognitive dysfunction training	Hospitalised patients fulfilling Schumacher's diagnostic criteria of MS. Cognitive impairment: said to include mild-moderate cognitive dysfunction based on neuropsychological testing	

Study	Intervention and comparison	Population	Comments
Manglani 2020[40] Associated papers: Schirda 2020[64] USA N=41 in these two groups	**Adaptive cognitive training:** group training sessions covering multiple domains including attention, processing speed, executive functions and working memory. Training involved didactics, group discussion and practice with training materials in the form of BrainHQ games and additional home practice. Adaptive process starting with building blocks of cognition and moving on to higher-order cognitive domains such as executive functioning. vs. **Waitlist control:** did not engage in any training	Diagnosis of MS Cognitive impairment: does not appear to be a specific inclusion criterion for cognitive impairment	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Indirectness as outcomes only reported at 4-weeks (end of intervention period)
Mantynen 2014[42] Associated papers: Rosti-Otajarvi 2013[62] and Rosti-Otajarvi 2013[62] Finland N=102	**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 vs. **Control:** no training	Clinically definite relapsing-remitting MS with EDSS <6.0 Cognitive impairment: 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)	
Pusswald 2014[58] Austria N=40	**Cognitive training:** computerised home-based attention training using Fresh Minder 2 software providing feedback to user + psychosocial group sessions covering other cognitive areas (coping	MS diagnosed by neurologist Cognitive impairment: does not appear to be a specific inclusion criterion for	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Indirectness as outcomes only reported at 5-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	methods and compensatory techniques, memory retraining and social skills training, etc.) vs. **Control:** no training	cognitive impairment	
Rahmani 2020[59] Iran N=60	**Cognitive rehabilitation – computer-based, manual-based or combination of manual and computer-based methods:** content of all three groups included memory, information processing speed, attention and executive functions, as well as psychoneurological skills such as linguistic functions and visual perception. Involved retraining of impaired functions, reorganising functions, promoting use of preserved functions and learning compensation strategies. Computer-based group performed through Captain's Log Training System and manual-based group performed through Pars Cognitive Rehabilitation package involving pen-paper programme. Combined group used a combination of both. vs. **Control – physical rehabilitation only or no intervention control:** two groups combined for purpose of this review as a single control group. Physical group receive physical rehabilitation only by a sports and health specialist. No	Relapsing-remitting MS with EDSS score up to 3.5 Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Reported multiple intervention and control groups which were combined as a single intervention and single control group for the purpose of this review as they were very similar interventions. Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper

Study	Intervention and comparison	Population	Comments
	intervention group received no intervention.		
Rilo 2018[61] Spain N=44	**Manual cognitive rehabilitation using REHACOP:** group integrative cognitive rehabilitation focused primarily on short-term memory. Based on principles of restoration, compensation and optimisation. Focus first on basic cognitive processes then adapting to more complex cognitive domains. Consists of paper and pen tasks on attention, learning and memory, language, executive functions, social cognition, social skills, activities of daily living and psychoeducation. vs. **Waitlist control:** no intervention	Clinically definite MS Cognitive impairment: does not appear to be a specific inclusion criterion for cognitive impairment	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper
Stuifbergen 2012[70] USA N=63	**Memory and Problem Solving Skills for people with Multiple Sclerosis (MAPSS-MS):** sessions teach the use of compensatory skills, retraining skills (home-based computer component) and environmental/lifestyle support for cognitive functioning. vs. **Waitlist control:** no intervention	Clinically definite MS for at least 6 months Cognitive impairment: responded 'sometimes' or more often to at least 5 problems on the Perceived Deficits Questionnaire	
Stuifbergen 2018[71] USA N=183	**Computer-assisted cognitive rehabilitation (MAPSS-MS intervention):** sessions teach the use of compensatory skills (related to attention, processing speed, memory, language, visuospatial and executive functioning),	Clinically definite MS for at least 6 months Cognitive impairment: score of at least 10 (indicating some problems in at least 5 areas) on	

Study	Intervention and comparison	Population	Comments
	retraining skills (home-based computer component – Lumosity software) and environmental/lifestyle support for cognitive functioning. Structured so most basic cognitive functions addressed first. vs. **Control – usual care + freely available games:** received their usual care + referral to MyBrainGames including games challenging processing speed, working memory attention and task switching ability. Weekly check-in calls with research staff.	Perceived Deficits Questionnaire	
General cognitive rehabilitation – multi-component (different types of strategies combined, e.g., computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) vs. control (psychoeducation and information-sharing only)			
Mani 2018[41] Iran N=34	**Group cognitive rehabilitation:** training covering multiple domains including processing speed, attention, executive function and working memory. Compensatory, problem-based and integrated approach consisting of psychoeducation elements and being taught strategies for various impairments with homework tasks aimed to improve understanding. vs. **Control – psychoeducation and information-sharing:** group sessions with sham intervention involving dynamic, interpersonal	Relapsing-remitting MS diagnosis Cognitive impairment: minimal cognitive impairment based on Addenbrooke's cognitive examination (scores >70 - patients with severe cognitive deficits not included)	

Study	Intervention and comparison	Population	Comments
	relationship training with discussions about daily life experiences of patients and sharing of scientific information about MS. Phone follow-ups twice weekly after last session encouraging them to use learned techniques in everyday lives		
General cognitive rehabilitation – multi-component (different types of strategies combined, e.g., computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) vs. non-specific cognitive rehabilitation programme			
Lamargue 2020[36] France N=35	**Specific cognitive rehabilitation programme (REACTIV):** focused on fundamental cognitive processes of information processing speed, attention, executive function, working memory and metacognition using various tasks and exercises including computer-based and pen and pencil exercises and rehabilitation games. Time also provided for focusing on difficulties in daily life and metacognitive deep-thinking. vs. **Non-specific cognitive intervention:** sessions focused on information about the disease, symptoms and management, relaxation, physical activity coaching and global cognitive stimulation (focus on semantic memory, autobiographical memory and verbal and visual episodic memory)	MS diagnosis for at least 6 months Cognitive impairment: mild objective cognitive impairment (at least 3 scores <1 SD on tests measuring information processing speed, attention, working memory and executive function); and complaining of discomfort in daily lives due to cognitive problems.	
General cognitive rehabilitation + outpatient rehabilitation – multi-component and tailored to individual deficits (different types of strategies combined, e.g., computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) vs. outpatient rehabilitation alone			

Study	Intervention and comparison	Population	Comments
Tesar 2005[72] Austria N=19	**Cognitive rehabilitation:** Rehacom computer training. Direct functional training of the two cognitive areas which were most severely affected and then teaching of compensation strategies to everyday life. vs. **Control:** no treatment	Patients with MS meeting the criteria of Posner plus a positive MRI scan Cognitive impairment: mild-moderate cognitive deficit (definition unclear)	
General cognitive rehabilitation - Goal Attainment Scaling goals for coping with cognitive challenges with cognitive rehabilitation, tailored to individual + usual rehabilitation vs. usual rehabilitation alone			
Hanssen 2016[31] Norway N=120	**Cognitive sessions + multidisciplinary rehabilitation:** cognitive rehabilitation involved guidance through the process of formulating Goal Attainment Scaling (GAS) goals for coping with cognitive problems in everyday life. To facilitate metacognitive awareness, cognitive strengths and symptoms were discussed with the patient and related to everyday challenges. Cognitive strengths and symptoms summarised in a form that contained general advice for coping with cognitive problems and sections in which the patient could enter goals and operationalize behaviours required to reach them. Included lectures, practical exercises and discussions. Also received multidisciplinary rehabilitation as described in the control group below. vs.	Inpatients with MS undergoing inpatient rehabilitation Cognitive impairment: subjective complaints about cognitive problems and motivation for working with cognitive problems to improve coping in everyday life were inclusion criteria	

Study	Intervention and comparison	Population	Comments
	Control – multidisciplinary rehabilitation only: offered neuropsychological assessment, including feedback, and otherwise participated in the ordinary 4-week rehabilitation program of individual follow-up by a multidisciplinary team. Physical activities and lectures about MS-related topics were offered daily. As part of the ordinary rehabilitation program, participants in the control group had the opportunity to consult a clinical psychologist and attend lectures on cognitive and psychological aspects of MS.		
Multi-domain skills training (e.g., computer or pen/pencil tasks) without additional strategies (e.g., computer training for multiple domains such as attention, memory and information processing speed – does not include other techniques such as teaching learning techniques or psychoeducation) vs. control			
Fillipi 2012[23] Associated papers: Parisi 2014[56] Italy N=20	**RehaCom cognitive training:** focus on executive function, attention and speed of information processing. 'Plan a Day', 'Divided Attention' and modified-PASAT task as part of the 'Divided Attention' session. vs. **Control:** no cognitive training	Relapsing-remitting MS with EDSS ≤4.0 Cognitive impairment: deficits in both PASAT (z-scores <-1.5) and Wisconsin Card Sorting Test (z scores <-1.5 in any of test measures)	
Gich 2015[26] Spain N=43	**MS Line! cognitive rehabilitation programme** written, manipulative and computer-based materials including logic and reasoning and working memory games, mathematical problems, crosswords and origami, among others. Also	Clinically definite MS Cognitive impairment: mild cognitive impairment as determined by the neuropsychological assessment (1.5 SD or more	

Study	Intervention and comparison	Population	Comments
	involved doing an exercise with family members daily for up to 5 min. vs. **Control:** no intervention	below the mean of normative data; cognitive impairment was defined as: mild, between one and three impaired cognitive tests; moderate, four to seven impaired tests; and severe, eight or more impaired tests)	
Mattioli 2010[46] Associated papers: Mattioli 2012[45]	**Intensive neuropsychological training:** attention, information processing and planning exercises for executive functions. Plan a day and divided attention components of the RehaCom package. vs. **Control:** no rehabilitation	Relapsing-remitting MS with EDSS <4.0 Cognitive impairment: scores fell below Z= -1.5 for the PASAT and T=35 for WCST	Medians only so not meta-analysed
Messinis 2017[49] Greece N=58	**Cognitive rehabilitation using RehaCom:** as most of those included were impaired in more than one cognitive domain but mostly on episodic memory, information processing speed/attention, and executive functions, the intervention was balanced over the 10-week period in order to train all domains equally. Included 'attention and concentration', 'divided attention', 'topological memory', 'verbal memory', 'logical reasoning' and 'shopping' modules/tasks. vs. **Control – usual care:** continued taking their prescribed medication	Relapsing-remitting MS diagnosis with EDSS score 0-5.0 Cognitive impairment: cognitive deficit on at least one domain of the Central Nervous System Vital Sign neuropsychological screening battery (performance between the 2nd and 8th percentile based on CNSVS demographically corrected normative data)	Indirectness as outcomes only reported at 10-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	and all other related treatments (e.g., physiotherapy, psychotherapy), and all other clinical or referral services were available to them as usual for the entire 10 weeks that the intervention group received cognitive training.		
Messinis 2020[48] Greece N=36	**Cognitive rehabilitation using RehaCom:** as most of those included were impaired in more than one cognitive domain but mostly on episodic memory, information processing speed/attention, and executive functions, the intervention was balanced over the 8-week period in order to train all domains equally. Included 'attention and concentration', 'divided attention', 'topological memory', 'verbal memory', 'logical reasoning' and 'shopping' modules/tasks. Trained on exercises in clinic beforehand to ensure understanding and caregivers present during at-home sessions. vs. **Control – usual care + sham computer exercises:** non-specific computerised activities including solving puzzles, reading and understanding magazines and newspapers, shopping games, brain teasers etc. were used at home in presence of caregiver. Also continued standard clinical care - taking their prescribed medication	Diagnosis of secondary progressive MS with EDSS score up to 7.0. Cognitive impairment: cognitive deficit on at least two domains of the Central Nervous System Vital Sign neuropsychological screening battery (performance 1.5 SD below healthy control group data)	Indirectness as outcomes only reported at 8-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	and all other related treatments (e.g., physiotherapy, psychotherapy), and all other clinical or referral services were available to them as usual for the entire period.		
Naeeni 2020[52] Iran N=60	**RehaCom cognitive software:** modules focused on working memory, attention, processing speed, response control, executive functions and spatial awareness and involved 'working memory', 'responsiveness', 'divided attention 2', 'attention and concentration', ''logical reasoning' and 'spatial awareness' modules. vs. **Control:** no intervention	MS diagnosis referred to specialised rehabilitation clinic (Brain and Cognition Clinic) Cognitive impairment: criteria not reported but selected participants from those referred to a Brain and Cognition Clinic	Indirectness as outcomes only reported at 10-weeks (5 weeks after the end of the 5-week intervention period)
Perez-Martin 2017[57] Spain – Canary Islands N=62	**Cognitive rehabilitation using computerised and pencil and paper tasks:** focused on attention, processing speed, memory and executive functions through computerized and paper and pencil tasks designed by the members of the research team in addition to work set as homework. Final booklet contained set of guidelines and general advice on the influence of habits and lifestyles on cognitive functions, practical exercises for working memory and the ability to concentrate as well as suggestions on planning and physical activity. vs.	MS diagnosis with EDSS up to 7.0. Cognitive impairment: subjective complaints about cognitive problems and objective cognitive impairment defined as a performance of 1.5 standard deviation lower than the mean in a control group in at least two cognitive tests (determined by the neuropsychological assessment).	

Study	Intervention and comparison	Population	Comments
	Waitlist control: only received feedback on their cognitive status and a booklet containing set of guidelines and general advice on influence of habits and lifestyles on cognitive functions. Group was contacted once a week.		
Multi-domain skills training tailored to individual (e.g., computer or pen/pencil tasks) without additional strategies (e.g., computer training for multiple domains such as attention, memory and information processing speed – does not include other techniques such as teaching learning techniques or psychoeducation) vs. control (no training)			
Shatil 2010[67]	**Cognitive training:** CogniFit Personal Coach (CPC), a home-based, computerised, individualised cognitive training program. vs. **Control:** no training	Diagnosis of relapsing remitting or relapsing progressive MS Cognitive impairment: 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	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper
Mattioli 2015[44] SMICT study Italy N=41	**Specific cognitive training tailored to individual:** RehaCom training with different modules depending on the impairments – 'Plan a Day' for executive function, memory package from same software for those with memory impairment and information processing training for those impaired in this domain. If multiple domains, time was split between them in the sessions. vs. **Non-specific intervention – psychoeducation with**	Relapsing-remitting MS diagnosis Cognitive impairment: impaired (age corrected z-score ≤1.5 SD to norms) in at least one test of the Italian version of the Rao's Brief Repeatable Battery	Median values only so could not be meta-analysed

Study	Intervention and comparison	Population	Comments
	no cognitive skills training: conversation about patient's disease perception, family and working life with aim of not exercising cognitive ability, avoiding treatment of depression or having any behavioural or psychoanalytical approach.		
Brain training apps/games vs. control			
Charvet 2015[9] USA N=20	**Adaptive cognitive remediation programme:** cognitive exercises using Lumosity platform. Study-specific portal and set of games that focused on the most common areas of impairment in MS, including speeded information processing and working memory. vs. **Control – ordinary computer games:** computer-based gaming program that would provide the experience of cognitive exercise associated with cognitive benefit but without the key components of the adaptive cognitive remediation programs (i.e., games not developed based on cognitive neuroscience principles to drive neural plasticity). Commercially available Hoyle puzzles and board games program.	Relapsing-remitting MS Cognitive impairment: included those seeking treatment for cognitive impairment due to MS as judged by referring neurologist	
Charvet 2017[10] USA N=135	**Adaptive cognitive training programme:** online adaptive cognitive training program developed by Posit Science Corporation. Research version of the BrainHQ program and	Diagnosis of MS Cognitive impairment: scoring one or more standard deviations below published	

Study	Intervention and comparison	Population	Comments
	portal dedicated to the study and a set of 15 exercises targeting speed, attention, working memory, and executive function through the visual and auditory domains. Vs. **Control – ordinary computer games:** computer-based gaming program that would provide the experience of cognitive exercise associated with cognitive benefit but without the key components of the adaptive cognitive remediation programs (i.e., games not developed based on cognitive neuroscience principles to drive neural plasticity). Commercially available Hoyle puzzles and board games program.	normative data on the Symbol Digit Modalities Test	
Chmelarova 2020[16] Czech Republic N=43	**Multi-domain cognitive programme:** Happy Neuron Brain Jogging computer programme at home. Primary goals of treatment plan included following cognitive functions: memory, attention and concentration, speed and information processing, executive functions, expression and speed comparison and self-orientation and perception. Specific tasks to be repeated given and then allowed choice of exercises in remaining time. vs. **Control – no training:** received no training but to control for placebo	MS diagnosis with EDSS score 0-6.0 Cognitive impairment: cognitive deficit at baseline was an inclusion criterion (definition not provided)	Indirectness as outcomes only reported at 8-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	effect they were repeatedly contacted for 2 months		
de Giglio 2015[19] Italy N=35	**Nintendo brain training game:** training in games of memory, attention, visuospatial processing, and calculations. The cognitive training was performed at home with the Italian version of the Dr Kawashima's Brain Training. The number of puzzles proposed increased through time. Games included Calculations and Voice Calculations, Reading Aloud, Low to High, Syllable Count, Head Count, Triangle Math and Time Lapse. vs. **Waitlist control:** no definition but assume continued usual care and received no additional intervention.	Relapsing-remitting MS Cognitive impairment: failure in at least 1 of the following tests: Stroop Test, PASAT 3-s presentation rate, and Symbol Digit Modalities Test (failure on PASAT and SDMT was defined as a score below the fifth percentile of normative data for the Italian population and failure on ST as an equivalent score below 3)	Indirectness as outcomes only reported at 8-weeks (end of intervention period)
de Giglio 2016[20] Italy N=24	**Nintendo brain training game:** training in games of memory, attention, visuospatial processing, and calculations. The cognitive training was performed at home with the Italian version of the Dr Kawashima's Brain Training. The number of puzzles proposed increased through time. Games included Calculations and Voice Calculations, Reading Aloud, Low to High, Syllable Count, Head Count, Triangle Math and Time Lapse. vs. **Waitlist control:** no definition but assume continued usual care and received no additional intervention.	Relapsing-remitting MS Cognitive impairment: specific deficits in working memory, information processing speed, or sustained attention (failure on at least one of the following tests: PASAT 3-second presentation rate, SDMT, and the Stroop Test - failure on the PASAT and SDMT was defined as a score lower than the 10th percentile of normative data from the Italian	Indirectness as outcomes only reported at 8-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
		population and failure on the ST as a score of less than 3)	
Janssen 2015[33] USA N=34	**Video-game training with cognitive-focused Space Fortress game:** Space Fortress game used to implement hybrid-variable priority training. Initial 10 training sessions required part-task training where game was divided into 14-part tasks focusing on different aspects of the game. Initially three full-emphasis games (games not altered from original Space Fortress format) followed by 14 part-task games and another three full-emphasis games. Subsequently 10 sessions using variable priority training which highlighted different aspects of the game with emphasis on each subscore to minimise overall cognitive load while integrating previously trained part-tasks. vs. **Waitlist control:** contacted every two weeks to ensure good health and compliance with study guidelines. Participants were requested to refrain from engaging in any other experimental trials.	MS diagnosis and score >1.0 on EDSS Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper
Vilou 2020[74] Greece N=47	**Computerised cognitive rehabilitation (BrainHQ):** cognitive rehabilitation intervention using the web-based BrainHQ platform. Used modules focusing on episodic	Relapsing-remitting MS Cognitive impairment: performed 1.5 Standard	Indirectness as outcomes only reported at 6-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	memory, attention and processing speed. Home-based programme and performed in native language. Activities pre-specified and given to participants in written form. vs. **Control:** no definition provided, assume received no additional intervention.	Deviation units below average on at least one of the neuropsychological measures administered	
Mental visual imagery vs. control (sham verbal/no intervention)			
Ernst 2016[22] France N=17	**Mental visual imagery programme:** based on the ability to mentally construct scenes and follows a goal-directed approach. Four-step approach involving Screening of basic visual imaging abilities based on Imagery and Perception Battery, external visualisation involving 10 names of objects to be imagined and described, construction phase involving figuring out complex scenes involving multiple characters and a self-visualisation step requiring participants to imagine themselves in a given scenario vs. **Control – verbal control programme:** Construct discussions about texts (extracted from websites) with the neuropsychologist's guidance, through steps of increasing difficulty in four-step process.	Relapsing-remitting MS with EDSS score up to 4.0 Cognitive impairment: impaired episodic future thought performance (mild-moderate cognitive impairment in attention and/or executive functions; mean number of internal details provided ≤19)	Indirectness as outcomes only reported at 6-8-weeks (end of intervention period)
Mindfulness vs. control (no intervention/waitlist control)			
De La Torre 2020[21]	**Mindfulness intervention +**	Relapsing-remitting MS	Cognitive impairment not explicitly stated to be an inclusion criterion,

Study	Intervention and comparison	Population	Comments
Spain N=60	**pharmacological treatment:** mindfulness group sessions based on Jon Kabat-Zinn's programme adapted for patients with depression. Focused on common problems and worries people with MS have such as functional independent living level, mood, uncertainty and work. Also, mindfulness sessions at home and written exercises. Assume usual pharmacological treatment continued for the pharmacological component mentioned in this group. vs. **Control:** no mindfulness intervention. Described as pharmacological treatment only and assume usual pharmacological treatment continued as no further details provided.	regardless of degree of functional deterioration Cognitive impairment: not explicitly stated to be an inclusion criterion, but possible that those selected by neuropsychologists were those who they thought would benefit most from attempt to improve cognitive abilities	but possible that those selected by neuropsychologists were those who they thought would benefit most from attempt to improve cognitive abilities.
Manglani 2020[40] Associated papers: Schirda 2020[64] USA N=41 in these two groups	**Mindfulness intervention:** group sessions including combination of didactics, group discussion and practice with training materials. Also, at-home practice requested. Based on Jon Kabat-Zinn's 8-week programme, 4-week programme used in this study was designed to provide training in the skills and principles of mindfulness in an abbreviated form vs.	Diagnosis of MS Cognitive impairment: does not appear to be a specific inclusion criterion for cognitive impairment	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Indirectness as outcomes only reported at 4-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	Waitlist control: did not engage in any training		
Mindfulness vs. general cognitive rehabilitation (different types of strategies combined, e.g., computer training for skills + teaching other strategies)			
Manglani 2020[40] Associated papers: Schirda 2020[64] USA N=40 in these two groups	**Mindfulness intervention:** group sessions including combination of didactics, group discussion and practice with training materials. Also, at-home practice requested. Based on Jon Kabat-Zinn's 8-week programme, 4-week programme used in this study was designed to provide training in the skills and principles of mindfulness in an abbreviated form vs. **Adaptive cognitive training:** group training sessions covering multiple domains including attention, processing speed, executive functions and working memory. Training involved didactics, group discussion and practice with training materials in the form of BrainHQ games and additional home practice. Adaptive process starting with building blocks of cognition and moving on to higher-order cognitive domains such as executive functioning.	Diagnosis of MS Cognitive impairment: does not appear to be a specific inclusion criterion for cognitive impairment	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Indirectness as outcomes only reported at 4-weeks (end of intervention period)
Mindfulness vs. medical treatment and counselling			
Nazaribadie 2020[53] Associated papers: Nazaribadie 2021[54] Iran	**Detached mindfulness:** performed in group sessions delivered by psychologists over 8 sessions, with one session weekly (60-70 min per session). Described as a meta-cognitive model of detached mindfulness.	Diagnosis of MS Cognitive impairment: information processing dysfunction (based on either PASAT test or Wisconsin Card	

Study	Intervention and comparison	Population	Comments
N=60	vs. **Control – outpatient visits with counselling:** visited outpatient clinic once weekly. Received medical treatment and counselling about MS complications, coping with these complications and socio-therapeutic factors. Pharmacological treatment consisted of interferon beta-1a weekly.	Sorting Test) an inclusion criterion	
Focus on information processing speed			
Cognitive rehabilitation focused on processing speed + occupational therapy vs. occupational therapy alone			
Azimian 2021[2] Iran N=71	**Cognitive-based rehabilitation focused on processing speed + usual occupational therapy:** usual occupational therapy involved several exercises for 30 min in 12 sessions (bending to sides in standing position, forward bending, toe standing, heel standing, heel cord stretch with bent knee, one leg standing, one leg standing with eyes closed, rotating the head in standing position or while walking, maintaining quadruped position, kneel standing and walking). Cognitive-based rehabilitation involved processing speed tasks for 4 weeks with at least two tasks in each session. vs. **Usual occupational therapy only:** usual occupational therapy involved several exercises for 30 min in 12 sessions (bending to sides in standing position, forward bending, toe standing,	Diagnosis of MS with EDSS score <5.0 Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper

Study	Intervention and comparison	Population	Comments
	heel standing, heel cord stretch with bent knee, one leg standing, one leg standing with eyes closed, rotating the head in standing position or while walking, maintaining quadruped position, kneel standing and walking).		
Cognitive rehabilitation software focused on processing speed vs. control			
Bove 2021[4] USA N=44	**Sensory and motor tasks designed to improve processing speed:** in-home, tablet-based video game-like digital treatment. Investigational medical device software developed by Akili Interactive. Uses Selective Stimulus Management Engine engaging patients in two simultaneous sensory and motor tasks and designed to engage frontal neural networks. vs. **Control – active control digital game:** administered on digital platform similar to intervention game. Aim is to connect letters on a grid and spell as many words as possible. Points earned by tracing words with two or more letters in any direction based on number of words formed, word length and use of uncommon letters with progressive difficulty. Active placebo control used to provide similar time on task and engagement.	Clinically isolated syndrome or MS diagnosis Cognitive impairment: SDMT z-scores between -2 and 1 (compared to a healthy population)	Indirectness as outcomes only reported at 6-weeks (end of intervention period)
Chiaravallot i 2018[12] USA N=21	**Speed of processing training:** computerised training sessions involving practice on three types of tasks presented on a	Clinically definite MS Cognitive impairment: impaired	Indirectness as outcomes only reported at 5-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	computer (simple speed of processing, divided attention and selective attention). Customised to each patient's ability and increases in difficulty based on performance. vs. **Control:** no treatment	processing speed at baseline (performance 1.5 SD below mean of published normative data on SDMT)	
Focus on information processing speed + working memory			
Visual n-back programs focused on working memory and processing speed vs. sham control			
Hancock 2015[30] Associated papers: Hancock 2014[29] USA N=71	**Processing speed and working memory training:** computerised cognitive training in homes using Posit Science InSight and Brain Twister (working memory) visual n-back programs. Processing tasks used were PositScience's Sweet Seeker and Road Tour. Detailed instructions on which modules to complete. Game continually challenged participants by increasing speed of stimuli presentation and making discriminations more difficult Vs. **Control – sham training group:** same programmes as in the intervention group used but sham control group tasks did not increase in difficulty and played a 0-back version of the game.	MS diagnosis Cognitive impairment: subjectively reported cognitive complaints was an inclusion criterion	Indirectness as outcomes only reported at 6-weeks (end of intervention period)
Focus on attention/working memory			
Computer-aided RehaCom training for attention (includes 'Divided Attention' or other tasks said to be focused on attention, with or without memory-specific modules) vs. active control			
Cerasa 2013[8] Italy	**Computer-assisted training:** computer-assisted training of several attention ability	Relapsing-remitting MS	Indirectness as outcomes only reported at 6-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
N=26	and information processing tasks using RehaCom software. Included 'divided attention' 'attention and concentration' and 'vigilance' modules. vs. **Control computer training:** visuomotor coordination task by using an in-house software with which they had to simply respond quickly and accurately to the appearance of target visual stimuli	Cognitive impairment: predominant deficits in either attention and/or information processing speed, working memory and/or executive function (mild-moderate as severe excluded)	
Computer aided training for attention/working memory vs. control (no training or control intervention not related to cognitive training)			
Blair 2021[3] Canada N=30	**Computer-assisted working memory training – CogMed:** online training involving eight exercises per day. Uses adaptive training approach where difficulty level is adjusted in real time based on performance. Each session involves various tasks targeting different aspects of working memory including visuospatial working memory and verbal working memory tasks. Reinforcement built into program in form of small weekly rewards. Each person had coach to provide feedback, structure and motivation. vs. **Control – usual treatment:** standard medical care.	Relapsing-remitting or progressive MS with EDSS score up to 7.0 Cognitive impairment: subjective reporting of cognitive difficulties and z-score <-1.5 on at least 2 of 3 measures (PASAT, SDMT and DKFES Color-Word Interference Test) and therefore characterised as having attention/working memory deficits.	
Campbell 2016[6] UK	**RehaCom cognitive rehabilitation:** divided attention, working memory and topological memory modules of	Clinically definite MS with EDSS score up to 6.5	

Study	Intervention and comparison	Population	Comments
N=38	RehaCom software. Difficulty tailored to individual's performance and increases automatically in line with progress. vs. **Control – natural history DVDs:** watched series of natural history DVDs of corresponding duration and frequency for intervention period.	Cognitive impairment: cognitive impairment defined as scores below 5th percentile for normative data adjusted for age, sex and years of formal education on one or more of Brief International Cognitive Assessment for MS tests	
High-intensity working memory training vs. distributed working memory training vs. control (no training)			
Vogt 2009[75] Switzerland N=45	**High intensity working memory training:** 45 min training four times weekly for 4 weeks vs. **Distributed working memory training:** 45 min training two times weekly for 8 weeks. vs. **Control:** no training during intervention period	Clinically definite MS Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Indirectness as outcomes only reported at 4-8-weeks (end of intervention period)
Attention Process Training (APT – focused, sustained, selective, alternating and divided attention, specific computer programme focused on attention) + multidisciplinary rehabilitation vs. multidisciplinary rehabilitation alone			
Grasso 2017[28] Italy N=34	**Cognitive training + multidisciplinary rehabilitation:** individualised, goal-oriented multidisciplinary inpatient programme performed, which for this group included cognitive training and standard physical rehabilitation (described below in control group). Cognitive training involved intensive computer-assisted cognitive rehabilitation for attention, information	MS diagnosis Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper

Study	Intervention and comparison	Population	Comments
	processing and executive functions. Based on Attention Processing Training program - group of hierarchically organised tasks that exercise different components of attention that are commonly impaired after brain injury including sustained, selective, alternating and divided attention. Tasks place increasing demands on complex attentional control and working memory systems. **Control – multidisciplinary rehabilitation without cognitive training:** individualised, goal-oriented multidisciplinary inpatient programme performed. Standard rehabilitation programme involved physiotherapy sessions (aimed at improving movements on paretic side and at upper-limb exercises as well as improving balance, standing, sitting and transferring).		
Reaction time tasks + usual rehab programme vs. active control (software aiming to improve similar cognitive functions of selective attention, cognitive flexibility and working memory but with no time component)			
Flachenecker 2017[25] Germany N=32	**Neuropsychological training with reaction time tasks + usual rehabilitation programme:** computerised training using reaction time tasks in software packages 'Reaktion' and 'Jeton' by Petra Rigling REHA Software. Each involves four programmes allowing demands to vary in terms of time constraint and difficulty to adapt tasks to performance of patient.	MS diagnosis and reported to be experiencing fatigue Cognitive impairment: abnormal results in neuropsychological testing of intensity of attention (T-values of mean reaction times <40).	Indirectness as outcomes only reported at 2-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	Also received usual, goal-oriented, specifically tailored rehabilitation programme vs. **Control – unspecific neuropsychological training without time components + usual rehabilitation programme:** computerised training using software packages 'Bilder', 'Garten', 'Mosaik', 'Partino' and 'Vario' by Petra Rigling REHA Software were used. Designed to improve distinct cognitive functions such as selective attention, cognitive flexibility and working memory. Training adjusted by neuropsychologist to possibilities and improvements of the patient. Also received usual, goal-oriented, specifically tailored rehabilitation programme		
Focus on memory (with or without attention components also included)			
Computer-aided training for memory (with or without attention components also included) vs. control (no training)			
Hildebrandt 2007[32] Germany N=42	**Computerised cognitive training:** CD with memory and working memory rehabilitation tasks (VILAT-G 1.0), including remembering lists and calculations. Increased in difficulty. vs. **Control:** no training	Relapsing-remitting MS Cognitive impairment: taking the results of all neuropsychological tests together 48% of control group and 47% of the treatment group showed some impairment	Indirectness as outcomes only reported at 6-weeks (end of intervention period)
Mendozzi 1998[47]	**Specific cognitive retraining programme:**	Relapsing-remitting or	Does not appear to have cognitive impairment as an inclusion criterion

Study	Intervention and comparison	Population	Comments
Italy N=40 for these two groups	training of memory and attention using RehaCom computer software. Two consecutive training periods in each session, one on memory task and another on attention task. Twelve difficulty levels. vs. **Control:** no cognitive training during intervention period.	secondary progressive chronic MS Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	but cognition is the focus of the paper Memory described as main focus of study
Computer-aided training for memory (with attention components also included) vs. active control			
Mendozzi 1998[47] Italy N=40 for these two groups	**Specific cognitive retraining programme:** training of memory and attention using RehaCom computer software. Two consecutive training periods in each session, one on memory task and another on attention task. Twelve difficulty levels. vs. **Non-specific cognitive retaining programme:** two similar training periods, one spent on visual tracking task and other on reaction-time task. Designed to train cognitive abilities other than memory.	Relapsing-remitting or secondary progressive chronic MS Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Memory described as main focus of study
Solari 2004[68] Italy N=82	**Computer-aided retraining of memory and attention:** individual training using RehaCom software including memory and attention modules. vs. **Control – visuo-constructional and visuo-motor coordination**	MS diagnosis Cognitive impairment: complained of poor attention or memory, confirmed by a score below the 80th percentile in at least two components of the Brief Repeatable	

Study	Intervention and comparison	Population	Comments
	retraining: using modules of RehaCom software. Designed as a sham intervention as they primarily train motor skills and adapted to minimise possible effects on attention and memory retraining.	Battery of Neuropsychological Tests	
Story Memory Technique (SMT) vs. control			
Chiaravalotti 2005[11] USA N=29	**Story Memory Technique:** participant learns the story memory technique, which involves participant being taught two interrelated skills: 1) to use visualisation to facilitate new learning and 2) to utilise context to learn new information (e.g., a story even if information is seemingly unrelated) vs. **Control – non-training orientated tasks:** met with therapist as in intervention group but engaged in non-training orientated tasks to control for contact.	Clinically definite MS Cognitive impairment: 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	Indirectness as outcomes only reported at 6-11 weeks (2-7 weeks after end of intervention period)
Chiaravalotti 2012[15] USA N=16	**Modified Story Memory Technique:** 10 sessions using this technique involving two related skills of imagery and context. First four sessions taught use of imagery to facilitate learning of verbal information and subsequent four sessions taught to use context to facilitate learning. Final two sessions involved applying this technique to real-life situations. vs. **Control – placebo intervention sessions** met with therapists as in	Clinically definite MS Cognitive impairment: new learning and memory abilities at least 1.5 SD lower than mean of healthy control group based on Selective Reminding Test	Indirectness as outcomes only reported at 5-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	intervention group but engaged in verbal tasks such as reading stories and answering questions to control for contact.		
Chiaravalotti 2013[14] USA MEMREHAB trial N=88	**Story Memory Technique:** participant learns the story memory technique, which involves participant being taught two interrelated skills: 1) to use visualisation to facilitate new learning and 2) to utilise context to learn new information (e.g., a story even if information is seemingly unrelated) vs. **Control – non-training orientated tasks:** met with therapist as in intervention group but engaged in non-training orientated tasks to control for contact.	Clinically definite MS Cognitive impairment: new learning impairment was an inclusion criterion	Indirectness as outcomes only reported at 5-weeks (end of intervention period) Included in previous guideline version
Chiaravalotti 2020[13] USA N=30	**Modified Story Memory Technique:** 10 sessions using this technique involving two related skills of imagery and context. First four sessions taught use of imagery to facilitate learning of verbal information and subsequent four sessions taught to use context to facilitate learning. Final two sessions involved applying this technique to real-life situations. vs. **Control – placebo intervention sessions:** met with therapists as in intervention group but engaged in verbal tasks such as reading stories and answering	Diagnosis of MS Cognitive impairment: new learning and memory impairment (1.5 SD+ compared to normative Open Trial Selective Reminding Test)	Indirectness as outcomes only reported at 5-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	questions to control for contact.		
Krch 2019[35] Mexico N=20	**Modified Story Memory Technique:** Spanish version translated from English by bilingual researcher. 10 sessions using this technique involving two related skills of imagery and context. First four sessions taught use of imagery to facilitate learning of verbal information and subsequent four sessions taught to use context to facilitate learning. Final two sessions involved applying this technique to real-life situations. vs. **Placebo control:** exposed to same stories and target words but not taught how to apply imagery and context to the material. Training sessions matched to treatment sessions in timing and frequency as well as presentation.	MS diagnosis Cognitive impairment: impaired new learning (measured by failing to achieve perfect recall on 2 consecutive trials by trial 7 on Open Trial administration of Selective Reminding Test)	Indirectness as outcomes only reported at 5-weeks (end of intervention period)
External compensatory strategies (e.g., lists, diaries and visual mnemonics) vs. control (feedback only with no further intervention)			
Lincoln 2002[39] UK N=240	**Cognitive assessment with feedback and cognitive rehabilitation:** training performed following detailed cognitive assessment and feedback included various techniques such as diaries, lists, and visual mnemonics, with techniques differing depending on deficits identified for each participant vs. **Control – two groups:**	Clinically definite, probably or lab-supported MS Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Median values only reported

Study	Intervention and comparison	Population	Comments
	• Screening only: underwent assessment only at screening and did not have additional, detailed cognitive assessment which was part of the intervention group • Assessment with feedback: underwent same detailed cognitive assessment as in the intervention group and received feedback on this with no cognitive rehabilitation intervention.		
External compensatory training vs. restitution training (internal ability to code, organise and retrieve information) vs. self-help control group			
Martin 2014[43] UK ReMIND study (MS subpopulation) N=39	**Compensatory memory training – external aids:** 10 sessions with homework giving opportunity to practice strategies learned in sessions. Participants in both of the intervention programmes (restitution and compensation) were taught the use of internal memory aids and errorless learning techniques. Compensation group were also taught how to use external memory aids (e.g., diaries). vs. **Restitution memory training – coding, organisation and retrieval of information:** 10	MS diagnosis verified by clinician Cognitive impairment: reporting of memory problems was an inclusion criterion	Median values only

Study	Intervention and comparison	Population	Comments
	sessions with homework giving opportunity to practice strategies learned in sessions. Participants in both of the intervention programmes (restitution and compensation) were taught the use of internal memory aids and errorless learning techniques. Those in the restitution group completed exercises to practice encoding and retrieval, and also included attention-retraining exercises, such as letter and number cancellation. Participants in the restitution group were also taught how to encode and retrieve specific information (e.g., remembering people's names by paying attention not only to the acoustic and orthographic presentation of the name but by creating a visual image of the name). vs. **Control – self-help control group:** not taught any memory strategies, but were taught relaxation techniques and ways in which they could cope with their condition.		
Group memory programme (various learning techniques, including internal and external aids) vs. control			
Carr 2014[7] UK N=48	**Group memory programme:** group sessions and homework including both restitution and compensation strategies. Training sessions covered attention training, internal memory strategies and external memory aids.	MS diagnosis Cognitive impairment: reporting memory problems in daily life was an inclusion criterion (though very severe memory	

Study	Intervention and comparison	Population	Comments
	vs. **Control – usual care:** received their usual care and all other rehabilitation (e.g., physiotherapy, occupational therapy) continued as usual.	problems excluded if may interfere with group session participation)	
Lincoln 2020[37] Associated papers: Lincoln 2020[38] UK CRAMMS study N=449	**Group cognitive programme mainly focusing on memory + usual care:** cognitive rehabilitation in group sessions, with homework, including restitution strategies to retrain attention and memory functions and strategies to improve encoding and retrieval. Compensation strategies taught included the use of internal mnemonics (such as chunking) and external devices (such as diaries and mobile phones) and ways of coping with attention and memory problems. Also received usual care as described in the control group. **Control – usual care:** usual care involved general advice from multiple sclerosis nurse specialists and occupational therapists on how to manage any cognitive difficulties. All participants were notified of information available on the webpages of multiple sclerosis charities.	Relapsing-remitting or progressive MS diagnosed for at least 3 months Cognitive impairment: reported having cognitive problems defined as >27 on the patient version of the Multiple Sclerosis Neuropsychological Screening Questionnaire and impaired on at least one of the Brief Repeatable Battery of Neuropsychological tests (defined as performance >1 SD below the mean of healthy controls, corrected for age and education)	
Mousavi 2018[51] Associated papers: Mousavi 2020[50]	**Group cognitive memory programme:** group-based programme involving training in compensatory strategies, explanations on different types of	MS diagnosis Cognitive impairment: Multiple sclerosis neuropsychological screening	Two comparator groups were combined into a single comparator group to compared with the intervention group for the purpose of this review.

Study	Intervention and comparison	Population	Comments
Iran N=60	internal and external memory aids, mnemonics, mental reviews and error-free learning. Memory problem adaptation methods offered based on individual difficulties and predetermined objectives. vs. **Control – placebo and control groups:** placebo group received body relaxation techniques during weekly sessions as well as usual care of offering information regarding cognitive problems. Control group were given ordinary information regarding cognitive problems in MS only (usual care).	questionnaire ≤ 27 and achieving 2 standard deviations lower than the healthy people on the scale of brief repeatable battery of neuropsychological test	
Shahpouri 2019[65] Iran N=66	**Tailored cognitive rehabilitation:** group cognitive rehabilitation with general aim of reinforcing/consolidation of previous cognitive abilities that have been impaired and reinforcing remaining abilities to compensate for those where there are impairments. Included attention, concentration, visual, auditory memory and autobiographical memory. Mnemonic approach which includes visual imagery, theological organization, and relational strategies including mnemonics of fiction, the clues about the first word, chain connection, and the technique of Preview, Question, Read, Self-recitation and Test. Involved explanations of disturbances in daily life	MS diagnosis with EDSS ≤5.5 Cognitive impairment: mild to moderate memorial impairment based on Everyday Memory Questionnaire Mild to moderate depression status based on second version of Beck depression inventory was also an inclusion criterion.	

Study	Intervention and comparison	Population	Comments
	and training of skills using techniques. vs. **Control – discussion only:** discussion of experiences and coping strategies only. Content of the sessions was different to intervention group and was not supporting cognitive rehabilitation.		
Behavioural intervention (Self-GEN trial) focused on teaching self-generation technique with metacognitive strategies vs. control (memory tasks but no self-generated learning and transfer instructions)			
Goverover 2018[27] USA Self-GEN trial N=35	**Self-generation learning programme focused on memory:** 6 sessions of individualised treatment. Items to be learned presented in provided and self-generated conditions (given list or filling in blanks), followed by immediate recall. Then asked which of these versions they remembered better and what helped them to remember it better. Recall results then presented to participants and researcher explained self-generation potential in memory and recall. First two stages repeated with different stimuli and participants asked how self-generation strategy can be used. Asked to complete journal summarising activity sessions and what was learned. vs. **Control – memory tasks with no self-generation element:** met with researcher as	Clinically definite MS Cognitive impairment: documented memory impairment based on selective memory test (SRT; those scoring at least 0.5 SD less than the mean of healthy control group)	Indirectness as outcomes only reported at 3-4-weeks (end of intervention period, measured within 1 week of completion)

Study	Intervention and comparison	Population	Comments
	in the intervention group and performed the same memory tasks but not exposed to self-generated learning and transfer instructions.		
Focus on executive function			
Executive function-specific training exercises vs. control			
Fink 2010[24] Germany N=50	**Cognitive intervention focused on executive function exercises:** textbook exercises for executive functioning and they met with a psychologist to receive feedback and to discuss the exercises vs. **Control – placebo:** performed tasks using RehaCom software where they had to respond quickly and accurately to visual stimuli. Had to call psychologist once weekly to report training time. vs. **Control – untrained group:** no training received	Relapsing-remitting MS with EDSS up to 7.0 Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Indirectness as outcomes only reported at 6-weeks (end of intervention period)
Sharifi 2019[66] Iran N=20	**Cognitive training focused on executive function:** computer-based cognitive rehabilitation using Captain's Log software. Two programs focused on executive functions used involving stimulus reaction/inhibition (red light and green light) and scanning reaction/inhibition (mouse hunt), each with 15 stages of increasing difficulty. vs.	MS diagnosis Cognitive impairment: does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper	Does not appear to have cognitive impairment as an inclusion criterion but cognition is the focus of the paper Indirectness as outcomes only reported at 6-weeks (end of intervention period)

Study	Intervention and comparison	Population	Comments
	Control: no training.		
Goal management programme vs. control (psychoeducation)			
Richard 2013[60] Canada N=28	**Goal management programme:** designed to be highly interactive, combining lectures on key topics with discussions relating to participants' experiences with in-class activities and homework. Group-based programme focused on information and activities to build skills in goal awareness, attentional control and self-regulation, while providing a socially supportive atmosphere to practice and discuss progress with these skills. vs. **Control – psychoeducation:** brain health workshop (psychoeducation). Designed to be highly interactive, combining lectures on key topics with discussions relating to participants' experiences with in-class activities and homework. Focused on increasing knowledge of brain function, cognition and MS, while providing social support and lifestyle recommendations. Differs from goal management training as though may increase awareness of potential deficits in cognition, they don't provide specific tools to help patients improve these deficits.	MS diagnosis with EDSS up to 8.0 Cognitive impairment: preliminary indication of functionally significant attention or executive deficits (e.g., from clinical presentation, chart information from the referring institutional clinic and/or patient self-report) and objective evidence of functionally significant attention or executive deficits (as determined by the baseline neuropsychological evaluation)	
<u>Focus on improving language</u>			
RehaCom verbal fluency training vs. control			
Arian Darestani 2020[1]	**Verbal fluency intervention:** RehaCom cognitive rehabilitation	MS diagnosis	Indirectness as outcomes only reported at 10-weeks (5 weeks after

Study	Intervention and comparison	Population	Comments
Iran N=60	software used. Unclear whether specific modules of this software used to target verbal fluency. vs. **Control:** no definition but assume no intervention	Cognitive impairment: included those referred to a Brain and Cognition Clinic	the end of 5-week intervention period)

See appendices for full evidence tables.

1.1.6 Summary of the effectiveness evidence

See the separate headings below for effectiveness evidence for the various comparisons included in the review. See appendices for full GRADE and/or GRADE-CERQual tables.

General cognitive rehabilitation (multi-component and multi-domain) vs. control

Table 3: Clinical evidence summary: General cognitive rehabilitation (multi-component and multi-domain) vs. control, 1-6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
Selective Reminding Test - Long-term storage follow up: 1-6 months	132 (2 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean selective Reminding Test - Long-term storage ranged from 44.2-53.9	MD 2.19 higher (2.48 lower to 6.86 higher)
Selective Reminding Test - Long-term storage - 1-6 months - Consistent long-term retrieval follow up: 1-6 months	132 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test - Long-term storage - 1-6 months - Consistent long-term retrieval ranged from 36.3-45.7	MD 3.06 higher (2.91 lower to 9.02 higher)
Selective Reminding Test - Long-term storage - 1-6 months - Delayed recall follow up: 1-6 months	233 (3 RCTs)	⊕⊕⊕◯ MODERATE a	-	The mean selective Reminding Test - Long-term storage - 1-6 months - Delayed recall ranged from 7.53-11.6	MD 0.4 higher (0.23 lower to 1.03 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
Selective Reminding Test - 1-6 months - Mean free recall follow up: 6 months	101 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean selective Reminding Test - 1-6 months - Mean free recall was 10.6	MD 0 (0.74 lower to 0.74 higher)
Selective Reminding Test - 1-6 months - Learning index follow up: 6 months	101 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean selective Reminding Test - 1-6 months - Learning index was 54.0	MD 6.7 higher (1.91 lower to 15.31 higher)
10/36 Spatial Recall Test - 1-6 months - Total score follow up: 1-6 months	233 (3 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean 10/36 Spatial Recall Test - 1-6 months - Total score ranged from 17.1-23.1	MD 1.15 higher (1.3 lower to 3.59 higher)
10/36 Spatial Recall Test - 1-6 months - Delayed recall follow up: 1-6 months	233 (3 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean 10/36 Spatial Recall Test - 1-6 months - Delayed recall ranged from 6.2-8.24	MD 0.06 higher (1.21 lower to 1.32 higher)
SDMT - 1-6 months - Similar at baseline follow up: 1-6 months	376 (4 RCTs)	⊕⊕◯◯ LOW a	-	The mean SDMT - 1-6 months - Similar at baseline ranged from 48.2-53.5	MD 1.65 higher (0.77 lower to 4.06 higher)
SDMT - 1-6 months - Larger difference at baseline (lower in intervention) follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean SDMT - 1-6 months - Larger difference at baseline (lower in intervention) was 47.52	MD 4.9 lower (12.6 lower to 2.8 higher)
PASAT (2 seconds) - 1-6 months follow up: 1-6 months	376 (4 RCTs)	⊕◯◯◯ VERY LOW a,c,f	-	The mean PASAT (2 seconds) - 1-6 months ranged from 20.8-42.1	MD 1.96 higher (4.31 lower to 8.23 higher)
PASAT (3 seconds) - 1-6 months follow up: 1-6 months	436 (5 RCTs)	⊕◯◯◯ VERY LOW a,e	-	The mean PASAT (3 seconds) - 1-6 months ranged from 19.56-52.7	MD 2.69 higher (1.37 higher to 4.01 higher)
COWAT - 5-6 months	342 (3 RCTs)	⊕⊕◯◯ LOW a	-	The mean COWAT - 5-6 months ranged from 24.2-28.1	MD 1.37 higher (0.77 lower to 3.52 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
follow up: 5-6 months					
Stroop test time - 5-6 months - Colour naming time follow up: 6 months	98 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean stroop test time - 5-6 months - Colour naming time was 77.0	MD 3.3 lower (10.45 lower to 3.85 higher)
Stroop test time - 5-6 months - Colour/word interference time follow up: 6 months	98 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean stroop test time - 5-6 months - Colour/word interference time was 116.0	MD 0.2 higher (13.03 lower to 13.43 higher)
Stroop test time - 5-6 months - General 'Stroop test' follow up: 5	60 (1 RCT)	⊕◯◯◯ VERY LOW a,e	-	The mean stroop test time - 5-6 months - General 'Stroop test' was 11.96	MD 2.83 lower (3.63 lower to 2.03 lower)
Stroop test - 3 months - Word-color test follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean stroop test - 3 months - Word-color test was 43.62	MD 1.05 lower (7.99 lower to 5.89 higher)
Stroop test - 3 months - Interference follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean stroop test - 3 months - Interference was 2.79	MD 2.55 higher (1.78 lower to 6.88 higher)
Trail Making Test time - 6 months - Part A follow up: 3-6 months	140 (2 RCTs)	⊕⊕⊕◯ MODERATE a	-	The mean trail Making Test time - 6 months - Part A ranged from 31.0-40.3	MD 1.62 higher (2.34 lower to 5.59 higher)
Trail Making Test time - 6 months - Part B follow up: 6 months	98 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean trail Making Test time - 6 months - Part B was 75.4	MD 3.7 higher (10.77 lower to 18.17 higher)
California Verbal Learning Test (CVLT) - 5 months - Total follow up: 5 months	244 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean california Verbal Learning Test (CVLT) - 5 months - Total ranged from 53.8-54.7	MD 2.93 higher (0.26 lower to 6.11 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
California Verbal Learning Test (CVLT) - 5 months - Delayed follow up: 5 months	244 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean california Verbal Learning Test (CVLT) - 5 months - Delayed ranged from 11.9-12.5	MD 0.4 higher (0.53 lower to 1.33 higher)
Hopkins Verbal Learning Test - Revised - 3 months - Learning follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean hopkins Verbal Learning Test - Revised - 3 months - Learning was 24.81	MD 0.33 lower (3.07 lower to 2.41 higher)
Hopkins Verbal Learning Test - Revised - 3 months - Recall follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean hopkins Verbal Learning Test - Revised - 3 months - Recall was 9.48	MD 0.77 lower (2.15 lower to 0.61 higher)
Brief Visuospatial Memory Test (BVMT) - 5 months - Total follow up: 5 months	244 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean brief Visuospatial Memory Test (BVMT) - 5 months - Total ranged from 20.7-24.6	MD 0.98 higher (0.65 lower to 2.61 higher)
Brief Visuospatial Memory Test (BVMT) - 5 months - Delayed follow up: 5 months	244 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean brief Visuospatial Memory Test (BVMT) - 5 months - Delayed ranged from 7.7-8.8	MD 0.5 higher (0.14 lower to 1.14 higher)
Digit Span - 3-6 months - Forward follow up: 6 months	101 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean digit Span - 3-6 months - Forward was 5.7	MD 0.1 higher (0.35 lower to 0.55 higher)
Digit Span - 3-6 months - Backward follow up: 3-6 months	143 (2 RCTs)	⊕⊕◯◯ LOW a,c	-	The mean digit Span - 3-6 months - Backward ranged from 4.5-6.24	MD 0.28 higher (0.21 lower to 0.76 higher)
Word List Generation - 1 month follow up: 1 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,c,g	-	The mean word List Generation - 1 month was 32.0	MD 1.9 higher (3.72 lower to 7.52 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
Wisconsin Card Sorting Test (time as described as benefits in intervention group?) 5 months follow up: 5 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,e	-	The mean wisconsin Card Sorting Test (time as described as benefits in intervention group?) 5 months was 13.29	MD 3.1 lower (4.09 lower to 2.11 lower)
Test of Attentional Performance (TAP) - Working Memory domain omissions - 6 months follow up: 6 months	101 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean test of Attentional Performance (TAP) - Working Memory domain omissions - 6 months was 2.9	MD 0.1 lower (1.1 lower to 0.9 higher)
Test of Attentional Performance (TAP) - Flexibility domain correct answers - 6 months follow up: 6 months	101 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean test of Attentional Performance (TAP) - Flexibility domain correct answers - 6 months was 96.0	MD 4.6 lower (8.8 lower to 0.4 lower)
Test of Attentional Performance (TAP) - Incompatibility domain correct answers - 6 months follow up: 6 months	101 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean test of Attentional Performance (TAP) - Incompatibility domain correct answers - 6 months was 56.5	MD 3.3 lower (7.41 lower to 0.81 higher)
Test of Attentional Performance (TAP) reaction time - 5 weeks - Alertness - simple follow up: 5 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c,h	-	The mean test of Attentional Performance (TAP) reaction time - 5 weeks - Alertness - simple was 269.8	MD 19.2 lower (34.64 lower to 3.76 lower)
Test of Attentional Performance (TAP) reaction	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c,h	-	The mean test of Attentional Performance (TAP) reaction time - 5	MD 21.5 lower (36.84 lower to 6.16 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
time - 5 weeks - Alertness - cued follow up: 5 weeks				weeks - Alertness - cued was 264.3	
Test of Attentional Performance (TAP) reaction time - 5 weeks - Divided Attention - acoustic follow up: 5 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c,h	-	The mean test of Attentional Performance (TAP) reaction time - 5 weeks - Divided Attention - acoustic was 605.1	MD 29.6 lower (99.47 lower to 40.27 higher)
Test of Attentional Performance (TAP) reaction time - 5 weeks - Divided Attention - visual follow up: 5 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c,h	-	The mean test of Attentional Performance (TAP) reaction time - 5 weeks - Divided Attention - visual was 865.5	MD 59.5 lower (105 lower to 14 lower)
Brief Test of Attention - 3 months follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean brief Test of Attention - 3 months was 15.1	MD 2.29 lower (4.69 lower to 0.11 higher)
Delis–Kaplan Executive Function System (D-KEFS) - 5 months - Descriptive follow up: 5 months	61 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean delis–Kaplan Executive Function System (D-KEFS) - 5 months - Descriptive was 41.7	MD 2.1 lower (7.02 lower to 2.82 higher)
Delis–Kaplan Executive Function System (D-KEFS) - 5 months - Sort follow up: 5 months	61 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean delis–Kaplan Executive Function System (D-KEFS) - 5 months - Sort was 10.9	MD 0.7 lower (1.94 lower to 0.54 higher)
Verbal fluency - 6 months - Letter M follow up: 6 months	101 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean verbal fluency - 6 months - Letter M was 12.5	MD 0.6 higher (1.06 lower to 2.26 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
Verbal fluency - 6 months - Animals follow up: 6 months	101 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean verbal fluency - 6 months - Animals was 19.0	MD 1.4 higher (0.81 lower to 3.61 higher)
Calibrated Ideational Fluency Assessment - 3 months - Animals follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean calibrated Ideational Fluency Assessment - 3 months - Animals was 22.24	MD 0.67 lower (4.63 lower to 3.29 higher)
Calibrated Ideational Fluency Assessment - 3 months - Supermarket follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean calibrated Ideational Fluency Assessment - 3 months - Supermarket was 21.1	MD 2.29 lower (6.45 lower to 1.87 higher)
Calibrated Ideational Fluency Assessment - 3 months - P-words follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean calibrated Ideational Fluency Assessment - 3 months - P-words was 30.57	MD 2.95 lower (8.58 lower to 2.68 higher)
MUSIC (unclear which outcome set/measure this is referring to) - 5 weeks - Verbal memory follow up: 5 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c,h	-	The mean MUSIC (unclear which outcome set/measure this is referring to) - 5 weeks - Verbal memory was 14.38	MD 0.12 higher (1.99 lower to 2.23 higher)
MUSIC (unclear which outcome set/measure this is referring to) - 5 weeks - Verbal retrieval follow up: 5 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c,h	-	The mean MUSIC (unclear which outcome set/measure this is referring to) - 5 weeks - Verbal retrieval was 5.88	MD 0.23 higher (0.69 lower to 1.15 higher)
MUSIC (unclear which outcome set/measure this is referring to) - 5 weeks - Verbal fluency	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c,h	-	The mean MUSIC (unclear which outcome set/measure this is referring to) - 5 weeks - Verbal fluency was 14.88	MD 0.16 lower (2.29 lower to 1.97 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
follow up: 5 weeks					
MUSIC (unclear which outcome set/measure this is referring to) - 5 weeks - Interferences follow up: 5 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c,h	-	The mean MUSIC (unclear which outcome set/measure this is referring to) - 5 weeks - Interferences was 12.01	MD 2.82 lower (6.73 lower to 1.09 higher)
Judgement of Line Orientation (JLO) - 5 months follow up: 5 months	61 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean judgement of Line Orientation (JLO) - 5 months was 27.4	MD 0.4 higher (1.66 lower to 2.46 higher)
Salthouse Perceptual Comparison Test (baseline values not equal) - 3 months follow up: 3 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,e	-	The mean Salthouse Perceptual Comparison Test (baseline values not equal) - 3 months was 27.38	MD 2 lower (7.03 lower to 3.03 higher)
Code (assessing processing speed) - 6 months follow up: 6 months	101 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean code (assessing processing speed) - 6 months was 49.2	MD 1.9 lower (6.33 lower to 2.53 higher)
DO80 (assesses language) - Total score - 6 months follow up: 6 months	101 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean DO80 (assesses language) - Total score - 6 months was 77.5	MD 0.4 higher (0.52 lower to 1.32 higher)
DO80 (assesses language) - Time - 6 months follow up: 6 months	101 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean DO80 (assesses language) - Time - 6 months was 143.3	MD 10.2 lower (31.05 lower to 10.65 higher)
Perceived Deficits Questionnaire - 6 months Scale from: 0 to 80	98 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean perceived Deficits Questionnaire - 6 months was 36.8	MD 8.9 lower (13.83 lower to 3.97 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
follow up: 6 months					
MS Neuropsychological Questionnaire - 5-6 months - Patient-reported Scale from: 0 to 60 follow up: 5-6 months	159 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean MS Neuropsychological Questionnaire - 5-6 months - Patient-reported ranged from 26.15-28.5	MD 1.47 lower (8.06 lower to 5.12 higher)
MS Neuropsychological Questionnaire - 5-6 months - Informant-reported follow up: 6 months	98 (1 RCT)	⊕⊕⊕◯ MODERATE a,c	-	The mean MS Neuropsychological Questionnaire - 5-6 months - Informant-reported was 20.7	MD 1.4 lower (5.76 lower to 2.96 higher)
PROMIS - Applied Cognition Abilities short form 8a - 5 months (scale 8-40) Scale from: 8 to 40 follow up: 5 months	183 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean PROMIS - Applied Cognition Abilities short form 8a - 5 months (scale 8-40) was 23.4	MD 2.2 higher (0.03 higher to 4.37 higher)
MSIS-29 - 6 months (scale usually 0-100) - Physical Scale from: 0 to 100 follow up: 6 months	98 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean MSIS-29 - 6 months (scale usually 0-100) - Physical was 26.7	MD 4.1 lower (11.02 lower to 2.82 higher)
MSIS-29 - 6 months (scale usually 0-100) - Psychological Scale from: 0 to 100 follow up: 6 months	98 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean MSIS-29 - 6 months (scale usually 0-100) - Psychological was 27.1	MD 2.2 lower (9.32 lower to 4.92 higher)
MS International Quality of Life Questionnaire - Index (mean of	101 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean MS International Quality of Life Questionnaire - Index (mean of 9 subdomains, scale 0-	MD 1.1 higher (4.63 lower to 6.83 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
9 subdomains, scale 0-100) - 6 months Scale from: 0 to 100 follow up: 6 months				100) - 6 months was 58.4	
WHO-BREF Quality of Life - 6 months (scale used unclear) - S1 Physical health follow up: 6 months	98 (1 RCT)	⨁⨁◯◯ LOW a,c	-	The mean WHO-BREF Quality of Life - 6 months (scale used unclear) - S1 Physical health was 13.6	MD 0.6 higher (0.34 lower to 1.54 higher)
WHO-BREF Quality of Life - 6 months (scale used unclear) - S2 Psychological follow up: 6 months	98 (1 RCT)	⨁⨁◯◯ LOW a,c	-	The mean WHO-BREF Quality of Life - 6 months (scale used unclear) - S2 Psychological was 13.7	MD 0.3 higher (0.71 lower to 1.31 higher)
WHO-BREF Quality of Life - 6 months (scale used unclear) - S3 Social relationship follow up: 6 months	98 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean WHO-BREF Quality of Life - 6 months (scale used unclear) - S3 Social relationship was 14.6	MD 0.1 lower (1.25 lower to 1.05 higher)
WHO-BREF Quality of Life - 6 months (scale used unclear) - S4 environment follow up: 6 months	98 (1 RCT)	⨁⨁◯◯ LOW a,c	-	The mean WHO-BREF Quality of Life - 6 months (scale used unclear) - S4 environment was 14.7	MD 0.5 higher (0.46 lower to 1.46 higher)
WHO Quality of Life and Satisfaction with life composite, z-score - 1 month (Positive and negative values indicate score relative to the mean score in a general	34 (1 RCT)	⨁◯◯◯ VERY LOW a,c,g	-	The mean WHO Quality of Life and Satisfaction with life composite, z-score - 1 month was 0.16	MD 0.1 lower (0.66 lower to 0.45 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
population, with the mean being 0 and every 1 unit positive or negative deviation from 0 indicating 1 standard deviation above or below, respectively, the mean score on that test in a general population) Scale from: -5 to 5 follow up: 1 months					
Memory span (t-score of various tests) - 6 months vs. baseline (Score of 50 represents the mean score on the test in a general population, with every increase or decrease of 10 units representing 1 standard deviation above or below the mean score in a general population, respectively) Scale from: 0 to 100 follow up: 6 months	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean memory span (t-score of various tests) - 6 months vs. baseline was 2.4	MD 0.6 lower (4.41 lower to 3.21 higher)
Verbal learning (t-score of various tests) - 6 months vs. baseline	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean verbal learning (t-score of various tests) - 6 months vs. baseline was 0.6	MD 1.6 higher (2.07 lower to 5.27 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
Scale from: 0 to 100 follow up: 6 months					
Visuo-spatial memory (t-score of various tests) - 6 months vs. baseline Scale from: 0 to 100 follow up: 6 months	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean visuo-spatial memory (t-score of various tests) - 6 months vs. baseline was 0.2	MD 2.5 higher (0.1 higher to 4.9 higher)
Visuo-motor speed (t-score of various tests) - 6 months vs. baseline Scale from: 0 to 100 follow up: 6 months	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean visuo-motor speed (t-score of various tests) - 6 months vs. baseline was -1.0	MD 1.5 higher (2.26 lower to 5.26 higher)
Visual perception (t-score of various tests) - 6 months vs. baseline Scale from: 0 to 100 follow up: 6 months	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean visual perception (t-score of various tests) - 6 months vs. baseline was 1.0	MD 1.2 higher (0.14 lower to 2.54 higher)
Sum of 11 tests (t-score of various tests) - 6 months vs. baseline Scale from: 0 to 100 follow up: 6 months	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean sum of 11 tests (t-score of various tests) - 6 months vs. baseline was -0.5	MD 2.1 higher (0.25 lower to 4.45 higher)
Information processing speed (unclear how measured) - 5 months follow up: 5 months	60 (1 RCT)	⨁◯◯◯ VERY LOW a,e	-	The mean information processing speed (unclear how measured) - 5 months was 1122.5	MD 81.1 lower (118.05 lower to 44.15 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
Wechsler Adult Intelligence Scale - Similarities test (t-score) - 6 months vs. baseline Scale from: 0 to 100 follow up: 6 months	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean wechsler Adult Intelligence Scale - Similarities test (t-score) - 6 months vs. baseline was 2.1	MD 0.6 lower (5.45 lower to 4.25 higher)
Wechsler Adult Intelligence Scale - Picture arrangement (t-score) - 6 months vs. baseline Scale from: 0 to 100 follow up: 6 months	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean wechsler Adult Intelligence Scale - Picture arrangement (t-score) - 6 months vs. baseline was 4.2	MD 0.5 lower (6.44 lower to 5.44 higher)
Fatigue - FSMC cognitive subscale - 6 months Scale from: 10 to 50 follow up: 6 months	98 (1 RCT)	⨁⨁◯◯ LOW a,c	-	The mean fatigue - FSMC cognitive subscale - 6 months was 33.6	MD 2.6 lower (6.39 lower to 1.19 higher)
Beck Depression Inventory - 1-6 months, mix of final value and change scores (scale usually 0-63) Scale from: 0 to 63 follow up: 1-6 months	164 (3 RCTs)	⨁◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory - 1-6 months, mix of final value and change scores (scale usually 0-63) ranged from 2.7 for change scores and 10.0-10.2 for final values	MD 1.38 lower (4.21 lower to 1.45 higher)
CES-D depression - 5 months (scale usually 0-60) Scale from: 0 to 60 follow up: 5 months	183 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean CES-D depression - 5 months (scale usually 0-60) was 11.5	MD 1.6 lower (3.46 lower to 0.26 higher)
State-Trait Anxiety Inventory	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean state-Trait Anxiety Inventory (STAI) - State - 6	MD 2.7 lower (9.17 lower to 3.77 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
(STAI) - State - 6 months vs. baseline (scale usually 20-80) Scale from: 20 to 80 follow up: 6 months				months vs. baseline (scale usually 20-80) was 1.6	
State-Trait Anxiety Inventory (STAI) - Trait - 6 months vs. baseline (scale usually 20-80) Scale from: 20 to 80 follow up: 6 months	32 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean state-Trait Anxiety Inventory (STAI) - Trait - 6 months vs. baseline (scale usually 20-80) was -0.6	MD 0.9 lower (6.39 lower to 4.59 higher)
Penn State Worry Questionnaire - 1 month (scale usually 16-80) Scale from: 16 to 80 follow up: 1 months	34 (1 RCT)	⨁◯◯◯ VERY LOW a,c,g	-	The mean penn State Worry Questionnaire - 1 month (scale usually 16-80) was 42.6	MD 5.9 higher (4.13 lower to 15.93 higher)
Difficulties in Emotional Regulation Scale (DERS) - 1 month (scale unclear) follow up: 1 months	34 (1 RCT)	⨁◯◯◯ VERY LOW a,c,g	-	The mean difficulties in Emotional Regulation Scale (DERS) - 1 month (scale unclear) was 75.0	MD 0.5 lower (13.25 lower to 12.25 higher)
MS Self-Efficacy Scale - Control subscale (scale 90-900) - 5 months Scale from: 90 to 900 follow up: 5 months	61 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean MS Self-Efficacy Scale - Control subscale (scale 90-900) - 5 months was 534.26	MD 23.46 higher (69.09 lower to 116.01 higher)
General self-efficacy scale (scale possibly 17-85) - 5 months Scale from: 17 to 85	183 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean general self-efficacy scale (scale possibly 17-85) - 5 months was 62.5	MD 1.5 higher (1.69 lower to 4.69 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
follow up: 5 months					
Multi-factorial Memory Questionnaire - Strategy subscale (scale 0-76 and indicates use of memory strategies) - 5 months Scale from: 0 to 76 follow up: 5 months	244 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean multi-factorial Memory Questionnaire - Strategy subscale (scale 0-76 and indicates use of memory strategies) - 5 months ranged from 39.6-41.15	MD 1.17 higher (1.68 lower to 4.01 higher)
Everyday Problems Test - Revised (activities of daily living performance, scale unclear) - 5 months follow up: 5 months	183 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean everyday Problems Test - Revised (activities of daily living performance, scale unclear) - 5 months was 23.1	MD 0.7 higher (0.62 lower to 2.02 higher)
Adherence follow up: 6 months	128 (1 RCT)	⊕⊕◯◯ LOW c	OR 1.62 (0.73 to 3.59)	Moderate	
				688 per 1,000	93 more per 1,000 (71 fewer to 200 more)

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

b. Downgraded by 1 increment as direction of point estimates varies between studies, which cannot be explained by prespecified subgroup analyses

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

d. Downgraded by 1 increment as statistical heterogeneity is present that cannot be explained by prespecified subgroup analyses, with I2>50%

e. Downgraded by 1 increment as the majority of the evidence comes from studies that did not appear to have cognitive impairment as an inclusion criterion

f. Downgraded by 2 increments as statistical heterogeneity is present that cannot be explained by prespecified subgroup analyses and point estimates vary widely, with I2 >80%

g. Downgraded by 1 increment as the majority of the evidence was reported at a time-point <3-month minimum specified in protocol

h. Downgraded by 2 increments as the majority of the evidence was reported at a time-point <3-month minimum specified in protocol and did not have cognitive impairment as an inclusion criterion

Table 4: Clinical evidence summary: General cognitive rehabilitation (multi-component and multi-domain) vs. control, >6 months – 1 year

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
SDMT - 8 months follow up: 8 months	183 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean SDMT - 8 months was 52.0	MD 2.6 higher (0.97 lower to 6.17 higher)
PASAT - 7-8 months - 2 seconds follow up: 8 months	183 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean PASAT - 7-8 months - 2 seconds was 33.4	MD 1.4 higher (2.28 lower to 5.08 higher)
PASAT - 7-8 months - 3 seconds follow up: 7-8 months	243 (2 RCTs)	⨁⨁◯◯ LOW a	-	The mean PASAT - 7-8 months - 3 seconds ranged from 19.46-45.9	MD 2.29 higher (0.77 higher to 3.8 higher)
COWAT - 8 months follow up: 8 months	183 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean COWAT - 8 months was 36.9	MD 2.6 higher (0.88 lower to 6.08 higher)
Stroop test time - 7 months - General 'Stroop test' follow up: 7 months	60 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean stroop test time - 7 months - General 'Stroop test' was 11.96	MD 2.19 lower (2.92 lower to 1.46 lower)
California Verbal Learning Test (CVLT) - 8 months - Total follow up: 8 months	183 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - 8 months - Total was 53.6	MD 2.5 higher (1.24 lower to 6.24 higher)
California Verbal Learning Test (CVLT) - 8 months - Delayed follow up: 8 months	183 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - 8 months - Delayed was 11.6	MD 0.8 higher (0.26 lower to 1.86 higher)
Brief Visuospatial Memory Test (BVMT) - 8 months - Total follow up: 8 months	183 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean brief Visuospatial Memory Test (BVMT) - 8 months - Total was 20.1	MD 1.8 higher (0.18 lower to 3.78 higher)
Brief Visuospatial Memory Test (BVMT) - 8	183 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean brief Visuospatial Memory Test (BVMT) - 8	MD 0.7 higher (0.08 lower to 1.48 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
months - Delayed follow up: 8 months				months - Delayed was 7.5	
Wisconsin Card Sorting Test (time as described benefits in intervention group?) 7 months follow up: 7 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean wisconsin Card Sorting Test (time as described benefits in intervention group?) 7 months was 13.33	MD 2.42 lower (3.5 lower to 1.34 lower)
Information processing speed (unclear how measured) - 7 months follow up: 7 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean information processing speed (unclear how measured) - 7 months was 1122.8	MD 51 lower (89.06 lower to 12.94 lower)
Perceived Deficits Questionnaire - 1 year Scale from: 0 to 80 follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean perceived Deficits Questionnaire - 1 year was 35.2	MD 7.3 lower (13.12 lower to 1.48 lower)
MS Neuropsychological Questionnaire - 1 year - Patient-reported Scale from: 0 to 60 follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MS Neuropsychological Questionnaire - 1 year - Patient-reported was 28.3	MD 6 lower (11 lower to 1 lower)
MS Neuropsychological Questionnaire - 1 year - Informant-reported Scale from: 0 to 60 follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MS Neuropsychological Questionnaire - 1 year - Informant-reported was 19.8	MD 1.2 lower (5.95 lower to 3.55 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
MSIS-29 - 1 year (scale usually 0-100) - Physical Scale from: 0 to 100 follow up: 1 years	78 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean MSIS-29 - 1 year (scale usually 0-100) - Physical was 24.2	MD 1.3 lower (8.03 lower to 5.43 higher)
MSIS-29 - 1 year (scale usually 0-100) - Psychological Scale from: 0 to 100 follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MSIS-29 - 1 year (scale usually 0-100) - Psychological was 22.5	MD 1.1 higher (6.7 lower to 8.9 higher)
WHO-BREF Quality of Life - 1 year (scale used unclear) - S1 Physical health follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean WHO-BREF Quality of Life - 1 year (scale used unclear) - S1 Physical health was 13.7	MD 0.7 higher (0.44 lower to 1.84 higher)
WHO-BREF Quality of Life - 1 year (scale used unclear) - S2 Psychological follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean WHO-BREF Quality of Life - 1 year (scale used unclear) - S2 Psychological was 13.6	MD 0.5 higher (0.69 lower to 1.69 higher)
WHO-BREF Quality of Life - 1 year (scale used unclear) - S3 Social relationship follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean WHO-BREF Quality of Life - 1 year (scale used unclear) - S3 Social relationship was 14.4	MD 0.1 higher (1.33 lower to 1.53 higher)
WHO-BREF Quality of Life - 1 year (scale used unclear) - S4 environment follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean WHO-BREF Quality of Life - 1 year (scale used unclear) - S4 environment was 14.4	MD 0.9 higher (0.17 lower to 1.97 higher)
PROMIS - Applied Cognition Abilities short form 8a - 8 months (scale	183 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PROMIS - Applied Cognition Abilities short form 8a - 8 months (scale 8-40) was 23.0	MD 2.6 higher (0.4 higher to 4.8 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
8-40) Scale from: 8 to 40 follow up: 8 months					
Fatigue - FSMC cognitive subscale - 1 year Scale from: 10 to 50 follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean fatigue - FSMC cognitive subscale - 1 year was 32.2	MD 2.6 lower (6.75 lower to 1.55 higher)
Beck Depression Inventory - 1 year (scale usually 0-63) Scale from: 0 to 63 follow up: 1 years	78 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean beck Depression Inventory - 1 year (scale usually 0-63) was 9.7	MD 1.1 higher (2.26 lower to 4.46 higher)
CES-D depression - 8 months (scale usually 0-60) Scale from: 0 to 60 follow up: 8 months	183 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean CES-D depression - 8 months (scale usually 0-60) was 10.5	MD 0.4 lower (2.15 lower to 1.35 higher)
General self-efficacy scale (scale possibly 17-85) - 8 months Scale from: 17 to 85 follow up: 8 months	183 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean general self-efficacy scale (scale possibly 17-85) - 8 months was 61.1	MD 2.6 higher (0.75 lower to 5.95 higher)
Multi-factorial Memory Questionnaire - Strategy subscale (scale 0-76 and indicates use of memory strategies) - 8 months Scale from: 0 to 76 follow up: 8 months	183 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean multi-factorial Memory Questionnaire - Strategy subscale (scale 0-76 and indicates use of memory strategies) - 8 months was 39.5	MD 0.7 higher (2.52 lower to 3.92 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1-6 months	Risk difference with General cog. rehab - multi-component
Everyday Problems Test - Revised (activities of daily living performance, scale unclear) - 8 months follow up: 8 months	183 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean everyday Problems Test - Revised (activities of daily living performance, scale unclear) - 8 months was 23.5	MD 0.7 higher (0.63 lower to 2.03 higher)

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

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

c. Downgraded by 1 increment as the majority of the evidence came from studies where cognitive impairment was not an inclusion criterion

General cognitive rehabilitation (multi-component and multi-domain) vs. psychoeducation + information-sharing

Table 5: Clinical evidence summary: General cognitive rehabilitation (multi-component and multi-domain) vs. psychoeducation + information-sharing, 3 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation + information sharing, 3 months	Risk difference with General cog. rehab - multi-component
Addenbrooke's cognitive examination - 3 months Scale from: 0 to 100 follow up: 3 months	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean addenbrooke's cognitive examination - 3 months was 86.4	MD 6.9 higher (2.74 higher to 11.06 higher)
Wisconsin Card Sorting Test (WCST) - categories completed - 3 months follow up: 3 months	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean wisconsin Card Sorting Test (WCST) - categories completed - 3 months was 2.4	MD 1.85 higher (0.64 higher to 3.06 higher)
Wisconsin Card Sorting Test (WCST) - errors - 3 months -	30 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean wisconsin Card Sorting Test (WCST) - errors - 3 months -	MD 8.04 lower (10.97 lower to 5.11 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation + information sharing, 3 months	Risk difference with General cog. rehab - multi-component
Perseverative errors follow up: 3 months				Perseverative errors was 12.2	
Wisconsin Card Sorting Test (WCST) - errors - 3 months - Non-perseverative errors follow up: 3 months	30 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean wisconsin Card Sorting Test (WCST) - errors - 3 months - Non-perseverative errors was 19.8	MD 4.72 lower (8.88 lower to 0.56 lower)
Wisconsin Card Sorting Test (WCST) - time - 3 months follow up: 3 months	30 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean wisconsin Card Sorting Test (WCST) - time - 3 months was 340.8	MD 32.7 lower (97.03 lower to 31.63 higher)
Behaviour Rating Inventory of Executive Function-Adult (BRIEF-A) Global Executive Function - 3 months Scale from: 0 to 150 follow up: 3 months	30 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean behaviour Rating Inventory of Executive Function-Adult (BRIEF-A) Global Executive Function - 3 months was 125.99	MD 28.58 lower (38.39 lower to 18.77 lower)
Memory Functioning Questionnaire (MFQ) - General rating (scale used unclear) - 3 months follow up: 3 months	30 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean memory Functioning Questionnaire (MFQ) - General rating (scale used unclear) - 3 months was 44.41	MD 6.87 higher (2.27 higher to 11.47 higher)
Weschler Memory Scale-Revised - 3 months - Visual memory (scale unclear) follow up: 3 months	30 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean Weschler Memory Scale-Revised - 3 months - Visual memory (scale unclear) was 12.0	MD 4.58 higher (2.1 higher to 7.06 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation + information sharing, 3 months	Risk difference with General cog. rehab - multi-component
Weschler Memory Scale-Revised - 3 months - Verbal memory (scale unclear) follow up: 3 months	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean ~Weschler Memory Scale-Revised - 3 months - Verbal memory (scale unclear) was 14.05	MD 5.27 higher (2.23 higher to 8.31 higher)

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

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

General cognitive rehabilitation (multi-component and multi-domain) vs. non-specific cognitive rehabilitation programme

Table 6: Clinical evidence summary: General cognitive rehabilitation (multi-component and multi-domain) vs. non-specific cognitive rehabilitation programme, 4 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 4 months	Risk difference with General cog. rehab - multi-component
SDMT - 4 months follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SDMT - 4 months was 57.2	MD 0.6 higher (5.8 lower to 7 higher)
Stroop test - time - 4 months - Colour naming follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean stroop test - time - 4 months - Colour naming was 66.5	MD 4.9 lower (11.3 lower to 1.5 higher)
Stroop test - time - 4 months - Word reading follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean stroop test - time - 4 months - Word reading was 48.5	MD 1.8 higher (5.75 lower to 9.35 higher)
Stroop test - time - 4 months - Interference follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean stroop test - time - 4 months - Interference was 38.2	MD 6.4 higher (7.88 lower to 20.68 higher)
Trail Making Test - time - 4 months - Part	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean trail Making Test - time -	MD 4.7 higher (2.4 lower to 11.8 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 4 months	Risk difference with General cog. rehab - multi-component
A follow up: 4 months				4 months - Part A was 30.2	
Trail Making Test - time - 4 months - Part B follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean trail Making Test - time - 4 months - Part B was 63.5	MD 6.1 higher (5.74 lower to 17.94 higher)
California Verbal Learning Test (CVLT) - correct answers - 4 months - Learning trials - List A follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a	-	The mean california Verbal Learning Test (CVLT) - correct answers - 4 months - Learning trials - List A was 65.7	MD 2 lower (7.59 lower to 3.59 higher)
California Verbal Learning Test (CVLT) - correct answers - 4 months - Learning trials - List B follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 4 months - Learning trials - List B was 8.3	MD 0.4 lower (1.89 lower to 1.09 higher)
California Verbal Learning Test (CVLT) - correct answers - 4 months - Immediate recall follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 4 months - Immediate recall was 13.5	MD 0.2 higher (1.27 lower to 1.67 higher)
California Verbal Learning Test (CVLT) - correct answers - 4 months - Delayed recall follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 4 months - Delayed recall was 14.1	MD 0.1 higher (1.06 lower to 1.26 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 4 months	Risk difference with General cog. rehab - multi-component
California Verbal Learning Test (CVLT) - correct answers - 4 months - Immediate cued recall follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 4 months - Immediate cued recall was 13.8	MD 0.5 higher (0.66 lower to 1.66 higher)
California Verbal Learning Test (CVLT) - correct answers - 4 months - Delayed cued recall follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 4 months - Delayed cued recall was 14.2	MD 0.2 lower (1.43 lower to 1.03 higher)
California Verbal Learning Test (CVLT) - correct answers - 4 months - Recognition follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 4 months - Recognition was 15.7	MD 0.3 lower (0.84 lower to 0.24 higher)
Alertness - Test of Attentional Performances subtest - reaction time - 4 months - Without warning follow up: 4 months	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean alertness - Test of Attentional Performances subtest - reaction time - 4 months - Without warning was 273.3	MD 23.2 lower (50.96 lower to 4.56 higher)
Alertness - Test of Attentional Performances subtest - reaction time - 4 months - With warning follow up: 4 months	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean alertness - Test of Attentional Performances subtest - reaction time - 4 months - With warning was 261.9	MD 13.7 lower (42.43 lower to 15.03 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 4 months	Risk difference with General cog. rehab - multi-component
Visual Scanning - Test of Attentional Performances subtest - correct answers - 4 months - With a target follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean visual Scanning - Test of Attentional Performances subtest - correct answers - 4 months - With a target was 40.3	MD 3.3 higher (0.93 lower to 7.53 higher)
Visual scanning - Test of Attentional Performances subtest - reaction time - 4 months - Without a target follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean visual scanning - Test of Attentional Performances subtest - reaction time - 4 months - Without a target was 5723.4	MD 736.7 higher (855.94 lower to 2329.34 higher)
Visual scanning - Test of Attentional Performances subtest - reaction time - 4 months - With a target follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,	-	The mean visual scanning - Test of Attentional Performances subtest - reaction time - 4 months - With a target was 3023.5	MD 301.8 higher (402.13 lower to 1005.73 higher)
Divided Attention (visual attention) - Test of Attentional Performances subtest - correct answers - 4 months - Simple task follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (visual attention) - Test of Attentional Performances subtest - correct answers - 4 months - Simple task was 16.2	MD 0.3 lower (1.2 lower to 0.6 higher)
Divided Attention (visual attention) - Test of Attentional Performances	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean divided Attention (visual attention) - Test of Attentional Performances subtest - correct	MD 0.2 higher (0.48 lower to 0.88 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 4 months	Risk difference with General cog. rehab - multi-component
subtest - correct answers - 4 months - Dual task follow up: 4 months				answers - 4 months - Dual task was 16.1	
Divided Attention (visual attention) - Test of Attentional Performances subtest - reaction time - 4 months - Simple task follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (visual attention) - Test of Attentional Performances subtest - reaction time - 4 months - Simple task was 856.9	MD 53.3 lower (126.19 lower to 19.59 higher)
Divided Attention (visual attention) - Test of Attentional Performances subtest - reaction time - 4 months - Dual task follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (visual attention) - Test of Attentional Performances subtest - reaction time - 4 months - Dual task was 800.5	MD 24.3 higher (44.63 lower to 93.23 higher)
Divided Attention (auditory attention) - Test of Attentional Performances subtest - correct answers - 4 months - Dual task follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (auditory attention) - Test of Attentional Performances subtest - correct answers - 4 months - Dual task was 15.3	MD 0.4 higher (0.37 lower to 1.17 higher)
Divided Attention (auditory attention) - Test of Attentional Performances subtest -	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (auditory attention) - Test of Attentional Performances subtest - reaction time - 4 months -	MD 34.7 lower (105.37 lower to 35.97 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 4 months	Risk difference with General cog. rehab - multi-component
reaction time - 4 months - Simple task follow up: 4 months				Simple task was 560.1	
Divided Attention (auditory attention) - Test of Attentional Performances subtest - reaction time - 4 months - Dual task follow up: 4 months	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean divided Attention (auditory attention) - Test of Attentional Performances subtest - reaction time - 4 months - Dual task was 583.6	MD 9.2 lower (75.76 lower to 57.36 higher)
N-back - Test of Attentional Performances subtest - reaction time - 4 months follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean n-back - Test of Attentional Performances subtest - reaction time - 4 months was 753.1	MD 49.8 lower (164.08 lower to 64.48 higher)
N-back - Test of Attentional Performances subtest - correct answers - 4 months follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean n-back - Test of Attentional Performances subtest - correct answers - 4 months was 13.2	MD 0.5 higher (0.39 lower to 1.39 higher)
Baddeley's Dual Task forward span - correct answers - 4 months follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean baddeley's Dual Task forward span - correct answers - 4 months was 5.5	MD 0.3 higher (0.37 lower to 0.97 higher)
Backward span - correct answers - 4 months follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean backward span - correct answers - 4 months was 3.7	MD 0.4 higher (0.26 lower to 1.06 higher)
Fluency - correct answers - 4 months - Semantic	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fluency - correct answers - 4 months - Semantic was 30.8	MD 1.2 lower (6.31 lower to 3.91 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 4 months	Risk difference with General cog. rehab - multi-component
follow up: 4 months					
Fluency - correct answers - 4 months - Phonemic follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fluency - correct answers - 4 months - Phonemic was 21.2	MD 0.6 lower (3.97 lower to 2.77 higher)
Rey complex figure (visuoconstruction and episodic memory) - correct answers - 4 months follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean rey complex figure (visuoconstruction and episodic memory) - correct answers - 4 months was 33.9	MD 0.6 higher (0.58 lower to 1.78 higher)
Rey complex figure (visuoconstruction and episodic memory) - time - 4 months follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean rey complex figure (visuoconstruction and episodic memory) - time - 4 months was 162.7	MD 29.5 higher (17.03 lower to 76.03 higher)
DO80 naming task - correct answers - 4 months follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean DO80 naming task - correct answers - 4 months was 78.1	MD 0.1 higher (1.04 lower to 1.24 higher)
Daily Cognitive Activities Questionnaire (scale 0-60) - 4 months Scale from: 0 to 60 follow up: 4 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean daily Cognitive Activities Questionnaire (scale 0-60) - 4 months was 49.2	MD 6.7 higher (3.64 lower to 17.04 higher)
Beck Depression Inventory (scale usually 0-63) - 4 months Scale from: 0 to 63 follow up: 4 years	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean beck Depression Inventory (scale usually 0-63) - 4 months was 9.5	MD 1 higher (3.64 lower to 5.64 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 4 months	Risk difference with General cog. rehab - multi-component
State-Trait Anxiety Inventory (STAI; scale usually 20-80) - 4 months - STAI-A (state?) Scale from: 20 to 80 follow up: 4 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean state-Trait Anxiety Inventory (STAI; scale usually 20-80) - 4 months - STAI-A (state?) was 32.2	MD 4.7 higher (3.74 lower to 13.14 higher)
State-Trait Anxiety Inventory (STAI; scale usually 20-80) - 4 months - STAI-B (trait?) Scale from: 20 to 80 follow up: 4 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean state-Trait Anxiety Inventory (STAI; scale usually 20-80) - 4 months - STAI-B (trait?) was 39.4	MD 3.1 higher (4 lower to 10.2 higher)
Modified Fatigue Impact Scale - Cognitive (scale usually 0-40) - 4 months Scale from: 0 to 40 follow up: 4 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - Cognitive (scale usually 0-40) - 4 months was 17.5	MD 0.3 lower (6.26 lower to 5.66 higher)
SF-36 quality of life (scale usually 0-100) - 4 months - Physical Scale from: 0 to 100 follow up: 4 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean SF-36 quality of life (scale usually 0-100) - 4 months - Physical was 55.8	MD 2.3 higher (10.13 lower to 14.73 higher)
SF-36 quality of life (scale usually 0-100) - 4 months - Psychological Scale from: 0 to 100 follow up: 4 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean SF-36 quality of life (scale usually 0-100) - 4 months - Psychological was 57.8	MD 2.1 higher (10.51 lower to 14.71 higher)

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

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

Table 7: Clinical evidence summary: General cognitive rehabilitation (multi-component and multi-domain) vs. non-specific cognitive rehabilitation programme, 8 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 8 months	Risk difference with General cog. rehab - multi-component
SDMT - 8 months follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean SDMT - 8 months was 59.4	MD 0.7 lower (7.43 lower to 6.03 higher)
Stroop test - time - 8 months - Colour naming follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean stroop test - time - 8 months - Colour naming was 64.1	MD 3.3 lower (10.14 lower to 3.54 higher)
Stroop test - time - 8 months - Word reading follow up: 8 months	35 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean stroop test - time - 8 months - Word reading was 48.4	MD 1.6 lower (6.83 lower to 3.63 higher)
Stroop test - time - 8 months - Interference follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean stroop test - time - 8 months - Interference was 40.2	MD 1.8 lower (12.03 lower to 8.43 higher)
Trail Making Test - time - 8 months - Part A follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean trail Making Test - time - 8 months - Part A was 28.3	MD 2.7 higher (3.63 lower to 9.03 higher)
Trail Making Test - time - 8 months - Part B follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean trail Making Test - time - 8 months - Part B was 57.2	MD 9.9 higher (4.22 lower to 24.02 higher)
California Verbal Learning Test (CVLT) - correct answers - 8 months - Learning trials - List A follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 8 months - Learning trials - List A was 66.1	MD 1.7 higher (3.04 lower to 6.44 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 8 months	Risk difference with General cog. rehab - multi-component
California Verbal Learning Test (CVLT) - correct answers - 8 months - Learning trials - List B follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 8 months - Learning trials - List B was 8.0	MD 0 (1.77 lower to 1.77 higher)
California Verbal Learning Test (CVLT) - correct answers - 8 months - Immediate recall follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 8 months - Immediate recall was 14.1	MD 0.6 lower (2.06 lower to 0.86 higher)
California Verbal Learning Test (CVLT) - correct answers - 8 months - Delayed recall follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 8 months - Delayed recall was 14.4	MD 0 (1.26 lower to 1.26 higher)
California Verbal Learning Test (CVLT) - correct answers - 8 months - Immediate cued recall follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 8 months - Immediate cued recall was 14.2	MD 0.4 higher (0.86 lower to 1.66 higher)
California Verbal Learning Test (CVLT) - correct answers - 8 months - Delayed cued recall follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - correct answers - 8 months - Delayed cued recall was 14.6	MD 0.2 higher (0.83 lower to 1.23 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 8 months	Risk difference with General cog. rehab - multi-component
California Verbal Learning Test (CVLT) - correct answers - 8 months - Recognition follow up: 8 months	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean california Verbal Learning Test (CVLT) - correct answers - 8 months - Recognition was 15.8	MD 0.1 lower (0.48 lower to 0.28 higher)
Alertness - Test of Attentional Performances subtest - reaction time - 8 months - Without warning follow up: 8 months	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean alertness - Test of Attentional Performances subtest - reaction time - 8 months - Without warning was 267.2	MD 16.1 lower (41.72 lower to 9.52 higher)
Alertness - Test of Attentional Performances subtest - reaction time - 8 months - With warning follow up: 8 months	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean alertness - Test of Attentional Performances subtest - reaction time - 8 months - With warning was 250.0	MD 12.5 higher (25.19 lower to 50.19 higher)
Visual scanning - Test of Attentional Performances subtest - correct answers - 8 months - Without target follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean visual scanning - Test of Attentional Performances subtest - correct answers - 8 months - Without target was 49.8	MD 0 (0.3 lower to 0.3 higher)
Visual scanning - Test of Attentional Performances subtest - correct answers - 8 months - With target follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean visual scanning - Test of Attentional Performances subtest - correct answers - 8 months - With target was 43.1	MD 2.3 lower (5.77 lower to 1.17 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 8 months	Risk difference with General cog. rehab - multi-component
Visual scanning - Test of Attentional Performances subtest - reaction time - 8 months - Without target follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean visual scanning - Test of Attentional Performances subtest - reaction time - 8 months - Without target was 5456.7	MD 139.9 higher (1016.11 lower to 1295.91 higher)
Visual scanning - Test of Attentional Performances subtest - reaction time - 8 months - With target follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean visual scanning - Test of Attentional Performances subtest - reaction time - 8 months - With target was 2789.2	MD 71 higher (443.59 lower to 585.59 higher)
Divided Attention (visual attention) - Test of Attentional Performances subtest - correct answers - 8 months - Simple task follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (visual attention) - Test of Attentional Performances subtest - correct answers - 8 months - Simple task was 15.8	MD 0.1 lower (1.31 lower to 1.11 higher)
Divided Attention (visual attention) - Test of Attentional Performances subtest - correct answers - 8 months - Dual task follow up: 8 months	35 (1 RCT)	⊕⊕◯◯ LOW	-	The mean divided Attention (visual attention) - Test of Attentional Performances subtest - correct answers - 8 months - Dual task was 15.8	MD 0.2 higher (0.5 lower to 0.9 higher)
Divided Attention (visual attention) - Test of Attentional	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (visual attention) - Test of Attentional Performances subtest - reaction	MD 19.9 lower (99.92 lower to 60.12 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 8 months	Risk difference with General cog. rehab - multi-component
Performances subtest - reaction time - 8 months - Simple task follow up: 8 months				time - 8 months - Simple task was 829.2	
Divided Attention (visual attention) - Test of Attentional Performances subtest - reaction time - 8 months - Dual task follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (visual attention) - Test of Attentional Performances subtest - reaction time - 8 months - Dual task was 811.2	MD 28.3 lower (94.19 lower to 37.59 higher)
Divided Attention (auditory attention) - Test of Attentional Performances subtest - correct answers - 8 months - Simple task follow up: 8 months	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean divided Attention (auditory attention) - Test of Attentional Performances subtest - correct answers - 8 months - Simple task was 15.8	MD 0 (0.27 lower to 0.27 higher)
Divided Attention (auditory attention) - Test of Attentional Performances subtest - correct answers - 8 months - Dual task follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean divided Attention (auditory attention) - Test of Attentional Performances subtest - correct answers - 8 months - Dual task was 15.3	MD 0.4 higher (0.49 lower to 1.29 higher)
Divided Attention (auditory attention) - Test of Attentional Performances	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean divided Attention (auditory attention) - Test of Attentional Performances subtest - reaction time - 8 months -	MD 3.1 lower (69.26 lower to 63.06 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 8 months	Risk difference with General cog. rehab - multi-component
subtest - reaction time - 8 months - Simple task follow up: 8 months				Simple task was 561.5	
Divided Attention (auditory attention) - Test of Attentional Performances subtest - reaction time - 8 months - Dual task follow up: 8 months	35 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean divided Attention (auditory attention) - Test of Attentional Performances subtest - reaction time - 8 months - Dual task was 588.4	MD 8.5 higher (60.38 lower to 77.38 higher)
N-back - Test of Attentional Performances subtest - reaction time - 8 months follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean n-back - Test of Attentional Performances subtest - reaction time - 8 months was 698.5	MD 50.2 lower (162.9 lower to 62.5 higher)
N-back - Test of Attentional Performances subtest - correct answers - 8 months follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean n-back - Test of Attentional Performances subtest - correct answers - 8 months was 13.6	MD 0.1 higher (0.83 lower to 1.03 higher)
Baddeley's Dual Task forward span - correct answers - 8 months follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean baddeley's Dual Task forward span - correct answers - 8 months was 5.4	MD 0.3 higher (0.4 lower to 1 higher)
Backward span - correct answers - 8 months follow up: 8 months	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean backward span - correct answers - 8 months was 4.2	MD 0.5 higher (0.16 lower to 1.16 higher)
Fluency - correct answers - 8 months -	35 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean fluency - correct answers - 8 months - Semantic was 5.5	MD 0.3 higher (0.37 lower to 0.97 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specific cognitive rehab programme, 8 months	Risk difference with General cog. rehab - multi-component
Semantic follow up: 8 months					
Fluency - correct answers - 8 months - Phonemic follow up: 8 months	35 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean fluency - correct answers - 8 months - Phonemic was 5.5	MD 0.3 higher (0.37 lower to 0.97 higher)
Rey complex figure (visuoconstruction and episodic memory) - correct answers - 8 months follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean rey complex figure (visuoconstruction and episodic memory) - correct answers - 8 months was 33.7	MD 1 higher (0.16 higher to 1.84 higher)
Rey complex figure (visuoconstruction and episodic memory) - time - 8 months follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean rey complex figure (visuoconstruction and episodic memory) - time - 8 months was 158.9	MD 14.1 higher (27.63 lower to 55.83 higher)
DO80 naming task - correct answers - 8 months follow up: 8 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean DO80 naming task - correct answers - 8 months was 78.7	MD 0.3 higher (0.56 lower to 1.16 higher)

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

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

General cognitive rehabilitation (multi-component and multi-domain) tailored to individual + outpatient rehabilitation vs. outpatient rehabilitation only

Table 8: Clinical evidence summary: General cognitive rehabilitation (multi-component and multi-domain) tailored to individual + outpatient rehabilitation vs. outpatient rehabilitation only, 3 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (outpatient rehabilitation only), 3 months	Risk difference with General cog. rehab + outpatient rehabilitation - multi-component and tailored to individual
Computer-aided card sorting - correct - 3 months follow up: 3 months	19 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean computer-aided card sorting - correct - 3 months was 53.9	MD 11.8 lower (27.87 lower to 4.27 higher)
Computer-aided card sorting - incorrect - 3 months follow up: 3 months	19 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean computer-aided card sorting - incorrect - 3 months was 16.8	MD 2.7 lower (5.62 lower to 0.22 higher)
Sustained attention - correct - 3 months follow up: 3 months	19 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean sustained attention - correct - 3 months was 53.9	MD 11.8 lower (27.87 lower to 4.27 higher)
Sustained attention - incorrect - 3 months follow up: 3 months	19 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean sustained attention - incorrect - 3 months was 51.2	MD 5 lower (18.62 lower to 8.62 higher)
Sustained attention - reaction time - 3 months follow up: 3 months	19 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean sustained attention - reaction time - 3 months was 46.8	MD 4.1 lower (11.86 lower to 3.66 higher)
Sustained attention - variation reaction time - 3 months follow up: 3 months	19 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean sustained attention - variation reaction time - 3 months was 50.7	MD 5.9 lower (14.73 lower to 2.93 higher)
Verbal learning test - 3 months follow up: 3 months	19 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean verbal learning test - 3 months was 50.4	MD 6.5 higher (5.54 lower to 18.54 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (outpatient rehabilitation only), 3 months	Risk difference with General cog. rehab + outpatient rehabilitation - multi-component and tailored to individual
Spatial construction - 3 months follow up: 3 months	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean spatial construction - 3 months was 10.4	MD 0.2 higher (2.06 lower to 2.46 higher)
Non-verbal learning test - 3 months follow up: 3 months	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean non-verbal learning test - 3 months was 48.3	MD 0.7 higher (11.5 lower to 12.9 higher)
Beck Depression Inventory (scale usually 0-63) - 3 months Scale from: 0 to 63 follow up: 3 months	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean beck Depression Inventory (scale usually 0-63) - 3 months was 8.3	MD 0 (4.23 lower to 4.23 higher)
Modified Fatigue Impact Scale (scale usually 0-84) - 3 months Scale from: 0 to 84 follow up: 3 months	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean modified Fatigue Impact Scale (scale usually 0-84) - 3 months was 31.7	MD 10.1 higher (5.49 lower to 25.69 higher)

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

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

c. General Modified Fatigue Impact Scale rather than specifically the cognitive subdomain

Goal Attainment Scaling (GAS) goals (multi-component cognitive rehabilitation tailored to individual) + usual rehabilitation vs. usual rehabilitation only

Table 9: Clinical evidence summary: Goal Attainment Scaling (GAS) goals (multi-component cognitive rehabilitation tailored to individual) + usual rehabilitation vs. usual rehabilitation only, 4 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual rehab alone), 4 months	Risk difference with General cog rehab - Goal Attainment Scaling (GAS) goals (multi-component cognitive rehab and tailored to individual) + usual rehab
Behaviour Rating Inventory of Executive Function – Adult (BRIEF-A) - General Executive Composite (T-score) - 4 months Scale from: 0 to 100 follow up: 4 months	102 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean behaviour Rating Inventory of Executive Function – Adult (BRIEF-A) - General Executive Composite (T-score) - 4 months was 56.7	MD 0.3 lower (4.84 lower to 4.24 higher)
BRIEF-A - Metacognition index (T-score) - 4 months Scale from: 0 to 100 follow up: 4 months	102 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean BRIEF-A - Metacognition index (T-score) - 4 months was 57.8	MD 0.4 higher (3.97 lower to 4.77 higher)
MSIS-29 psychological subscale (Norwegian version, scale reported 9-45) - 4 months Scale from: 9 to 45 follow up: 4 months	102 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 psychological subscale (Norweigian version, scale reported 9-45) - 4 months was 19.9	MD 1.6 lower (4.44 lower to 1.24 higher)
Hopkins Symptom Checklist- 25 (measures psychological health; scale 1-4) - 4 months	102 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean hopkins Symptom Checklist-25 (measures psychological health; scale 1-4) - 4 months was 1.74	MD 0.14 lower (0.33 lower to 0.05 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual rehab alone), 4 months	Risk difference with General cog rehab - Goal Attainment Scaling (GAS) goals (multi-component cognitive rehab and tailored to individual) + usual rehab
Scale from: 1 to 4 follow up: 4 months					

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

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

Table 10: Clinical evidence summary: Goal Attainment Scaling (GAS) goals (multi-component cognitive rehabilitation tailored to individual) + usual rehabilitation vs. usual rehabilitation only, 7 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual rehab alone), 7 months	Risk difference with General cog rehab - Goal Attainment Scaling (GAS) goals (multi-component cognitive rehab and tailored to individual) + usual rehab
Behaviour Rating Inventory of Executive Function – Adult (BRIEF-A) - General Executive Composite (T-score) - 7 months Scale from: 0 to 100 follow up: 7 months	102 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean behaviour Rating Inventory of Executive Function – Adult (BRIEF-A) - General Executive Composite (T-score) - 7 months was 55.2	MD 1.1 higher (3.43 lower to 5.63 higher)
BRIEF-A - Metacognition index (T-score) - 7 months Scale from: 0	102 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean BRIEF-A - Metacognition index (T-score) - 7 months was 56.7	MD 1 higher (3.43 lower to 5.43 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual rehab alone), 7 months	Risk difference with General cog rehab - Goal Attainment Scaling (GAS) goals (multi-component cognitive rehab and tailored to individual) + usual rehab
to 100 follow up: 7 months					
MSIS-29 psychological subscale (Norwegian version, scale reported 9-45) - 7 months Scale from: 9 to 45 follow up: 7 months	102 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean MSIS-29 psychological subscale (Norweigian version, scale reported 9-45) - 7 months was 20.6	MD 2.3 lower (5.27 lower to 0.67 higher)
Hopkins Symptom Checklist- 25 (measures psychological health; scale 1-4) - 7 months Scale from: 1 to 4 follow up: 7 months	102 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean hopkins Symptom Checklist-25 (measures psychological health; scale 1-4) - 7 months was 1.65	MD 0.03 lower (0.23 lower to 0.17 higher)

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

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

Multi-domain cognitive rehabilitation (pen/paper or computer tasks with no additional teaching strategies) vs. control

See also summary of evidence from an additional paper (Mattioli 2010/2012) comparing computer tasks to control that reported results only as medians.

Table 11: Clinical evidence summary: Multi-domain cognitive rehabilitation (pen/paper tasks or computer tasks with no additional teaching strategies) vs control, 2-6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
SDMT - 2-6 months (mix final values and change from baseline) - Similar at baseline or change from baseline reported follow up: 2-6 months	189 (4 RCTs)	⨁◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 2-6 months (mix final values and change from baseline) - Similar at baseline or change from baseline reported ranged from -0.19 to -1.70 for change scores and 32.92 to 37.43 for final values	MD 5.57 higher (3.69 higher to 7.45 higher)
SDMT - 3 months - Larger difference at baseline (lower in intervention) follow up: 3 months	82 (2 RCTs)	⨁◯◯◯ VERY LOW a,d	-	The mean SDMT - 3 months - Larger difference at baseline (lower in intervention) ranged from 34.8-47.93	MD 1.57 lower (7 lower to 3.86 higher)
PASAT - 2 seconds - 3 months follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean PASAT - 2 seconds - 3 months was 4.90	MD 12.8 higher (1.83 higher to 23.77 higher)
PASAT - 3 seconds - 2.5-6 months (mix of final values and change from baseline) follow up: 2.5-6 months	177 (4 RCTs)	⨁◯◯◯ VERY LOW a,d,e	-	The mean PASAT - 3 seconds - 2.5-6 months (mix of final values and change from baseline) ranged from 0.35-0.35 for change scores and 9.7-36.54 for final values	MD 4.76 higher (0.53 lower to 10.05 higher)
Contralled Oral Word Association Test (COWAT) - 3 months - Phonemic cues follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean contralled Oral Word Association Test (COWAT) - 3 months - Phonemic cues was 30.0	MD 4.4 higher (5.42 lower to 14.22 higher)
Contralled Oral Word Association Test (COWAT) - 3 months -	20 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean contralled Oral Word Association Test (COWAT) - 3 months	MD 2.6 higher (6.55 lower to 11.75 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
Semantic cues follow up: 3 months				- Semantic cues was 35.0	
Contralled Oral Word Association Test (COWAT) - 3 months - Animals follow up: 3 months	62 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean contralled Oral Word Association Test (COWAT) - 3 months - Animals was 19.63	MD 0.4 lower (2.89 lower to 2.09 higher)
Wisconsin Card Sorting Test (WCST) - 3 months - Total errors follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean wisconsin Card Sorting Test (WCST) - 3 months - Total errors was 41.3	MD 13.3 lower (28.07 lower to 1.47 higher)
Wisconsin Card Sorting Test (WCST) - 3 months - Perseverative errors follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean wisconsin Card Sorting Test (WCST) - 3 months - Perseverative errors was 39.8	MD 14.3 lower (32.66 lower to 4.06 higher)
Wisconsin Card Sorting Test (WCST) - 3 months - Perseverative responses follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean wisconsin Card Sorting Test (WCST) - 3 months - Perseverative responses was 29.0	MD 10.9 lower (23.62 lower to 1.82 higher)
Delis-Kaplan Executive Function System (D-KEFS) - card sorting test - 2.5 months - Verbal follow up: 2.5 months	54 (1 RCT)	⨁◯◯◯ VERY LOW a,c,d	-	The mean delis-Kaplan Executive Function System (D-KEFS) - card sorting test - 2.5 months - Verbal was 24.85	MD 4.61 higher (1.14 lower to 10.36 higher)
Delis-Kaplan Executive Function System (D-KEFS) - card sorting test -	54 (1 RCT)	⨁◯◯◯ VERY LOW a,c,d	-	The mean delis-Kaplan Executive Function System (D-KEFS) - card sorting test - 2.5 months - Non-verbal was 6.46	MD 1.39 higher (0.03 lower to 2.81 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
2.5 months - Non-verbal follow up: 2.5 months					
Word List Generation Test - 6 months (change from baseline) - Word List Generation Test - 6 months (change from baseline) follow up: 6 months	41 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean word List Generation Test - 6 months (change from baseline) - Word List Generation Test - 6 months (change from baseline) was 1.13	MD 3.6 higher (0.83 higher to 6.37 higher)
10/36 SPART (Spatial Recall Test) - 3-6 months (mix final values and change from baseline) - Total follow up: 3-6 months	103 (2 RCTs)	⨁◯◯◯ VERY LOW a,d,f	-	The mean 10/36 SPART (Spatial Recall Test) - 3-6 months (mix final values and change from baseline) - Total was -1.20 for change scores and 21.38 for final values	MD 3.46 higher (0.69 lower to 7.6 higher)
10/36 SPART (Spatial Recall Test) - 3-6 months (mix final values and change from baseline) - Long-term retrieval follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean 10/36 SPART (Spatial Recall Test) - 3-6 months (mix final values and change from baseline) - Long-term retrieval was 16.3	MD 0.3 lower (4.43 lower to 3.83 higher)
10/36 SPART (Spatial Recall Test) - 3-6 months (mix final values and change from baseline) - Delayed recall follow up: 3-6 months	123 (3 RCTs)	⨁◯◯◯ VERY LOW a,d,e	-	The mean 10/36 SPART (Spatial Recall Test) - 3-6 months (mix final values and change from baseline) - Delayed recall ranged from -0.23 to -0.23 for change scores and 5.7 to 7.63 for final values	MD 0.67 higher (0.9 lower to 2.23 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
Selective Reminding Test (SRT) 2.5-6 months (change from baseline) - Total follow up: 6 months	41 (1 RCT)	⊕◯◯◯ VERY LOW a	-	The mean selective Reminding Test (SRT) 2.5-6 months (change from baseline) - Total was 1.19	MD 1.63 higher (2.76 lower to 6.02 higher)
Selective Reminding Test (SRT) 2.5-6 months (mix final values and change from baseline) - Long-term storage follow up: 2.5-6 months	159 (4 RCTs)	⊕⊕◯◯ LOW a	-	The mean selective Reminding Test (SRT) 2.5-6 months (mix final values and change from baseline) - Long-term storage ranged from -0.05 to -0.05 for change scores and 25.2-36.38 for final values	MD 6.18 higher (3.36 higher to 8.99 higher)
Selective Reminding Test (SRT) 2.5-6 months (mix final values and change from baseline) - Delayed retrieval follow up: 2.5-6 months	159 (4 RCTs)	⊕◯◯◯ VERY LOW a,d	-	The mean selective Reminding Test (SRT) 2.5-6 months (mix final values and change from baseline) - Delayed retrieval ranged from 0.20 for change scores and 5.7-7.12 for final values	MD 1.15 higher (0.6 higher to 1.7 higher)
Selective Reminding Test (SRT) 2.5-6 months (mix final values and change from baseline) - Consistent long-term retrieval follow up: 3-6 months	123 (3 RCTs)	⊕◯◯◯ VERY LOW a,d	-	The mean selective Reminding Test (SRT) 2.5-6 months (mix final values and change from baseline) - Consistent long-term retrieval ranged from 0.23-0.23 for change scores and 16.3-24.53 for final values	MD 5.11 higher (0.49 lower to 10.7 higher)
Brief Visuospatial Memory Test-Revised	94 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean brief Visuospatial Memory Test-Revised (BVMT-R) - 2-2.5	MD 3.52 higher (2.26 higher to 4.78 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
(BVMT-R) - 2-2.5 months (mix of final values and change from baseline) follow up: 2-2.5 months				months (mix of final values and change from baseline) was 0.29 for change scores and 20.8 for final values	
Trail Making Test - 2.5-6 months (mix of final values and change from baseline) - Part A follow up: 2.5-6 months	77 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean trail Making Test - 2.5-6 months (mix of final values and change from baseline) - Part A was 0.01 for change scores and 68.88 for final values	MD 11.59 lower (18.85 lower to 4.33 lower)
Trail Making Test - 2.5-6 months (change from baseline) - Part B, similar at baseline or change from baseline reported follow up: 6 months	41 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean trail Making Test - 2.5-6 months (change from baseline) - Part B, similar at baseline or change from baseline reported was -0.83	MD 13.97 lower (34.4 lower to 6.46 higher)
Trail Making Test - 2.5-6 months - Part B, larger difference at baseline (higher in intervention) follow up: 2.5 months	36 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean trail Making Test - 2.5-6 months - Part B, larger difference at baseline (lower in intervention) was 110.96	MD 2.32 higher (20.39 lower to 25.03 higher)
Stroop neuropsychological screening test - 2.5 months follow up: 2.5 months	58 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean stroop neuropsychological screening test - 2.5 months was 57.6	MD 5.9 higher (1.23 lower to 13.03 higher)
Test of Everyday Attention (TEA) median - 3	20 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean test of Everyday Attention (TEA) median - 3	MD 137.5 higher (8.81 higher to 266.19 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
months - Auditory stimulus follow up: 3 months				months - Auditory stimulus was 612.8	
Test of Everyday Attention (TEA) median - 3 months - Visual stimulus follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean test of Everyday Attention (TEA) median - 3 months - Visual stimulus was 1048.7	MD 89.6 lower (234.87 lower to 55.67 higher)
Test of Everyday Attention (TEA) errors/omissions - 3 months - Total omitted stimuli follow up: 3 months	20 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean test of Everyday Attention (TEA) errors/omissions - 3 months - Total omitted stimuli was 4.6	MD 0.1 lower (2.27 lower to 2.07 higher)
Test of Everyday Attention (TEA) errors/omissions - 3 months - Total errors follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean test of Everyday Attention (TEA) errors/omissions - 3 months - Total errors was 6.1	MD 1.3 lower (5.93 lower to 3.33 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Visual attention follow up: 2.5 months	54 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Visual attention was 76.73	MD 11.58 higher (0.61 lower to 23.77 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory attention follow up: 2.5 months	54 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory attention was 71.27	MD 13.31 higher (1.71 higher to 24.91 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months	54 (1 RCT)	⊕◯◯◯ VERY LOW a,c.d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months	MD 11.61 higher (1.37 higher to 21.85 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
- Visual response control follow up: 2.5 months				- Visual response control was 86.58	
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory response control follow up: 2.5 months	54 (1 RCT)	⨁◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory response control was 71.19	MD 15.73 higher (5.68 higher to 25.78 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Visual comprehension follow up: 2.5 months	54 (1 RCT)	⨁◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Visual comprehension was 80.81	MD 12.96 higher (0.63 higher to 25.29 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory comprehension follow up: 2.5 months	54 (1 RCT)	⨁◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory comprehension was 75.69	MD 11.58 higher (1.19 lower to 24.35 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Visual persistence attention follow up: 2.5 months	54 (1 RCT)	⨁◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Visual persistence attention was 88.19	MD 11.58 higher (0.15 higher to 23.01 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory persistence attention follow up: 2.5 months	54 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory persistence attention was 88.62	MD 14.26 higher (7.55 higher to 20.97 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Visual sensory-motor attention follow up: 2.5 months	54 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Visual sensory-motor attention was 83.77	MD 12.31 higher (2.8 higher to 21.82 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory sensory-motor attention follow up: 2.5 months	54 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Auditory sensory-motor attention was 97.88	MD 9.7 higher (1.34 lower to 20.74 higher)
Integrated Auditory Visual-2 (IVA-2) - 2.5 months - Fine motor hyperactivity follow up: 2.5 months	54 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean integrated Auditory Visual-2 (IVA-2) - 2.5 months - Fine motor hyperactivity was 65.46	MD 12.16 higher (3.6 lower to 27.92 higher)
Digit span (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) - Forward follow up: 6 months	41 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean digit span (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) - Forward was 0.51	MD 0.43 higher (0.34 lower to 1.2 higher)
Digit span (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) - Backward follow up: 6 months	41 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean digit span (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) - Backward was -0.03	MD 0.92 higher (0.2 lower to 2.04 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
Block design (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) - Block design (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) follow up: 6 months	41 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean block design (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) - Block design (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) was 3.89	MD 4.35 higher (1.01 lower to 9.71 higher)
Letter-number sequencing (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) - Letter-number sequencing (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) follow up: 6 months	41 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean letter-number sequencing (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) - Letter-number sequencing (Weschler Adult Intelligence Scale III) - 6 months (change from baseline) was 0.15	MD 1.48 higher (0.06 higher to 2.9 higher)
Judgement of line orientation - 2.5 months follow up: 2.5 months	54 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean judgement of line orientation - 2.5 months was 18.77	MD 1.92 higher (0.24 higher to 3.6 higher)
Boston Naming Test - 6 months (change from baseline) - Boston Naming Test - 6 months	41 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean boston Naming Test - 6 months (change from baseline) - Boston Naming Test - 6 months (change from baseline) was 0.59	MD 2.58 higher (1.16 higher to 4 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
(change from baseline) follow up: 6 months					
FAS test (verbal fluency) - 3-6 months (change from baseline) - Similar at baseline or change from baseline values reported follow up: 6 months	41 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean FAS test (verbal fluency) - 3-6 months (change from baseline) - Similar at baseline or change from baseline values reported was 5.54	MD 1.55 higher (3.48 lower to 6.58 higher)
FAS test (verbal fluency) - 3-6 months - Larger difference at baseline (lower in intervention group) follow up: 3 months	62 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean FAS test (verbal fluency) - 3-6 months - Larger difference at baseline (lower in intervention group) was 33.13	MD 0.9 lower (6.1 lower to 4.3 higher)
Verbal fluency test - 2.5 months - Phonemic follow up: 2.5 months	58 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean verbal fluency test - 2.5 months - Phonemic was 29.95	MD 3.18 higher (0.7 lower to 7.06 higher)
Verbal fluency test - 2.5 months - Semantic follow up: 2.5 months	58 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean verbal fluency test - 2.5 months - Semantic was 39.58	MD 3.98 higher (0.78 lower to 8.74 higher)
Greek Verbal Learning Test - 2 months (change from baseline) follow up: 2 months	36 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean Greek Verbal Learning Test - 2 months (change from baseline) was -0.94	MD 9.04 higher (6.15 higher to 11.93 higher)
MS Neuropsychological Questionnaire	62 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean MS Neuropsychological Questionnaire (MNSQ, scale	MD 1.76 lower (7.65 lower to 4.13 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
(MNSQ, scale usually 0-60) - 3 months Scale from: 0 to 60 follow up: 3 months				usually 0-60) - 3 months was 25.63	
MSQoL-54 (scale usually 0-100) - 2.5 months - Physical composite Scale from: 0 to 100 follow up: 3 months	62 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean mSQoL-54 (scale usually 0-100) - 2.5 months - Physical composite was 63.24	MD 10.25 lower (19.3 lower to 1.2 lower)
MSQoL-54 (scale usually 0-100) - 3 months - Mental composite Scale from: 0 to 100 follow up: 3 months	62 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean mSQoL-54 (scale usually 0-100) - 3 months - Mental composite was 67.32	MD 10.93 lower (19.86 lower to 2 lower)
MS quality of life (scale unclear) - 3 months follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,d	-	The mean MS quality of life (scale unclear) - 3 months was 157.56	MD 30.88 higher (1.83 lower to 63.59 higher)
EQ-5D visual analogue (scale usually 0-100) - 2 months (change from baseline) Scale from: 0 to 100 follow up: 2 months	36 (1 RCT)	⨁◯◯◯ VERY LOW a,c,d	-	The mean EQ-5D visual analogue (scale usually 0-100) - 2 months (change from baseline) was -1.17	MD 10.59 higher (6.38 higher to 14.8 higher)
Beck Depression Inventory-Fast Screen (scale usually 0-21) - 2 months	36 (1 RCT)	⨁◯◯◯ VERY LOW a,c,d	-	The mean beck Depression Inventory-Fast Screen (scale usually 0-21) - 2 months	MD 2.86 lower (4.57 lower to 1.15 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
(change from baseline) Scale from: 0 to 21 follow up: 2 months				(change from baseline) was 0.29	
Montgomery-Asberg Depression Scale (scale usually 0-60) - 3 months Scale from: 0 to 60 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean montgomery-Asberg Depression Scale (scale usually 0-60) - 3 months was 14.7	MD 8.8 lower (15.35 lower to 2.25 lower)
HADS - 3-6 months (scale usually 0-21, mix of final values and change from baseline) - Anxiety Scale from: 0 to 21 follow up: 3-6 months	103 (2 RCTs)	⊕◯◯◯ VERY LOW a,d	-	The mean HADS - 3-6 months (scale usually 0-21, mix of final values and change from baseline) - Anxiety was -0.53 for change scores and 7.41 for final values	MD 1.63 lower (2.9 lower to 0.36 lower)
HADS - 3-6 months (scale usually 0-21, mix of final values and change from baseline) - Depression Scale from: 0 to 21 follow up: 3-6 months	103 (2 RCTs)	⊕◯◯◯ VERY LOW a,d	-	The mean HADS - 3-6 months (scale usually 0-21, mix of final values and change from baseline) - Depression was 0.58 for change scores and 6.13 for final values	MD 1.08 lower (2.33 lower to 0.16 higher)
Modified Fatigue Impact Scale (MFIS) - cognitive (scale usually 0-40) - 2 months (change from baseline) Scale from: 0 to 40	36 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact Scale (MFIS) - cognitive (scale usually 0-40) - 2 months (change from baseline) was -0.88	MD 4.8 lower (6.52 lower to 3.08 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-6 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
follow up: 2 months					
Fatigue Severity Scale (FSS, scale usually 9-63) - 3 months Scale from: 9 to 63 follow up: 3 months	62 (1 RCT)	⨁◯◯◯ VERY LOW a,d,g	-	The mean fatigue Severity Scale (FSS, scale usually 9-63) - 3 months was 29.21	MD 1.3 higher (9.19 lower to 11.79 higher)

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

b. Downgraded by 1 increment as point estimates vary in size of effect, which cannot be explained by prespecified subgrouping analyses

c. Downgraded by 1 increment as the majority of the evidence came from studies reporting the results at a time-point <3-month minimum specified in the protocol

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

e. Downgraded by 1 increment as statistical heterogeneity is present that cannot be explained by prespecified subgroup analyses, with I2 >50%

f. Downgraded by 2 increments as statistical heterogeneity is present, with I2 ≥80% and point estimates differing in size of the effect and that could not be explained by prespecified subgroup analyses

g. Downgraded by 1 increment as general Fatigue Severity Scale reported rather than specifically cognitive fatigue

Table 12: Clinical evidence summary: Multi-domain cognitive rehabilitation (pen/paper tasks or computer tasks with no additional teaching strategies) vs control, 9 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-9 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
PASAT - 9 months - 2 seconds follow up: 9 months	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean PASAT - 9 months - 2 seconds was 6.8	MD 11.2 higher (0.01 lower to 22.41 higher)
PASAT - 9 months - 3 seconds	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean PASAT - 9 months - 3 seconds was 15.2	MD 14.3 higher (1.06 lower to 29.66 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-9 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
follow up: 9 months					
SDMT - 9 months follow up: 9 months	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean SDMT - 9 months was 34.7	MD 0.3 higher (12.92 lower to 13.52 higher)
Controlled Oral Word Association (COWA) - 9 months - Phonemic cues follow up: 9 months	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean controlled Oral Word Association (COWA) - 9 months - Phonemic cues was 31.1	MD 0.2 higher (7.99 lower to 8.39 higher)
Controlled Oral Word Association (COWA) - 9 months - Semantic cues follow up: 9 months	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean controlled Oral Word Association (COWA) - 9 months - Semantic cues was 31.5	MD 7.3 higher (1.89 lower to 16.49 higher)
Wisconsin Card Sorting Test (WCST) - 9 months - Total errors follow up: 9 months	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean wisconsin Card Sorting Test (WCST) - 9 months - Total errors was 49.5	MD 19.4 lower (36.03 lower to 2.77 lower)
Wisconsin Card Sorting Test (WCST) - 9 months - Perseverative errors follow up: 9 months	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean wisconsin Card Sorting Test (WCST) - 9 months - Perseverative errors was 36.8	MD 12.6 lower (29.56 lower to 4.36 higher)
Wisconsin Card Sorting Test (WCST) - 9 months - Perseverative responses follow up: 9 months	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean wisconsin Card Sorting Test (WCST) - 9 months - Perseverative responses was 32.22	MD 12.42 lower (23.77 lower to 1.07 lower)
Selective Reminding Test (SRT) - 9 months - Long-	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean selective Reminding Test (SRT) - 9 months -	MD 5.6 higher (5.16 lower to 16.36 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-9 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
term storage follow up: 9 months				Long-term storage was 30.2	
Selective Reminding Test (SRT) - 9 months - Consistent long-term retrieval follow up: 9 months	18 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean selective Reminding Test (SRT) - 9 months - Consistent long-term retrieval was 21.1	MD 2.2 higher (11.85 lower to 16.25 higher)
Selective Reminding Test (SRT) - 9 months - Delayed retrieval follow up: 9 months	18 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean selective Reminding Test (SRT) - 9 months - Delayed retrieval was 6.3	MD 1.4 higher (0.84 lower to 3.64 higher)
10/36 SPART (Spatial Recall Test) - 9 months - Long-term retrieval follow up: 9 months	18 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 10/36 SPART (Spatial Recall Test) - 9 months - Long-term retrieval was 15.5	MD 0.4 lower (4.18 lower to 3.38 higher)
10/36 SPART (Spatial Recall Test) - 9 months - Delayed recall follow up: 9 months	18 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 10/36 SPART (Spatial Recall Test) - 9 months - Delayed recall was 5.4	MD 0.2 lower (2.37 lower to 1.97 higher)
Test of Everyday Attention (TEA) median - 9 months - Auditory stimulus follow up: 9 months	18 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean test of Everyday Attention (TEA) median - 9 months - Auditory stimulus was 500.1	MD 172.6 higher (40.85 lower to 386.05 higher)
Test of Everyday Attention (TEA) median - 9 months - Visual stimulus	18 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean test of Everyday Attention (TEA) median - 9 months - Visual stimulus was 734.5	MD 228.2 higher (68.89 lower to 525.29 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 2-9 months	Risk difference with Multi-domain cog. rehab (pen/paper or computer tasks, with no additional teaching strategies)
follow up: 9 months					
MS quality of life (scale unclear) - 9 months follow up: 9 months	19 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean MS quality of life (scale unclear) - 9 months was 171.13	MD 27.37 higher (6.15 lower to 60.89 higher)
Montgomery-Asberg Depression Scale (scale usually 0-60) - 9 months Scale from: 0 to 60 follow up: 9 months	18 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean montgomery-Asberg Depression Scale (scale usually 0-60) - 9 months was 17.1	MD 9.8 lower (19.15 lower to 0.45 lower)

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

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

Multi-domain cognitive rehabilitation tailored to individual (CogniFit – computer tasks, with no additional teaching strategies) vs. control

See also summary of evidence from an additional paper (Mattioli 2014) comparing computer tasks to control, in this case consisting of psychoeducation with no cognitive training, that reported results only as medians.

Table 13: Clinical evidence summary: Multi-domain cognitive rehabilitation tailored to individual (CogniFit – computer tasks, with no additional teaching strategies) vs control, 3 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 3 months	Risk difference with Multi-domain cog. rehab tailored to individual (CogniFit - computer tasks, with no additional teaching strategies)
Divided attention - 3 months	46 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean divided attention - 3 months was 2.41	MD 0.04 lower (0.47 lower to 0.39 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 3 months	Risk difference with Multi-domain cog. rehab tailored to individual (CogniFit - computer tasks, with no additional teaching strategies)
follow up: 3 months					
Avoiding distractions - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean avoiding distractions - 3 months was 0.67	MD 0.03 higher (0.31 lower to 0.37 higher)
Hand-eye coordination - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean hand-eye coordination - 3 months was 0.562	MD 0.3 lower (0.81 lower to 0.2 higher)
General memory - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean general memory - 3 months was 0.56	MD 0.57 higher (0.01 higher to 1.13 higher)
Naming - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean naming - 3 months was 0.54	MD 0.14 higher (0.27 lower to 0.55 higher)
Response time - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean response time - 3 months was 0.51	MD 0.12 lower (0.53 lower to 0.29 higher)
Shifting attention - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean shifting attention - 3 months was 0.48	MD 0.11 lower (0.56 lower to 0.34 higher)
Spatial perception - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean spatial perception - 3 months was 0.54	MD 0.08 lower (0.47 lower to 0.31 higher)
Time estimation - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean time estimation - 3 months was 0.34	MD 0.28 higher (0.19 lower to 0.75 higher)
Visual working memory - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean visual working memory - 3 months was 0.65	MD 0.5 higher (0.04 lower to 1.04 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 3 months	Risk difference with Multi-domain cog. rehab tailored to individual (CogniFit - computer tasks, with no additional teaching strategies)
Visual scanning - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean visual scanning - 3 months was 0.57	MD 0.04 lower (0.53 lower to 0.45 higher)
Verbal auditory working memory - 3 months follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean verbal auditory working memory - 3 months was 0.53	MD 0.56 higher (0.03 higher to 1.09 higher)

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

b. Downgraded by 1 increment as cognitive impairment does not appear to be an inclusion criterion

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

Brain training apps/games (targeting general cognitive function/multiple domains) vs. control

Table 14: Clinical evidence summary: Brain training apps/games (targeting general cognitive function/multiple domains) vs. control, 1.5-3 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1.5-3 months	Risk difference with Brain training apps/games
Trail Making Test - 1.5-2 months - Part A, difference at baseline (higher in intervention) follow up: 1.5 months	47 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean trail Making Test - 1.5-2 months - Part A, difference at baseline (higher in intervention) was 43.4	MD 5.2 lower (16.92 lower to 6.52 higher)
Trail Making Test - 1.5-2 months - Part A, difference at baseline (lower in intervention) follow up: 2 months	43 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean trail Making Test - 1.5-2 months - Part A, difference at baseline (lower in intervention) was 58.2	MD 18.2 lower (37.27 lower to 0.87 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1.5-3 months	Risk difference with Brain training apps/games
Trail Making Test - 1.5-2 months - Part B, difference at baseline (higher in intervention) follow up: 1.5 months	47 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean trail Making Test - 1.5-2 months - Part B, difference at baseline (higher in intervention) was 82.5	MD 9.1 lower (23.73 lower to 5.53 higher)
Trail Making Test - 1.5-2 months - Part B, difference at baseline (lower in intervention) follow up: 2 months	43 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean trail Making Test - 1.5-2 months - Part B, difference at baseline (lower in intervention) was 121.1	MD 39.8 lower (74.24 lower to 5.36 lower)
Stroop test - 1.5-2 months - General 'Stroop Test' follow up: 2 months	58 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean stroop test - 1.5-2 months - General 'Stroop Test' ranged from 23.38-24.9	MD 4.03 higher (0.21 higher to 7.85 higher)
Stroop test - 1.5-2 months - Color follow up: 1.5 months	47 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean stroop test - 1.5-2 months - Color was 57.5	MD 7.7 higher (4.08 lower to 19.48 higher)
Stroop test - 1.5-2 months - Color-Word follow up: 1.5 months	47 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean stroop test - 1.5-2 months - Color-Word was 38.8	MD 5.2 higher (2.26 lower to 12.66 higher)
PASAT - 2 months - 2 seconds follow up: 2 months	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT - 2 months - 2 seconds was 34.14	MD 1.22 higher (8.69 lower to 11.13 higher)
PASAT - 2 months - 3 seconds follow up: 2 months	86 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT - 2 months - 3 seconds ranged from 31.69-45.64	MD 5.91 higher (1.6 higher to 10.22 higher)
PASAT - 3 months (z-score) - 2 second Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean PASAT - 3 months (z-score) - 2 second was -0.48	MD 0.2 higher (0.78 lower to 1.18 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1.5-3 months	Risk difference with Brain training apps/games
PASAT - 3 months (z-score) - 3 second Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean PASAT - 3 months (z-score) - 3 second was -0.32	MD 0.56 higher (0.26 lower to 1.38 higher)
SDMT - 1.5-2 months follow up: 1.5-2 months	133 (4 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 1.5-2 months ranged from 38.59-44.50	MD 7.17 higher (3.15 higher to 11.2 higher)
Selective Reminding Test (SRT) - 2 months - Long-term storage follow up: 2 months	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test (SRT) - 2 months - Long-term storage was 51.14	MD 3.78 lower (15.71 lower to 8.15 higher)
Selective Reminding Test (SRT) - 2 months - Consecutive long-term retrieval follow up: 2 months	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test (SRT) - 2 months - Consecutive long-term retrieval was 38.29	MD 1.14 higher (14.3 lower to 16.58 higher)
Selective Reminding Test (SRT) - 2 months - Delayed recall follow up: 2 months	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test (SRT) - 2 months - Delayed recall was 9.29	MD 0.43 lower (2.08 lower to 1.22 higher)
Selective Reminding Test (SRT) - 3 months (z-score) - Learning trials Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean selective Reminding Test (SRT) - 3 months (z-score) - Learning trials was -0.24	MD 0.37 higher (0.65 lower to 1.39 higher)
Selective Reminding Test (SRT) - 3 months (z-score) - Delay Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean selective Reminding Test (SRT) - 3 months (z-score) - Delay was 0.3	MD 0.29 higher (0.83 lower to 1.41 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1.5-3 months	Risk difference with Brain training apps/games
10/36 SPART (Spatial Recall Test) - 2 months - Correct follow up: 2 months	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 10/36 SPART (Spatial Recall Test) - 2 months - Correct was 19.43	MD 2.57 higher (2.22 lower to 7.36 higher)
10/36 SPART (Spatial Recall Test) - 2 months - Delayed follow up: 2 months	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 10/36 SPART (Spatial Recall Test) - 2 months - Delayed was 6.5	MD 1.07 higher (1.38 lower to 3.52 higher)
Word List Generation Test - 2 months follow up: 2 months	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean word List Generation Test - 2 months was 28.07	MD 0.28 lower (6.09 lower to 5.53 higher)
Brief Visuospatial Memory Test-Revised (BVMT-R) - 1.5 months follow up: 1.5 months	47 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean brief Visuospatial Memory Test-Revised (BVMT-R) - 1.5 months was 22.5	MD 5 higher (0.45 lower to 10.45 higher)
Brief Visuospatial Memory Test-Revised (BVMT-R) - 3 months (z-score) - Learning trials Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean brief Visuospatial Memory Test-Revised (BVMT-R) - 3 months (z-score) - Learning trials was -0.25	MD 0.1 higher (1.31 lower to 1.51 higher)
Brief Visuospatial Memory Test-Revised (BVMT-R) - 3 months (z-score) - Delay Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean brief Visuospatial Memory Test-Revised (BVMT-R) - 3 months (z-score) - Delay was -0.33	MD 0.16 higher (1.22 lower to 1.54 higher)
Greek Verbal Learning Test - 1.5 months	47 (1 RCT)	⊕◯◯◯ VERY	-	The mean Greek Verbal Learning Test	MD 9.3 higher (0.38 higher to 18.22 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1.5-3 months	Risk difference with Brain training apps/games
follow up: 1.5 months		LOW a,b,c		- 1.5 months was 54.4	
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Immediate memory (scale 40-152?) Scale from: 40 to 152 follow up: 2 months	43 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Immediate memory (scale 40-152?) was 97.7	MD 10.1 higher (0.45 lower to 20.65 higher)
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Visuospatial/constructional (scale 50-131?) Scale from: 50 to 131 follow up: 2 months	43 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Visuospatial/constructional (scale 50-131?) was 101.9	MD 5 higher (3.39 lower to 13.39 higher)
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Language (scale 40-134?) Scale from: 40 to 134 follow up: 2 months	43 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Language (scale 40-134?) was 100.4	MD 7.3 higher (0.6 lower to 15.2 higher)
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Attention (scale 40-150?)	43 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Attention (scale 40-150?) was 81.4	MD 13.4 higher (3.89 higher to 22.91 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1.5-3 months	Risk difference with Brain training apps/games
Scale from: 40 to 150 follow up: 2 months					
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Delayed memory (scale 40-133?) Scale from: 40 to 133 follow up: 2 months	43 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Delayed memory (scale 40-133?) was 98.8	MD 9.6 higher (0.16 higher to 19.04 higher)
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Total score (scale 40-160?) Scale from: 40 to 160 follow up: 2 months	43 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean repeatable Battery for the Assessment of Neuropsychological Status (RBANS) - 2 months - Total score (scale 40-160?) was 94.9	MD 12.4 higher (2.35 higher to 22.45 higher)
Wechsler adult intelligence scale IV (WAIS-IV) Letter-Number Sequencing - 3 months (z-score) Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean Wechsler adult intelligence scale IV (WAIS-IV) Letter-Number Sequencing - 3 months (z-score) was -0.04	MD 0 (0.64 lower to 0.64 higher)
Visual span (Corsi block tapping test) - 3 months (z-score) Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean visual span (Corsi block tapping test) - 3 months (z-score) was -0.52	MD 0.26 higher (0.33 lower to 0.85 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1.5-3 months	Risk difference with Brain training apps/games
DelisKaplan executive function system (DKEFS) - 3 months (z-score) - Trail 5 Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean delisKaplan executive function system (DKEFS) - 3 months (z-score) - Trail 5 was 0.63	MD 0.01 higher (0.3 lower to 0.32 higher)
DelisKaplan executive function system (DKEFS) - 3 months (z-score) - Trails 2/3 combo Scale from: -5 to 5 follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean delisKaplan executive function system (DKEFS) - 3 months (z-score) - Trails 2/3 combo was 0.0	MD 0.27 higher (0.57 lower to 1.11 higher)
General cognitive composite (average of multiple cognitive tests) - 3 months change from baseline (z-score) Scale from: -5 to 5 follow up: 3 months	155 (2 RCTs)	⊕◯◯◯ VERY LOW a, c,d	-	The mean general cognitive composite (average of multiple cognitive tests) - 3 months change from baseline (z-score) ranged from -0.14 to 0.09	MD 0.32 higher (0.09 lower to 0.74 higher)
Self-reported improvement in cognition - 3 months follow up: 3 months	135 (1 RCT)	⊕⊕⊕◯ MODERATE a	OR 2.90 (1.43 to 5.91)	Moderate 312 per 1,000	256 more per 1,000 (81 more to 416 more)
Modified Fatigue Impact Scale (MFIS) - Cognitive (scale usually 0-40) - 2 months Scale from: 0 to 40 follow up: 2 months	43 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean modified Fatigue Impact Scale (MFIS) - Cognitive (scale usually 0-40) - 2 months was 18.06	MD 7 lower (12.03 lower to 1.97 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 1.5-3 months	Risk difference with Brain training apps/games
MSQoL-54 (scale usually 0-100) - 2 months - Physical composite Scale from: 0 to 100 follow up: 2 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean mSQoL-54 (scale usually 0-100) - 2 months - Physical composite was 62.72	MD 0.02 lower (9.12 lower to 9.08 higher)
MSQoL-54 (scale usually 0-100) - 2 months - Mental health composite Scale from: 0 to 100 follow up: 2 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean mSQoL-54 (scale usually 0-100) - 2 months - Mental health composite was 54.03	MD 7.47 higher (2.38 lower to 17.32 higher)
Adherence (varying definitions) - Compliant to study requirements follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 1.29 (0.14 to 11.54)	Moderate	
				778 per 1,000	41 more per 1,000 (449 fewer to 198 more)
Adherence (varying definitions) - At least 6 compliant weeks (50% of target) follow up: 3 months	135 (1 RCT)	⊕⊕◯◯ LOW a,c	OR 0.38 (0.17 to 0.81)	Moderate	
				787 per 1,000	203 fewer per 1,000 (401 fewer to 37 fewer)
Adherence (varying definitions) - Meeting or exceeding 30 h of training (50% of target) follow up: 3 months	135 (1 RCT)	⊕⊕◯◯ LOW a,c	OR 0.40 (0.18 to 0.86)	Moderate	
				787 per 1,000	191 fewer per 1,000 (388 fewer to 26 fewer)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies that reported the outcome at a time-point <3-month minimum specified in the protocol

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

d. Downgraded by 1 increment as statistical heterogeneity is present, with I2>50%, that cannot be explained by prespecified subgroup analyses

Mental visual imagery vs. control (sham verbal control)

Table 15: Clinical evidence summary: Mental visual imagery vs. control (sham verbal control), 6-8 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (sham verbal control), 6-8 weeks	Risk difference with Mental visual imagery
Number of details provided (Measure of mental visualisation ability) follow up: 6-8 weeks	17 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean number of details provided (Measure of mental visualisation ability) was 6.41	MD 0.55 lower (2.71 lower to 1.61 higher)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

Mindfulness vs. control

Table 16: Mindfulness vs. control (waitlist control), 4 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control), 4 weeks	Risk difference with Mindfulness
SDMT - 4 weeks follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 4 weeks was 61.1	MD 7.6 lower (18.11 lower to 2.91 higher)
PASAT - 4 weeks - 2 seconds follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT - 4 weeks - 2 seconds was 42.1	MD 3.8 lower (11.32 lower to 3.72 higher)
PASAT - 4 weeks - 3 seconds follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT - 4 weeks - 3 seconds was 52.7	MD 4.4 lower (11.4 lower to 2.6 higher)
Selective Reminding Test - 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY	-	The mean selective Reminding Test - 4	MD 6.7 higher (6.16 lower to 19.56 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control), 4 weeks	Risk difference with Mindfulness
- Long-term storage follow up: 4 weeks		LOW a,b,c		weeks - Long-term storage was 44.2	
Selective Reminding Test - 4 weeks - Consistent long-term retrieval follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test - 4 weeks - Consistent long-term retrieval was 36.3	MD 9.1 higher (6.74 lower to 24.94 higher)
Selective Reminding Test - 4 weeks - Delayed recall follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test - 4 weeks - Delayed recall was 7.53	MD 1.22 higher (1.16 lower to 3.6 higher)
Word List Generation - 4 weeks follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean word List Generation - 4 weeks was 32.0	MD 2.2 lower (7.73 lower to 3.33 higher)
10/36 SPART (Spatial Recall Test) - 4 weeks - Immediate follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 10/36 SPART (Spatial Recall Test) - 4 weeks - Immediate was 23.1	MD 1.2 lower (5.18 lower to 2.78 higher)
10/36 SPART (Spatial Recall Test) - 4 weeks - Delayed follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 10/36 SPART (Spatial Recall Test) - 4 weeks - Delayed was 8.24	MD 0.99 lower (2.69 lower to 0.71 higher)
Beck Depression Inventory (scale usually 0-63) - 4 weeks Scale from: 0 to 63 follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory (scale usually 0-63) - 4 weeks was 10.2	MD 2.07 lower (8.13 lower to 3.99 higher)
Penn State Worry Questionnaire (scale usually 16-80) - 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean Penn State Worry Questionnaire (scale usually 16-80) - 4 weeks was 42.6	MD 1.3 higher (10.16 lower to 12.76 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control), 4 weeks	Risk difference with Mindfulness
Scale from: 16 to 80 follow up: 4 weeks					
Difficulties in Emotion Regulation (DERS, scale unclear) - 4 weeks follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean difficulties in Emotion Regulation (DERS, scale unclear) - 4 weeks was 75.0	MD 6.2 lower (19.04 lower to 6.64 higher)
WHO Quality of Life and Satisfaction With Life Scale composite (z-score) - 4 weeks Scale from: -5 to 5 follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean WHO Quality of Life and Satisfaction With Life Scale composite (z-score) - 4 weeks was 0.16	MD 0.29 higher (0.26 lower to 0.84 higher)

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

b. Downgraded by 1 increment as the majority of the evidence was from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

Table 17: Mindfulness vs. control (pharma only), 12 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (pharma only), 12 months	Risk difference with Mindfulness
SDMT - 12 months follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SDMT - 12 months was 33.43	MD 7.54 higher (0.18 higher to 14.9 higher)
PASAT - 12 months - 2 seconds follow up: 12 months	60 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean PASAT - 12 months - 2 seconds was 23.1	MD 12.4 higher (5.93 higher to 18.87 higher)
PASAT - 12 months - 3 seconds follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PASAT - 12 months - 3 seconds was 26.23	MD 10.97 higher (4.85 higher to 17.09 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (pharma only), 12 months	Risk difference with Mindfulness
COWAT verbal fluency test - 12 months - Words (FAS) follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean COWAT verbal fluency test - 12 months - Words (FAS) was 30.37	MD 6.76 higher (0.57 higher to 12.95 higher)
COWAT verbal fluency test - 12 months - Names of animals follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,	-	The mean COWAT verbal fluency test - 12 months - Names of animals was 15.73	MD 2.3 higher (0.74 lower to 5.34 higher)
Wechsler Memory Scale - III Spanish Version - 12 months - Attention follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Memory Scale - III Spanish Version - 12 months - Attention was 4.87	MD 0.16 higher (0.89 lower to 1.21 higher)
Wechsler Memory Scale - III Spanish Version - 12 months - Long-term memory follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Memory Scale - III Spanish Version - 12 months - Long-term memory was 6.1	MD 1.77 higher (0.1 higher to 3.44 higher)
Wechsler Memory Scale - III Spanish Version - 12 months - Short-term memory follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Memory Scale - III Spanish Version - 12 months - Short-term memory was 27.17	MD 2.26 higher (1.88 lower to 6.4 higher)
Wechsler Memory Scale - III Spanish Version - 12 months - Recognition follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Memory Scale - III Spanish Version - 12 months - Recognition was 20.0	MD 2.23 higher (0.16 higher to 4.3 higher)
Wechsler Memory Scale - III Spanish Version - 12 months - Learning follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Memory Scale - III Spanish Version - 12 months - Learning was 3.37	MD 0.6 higher (0.32 lower to 1.52 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (pharma only), 12 months	Risk difference with Mindfulness
Beck Depression Inventory (scale usually 0-63) - 12 months Scale from: 0 to 63 follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Beck Depression Inventory (scale usually 0-63) - 12 months was 18.67	MD 4.67 lower (9.34 lower to 0.00)
State-Trait Anxiety Inventory (unclear if state or trait subscale or both combined, scale usually 20-80 for each subscale) - 12 months follow up: 12 months	60 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean state-Trait Anxiety Inventory (unclear if state or trait subscale or both combined, scale usually 20-80 for each subscale) - 12 months was 41.77	MD 2.8 lower (14.57 lower to 8.97 higher)
FIM + FAM composite (functional independence and assessment measures, scale used unclear) - 12 months follow up: 12 months	60 (1 RCT)	⊕◯◯◯ VERY LOW a,	-	The mean FIM + FAM composite (functional independence and assessment measures, scale used unclear) - 12 months was 53.25	MD 3.08 lower (12.02 lower to 5.86 higher)

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

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

Mindfulness vs. general cognitive rehabilitation (multi-component and multi-domain)

Table 18: Mindfulness vs. general cognitive rehabilitation (multi-component and multi-domain) 4 weeks

Outcomes	№ of 122articip ant (studies) Follow up	Certainty of the evidence (GRADE)	Relativ e effect (95% CI)	Anticipated absolute effects	
				Risk with general cogn. Rehab (multi-component), 4 weeks	Risk difference with Mindfulness
SDMT – 4 weeks follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT – 4 weeks was 53.2	MD 0.3 higher (9.53 lower to 10.13 higher)
PASAT – 4 weeks – 2 seconds follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT – 4 weeks – 2 seconds was 38.9	MD 0.6 lower (7.76 lower to 6.56 higher)
PASAT – 4 weeks – 3 seconds follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT – 4 weeks – 3 seconds was 51.1	MD 2.8 lower (10.22 lower to 4.62 higher)
Selective Reminding Test – 4 weeks – Long-term storage follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test – 4 weeks – Long-term storage was 43.3	MD 7.6 higher (3.42 lower to 18.62 higher)
Selective Reminding Test – 4 weeks – Consistent long-term retrieval follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test – 4 weeks – Consistent long-term retrieval was 33.3	MD 12.1 higher (1.7 lower to 25.9 higher)
Selective Reminding Test – 4 weeks – Delayed recall follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective Reminding Test – 4 weeks – Delayed recall was 7.59	MD 1.16 higher (0.91 lower to 3.23 higher)
Word List Generation – 4 weeks follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean word List Generation – 4 weeks was 33.9	MD 4.1 lower (9.78 lower to 1.58 higher)
10/36 SPART (Spatial Recall Test) – 4 weeks – Immediate follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 10/36 SPART (Spatial Recall Test) – 4 weeks – Immediate was 20.3	MD 1.6 higher (3.02 lower to 6.22 higher)

Outcomes	№ of 122articip ant (studies) Follow up	Certainty of the evidence (GRADE)	Relativ e effect (95% CI)	Anticipated absolute effects	
				Risk with general cogn. Rehab (multi-component), 4 weeks	Risk difference with Mindfulness
10/36 SPART (Spatial Recall Test) – 4 weeks – Delayed follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 10/36 SPART (Spatial Recall Test) – 4 weeks – Delayed was 6.88	MD 0.37 higher (1.6 lower to 2.34 higher)
Beck Depression Inventory (scale usually 0-63) – 4 weeks Scale from: 0 to 63 follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory (scale usually 0-63) – 4 weeks was 11.4	MD 3.27 lower (9.03 lower to 2.49 higher)
Penn State Worry Questionnaire (scale usually 16-80) – 4 weeks Scale from: 16 to 80 follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean penn State Worry Questionnaire (scale usually 16-80) – 4 weeks was 48.5	MD 4.6 lower (14.28 lower to 5.08 higher)
Difficulties in Emotion Regulation (DERS, scale unclear) – 4 weeks follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean difficulties in Emotion Regulation (DERS, scale unclear) – 4 weeks was 74.5	MD 5.7 lower (18.61 lower to 7.21 higher)
WHO Quality of Life and Satisfaction With Life Scale composite (z-score) – 4 weeks Scale from: -5 to 5 follow up: 4 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean WHO Quality of Life and Satisfaction With Life Scale composite (z-score) – 4 weeks was 0.056	MD 0.39 higher (0.16 lower to 0.95 higher)
Adherence – completing all four weekly sessions follow up: 4 weeks	40 (1 RCT)	⊕⊕◯◯ LOW a,c	OR 4.85 (0.86 to 27.22)	Moderate	
				650 per 1,000	250 more per 1,000 (35 fewer to 331 more)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies that reported the outcome at a time-point <3-month minimum specified in the protocol

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

Mindfulness vs. medical treatment and counselling

Table 19: Mindfulness vs. medical treatment + counselling, 16 weeks (8 weeks after end of intervention)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with medical treatment and counselling, 8 weeks	Risk difference with Mindfulness
Wechsler Adult Intelligence Scale-Revised (WAIS-R) - symbol coding test follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Adult Intelligence Scale-Revised (WAIS-R) - symbol coding test was 45.88	MD 4.52 higher (0.84 lower to 9.88 higher)
Wechsler Adult Intelligence Scale-Revised (WAIS-R) - digit span test follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Adult Intelligence Scale-Revised (WAIS-R) - digit span test was 5.38	MD 1.36 higher (0.62 higher to 2.1 higher)
Rey Complex Figure Test - recall follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Rey Complex Figure Test - recall was 22.8	MD 1.97 higher (0.39 lower to 4.33 higher)
PASAT 3 seconds follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PASAT 3 seconds was 33.61	MD 10.5 higher (3.67 higher to 17.33 higher)
PASAT 2 seconds follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PASAT 2 seconds was 31.69	MD 6.19 higher (1.29 higher to 11.09 higher)
Wisconsin Card Sorting Test - category follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wisconsin Card Sorting Test - category was 3.53	MD 0.65 higher (0.45 lower to 1.75 higher)
Wisconsin Card Sorting Test - perseveration follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wisconsin Card Sorting Test - perseveration was 8.0	MD 4.78 lower (7.1 lower to 2.46 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with medical treatment and counselling, 8 weeks	Risk difference with Mindfulness
Wisconsin Card Sorting Test - conception responses follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wisconsin Card Sorting Test - conception responses was 3.92	MD 0.89 higher (0.42 lower to 2.2 higher)
Wisconsin Card Sorting Test - total correct follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wisconsin Card Sorting Test - total correct was 31.69	MD 6.19 higher (1.29 higher to 11.09 higher)
Wisconsin Card Sorting Test - number of errors follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wisconsin Card Sorting Test - number of errors was 25.26	MD 7.41 lower (12.06 lower to 2.76 lower)
Wisconsin Card Sorting Test - other errors follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wisconsin Card Sorting Test - other errors was 16.61	MD 2.65 lower (5.97 lower to 0.67 higher)
Wisconsin Card Sorting Test - first trial follow up: 16 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wisconsin Card Sorting Test - first trial was 13.84	MD 3.96 lower (10.37 lower to 2.45 higher)
Hamilton Anxiety Scale (scale 0-56) follow up: 16 weeks	53 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean Hamilton Anxiety Scale (scale 0-56) was 13.0	MD 6.56 lower (9.27 lower to 3.85 lower)

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

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

Information processing speed: cognitive rehabilitation software focused on processing speed + occupational therapy vs. occupational therapy only

Table 20: Cognitive rehabilitation focused on processing speed + occupational therapy vs. occupational therapy only, 3 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with occupational therapy alone, 3 months	Risk difference with Info processing speed: cogn. rehab focused on processing speed + occupational therapy
SDMT - 3 months follow up: 3 months	64 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SDMT - 3 months was 32.75	MD 3.44 higher (1.87 lower to 8.75 higher)
PASAT - 3 months follow up: 3 months	64 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PASAT - 3 months was 31.97	MD 5.93 higher (0.54 lower to 12.4 higher)

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

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

Information processing speed: cognitive rehabilitation software focused on processing speed vs. control (active game or no intervention)

Table 21: Cognitive rehabilitation software focused on processing speed vs. control (active game or no intervention), 5-6 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (active game or no intervention), 5-6 weeks	Risk difference with Info processing speed: cogn. rehab software focused on processing speed
SDMT - 6 weeks (change from baseline) follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 6 weeks (change from baseline) was 3.55	MD 2.55 higher (1.31 lower to 6.41 higher)
Wechsler Adult Intelligence Scale-III - Digit Symbol Coding Subtest - 5 weeks follow up: 5 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean wechsler Adult Intelligence Scale-III - Digit Symbol Coding Subtest - 5 weeks was 5.44	MD 2.06 higher (0.16 lower to 4.28 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (active game or no intervention), 5-6 weeks	Risk difference with Info processing speed: cogn. rehab software focused on processing speed
PASAT - 6 weeks (change from baseline) follow up: 6 weeks	37 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PASAT - 6 weeks (change from baseline) was 2.53	MD 0.19 higher (3.9 lower to 4.28 higher)
Brief Visuospatial Memory Test-Revised (BVMT-R) - 6 weeks (change from baseline) follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean brief Visuospatial Memory Test-Revised (BVMT-R) - 6 weeks (change from baseline) was 3.25	MD 2.55 lower (5.57 lower to 0.47 higher)
California Verbal Learning Test-II (CVLT-II) - 6 weeks (change from baseline) - Number correct follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test-II (CVLT-II) - 6 weeks (change from baseline) - Number correct was 5.2	MD 3.15 lower (8.65 lower to 2.35 higher)
California Verbal Learning Test-II (CVLT-II) - 5-weeks - Learning slope follow up: 5 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test-II (CVLT-II) - 5-weeks - Learning slope was 0.99	MD 0.18 higher (0.25 lower to 0.61 higher)
California Verbal Learning Test-II (CVLT-II) - 5 weeks - Short-delay free recall follow up: 5 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test-II (CVLT-II) - 5 weeks - Short-delay free recall was 6.65	MD 2.1 higher (1.24 lower to 5.44 higher)
Letter comparison (perceptual speed) - 5 weeks follow up: 5 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean letter comparison (perceptual speed) - 5 weeks was 6.78	MD 1.35 higher (0.81 lower to 3.51 higher)
Pattern comparison (perceptual	21 (1 RCT)	⊕◯◯◯ VERY	-	The mean pattern comparison	MD 1.65 higher (1.68 lower to 4.98 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (active game or no intervention), 5-6 weeks	Risk difference with Info processing speed: cogn. rehab software focused on processing speed
speed) - 5 weeks follow up: 5 weeks		LOW a,b,c		(perceptual speed) - 5 weeks was 12.06	
Perceived Deficits Questionnaire (5-item, scale usually 0-80) - 6 weeks (change from baseline) follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean perceived Deficits Questionnaire (5-item, scale usually 0-80) - 6 weeks (change from baseline) was -1.64	MD 1.07 higher (0.1 lower to 2.24 higher)
Timed Instrumental Activities of Daily Living Test (TIADL - z-score for speed and accuracy combined) - 5 weeks Scale from: -5 to 5 follow up: 5 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean timed Instrumental Activities of Daily Living Test (TIADL - z-score for speed and accuracy combined) - 5 weeks was -0.23	MD 0.61 higher (0.09 higher to 1.13 higher)
CES-D depression (scale usually 0-60) - 6 weeks (change from baseline) Scale from: 0 to 60 follow up: 6 weeks	38 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean CES-D depression (scale usually 0-60) - 6 weeks (change from baseline) was -0.9	MD 2.01 higher (1.72 lower to 5.74 higher)
State-Trait Anxiety Index - State subscore (STAI-S; scale usually 20-80) - 6 weeks (change from baseline) Scale from: 20 to 80 follow up: 6 weeks	38 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean state-Trait Anxiety Index - State subscore (STAI-S; scale usually 20-80) - 6 weeks (change from baseline) was 0.21	MD 0.16 lower (4.69 lower to 4.37 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (active game or no intervention), 5-6 weeks	Risk difference with Info processing speed: cogn. rehab software focused on processing speed
State-Trait Anxiety Index - Trait subscore (STAI-T; scale usually 20-80) - 6 weeks (change from baseline) Scale from: 20 to 80 follow up: 6 weeks	34 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean state-Trait Anxiety Index - Trait subscore (STAI-T; scale usually 20-80) - 6 weeks (change from baseline) was 0.35	MD 0.42 higher (2.1 lower to 2.94 higher)
Modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 6 weeks (change from baseline) Scale from: 0 to 84 follow up: 6 weeks	38 (1 RCT)	⨁◯◯◯ VERY LOW a,b,d	-	The mean modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 6 weeks (change from baseline) was -2.95	MD 1.84 lower (6.98 lower to 3.3 higher)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

d. Downgraded by 1 increment as general Modified Fatigue Impact Scale reported rather than specifically the cognitive subdomain

Information processing speed + working memory: n-back training focused on processing speed + working memory vs. sham training (n-back with no increasing difficulty)

Table 22: N-back training focused on processing speed + working memory vs. sham training (n-back training with no increasing difficulty), 6 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with sham training (n-back with no increasing difficulty), 6 weeks	Risk difference with Info processing speed + working memory: n-back training focused on processing speed + working memory
SDMT - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 6 weeks was 50.5	MD 0.35 higher (7.99 lower to 8.69 higher)
PASAT - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT - 6 weeks was 76.67	MD 11.38 higher (2.25 lower to 25.01 higher)
Stroop Test - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean stroop Test - 6 weeks was 32.16	MD 3.44 higher (1.23 lower to 8.11 higher)
COWAT - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean COWAT - 6 weeks was 37.95	MD 4.2 higher (4.93 lower to 13.33 higher)
Letter-Number Sequencing - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean letter-Number Sequencing - 6 weeks was 10.95	MD 0.2 higher (1.45 lower to 1.85 higher)
Digits backwards - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean digits backwards - 6 weeks was 5.1	MD 0.05 lower (1.29 lower to 1.19 higher)
Raven's Advanced Progressive Matrices (test of fluid intelligence) - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean raven's Advanced Progressive Matrices (test of fluid intelligence) - 6 weeks was 10.44	MD 1.12 lower (3.56 lower to 1.32 higher)
Brief Visuospatial Memory Test (BVMT) - Trials 1-3 - 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean brief Visuospatial Memory Test (BVMT) - Trials 1-3 - 6 weeks was 20.05	MD 1.4 higher (2.27 lower to 5.07 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with sham training (n-back with no increasing difficulty), 6 weeks	Risk difference with Info processing speed + working memory: n-back training focused on processing speed + working memory
follow up: 6 weeks					
Conners' Continuous Performance Task Commissions - Speed (measures sustained attention and response inhibition, T-score) - 6 weeks Scale from: 0 to 100 follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean conners' Continuous Performance Task Commissions - Speed (measures sustained attention and response inhibition, T-score) - 6 weeks was 49.5	MD 1.5 lower (8.41 lower to 5.41 higher)
Auditory Verbal Learning Task (AVLT) - Trials 1-5 - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean auditory Verbal Learning Task (AVLT) - Trials 1-5 - 6 weeks was 45.95	MD 6.7 higher (0.15 higher to 13.25 higher)
MSQoL-54 (scale usually 0-100) - 6 weeks Scale from: 0 to 100 follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean mSQoL-54 (scale usually 0-100) - 6 weeks was 75.45	MD 4.95 lower (13.62 lower to 3.72 higher)
State-Trait Anxiety Inventory - State subscale (STAI; scale usually 20-80) - 6 weeks Scale from: 20 to 80 follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean state-Trait Anxiety Inventory - State subscale (STAI; scale usually 20-80) - 6 weeks was 44.33	MD 1.27 higher (2.46 lower to 5 higher)
State-Trait Anxiety Inventory - Trait subscale	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean state-Trait Anxiety Inventory - Trait subscale (STAI;	MD 0.86 higher (2.5 lower to 4.22 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with sham training (n-back with no increasing difficulty), 6 weeks	Risk difference with Info processing speed + working memory: n-back training focused on processing speed + working memory
(STAI; scale usually 20-80) - 6 weeks Scale from: 20 to 80 follow up: 6 weeks				scale usually 20-80) - 6 weeks was 44.64	
Beck Depression Inventory-Fast Screen (scale usually 0-21) - 6 weeks Scale from: 0 to 21 follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory-Fast Screen (scale usually 0-21) - 6 weeks was 2.6	MD 1.4 higher (0.23 lower to 3.03 higher)
Modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 6 weeks Scale from: 0 to 84 follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 6 weeks was 43.6	MD 0.35 higher (10.95 lower to 11.65 higher)
Adherence - % training completed (objective report) - 6 weeks follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean adherence - % training completed (objective report) - 6 weeks was 95.3	MD 0.67 lower (8.29 lower to 6.95 higher)
Satisfaction - proportion very satisfied with overall study experience follow up: 6 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 0.26 (0.06 to 1.21)	Moderate	
				850 per 1,000	254 fewer per 1,000 (596 fewer to 23 more)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

d. Downgraded by 1 increment as general Modified Fatigue Impact Scale reported rather than specifically the cognitive subdomain

Attention/working memory: computer-aided RehaCom training (attention and information processing) vs. active control

Table 23: Computer-aided RehaCom training (attention and information processing) vs. active control, 6 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control, 6 weeks	Risk difference with Attention/working memory: computer-aided RehaCom training (attention and info processing)
SDMT - 6 weeks follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 6 weeks was 37.3	MD 1.39 higher (6.11 lower to 8.89 higher)
PASAT 3 seconds - 6 weeks follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT 3 seconds - 6 weeks was 41.0	MD 0.23 higher (8.64 lower to 9.1 higher)
Selective reminding test (SRT) - 6 weeks - Long-term storage follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective reminding test (SRT) - 6 weeks - Long-term storage was 29.9	MD 7 higher (2.12 lower to 16.12 higher)
Selective reminding test (SRT) - 6 weeks - Consistent long-term retrieval follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective reminding test (SRT) - 6 weeks - Consistent long-term retrieval was 17.1	MD 7.76 higher (0.16 higher to 15.36 higher)
Selective reminding test (SRT) - 6 weeks - Delayed recall follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean selective reminding test (SRT) - 6 weeks - Delayed recall was 6.2	MD 0.91 higher (1.53 lower to 3.35 higher)
10/36 SPART (Spatial Recall Test) - 6 weeks - Immediate follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 10/36 SPART (Spatial Recall Test) - 6 weeks - Immediate was 24.3	MD 5.88 lower (10.12 lower to 1.64 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control, 6 weeks	Risk difference with Attention/working memory: computer-aided RehaCom training (attention and info processing)
10/36 SPART (Spatial Recall Test) - 6 weeks - Delayed follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 10/36 SPART (Spatial Recall Test) - 6 weeks - Delayed was 8.3	MD 2.72 lower (4.51 lower to 0.93 lower)
Word List Generation - 6 weeks follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean word List Generation - 6 weeks was 20.6	MD 0.2 higher (4.52 lower to 4.92 higher)
Stroop Test - 6 weeks follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean stroop Test - 6 weeks was 16.5	MD 2.91 higher (1.33 lower to 7.15 higher)
Trail Making Test - 6 weeks - Part A follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean trail Making Test - 6 weeks - Part A was 40.9	MD 3.93 higher (7.15 lower to 15.01 higher)
Trail Making Test - 6 weeks - Part B follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean trail Making Test - 6 weeks - Part B was 121.1	MD 0.2 lower (30.99 lower to 30.59 higher)
Trail Making Test - 6 weeks - Part B-A follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW b,c	-	The mean trail Making Test - 6 weeks - Part B-A was 76.9	MD 0.82 lower (27.3 lower to 25.66 higher)
State-Trait Anxiety Inventory Y1 (State?; scale usually 20-80) - 6 weeks Scale from: 20 to 80 follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean state-Trait Anxiety Inventory Y1 (State?; scale usually 20-80) - 6 weeks was 41.0	MD 4.4 lower (12.67 lower to 3.87 higher)
State-Trait Anxiety Inventory Y2 (Trait?; scale usually 20-80) - 6 weeks Scale from: 20 to 80	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean state-Trait Anxiety Inventory Y2 (Trait?; scale usually 20-80) - 6 weeks was 46.0	MD 10.5 lower (18.67 lower to 2.33 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control, 6 weeks	Risk difference with Attention/working memory: computer-aided RehaCom training (attention and info processing)
follow up: 6 weeks					
Beck Depresion Inventory-II (scale usually 0-63) - 6 weeks Scale from: 0 to 63 follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory-II (scale usually 0-63) - 6 weeks was 12.8	MD 8.47 lower (16.65 lower to 0.29 lower)
Fatigue severity scale (FSS; scale likely 1-7) - 6 weeks Scale from: 1 to 7 follow up: 6 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean fatigue severity scale (FSS; scale likely 1-7) - 6 weeks was 4.22	MD 1.39 lower (2.78 lower to 0.00 lower)

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

b. Downgraded by 1 increment as the majority of the evidence comes from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

d. Downgraded by 1 increment as general Fatigue Severity Scale reported rather than cognitive fatigue specifically

Attention/working memory: computer-aided training for attention/working memory vs. control

Table 24: Computer-aided training for attention/working memory vs. control, 18 weeks – 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 18 weeks - 6 months	Risk difference with Attention/working memory: computer-aided training of attention/working memory
SDMT - 18 weeks - 6 months (mix of final values and change from baseline) follow up: 18 weeks - 6 months	53 (2 RCTs)	⨁⨁◯◯ LOW a,b	-	The mean SDMT - 18 weeks - 6 months (mix of final values and change from baseline) was 4.57 for change scores and 40.64 for final values	MD 1.14 lower (4.82 lower to 2.54 higher)
PASAT - 6 months follow up: 6 months	22 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean PASAT - 6 months was 33.91	MD 1.27 higher (8.32 lower to 10.86 higher)
California Verbal Learning Test-II (CVLT-II) Total Immediate Recall - 18-weeks - 6 months follow up: 18 weeks - 6 months	53 (2 RCTs)	⨁⨁◯◯ LOW a,b	-	The mean california Verbal Learning Test-II (CVLT-II) Total Immediate Recall - 18-weeks - 6 months was 7.5 for change scores and 45.0 for final values	MD 0.12 lower (5.19 lower to 4.95 higher)
Brief Visuospatial Memory Test – Revised (BVMT-R) Total Immediate Recall - 18 weeks - 6 months follow up: 18 weeks - 6 months	53 (2 RCTs)	⨁⨁◯◯ LOW a,b	-	The mean brief Visuospatial Memory Test – Revised (BVMT-R) Total Immediate Recall - 18 weeks - 6 months was 4.14 for change scores and 17.64 for final values	MD 2.88 higher (0.46 lower to 6.22 higher)
Wechsler Memory Scale-III Spatial Span - 6 months - Forward	22 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean Wechsler Memory Scale-III Spatial Span - 6 months - Forward was 6.82	MD 0.73 lower (1.86 lower to 0.4 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 18 weeks - 6 months	Risk difference with Attention/working memory: computer-aided training of attention/working memory
follow up: 6 months					
Wechsler Memory Scale-III Spatial Span - 6 months - Backward follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Memory Scale-III Spatial Span - 6 months - Backward was 6.45	MD 0.27 lower (1.6 lower to 1.06 higher)
Wechsler Adult Intelligence Scale-III - 6 months - Arithmetic follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Adult Intelligence Scale-III - 6 months - Arithmetic was 11.0	MD 1 higher (1.35 lower to 3.35 higher)
Wechsler Adult Intelligence Scale-III - 6 months - Letter-Number Sequencing follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Adult Intelligence Scale-III - 6 months - Letter-Number Sequencing was 7.82	MD 0.63 higher (1.84 lower to 3.1 higher)
Wechsler Adult Intelligence Scale-III - 6 months - Digit span forward follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Adult Intelligence Scale-III - 6 months - Digit span forward was 8.82	MD 0.36 higher (1.33 lower to 2.05 higher)
Wechsler Adult Intelligence Scale-III - 6 months - Digit span backward follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Wechsler Adult Intelligence Scale-III - 6 months - Digit span backward was 5.36	MD 0.73 higher (0.61 lower to 2.07 higher)
Delis-Kaplan Executive Function System (DKEFS) - Color-Word Interference - 6 months follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean delis-Kaplan Executive Function System (DKEFS) - Color-Word Interference - 6 months was 29.73	MD 1.46 lower (8.37 lower to 5.45 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 18 weeks - 6 months	Risk difference with Attention/working memory: computer-aided training of attention/working memory
N-back test errors - 18 weeks - 0-back errors follow up: 18 weeks	31 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean n-back test errors - 18 weeks - 0-back errors was 2.64	MD 0.11 lower (2.28 lower to 2.06 higher)
N-back test errors - 18 weeks - 1-back errors follow up: 18 weeks	31 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean n-back test errors - 18 weeks - 1-back errors was 2.14	MD 0.92 higher (0.95 lower to 2.79 higher)
N-back test errors - 18 weeks - 2-back errors follow up: 18 weeks	31 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean n-back test errors - 18 weeks - 2-back errors was 5.29	MD 0.53 lower (3.92 lower to 2.86 higher)
Multiple Sclerosis Neuropsychological Screening Questionnaire (MSNQ; scale usually 0-60) - 18 weeks - 6 months Scale from: 0 to 60 follow up: 18 weeks - 6 months	53 (2 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean multiple Sclerosis Neuropsychological Screening Questionnaire (MSNQ; scale usually 0-60) - 18 weeks - 6 months ranged from 28.93-29.91	MD 0.47 lower (7.86 lower to 6.91 higher)
Cognitive Failure Questionnaire (CFQ; scale usually 0-100) - 6 months Scale from: 0 to 100 follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cognitive Failure Questionnaire (CFQ; scale usually 0-100) - 6 months was 36.45	MD 6.81 higher (11.97 lower to 25.59 higher)
Dysexecutive questionnaire (DEX; scale usually 0-80) - 6 months Scale from: 0 to 80	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean dysexecutive questionnaire (DEX; scale usually 0-80) - 6 months was 20.55	MD 2.54 higher (9.71 lower to 14.79 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 18 weeks - 6 months	Risk difference with Attention/working memory: computer-aided training of attention/working memory
follow up: 6 months					
Perceived Deficits Questionnaire (PDQ; scale usually 0-80) - 6 months Scale from: 0 to 80 follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean perceived Deficits Questionnaire (PDQ; scale usually 0-80) - 6 months was 30.73	MD 7.09 higher (9.96 lower to 24.14 higher)
SF-36 quality of life (unclear which subscale or composite of physical and mental health) - 6 months Scale from: 0 to 100 follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 quality of life (unclear which subscale or composite of physical and mental health) - 6 months was 44.55	MD 11.9 higher (4.06 lower to 27.86 higher)
EQ-5D (scale 0-1) - 18 weeks Scale from: 0 to 1 follow up: 18 weeks	31 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean EQ-5D (scale 0-1) - 18 weeks was 0.57	MD 0.04 lower (0.21 lower to 0.13 higher)
Functional Assessment of MS (FAMS; scale usually 0-176) - 18 weeks Scale from: 0 to 176 follow up: 18 weeks	31 (1 RCT)	⊕◯◯◯ VERY LOW a,	-	The mean functional Assessment of MS (FAMS; scale usually 0-176) - 18 weeks was 101.0	MD 12 lower (34.47 lower to 10.47 higher)
Beck Depression Inventory-Fast Screen (scale usually 0-21) - 6 months Scale from: 0 to 63	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean beck Depression Inventory-Fast Screen (scale usually 0-21) - 6 months was 2.73	MD 0.09 lower (2.93 lower to 2.75 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 18 weeks - 6 months	Risk difference with Attention/working memory: computer-aided training of attention/working memory
follow up: 6 months					
Hospital Anxiety and Depression Scale (HADS; scale usually 0-21) - 18 weeks - 6 months - Anxiety Scale from: 0 to 21 follow up: 18 weeks - 6 months	53 (2 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean hospital Anxiety and Depression Scale (HADS; scale usually 0-21) - 18 weeks - 6 months - Anxiety ranged from 6.09-6.86	MD 1.39 higher (1.14 lower to 3.91 higher)
Hospital Anxiety and Depression Scale (HADS; scale usually 0-21) - 18 weeks - 6 months - Depression Scale from: 0 to 21 follow up: 18 weeks - 6 months	53 (2 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean hospital Anxiety and Depression Scale (HADS; scale usually 0-21) - 18 weeks - 6 months - Depression ranged from 4.91-8.79	MD 0.25 higher (1.72 lower to 2.21 higher)
Fatigue Severity Scale (FSS; scale likely 1-7) - 6 months Scale from: 1 to 7 follow up: 6 months	22 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (FSS; scale likely 1-7) - 6 months was 5.18	MD 0.29 lower (1.82 lower to 1.24 higher)
Fatigue Severity Scale (9-63 scale) - 18 weeks Scale from: 9 to 63 follow up: 18 weeks	31 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63 scale) - 18 weeks was 49.29	MD 3.24 higher (6.54 lower to 13.02 higher)
Patient Activation Measure-13	31 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean patient Activation Measure-13 (PAM-13;	MD 3.31 lower (14.44 lower to 7.82 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 18 weeks - 6 months	Risk difference with Attention/working memory: computer-aided training of attention/working memory
(PAM-13; measures engagement in health; scale usually 0-100) - 18 weeks Scale from: 0 to 100 follow up: 18 weeks				measures engagement in health; scale usually 0-100) - 18 weeks was 62.1	
Unidimensional Self-Efficacy scale for MS (USE-MS; scale unclear) - 18 weeks follow up: 18 weeks	31 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean unidimensional Self-Efficacy scale for MS (USE-MS; scale unclear) - 18 weeks was 19.31	MD 2.84 lower (8.14 lower to 2.46 higher)

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

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

c. Downgraded by 1 increment as general Fatigue Severity Scale reported rather than specifically cognitive fatigue

Attention/working memory: high-intensity working memory training vs. distributed working memory training

Table 25: High-intensity working memory training vs. distributed working memory training, 4-8 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with distributed working memory training, 4-8 weeks	Risk difference with Attention/working memory: high-intensity working memory training
SDMT - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 4-8 weeks was 62.22	MD 8.35 lower (19.45 lower to 2.75 higher)
PASAT - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT - 4-8 weeks was 53.61	MD 3.2 lower (8.13 lower to 1.73 higher)
Corsi blocks - 4-8 weeks - Backward	30 (1 RCT)	⊕◯◯◯ VERY	-	The mean corsi blocks - 4-8 weeks - Backward was 9.33	MD 0.46 lower (1.76 lower to 0.84 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with distributed working memory training, 4-8 weeks	Risk difference with Attention/working memory: high-intensity working memory training
follow up: 4-8 weeks		LOW a,b,c			
Corsi blocks - 4-8 weeks - Forward follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean corsi blocks - 4-8 weeks - Forward was 7.73	MD 0.53 lower (1.94 lower to 0.88 higher)
Digit Span - 4-8 weeks - Backward follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean digit Span - 4-8 weeks - Backward was 7.41	MD 0.46 higher (1.03 lower to 1.95 higher)
Digit Span - 4-8 weeks - Forward follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean digit Span - 4-8 weeks - Forward was 9.33	MD 0.46 lower (1.76 lower to 0.84 higher)
2-back number correct - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back number correct - 4-8 weeks was 57.33	MD 2.26 lower (5.15 lower to 0.63 higher)
2-back omissions - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 2-back omissions - 4-8 weeks was 0.06	MD 0.34 higher (0.05 lower to 0.73 higher)
2-back reaction time - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back reaction time - 4-8 weeks was 666.4	MD 101.26 higher (67.23 lower to 269.75 higher)
Faces Symbol Test - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean faces Symbol Test - 4-8 weeks was 2.13	MD 0.41 higher (0.11 lower to 0.93 higher)
Functional Assessment of MS (FAMS; scale usually 0-176) - 4-8 weeks Scale from: 0 to 176 follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean functional Assessment of MS (FAMS; scale usually 0-176) - 4-8 weeks was 134.2	MD 15.59 lower (35.23 lower to 4.05 higher)
Allgemeine Depressionsskala (scale unclear) - 4-8	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean allgemeine Depressionsskala	MD 1.95 higher (5.25 lower to 9.15 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with distributed working memory training, 4-8 weeks	Risk difference with Attention/working memory: high-intensity working memory training
weeks follow up: 4-8 weeks				(scale unclear) - 4-8 weeks was 10.26	
Fatigue Scale for Motor and Cognitive Functions (FSMC; scale usually 20-100) - 4-8 weeks Scale from: 20 to 100 follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean fatigue Scale for Motor and Cognitive Functions (FSMC; scale usually 20-100) - 4-8 weeks was 58.0	MD 3.73 higher (11.04 lower to 18.5 higher)
Modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 4-8 weeks Scale from: 0 to 84 follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,e	-	The mean modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 4-8 weeks was 34.23	MD 0.1 lower (13.37 lower to 13.17 higher)

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

b. Downgraded by 1 increment as the majority of the evidence comes from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

d. Downgraded by 1 increment as reported general FSMC score and not specifically the cognitive subdomain

e. Downgraded by 1 increment as reported general MFIS score and not specifically the cognitive subdomain

Attention/working memory: high-intensity working memory training vs. control (no training)

Table 26: High-intensity working memory training vs. control (no training), 4 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 4 weeks	Risk difference with Attention/working memory: high-intensity working memory training
SDMT - 4 weeks follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 4 weeks was 58.67	MD 4.8 lower (17.06 lower to 7.46 higher)
PASAT - 4 weeks follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT - 4 weeks was 48.53	MD 1.88 higher (5.02 lower to 8.78 higher)
Corsi blocks - 4 weeks - Backward follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean corsi blocks - 4 weeks - Backward was 8.13	MD 0.74 higher (0.62 lower to 2.1 higher)
Corsi blocks - 4 weeks - Forward follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean corsi blocks - 4 weeks - Forward was 8.8	MD 1.6 lower (2.88 lower to 0.32 lower)
Digit Span - 4 weeks - Backward follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean digit Span - 4 weeks - Backward was 6.4	MD 1.47 higher (0.1 lower to 3.04 higher)
Digit Span - 4 weeks - Forward follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean digit Span - 4 weeks - Forward was 6.73	MD 2.14 higher (0.83 higher to 3.45 higher)
2-back number correct - 4 weeks follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back number correct - 4 weeks was 55.27	MD 0.2 lower (3.04 lower to 2.64 higher)
2-back omissions - 4 weeks follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back omissions - 4 weeks was 0.53	MD 0.13 lower (0.81 lower to 0.55 higher)
2-back reaction time - 4 weeks follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back reaction time - 4 weeks was 762.07	MD 5.59 higher (184.07 lower to 195.25 higher)
Faces Symbol Test - 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY	-	The mean faces Symbol Test - 4 weeks was 2.49	MD 0.05 higher (0.54 lower to 0.64 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 4 weeks	Risk difference with Attention/working memory: high-intensity working memory training
follow up: 4 weeks		LOW a,b,c			
Functional Assessment of MS (FAMS; scale usually 0-176) - 4 weeks Scale from: 0 to 176 follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean functional Assessment of MS (FAMS; scale usually 0-176) - 4 weeks was 122.93	MD 4.32 lower (28.25 lower to 19.61 higher)
Allgemeine Depressionsskala (scale unclear) - 4 weeks follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean allgemeine Depressionsskala (scale unclear) - 4 weeks was 12.86	MD 0.65 lower (8.96 lower to 7.66 higher)
Fatigue Scale for Motor and Cognitive Functions (FSMC; scale usually 20-100) - 4 weeks Scale from: 20 to 100 follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean fatigue Scale for Motor and Cognitive Functions (FSMC; scale usually 20-100) - 4 weeks was 65.06	MD 3.33 lower (16.16 lower to 9.5 higher)
Modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 4 weeks Scale from: 0 to 84 follow up: 4 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,e	-	The mean modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 4 weeks was 37.53	MD 3.4 lower (12.92 lower to 6.12 higher)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies that reported the outcome at a time-point <3-month minimum specified in the protocol

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

d. Downgraded by 1 increment as general FSMC score reported rather than the cognitive subdomain

e. Downgraded by 1 increment as general MFIS score reported rather than the cognitive subdomain

Attention/working memory: distributed working memory training vs. control (no training)

Table 27: Distributed working memory training vs. control (no training), 4-8 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 4-8 weeks	Risk difference with Attention/working memory: distributed working memory training
SDMT - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT - 4-8 weeks was 58.67	MD 3.55 higher (9.17 lower to 16.27 higher)
PASAT - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT - 4-8 weeks was 48.53	MD 5.08 higher (1.23 lower to 11.39 higher)
Corsi blocks - 4-8 weeks - Backward follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean corsi blocks - 4-8 weeks - Backward was 8.13	MD 1.2 higher (0 to 2.4 higher)
Corsi blocks - 4-8 weeks - Forward follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean corsi blocks - 4-8 weeks - Forward was 8.8	MD 0.4 lower (1.39 lower to 0.59 higher)
Digit Span - 4-8 weeks - Backward follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean digit Span - 4-8 weeks - Backward was 6.4	MD 1.01 higher (0.32 lower to 2.34 higher)
Digit Span - 4-8 weeks - Forward follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean digit Span - 4-8 weeks - Forward was 6.73	MD 1 higher (0.28 lower to 2.28 higher)
2-back number correct - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back number correct - 4-8 weeks was 55.27	MD 2.06 higher (0.8 lower to 4.92 higher)
2-back omissions - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back omissions - 4-8 weeks was 0.53	MD 0.47 lower (1.05 lower to 0.11 higher)
2-back reaction time - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back reaction time - 4-8 weeks was 762.07	MD 95.67 lower (258.27 lower to 66.93 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 4-8 weeks	Risk difference with Attention/working memory: distributed working memory training
Faces Symbol Test - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,	-	The mean faces Symbol Test - 4-8 weeks was 2.49	MD 0.36 lower (0.95 lower to 0.23 higher)
Functional Assessment of MS (FAMS; scale usually 0-176) - 4-8 weeks Scale from: 0 to 176 follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean functional Assessment of MS (FAMS; scale usually 0-176) - 4-8 weeks was 122.93	MD 11.27 higher (7.79 lower to 30.33 higher)
Allgemeine Depressionsskala (scale unclear) - 4-8 weeks follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean allgemeine Depressionsskala (scale unclear) - 4-8 weeks was 12.86	MD 2.6 lower (9.28 lower to 4.08 higher)
Fatigue Scale for Motor and Cognitive Functions (FSMC; scale usually 20-100) - 4-8 weeks Scale from: 20 to 100 follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean fatigue Scale for Motor and Cognitive Functions (FSMC; scale usually 20-100) - 4-8 weeks was 65.06	MD 7.06 lower (21.06 lower to 6.94 higher)
Modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 4-8 weeks Scale from: 0 to 84 follow up: 4-8 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,e	-	The mean modified Fatigue Impact Scale (MFIS; scale usually 0-84) - 4-8 weeks was 37.53	MD 3.3 lower (13.91 lower to 7.31 higher)

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

b. Downgraded by 1 increment as the majority of the evidence comes from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

d. Downgraded by 1 increment as general FSMC score reported rather than the cognitive subdomain

e. Downgraded by 1 increment as general MFIS score reported rather than the cognitive subdomain

Attention/working memory: Attention Processing Training (APT) + multidisciplinary rehabilitation vs. multidisciplinary rehabilitation only

Table 28: Attention Processing Training (APT) + multidisciplinary rehabilitation vs. multidisciplinary rehabilitation only, 3-6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with multidisciplinary rehabilitation only, 3-6 months	Risk difference with Attention/working memory: Attention Processing Training (APT) + multidisciplinary rehabilitation
SDMT - 6 months follow up: 6 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SDMT - 6 months was 19.1	MD 0.8 lower (5.51 lower to 3.91 higher)
PASAT - 6 months - 2 seconds follow up: 6 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PASAT - 6 months - 2 seconds was 17.7	MD 0.8 lower (5.17 lower to 3.57 higher)
PASAT - 6 months - 3 seconds follow up: 6 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PASAT - 6 months - 3 seconds was 17.9	MD 0.6 higher (5.41 lower to 6.61 higher)
Selective Reminding Test (SRT) - 6 months - Long-term storage follow up: 6 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean selective Reminding Test (SRT) - 6 months - Long-term storage was 33.7	MD 0.4 higher (4.35 lower to 5.15 higher)
Selective Reminding Test (SRT) - 6 months - Delayed recall follow up: 6 months	34 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean selective Reminding Test (SRT) - 6 months - Delayed recall was 3.1	MD 1 higher (1.11 lower to 3.11 higher)
10/36 SPART (Spatial Recall Test) - 6 months - Immediate follow up: 6 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 10/36 SPART (Spatial Recall Test) - 6 months - Immediate was 14.7	MD 1 lower (5.64 lower to 3.64 higher)
10/36 SPART (Spatial Recall Test) - 6	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 10/36 SPART (Spatial Recall Test) - 6	MD 0.1 higher (3.23 lower to 3.43 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with multidisciplinary rehabilitation only, 3-6 months	Risk difference with Attention/working memory: Attention Processing Training (APT) + multidisciplinary rehabilitation
months - Delayed follow up: 6 months				months - Delayed was 6.0	
Word List Generation - 6 months follow up: 6 months	34 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean word List Generation - 6 months was 14.1	MD 1.6 lower (7.65 lower to 4.45 higher)
Stroop Test - 6 months follow up: 6 months	34 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean stroop Test - 6 months was 30.5	MD 9 lower (16.21 lower to 1.79 lower)
Montgomery and Asberg Depression Rating Scale (scale possibly 0-60) - 3 months Scale from: 0 to 60 follow up: 3 months	34 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean montgomery and Asberg Depression Rating Scale (scale possibly 0-60) - 3 months was 20.44	MD 3.71 lower (9.46 lower to 2.04 higher)
Barthel Index (measure of activities of daily living; scale 0-100) - 3 months Scale from: 0 to 100 follow up: 3 months	34 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean Barthel Index (measure of activities of daily living; scale 0-100) - 3 months was 44.22	MD 5.22 lower (18.81 lower to 8.37 higher)

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

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

Attention/working memory: reaction time tasks + usual rehabilitation vs. active control (cognitive software with no time component)

Table 29: Reaction time tasks + usual rehabilitation vs. active control (cognitive software with no time component), 2 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control (cognitive software with no time component), 2 weeks	Risk difference with Attention/working memory: reaction time tasks + usual rehab
Alertness - T-value indicating normal results (≥40), 2 weeks follow up: 2 weeks	30 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	OR 3.00 (0.68 to 13.31)	Moderate	
				375 per 1,000	268 more per 1,000 (85 fewer to 514 more)
WEIMuS score indicating fatigue (≥32), 2 weeks follow up: 2 weeks	30 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c,d	OR 0.34 (0.08 to 1.52)	Moderate	
				688 per 1,000	260 fewer per 1,000 (538 fewer to 82 more)
Adherence - completed training sessions of 10 h total, 2 weeks follow up: 2 weeks	30 (1 RCT)	⨁⨁◯◯ LOW c	OR 2.50 (0.55 to 11.41)	Moderate	
				500 per 1,000	214 more per 1,000 (145 fewer to 419 more)

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

b. Downgraded by 1 increment as the majority of the evidence comes from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

d. Downgraded by 1 increment as a general fatigue scale used rather than one specific to cognitive fatigue

Memory: computer-aided training for memory (with or without attention components) vs. control (no training)

Table 30: Computer-aided training for memory (with or without attention components) vs. control (no training), 6-14 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 6-14 weeks	Risk difference with Memory: computer-aided training for memory (with or without attention)
California Verbal Learning Test (CVLT) - 6 weeks - Learning trials follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test (CVLT) - 6 weeks - Learning trials was 11.3	MD 0.99 higher (0.27 lower to 2.25 higher)
California Verbal Learning Test (CVLT) - 6 weeks - Short delay free recall follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test (CVLT) - 6 weeks - Short delay free recall was 11.32	MD 1.86 higher (0.12 lower to 3.84 higher)
California Verbal Learning Test (CVLT) - 6 weeks - Short delay cued recall follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test (CVLT) - 6 weeks - Short delay cued recall was 12.48	MD 0.99 higher (0.85 lower to 2.83 higher)
California Verbal Learning Test (CVLT) - 6 weeks - Long delay free recall follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test (CVLT) - 6 weeks - Long delay free recall was 12.16	MD 1.08 higher (0.95 lower to 3.11 higher)
California Verbal Learning Test (CVLT) - 6 weeks - Long delay cued recall follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test (CVLT) - 6 weeks - Long delay cued recall was 12.96	MD 0.35 higher (1.49 lower to 2.19 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 6-14 weeks	Risk difference with Memory: computer-aided training for memory (with or without attention)
PASAT (MSFC) - 6 weeks follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean PASAT (MSFC) - 6 weeks was 0.01	MD 0.01 higher (0.57 lower to 0.59 higher)
Object alternation reaction time - 6 weeks follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean object alternation reaction time - 6 weeks was 744.0	MD 76 higher (102.65 lower to 254.65 higher)
Object alternation errors - 6 weeks follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean object alternation errors - 6 weeks was 2.16	MD 0.98 lower (2.42 lower to 0.46 higher)
Alertness - 6 weeks - Without cueing follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean alertness - 6 weeks - Without cueing was 233.0	MD 15 higher (29.09 lower to 59.09 higher)
Alertness - 6 weeks - With cueing follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean alertness - 6 weeks - With cueing was 223.0	MD 11 higher (32.91 lower to 54.91 higher)
Spatial span (Corsi) % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean spatial span (Corsi) % change - ~14 weeks was -1.1	MD 26.5 higher (14.88 higher to 38.12 higher)
Paired associates % change - ~14 weeks - Easy follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean paired associates % change - ~14 weeks - Easy was 1.1	MD 9.2 higher (0.87 lower to 19.27 higher)
Paired associates % change - ~14 weeks - Hard follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean paired associates % change - ~14 weeks - Hard was 2.21	MD 56.79 higher (9.25 higher to 104.33 higher)
Short story recall % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean short story recall % change - ~14 weeks was 22.9	MD 14.7 higher (8.16 lower to 37.56 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 6-14 weeks	Risk difference with Memory: computer-aided training for memory (with or without attention)
Visual reproduction % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean visual reproduction % change - ~14 weeks was -0.7	MD 49.8 higher (26.52 higher to 73.08 higher)
Luria-Nebraska neuropsychological battery memory scale % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean Luria-Nebraska neuropsychological battery memory scale % change - ~14 weeks was -0.6	MD 3.1 higher (1.47 higher to 4.73 higher)
Signal detection hits % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean signal detection hits % change - ~14 weeks was 6.4	MD 2.1 higher (8.08 lower to 12.28 higher)
Signal detection reaction time % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean signal detection reaction time % change - ~14 weeks was 1.7	MD 7.7 higher (1.5 higher to 13.9 higher)
Recognition memory % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean recognition memory % change - ~14 weeks was -0.4	MD 5.9 higher (1 higher to 10.8 higher)
Digit Span % change - ~14 weeks - Forward follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean digit Span % change - ~14 weeks - Forward was -6.35	MD 24.15 higher (10.5 higher to 37.8 higher)
Digit Span % change - ~14 weeks - Backward follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean digit Span % change - ~14 weeks - Backward was -5.75	MD 16.55 higher (1.3 lower to 34.4 higher)
SF-12 quality of life (scale usually 0-100) - 6 weeks - Bodily score	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SF-12 quality of life (scale usually 0-100) - 6 weeks - Bodily score was 41.1	MD 2.5 lower (9.91 lower to 4.91 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 6-14 weeks	Risk difference with Memory: computer-aided training for memory (with or without attention)
Scale from: 0 to 100 follow up: 6 weeks					
SF-12 quality of life (scale usually 0-100) - 6 weeks - Mental score Scale from: 0 to 100 follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SF-12 quality of life (scale usually 0-100) - 6 weeks - Mental score was 47.8	MD 0.7 higher (6.68 lower to 8.08 higher)
Beck Depression Inventory (scale usually 0-63) - 6 weeks Scale from: 0 to 63 follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory (scale usually 0-63) - 6 weeks was 11.0	MD 0.7 lower (5.79 lower to 4.39 higher)
Fatigue Severity Scale (FSS; scale usually 9-63) - 6 weeks Scale from: 9 to 63 follow up: 6 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,,b,c,d	-	The mean fatigue Severity Scale (FSS; scale usually 9-63) - 6 weeks was 36.8	MD 0.7 higher (8.42 lower to 9.82 higher)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

d. MIDs used to assess imprecision were ±1.05

d Downgraded by 1 increment as general Fatigue Severity Scale reported rather than cognitive fatigue specifically

Memory: computer-aided RehaCom memory (and attention) training vs. active control

Table 31: Computer-aided RehaCom memory (and attention) training vs. active control, 14-16 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control, 14-16 weeks	Risk difference with Memory: computer-aided RehaCom memory (and attention) training
SDMT - % change from baseline - 16 weeks follow up: 16 weeks	77 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean SDMT - % change from baseline - 16 weeks was 16.9	MD 1.5 lower (15.78 lower to 12.78 higher)
PASAT 2 seconds - % change from baseline - 16 weeks follow up: 16 weeks	77 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean PASAT 2 seconds - % change from baseline - 16 weeks was 38.5	MD 22.1 lower (58.09 lower to 13.89 higher)
Selective reminding test (SRT) - 16 weeks - Consistent long-term retrieval, % change from baseline follow up: 16 weeks	77 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean selective reminding test (SRT) - 16 weeks - Consistent long-term retrieval, % change from baseline was 143.2	MD 16.8 higher (116.77 lower to 150.37 higher)
Selective reminding test (SRT) - 16 weeks - Delayed recall, % change from baseline follow up: 16 weeks	77 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean selective reminding test (SRT) - 16 weeks - Delayed recall, % change from baseline was 44.3	MD 34.5 lower (68.41 lower to 0.59 lower)
10/36 SPART (Spatial Recall Test) - 16 weeks - Immediate, % change from baseline follow up: 16 weeks	77 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean 10/36 SPART (Spatial Recall Test) - 16 weeks - Immediate, % change from baseline was 26.6	MD 9.2 lower (36.48 lower to 18.08 higher)
10/36 SPART (Spatial Recall Test) - 16 weeks - Delayed, %	77 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean 10/36 SPART (Spatial Recall Test) - 16 weeks - Delayed, %	MD 65.1 lower (117.2 lower to 13 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control, 14-16 weeks	Risk difference with Memory: computer-aided RehaCom memory (and attention) training
change from baseline follow up: 16 weeks				change from baseline was 77.1	
Word List Generation - 16 weeks - % change from baseline follow up: 16 weeks	77 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean word List Generation - 16 weeks - % change from baseline was 0.0	MD 31.7 higher (13.7 higher to 49.7 higher)
Spatial span (Corsi) % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean spatial span (Corsi) % change - ~14 weeks was 14.7	MD 10.7 higher (3.13 lower to 24.53 higher)
Digit span % change - ~14 weeks - Forward follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean digit span % change - ~14 weeks - Forward was 0.0	MD 17.8 higher (5.17 higher to 30.43 higher)
Digit span % change - ~14 weeks - Backward follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean digit span % change - ~14 weeks - Backward was -12.5	MD 23.3 higher (7.72 higher to 38.88 higher)
Paired associates % change - ~14 weeks - Easy follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean paired associates % change - ~14 weeks - Easy was 1.9	MD 8.4 higher (1.82 lower to 18.62 higher)
Paired associates % change - ~14 weeks - Hard follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean paired associates % change - ~14 weeks - Hard was 21.6	MD 37.4 higher (5.83 lower to 80.63 higher)
Short story recall % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean short story recall % change - ~14 weeks was 1.55	MD 36.05 higher (18.27 higher to 53.83 higher)
Visual reproduction % change - ~14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean visual reproduction % change - ~14 weeks was 46.9	MD 2.2 higher (37.79 lower to 42.19 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control, 14-16 weeks	Risk difference with Memory: computer-aided RehaCom memory (and attention) training
follow up: 14 weeks					
Luria-Nebraska neuropsychological battery memory scale % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Luria-Nebraska neuropsychological battery memory scale % change - ~14 weeks was 0.4	MD 2.1 higher (0.3 higher to 3.9 higher)
Recognition memory % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean recognition memory % change - ~14 weeks was 6.8	MD 1.3 lower (7.59 lower to 4.99 higher)
Signal detection hits % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean signal detection hits % change - ~14 weeks was 3.8	MD 4.7 higher (4.87 lower to 14.27 higher)
Signal detection reaction time % change - ~14 weeks follow up: 14 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean signal detection reaction time % change - ~14 weeks was 4.5	MD 4.9 higher (1.04 lower to 10.84 higher)
Improvement >20% in at least 5 of Brief Repeatable Battery of Neuropsychological Tests (BRBNT) - 16 weeks follow up: 16 weeks	77 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 0.77 (0.31 to 1.88)	Moderate 541 per 1,000	65 fewer per 1,000 (273 fewer to 148 more)
MSQoL-54 (scale usually 0-100)- 16 weeks - Physical composite Scale from: 0 to 100 follow up: 16 weeks	77 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean mSQoL-54 (scale usually 0-100)- 16 weeks - Physical composite was 22.7	MD 7.1 lower (33.74 lower to 19.54 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control, 14-16 weeks	Risk difference with Memory: computer-aided RehaCom memory (and attention) training
MSQoL-54 (scale usually 0-100)- 16 weeks - Mental health composite Scale from: 0 to 100 follow up: 16 weeks	77 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean mSQoL-54 (scale usually 0-100)- 16 weeks - Mental health composite was 55.9	MD 13.2 lower (72.94 lower to 46.54 higher)
Chicago Mood Depression Inventory (scale unclear) % change - 16 weeks follow up: 16 weeks	77 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean Chicago Mood Depression Inventory (scale unclear) % change - 16 weeks was -5.3	MD 0.3 lower (1.74 lower to 1.14 higher)

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

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

Memory: Story Memory Technique vs. control

Table 32: Story Memory Technique vs. control, 5-11 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 5-11 weeks	Risk difference with Memory: Story Memory Technique
SDMT z-score - processing speed - 5 weeks Scale from: -5 to 5 follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SDMT z-score - processing speed - 5 weeks was -1.0	MD 0.15 lower (0.73 lower to 0.43 higher)
Hopkins Verbal Learning Test-Revised (HVLT-R) - 5-11 weeks (mix of change from baseline and final values)	48 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean hopkins Verbal Learning Test-Revised (HVLT-R) - 5-11 weeks (mix of change from baseline and final values) was 0.57 for change scores and	MD 2.99 higher (0.55 higher to 5.43 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 5-11 weeks	Risk difference with Memory: Story Memory Technique
follow up: 5-11 weeks				21.935 for final values	
% with improvement on HVLT-R - 6 weeks follow up: 6 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	RR 1.60 (0.69 to 3.69)	Moderate	
				357 per 1,000	214 more per 1,000 (111 fewer to 960 more)
California Verbal Learning Test (CVLT) learning slope - 5 weeks follow up: 5 weeks	114 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean California Verbal Learning Test (CVLT) learning slope - 5 weeks ranged from 1.28-1.54	MD 0.27 higher (0.03 lower to 0.57 higher)
CVLT total learning (T-score) - 5 weeks Scale from: 0 to 100 follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean CVLT total learning (T-score) - 5 weeks was 45.24	MD 4.89 higher (0.51 lower to 10.29 higher)
>10% improvement on California Verbal Learning Test (CVLT) - 5 weeks - Short-delay recall follow up: 5 weeks	16 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	OR 9.00 (0.94 to 86.52)	Moderate	
				250 per 1,000	500 more per 1,000 (11 fewer to 716 more)
>10% improvement on California Verbal Learning Test (CVLT) - 5 weeks - Learning slope follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	OR 2.85 (1.19 to 6.85)	Moderate	
				366 per 1,000	256 more per 1,000 (41 more to 432 more)
Objective everyday memory (Rivermead Behavioural Memory Test Story Memory) - 5 weeks follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean objective everyday memory (Rivermead Behavioural Memory Test Story Memory) - 5 weeks was 1.25	MD 0.32 higher (0.05 higher to 0.59 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 5-11 weeks	Risk difference with Memory: Story Memory Technique
Working memory - Letter-Number Sequencing scaled score - 5 weeks follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean working memory - Letter-Number Sequencing scaled score - 5 weeks was 10.49	MD 0.73 higher (0.63 lower to 2.09 higher)
Attention - Digit Span scale score - 5 weeks follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean attention - Digit Span scale score - 5 weeks was 10.27	MD 0.24 higher (0.87 lower to 1.35 higher)
Memory Functioning Questionnaire Spanish version- 5 weeks Scale from: 31 to 217 follow up: 5 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean memory Functioning Questionnaire Spanish version- 5 weeks was 110.8	MD 4.9 lower (12.91 lower to 3.11 higher)
Awareness of Cognitive Deficits Questionnaire (AQ; scale possibly 17-85) - 5 weeks Scale from: 17 to 85 follow up: 5 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean awareness of Cognitive Deficits Questionnaire (AQ; scale possibly 17-85) - 5 weeks was 11.58	MD 4.26 higher (0.41 higher to 8.11 higher)
Functional Assessment of Multiple Sclerosis - General Contentment (FAMS; scale usually 0-28, subjective everyday cognition and emotional functioning) - 5 weeks Scale from: 0 to 28 follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean functional Assessment of Multiple Sclerosis - General Contentment (FAMS; scale usually 0-28, subjective everyday cognition and emotional functioning) - 5 weeks was 15.43	MD 4.45 higher (0.33 lower to 9.23 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 5-11 weeks	Risk difference with Memory: Story Memory Technique
Frontal Systems Behaviour Scale (FrSBe; reported by significant others) - 5 weeks - Apathy (scale unclear, possibly 14-70) Scale from: 14 to 70 follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean frontal Systems Behaviour Scale (FrSBe; reported by significant others) - 5 weeks - Apathy (scale unclear, possibly 14-70) was 30.125	MD 4.02 higher (0.59 lower to 8.63 higher)
Frontal Systems Behaviour Scale (FrSBe; reported by significant others) - 5 weeks - Executive dysfunction (scale unclear, possibly 17-85) Scale from: 17 to 85 follow up: 5 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean frontal Systems Behaviour Scale (FrSBe; reported by significant others) - 5 weeks - Executive dysfunction (scale unclear, possibly 17-85) was 38.109	MD 4.13 higher (0.92 lower to 9.19 higher)
Frontal Systems Behaviour Scale (FrSBe; reported by significant others) - 5 weeks - Disinhibition after illness (scale unclear, possibly 15-75) Scale from: 15 to 75 follow up: 5 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean frontal Systems Behaviour Scale (FrSBe; reported by significant others) - 5 weeks - Disinhibition after illness (scale unclear, possibly 15-75) was 24.82	MD 2.65 higher (0.4 higher to 4.89 higher)
State-Trait Anxiety Inventory (STAI) T-score - 5 weeks -	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean state-Trait Anxiety Inventory (STAI) T-score - 5 weeks -	MD 3.01 lower (9.66 lower to 3.64 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 5-11 weeks	Risk difference with Memory: Story Memory Technique
State score Scale from: 0 to 100 follow up: 5 weeks				State score was 54.24	
State-Trait Anxiety Inventory (STAI) T-score - 5 weeks - Trait score Scale from: 0 to 100 follow up: 5 weeks	86 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean state-Trait Anxiety Inventory (STAI) T-score - 5 weeks - Trait score was 59.06	MD 4.29 lower (10.86 lower to 2.28 higher)
Chicago Multidimensional Depression Inventory T-score - 5 weeks Scale from: 0 to 100 follow up: 5 weeks	86 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean Chicago Multidimensional Depression Inventory T-score - 5 weeks was 56.39	MD 1.34 lower (7.4 lower to 4.72 higher)
Satisfaction with Life Scale (scale usually 5-35) - 5 weeks Scale from: 5 to 35 follow up: 5 weeks	20 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean satisfaction with Life Scale (scale usually 5-35) - 5 weeks was 20.31	MD 3.28 higher (0.16 higher to 6.4 higher)
Patient Competency Rating Scale (PCRS; scale usually 30-150) - 5 weeks - Patient-reported Scale from: 30 to 150 follow up: 5 weeks	20 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean patient Competency Rating Scale (PCRS; scale usually 30-150) - 5 weeks - Patient-reported was 97.665	MD 0.67 higher (6.92 lower to 8.26 higher)
Patient Competency Rating Scale (PCRS; scale usually 30-150) - 5 weeks - Family-reported Scale from: 30	20 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean patient Competency Rating Scale (PCRS; scale usually 30-150) - 5 weeks - Family-reported was 104.79	MD 2.38 lower (5.19 lower to 0.43 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 5-11 weeks	Risk difference with Memory: Story Memory Technique
to 150 follow up: 5 weeks					

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

b. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

Table 33: Story Memory Technique vs. control, 7 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, ~7 months	Risk difference with Memory: Story Memory Technique
SDMT z-score - processing speed - ~7 months Scale from: -5 to 5 follow up: 7 months	78 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SDMT z-score - processing speed - ~7 months was 0.97	MD 1.97 lower (2.58 lower to 1.36 lower)
California Verbal Learning Test (CVLT) learning slope z-score - ~7 months Scale from: -5 to 5 follow up: 7 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) learning slope z-score - ~7 months was 1.0	MD 0.11 higher (0.15 lower to 0.37 higher)
CVLT total learning (T-score) - ~7 months Scale from: 0 to 100 follow up: 7 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean CVLT total learning (T-score) - ~7 months was 35.94	MD 6.85 higher (0.31 lower to 14.01 higher)
Objective everyday memory (Rivermead Behavioural Memory Test Story Memory) - ~7 months -	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean objective everyday memory (Rivermead Behavioural Memory Test Story Memory) - ~7 months - Immediate Profile Score was 1.43	MD 0.09 lower (0.47 lower to 0.29 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, ~7 months	Risk difference with Memory: Story Memory Technique
Immediate Profile Score follow up: 7 months					
Objective everyday memory (Rivermead Behavioural Memory Test Story Memory) - ~7 months - Delayed Profile Score follow up: 7 months	78 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean objective everyday memory (Rivermead Behavioural Memory Test Story Memory) - ~7 months - Delayed Profile Score was 1.48	MD 0.03 higher (0.29 lower to 0.35 higher)
Working memory - Letter-Number Sequencing scaled score - ~7 months follow up: 7 months	78 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean working memory - Letter-Number Sequencing scaled score - ~7 months was 10.37	MD 0 (1.35 lower to 1.35 higher)
Attention - Digit Span scale score - ~7 months follow up: 7 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean attention - Digit Span scale score - ~7 months was 10.4	MD 0.23 higher (1.01 lower to 1.47 higher)
FAMS General Contentment (scale usually 0-28, subjective everyday cognition and emotional functioning) - ~7 months Scale from: 0 to 28 follow up: 7 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean FAMS General Contentment (scale usually 0-28, subjective everyday cognition and emotional functioning) - ~7 months was 14.48	MD 2.69 higher (0.22 lower to 5.6 higher)
FrSBe T-score (reported by significant others) - ~7 months - Apathy Scale from: 0 to 100 follow up: 7 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean frSBe T-score (reported by significant others) - ~7 months - Apathy was 63.88	MD 3.93 higher (6.13 lower to 13.99 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, ~7 months	Risk difference with Memory: Story Memory Technique
FrSBe T-score (reported by significant others) - ~7 months - Executive dysfunction Scale from: 0 to 100 follow up: 7 months	78 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean frSBe T-score (reported by significant others) - ~7 months - Executive dysfunction was 60.75	MD 1.06 lower (8.43 lower to 6.31 higher)
State-Trait Anxiety Inventory (STAI) T-score - ~7 months - State score Scale from: 0 to 100 follow up: 7 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean state-Trait Anxiety Inventory (STAI) T-score - ~7 months - State score was 53.44	MD 3.61 lower (9.53 lower to 2.31 higher)
State-Trait Anxiety Inventory (STAI) T-score - ~7 months - Trait score Scale from: 0 to 100 follow up: 7 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean state-Trait Anxiety Inventory (STAI) T-score - ~7 months - Trait score was 56.22	MD 1.5 lower (8.13 lower to 5.13 higher)
Chicago Multidimensional Depression Inventory T-score - ~7 months Scale from: 0 to 100 follow up: 7 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean Chicago Multidimensional Depression Inventory T-score - ~7 months was 56.48	MD 2.04 lower (8.1 lower to 4.02 higher)

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

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

Memory: group memory programme (various learning techniques) vs. control

Table 34: Group memory programme (various learning techniques) vs. control, 3-6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 3-6 months	Risk difference with Memory: Group memory programme (various learning techniques)
SDMT - 6 months - SDMT - 6 months follow up: 6 months	401 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SDMT - 6 months - SDMT - 6 months was 40.7	MD 1.3 higher (0.6 lower to 3.2 higher)
Selective Reminding Test (SRT) - 6 months - Total follow up: 6 months	402 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean selective Reminding Test (SRT) - 6 months - Total was 43.5	MD 1.6 higher (0.1 higher to 3.1 higher)
Selective Reminding Test (SRT) - 6 months - Delay follow up: 6 months	402 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean selective Reminding Test (SRT) - 6 months - Delay was 6.5	MD 0.2 higher (0.2 lower to 0.6 higher)
10/36 SPART (Spatial Recall Test) - 6 months - Total follow up: 6 months	399 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean 10/36 SPART (Spatial Recall Test) - 6 months - Total was 19.8	MD 0.6 lower (1.5 lower to 0.3 higher)
10/36 SPART (Spatial Recall Test) - 6 months - Delay follow up: 6 months	399 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean 10/36 SPART (Spatial Recall Test) - 6 months - Delay was 6.6	MD 0 (0.4 lower to 0.4 higher)
PASAT - 6 months - Easy follow up: 6 months	395 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean PASAT - 6 months - Easy was 35.7	MD 0 (2.4 lower to 2.4 higher)
PASAT - 6 months - Hard follow up: 6 months	395 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean PASAT - 6 months - Hard was 19.3	MD 0.3 lower (2.9 lower to 2.3 higher)
Trail Making Test (B-A) - 6 months - Trail Making Test (B-A) - 6 months follow up: 6 months	397 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean trail Making Test (B-A) - 6 months - Trail Making Test (B-A) - 6 months was 62.3	MD 0.3 lower (6.8 lower to 6.2 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 3-6 months	Risk difference with Memory: Group memory programme (various learning techniques)
Word fluency - 6 months - Word fluency - 6 months follow up: 6 months	401 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean word fluency - 6 months - Word fluency - 6 months was 27.2	MD 0 (1.3 lower to 1.3 higher)
Working memory (possibly Wechsler Memory Scale–III) - 13 weeks follow up: 13 weeks	60 (1 RCT)	⨁⨁⨁◯ MODERATE b	-	The mean working memory (possibly Wechsler Memory Scale–III) - 13 weeks was 20.65	MD 2.2 higher (0.5 higher to 3.9 higher)
Doors and people (overall age-scaled score) - 6 months - Doors and people (overall age-scaled score) - 6 months follow up: 6 months	402 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean doors and people (overall age-scaled score) - 6 months - Doors and people (overall age-scaled score) - 6 months was 9.1	MD 0.4 higher (0.1 lower to 0.9 higher)
Digit Span Test for attention assessment - 3 months follow up: 3 months	56 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean digit Span Test for attention assessment - 3 months was 11.54	MD 0.46 higher (0.95 lower to 1.87 higher)
Everyday Memory Questionnaire (EMQ; scale 0-140) - 3-6 months - Self-report Scale from: 0 to 140 follow up: 3-6 months	489 (3 RCTs)	⨁⨁◯◯ LOW a	-	The mean everyday Memory Questionnaire (EMQ; scale 0-140) - 3-6 months - Self-report ranged from 25.8-112.57	MD 5.48 lower (8.69 lower to 2.28 lower)
Everyday Memory Questionnaire (EMQ; scale 0-140) - 4-6 months - Carer-report	374 (2 RCTs)	⨁⨁◯◯ LOW a	-	The mean everyday Memory Questionnaire (EMQ; scale 0-140) - 4-6 months - Carer-report ranged from 20.2-38.6	MD 4.02 lower (7.3 lower to 0.75 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 3-6 months	Risk difference with Memory: Group memory programme (various learning techniques)
Scale from: 0 to 140 follow up: 4-6 months					
Everyday Memory Questionnaire (scale 0-175) - 13 weeks Scale from: 0 to 175 follow up: 13 weeks	60 (1 RCT)	⨁⨁⨁◯ MODERATE b	-	The mean everyday Memory Questionnaire (scale 0-175) - 13 weeks was 120.9	MD 0.3 higher (0.52 lower to 1.12 higher)
Prospective and Retrospective Memory Questionnaire (scale 16-80?) - 3 months Scale from: 16 to 80 follow up: 3 months	56 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean prospective and Retrospective Memory Questionnaire (scale 16-80?) - 3 months was 45.57	MD 9.46 lower (14.07 lower to 4.85 lower)
MSIS-29 quality of life (scale 0-100) - 6 months - Psychological Scale from: 0 to 100 follow up: 6 months	404 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean MSIS-29 quality of life (scale 0-100) - 6 months - Psychological was 24.1	MD 0.9 lower (1.7 lower to 0.1 lower)
MSIS-29 quality of life (scale 0-100) - 6 months - Physical Scale from: 0 to 100 follow up: 6 months	402 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean MSIS-29 quality of life (scale 0-100) - 6 months - Physical was 53.0	MD 0.6 lower (2.2 lower to 1 higher)
MSIS-29 quality of life (scale 29-145) - 4 months follow up: 4 months	37 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean MSIS-29 quality of life (scale 29-145) - 4 months was 69.0	MD 8.2 higher (9.92 lower to 26.32 higher)
MSQoL-54 - 3 months - Physical health Scale from: 0	56 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean mSQoL-54 - 3 months - Physical health was 56.25	MD 10.28 higher (2.97 higher to 17.59 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 3-6 months	Risk difference with Memory: Group memory programme (various learning techniques)
to 100 follow up: 3 months					
MSQoL-54 - 3 months - Mental health Scale from: 0 to 100 follow up: 3 months	56 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean mSQoL-54 - 3 months - Mental health was 50.9	MD 16.87 higher (8.9 higher to 24.84 higher)
EQ-5D visual analogue (scale 0-100) - 6 months - EQ-5D visual analogue (scale 0-100) - 6 months Scale from: 0 to 100 follow up: 6 months	411 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean EQ-5D visual analogue (scale 0-100) - 6 months - EQ-5D visual analogue (scale 0-100) - 6 months was 59.9	MD 2.6 higher (0.9 lower to 6.1 higher)
General Health Questionnaire (GHQ-28; scale 0-84) - 4 months Scale from: 0 to 84 follow up: 4 months	37 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean general Health Questionnaire (GHQ-28; scale 0-84) - 4 months was 22.7	MD 1 higher (5.82 lower to 7.82 higher)
GHQ-30 (scale 0-90) - 6 months - GHQ-30 (scale 0-90) - 6 months Scale from: 0 to 90 follow up: 6 months	395 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean GHQ-30 (scale 0-90) - 6 months - GHQ-30 (scale 0-90) - 6 months was 37.8	MD 3.4 lower (5.9 lower to 0.9 lower)
Beck Depression Inventory (scale usually 0-63) - 3 months Scale from: 0 to 63 follow up: 3 months	56 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean beck Depression Inventory (scale usually 0-63) - 3 months was 20.64	MD 9.64 lower (12.94 lower to 6.34 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 3-6 months	Risk difference with Memory: Group memory programme (various learning techniques)
Fatigue Severity Scale (FSS; scale likely 1-7) - 6 months - Fatigue Severity Scale (FSS; scale likely 1-7) - 6 months Scale from: 1 to 7 follow up: 6 months	399 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean fatigue Severity Scale (FSS; scale likely 1-7) - 6 months - Fatigue Severity Scale (FSS; scale likely 1-7) - 6 months was 1.1	MD 0.1 lower (0.3 lower to 0.1 higher)
Carer Strain Index (scale possibly 0-13) - 6 months - Carer Strain Index (scale possibly 0-13) - 6 months Scale from: 0 to 13 follow up: 6 months	327 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean carer Strain Index (scale possibly 0-13) - 6 months - Carer Strain Index (scale possibly 0-13) - 6 months was 6.8	MD 0.9 lower (2.2 lower to 0.4 higher)
Any employment - 6 months - Any employment - 6 months follow up: 6 months	411 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 0.88 (0.55 to 1.41)	305 per 1,000	26 fewer per 1,000 (111 fewer to 77 more)

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

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

c. Downgraded by 1 increment as general Fatigue Severity Scale reported rather than cognitive fatigue specifically

Table 35: Group memory programme (various learning techniques) vs. control, 8-12 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 8-12 months	Risk difference with Memory: Group memory programme (various learning techniques)
SDMT - 12 months - SDMT - 12 months follow up: 12 months	375 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean SDMT - 12 months - SDMT - 12 months was 39.9	MD 0.4 higher (1.7 lower to 2.5 higher)
Selective Reminding Test (SRT) - 12 months - Total follow up: 12 months	376 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean selective Reminding Test (SRT) - 12 months - Total was 46.5	MD 0.6 higher (0.9 lower to 2.1 higher)
Selective Reminding Test (SRT) - 12 months - Delay follow up: 12 months	376 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean selective Reminding Test (SRT) - 12 months - Delay was 7.1	MD 0.4 higher (0.1 higher to 0.7 higher)
10/36 SPART (Spatial Recall Test) - 12 months - Total follow up: 12 months	376 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean 10/36 SPART (Spatial Recall Test) - 12 months - Total was 20.4	MD 0.1 lower (1 lower to 0.8 higher)
10/36 SPART (Spatial Recall Test) - 12 months - Delay follow up: 12 months	376 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean 10/36 SPART (Spatial Recall Test) - 12 months - Delay was 7.0	MD 0.1 lower (0.5 lower to 0.3 higher)
PASAT - 12 months - Easy follow up: 12 months	374 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean PASAT - 12 months - Easy was 36.5	MD 0.6 lower (3.1 lower to 1.9 higher)
PASAT - 12 months - Hard follow up: 12 months	374 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean PASAT - 12 months - Hard was 19.2	MD 1.9 lower (4.8 lower to 1 higher)
Trail Making Test (B-A) - 12 months - Trail Making Test (B-A) - 12 months follow up: 12 months	370 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean trail Making Test (B-A) - 12 months - Trail Making Test (B-A) - 12 months was 63.0	MD 3.2 lower (10 lower to 3.6 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 8-12 months	Risk difference with Memory: Group memory programme (various learning techniques)
Word fluency - 12 months - Word fluency - 12 months follow up: 12 months	375 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean word fluency - 12 months - Word fluency - 12 months was 28.3	MD 0.2 lower (1.5 lower to 1.1 higher)
Doors and people (overall age-scaled score) - 12 months - Doors and people (overall age-scaled score) - 12 months follow up: 12 months	374 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean doors and people (overall age-scaled score) - 12 months - Doors and people (overall age-scaled score) - 12 months was 9.9	MD 0.6 higher (0 to 1.2 higher)
Everyday Memory Questionnaire (EMQ; scale 0-140) - 8-12 months - Self-report Scale from: 0 to 140 follow up: 8-12 months	409 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean everyday Memory Questionnaire (EMQ; scale 0-140) - 8-12 months - Self-report ranged from 26.9-43.1	MD 4.85 lower (8.1 lower to 1.6 lower)
Everyday Memory Questionnaire (EMQ; scale 0-140) - 8-12 months - Carer-report Scale from: 0 to 140 follow up: 8-12 months	336 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean everyday Memory Questionnaire (EMQ; scale 0-140) - 8-12 months - Carer-report ranged from 21.6-38.5	MD 5.13 lower (9.1 lower to 1.16 lower)
General Health Questionnaire (GHQ; scale 0-84) - 8 months Scale from: 0 to 84 follow up: 8 months	33 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean general Health Questionnaire (GHQ; scale 0-84) - 8 months was 25.3	MD 6.9 lower (13.19 lower to 0.61 lower)
MSIS-29 quality of life (scale 29-145) - 8 months	31 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 quality of life (scale 29-145) - 8 months was 74.6	MD 6.3 lower (25.16 lower to 12.56 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 8-12 months	Risk difference with Memory: Group memory programme (various learning techniques)
Scale from: 29 to 145 follow up: 8 months					
MSIS-29 quality of life (scale 0-100) - 12 months - Psychological Scale from: 0 to 100 follow up: 12 months	387 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean MSIS-29 quality of life (scale 0-100) - 12 months - Psychological was 23.4	MD 0.6 lower (1.5 lower to 0.3 higher)
MSIS-29 quality of life (scale 0-100) - 12 months - Physical Scale from: 0 to 100 follow up: 12 months	387 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean MSIS-29 quality of life (scale 0-100) - 12 months - Physical was 52.5	MD 0.1 lower (1.8 lower to 1.6 higher)
EQ-5D visual analogue (scale 0-100) - 12 months - EQ-5D visual analogue scale (0-100) - 12 months Scale from: 0 to 100 follow up: 12 months	382 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean EQ-5D visual analogue (scale 0-100) - 12 months - EQ-5D visual analogue scale (0-100) - 12 months was 59.7	MD 2.6 higher (0.9 lower to 6.1 higher)
GHQ-30 (scale 0-90) - 12 months - GHQ-30 (scale 0-90) - 12 months Scale from: 0 to 90 follow up: 12 months	376 (1 RCT)	⨁⨁◯◯ LOW a	-	The mean GHQ-30 (scale 0-90) - 12 months - GHQ-30 (scale 0-90) - 12 months was 38.3	MD 3.4 lower (6.2 lower to 0.6 lower)
Fatigue Severity Scale (FSS; scale likely 1-7) - 12 months - Fatigue Severity Scale (FSS; scale	378 (1 RCT)	⨁◯◯◯ VERY LOW a,c	-	The mean fatigue Severity Scale (FSS; scale likely 1-7) - 12 months - Fatigue Severity Scale (FSS; scale likely 1-7) - 12 months was 1.2	MD 0.3 lower (0.5 lower to 0.1 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control, 8-12 months	Risk difference with Memory: Group memory programme (various learning techniques)
likely 1-7) - 12 months Scale from: 1 to 7 follow up: 12 months					
Carer Strain Index (scale possibly 0-13) - 12 months - Carer Strain Index (scale possibly 0-13) - 12 months Scale from: 0 to 13 follow up: 12 months	300 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean carer Strain Index (scale possibly 0-13) - 12 months - Carer Strain Index (scale possibly 0-13) - 12 months was 6.2	MD 0.4 lower (1.6 lower to 0.8 higher)
Any employment - 12 months - Any employment - 12 months follow up: 12 months	382 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 0.99 (0.60 to 1.63)	289 per 1,000	2 fewer per 1,000 (93 fewer to 110 more)

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

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

c. Downgraded by 1 increment as general Fatigue Severity Scale reported rather than cognitive fatigue specifically

Memory: behaviour intervention (self-generated learning) vs. control (memory tasks with no self-generated learning taught)

Table 36: Behaviour intervention (self-generated learning) vs. control (memory tasks with no self-generated learning taught), 3-4 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (memory tasks with no self-generated learning taught), 3-4 weeks	Risk difference with Memory: behaviour intervention (self-generated learning)
California Verbal Learning Test-	35 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test-II (CVLT-II) - 3-	MD 1.4 higher (5.62 lower to 8.42 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (memory tasks with no self-generated learning taught), 3-4 weeks	Risk difference with Memory: behaviour intervention (self-generated learning)
II (CVLT-II) - 3-4 weeks - Five trials sum follow up: 3-4 weeks				4 weeks - Five trials sum was 52.1	
California Verbal Learning Test-II (CVLT-II) - 3-4 weeks - Long delay follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean california Verbal Learning Test-II (CVLT-II) - 3-4 weeks - Long delay was 10.5	MD 1.1 higher (1.39 lower to 3.59 higher)
Contextual Memory Test (CMT) - 3-4 weeks - Immediate follow up: 3-4 weeks	35 (1 RCT)	⨁⨁⨁◯ MODERATE a	-	The mean contextual Memory Test (CMT) - 3-4 weeks - Immediate was 11.6	MD 4 higher (2.23 higher to 5.77 higher)
Contextual Memory Test (CMT) - 3-4 weeks - Delay follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean contextual Memory Test (CMT) - 3-4 weeks - Delay was 10.38	MD 3.62 higher (1.43 higher to 5.81 higher)
Memory for Intentions Test (MIST) - 3-4 weeks follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean memory for Intentions Test (MIST) - 3-4 weeks was 48.8	MD 14.6 higher (2.77 lower to 31.97 higher)
Verbal fluency test (total across three letters) - 3-4 weeks follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean verbal fluency test (total across three letters) - 3-4 weeks was 35.7	MD 4.55 higher (4.97 lower to 14.07 higher)
Actual Reality™ Task (AR) - 3-4 weeks - Total errors follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean actual Reality™ Task (AR) - 3-4 weeks - Total errors was 6.8	MD 2.4 lower (5.09 lower to 0.29 higher)
Actual Reality™ Task (AR) - 3-4 weeks - Cognitive score (scale usually 0-20)	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean actual Reality™ Task (AR) - 3-4 weeks - Cognitive score (scale usually 0-20) was 4.1	MD 0.9 lower (2.99 lower to 1.19 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (memory tasks with no self-generated learning taught), 3-4 weeks	Risk difference with Memory: behaviour intervention (self-generated learning)
Scale from: 0 to 20 follow up: 3-4 weeks					
Memory Functioning Questionnaire (MFQ; scale usually 64-448) - 3-4 weeks Scale from: 64 to 448 follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean memory Functioning Questionnaire (MFQ; scale usually 64-448) - 3-4 weeks was 209.4	MD 40.8 higher (6.05 higher to 75.55 higher)
Functional Behavioural Profile (FBP; scale possibly 0-108) - 3-4 weeks Scale from: 0 to 108 follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean functional Behavioural Profile (FBP; scale possibly 0-108) - 3-4 weeks was 88.9	MD 12.6 higher (3.32 higher to 21.88 higher)
Functional Assessment of MS (FAMS; scale usually 0-176) - 3-4 weeks Scale from: 0 to 176 follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean functional Assessment of MS (FAMS; scale usually 0-176) - 3-4 weeks was 98.3	MD 4.2 higher (15.58 lower to 23.98 higher)
Self-awareness of cognitive deficits questionnaire (AQ; scale usually 17-85) - 3-4 weeks Scale from: 17 to 85 follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean self-awareness of cognitive deficits questionnaire (AQ; scale usually 17-85) - 3-4 weeks was 4.8	MD 3.8 higher (0.54 lower to 8.14 higher)
Self-regulation skills interview (self-awareness and strategy use; scale	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean self-regulation skills interview (self-awareness and strategy use; scale	MD 2.5 lower (6.94 lower to 1.94 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (memory tasks with no self-generated learning taught), 3-4 weeks	Risk difference with Memory: behaviour intervention (self-generated learning)
unclear) - 3-4 weeks follow up: 3-4 weeks				unclear) - 3-4 weeks was 31.7	
State-Trait Anxiety Inventory (STAI) - Trait score (scale usually 20-80) - 3-4 weeks Scale from: 20 to 80 follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean state-Trait Anxiety Inventory (STAI) - Trait score (scale usually 20-80) - 3-4 weeks was 41.4	MD 2 lower (9.65 lower to 5.65 higher)
Chicago Multiscale Depression Inventory (CDMI; scale possibly 42-210) - 3-4 weeks Scale from: 42 to 210 follow up: 3-4 weeks	35 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean Chicago Multiscale Depression Inventory (CDMI; scale possibly 42-210) - 3-4 weeks was 63.2	MD 10.1 lower (20.46 lower to 0.26 higher)
Satisfaction with Life Scale (scale usually 5-35) - 3-4 weeks Scale from: 5 to 35 follow up: 3-4 weeks	35 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean satisfaction with Life Scale (scale usually 5-35) - 3-4 weeks was 18.3	MD 0.89 higher (8.29 lower to 10.07 higher)

a. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

Executive function: executive function-specific training vs. control (no training)

Table 37: Executive function-specific training vs. control (no training), 6 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 6 weeks	Risk difference with Executive function: executive function specific training
California Verbal Learning Test (CVLT) - Learning - 6 weeks follow up: 6 weeks	26 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test (CVLT) - Learning - 6 weeks was 11.5	MD 0.6 higher (1.03 lower to 2.23 higher)
Wisconsin Card Sorting Test (WCST) - Number of categories - 6 weeks follow up: 6 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean wisconsin Card Sorting Test (WCST) - Number of categories - 6 weeks was 3.5	MD 1.3 higher (0.48 higher to 2.12 higher)
Wisconsin Card Sorting Test (WCST) - Total errors - 6 weeks follow up: 6 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean wisconsin Card Sorting Test (WCST) - Total errors - 6 weeks was 40.7	MD 21.7 lower (24.82 lower to 18.58 lower)
Preference Shifting trials to criterion - 6 weeks follow up: 6 weeks	26 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean preference Shifting trials to criterion - 6 weeks was 40.8	MD 7.8 lower (23.86 lower to 8.26 higher)
Preference Shifting reaction time - 6 weeks follow up: 6 weeks	26 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean preference Shifting reaction time - 6 weeks was 697.0	MD 59 lower (190.39 lower to 72.39 higher)
Response Shifting trials to criterion - 6 weeks follow up: 6 weeks	26 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean response Shifting trials to criterion - 6 weeks was 39.8	MD 9.5 higher (9.41 lower to 28.41 higher)
Response Shifting reaction time - 6 weeks follow up: 6 weeks	26 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean response Shifting reaction time - 6 weeks was 727.0	MD 71 lower (227.25 lower to 85.25 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 6 weeks	Risk difference with Executive function: executive function specific training
2-back commissions - 6 weeks follow up: 6 weeks	26 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back commissions - 6 weeks was 3.0	MD 1.2 higher (3.6 lower to 6 higher)
2-back omissions - 6 weeks follow up: 6 weeks	26 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back omissions - 6 weeks was 2.5	MD 1 lower (2.24 lower to 0.24 higher)
2-back reaction time - 6 weeks follow up: 6 weeks	26 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back reaction time - 6 weeks was 604.0	MD 15 lower (143.81 lower to 113.81 higher)

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

b. Downgraded by 1 increment as the majority of the evidence comes from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

Table 38: Executive function-specific training vs. control (no training), 12 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 12 months	Risk difference with Executive function: executive function specific training
California Verbal Learning Test (CVLT) - Learning - 12 months follow up: 12 months	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - Learning - 12 months was 12.3	MD 0.8 lower (3.01 lower to 1.41 higher)
Preference Shifting trials to criterion - 12 months follow up: 12 months	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean preference Shifting trials to criterion - 12 months was 37.8	MD 21.4 higher (5.04 lower to 47.84 higher)
Preference Shifting reaction time - 12 months follow up: 12 months	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean preference Shifting reaction time - 12 months was 600.0	MD 85 higher (113.88 lower to 283.88 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no training), 12 months	Risk difference with Executive function: executive function specific training
Response Shifting trials to criterion - 12 months follow up: 12 months	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean response Shifting trials to criterion - 12 months was 51.9	MD 11.5 lower (44.35 lower to 21.35 higher)
Response Shifting reaction time - 12 months follow up: 12 months	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean response Shifting reaction time - 12 months was 675.0	MD 9 higher (254.11 lower to 272.11 higher)
2-back commissions - 12 months follow up: 12 months	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 2-back commissions - 12 months was 4.3	MD 0.1 lower (4.75 lower to 4.55 higher)
2-back omissions - 12 months follow up: 12 months	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 2-back omissions - 12 months was 1.7	MD 0.1 lower (1.34 lower to 1.14 higher)
2-back reaction time - 12 months follow up: 12 months	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 2-back reaction time - 12 months was 582.0	MD 103 higher (105.78 lower to 311.78 higher)

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

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

c. MIDs used to assess imprecision were ±0.98

Executive function: executive function-specific training vs. active control (responding quickly to visual stimuli)

Table 39: Executive function-specific training vs. active control (responding quickly to visual stimuli), 6 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control (responding quickly to visual stimuli), 6 weeks	Risk difference with Executive function: executive function specific training
California Verbal Learning Test (CVLT) -	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test (CVLT) -	MD 0.6 higher (0.79 lower to 1.99 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control (responding quickly to visual stimuli), 6 weeks	Risk difference with Executive function: executive function specific training
Learning - 6 weeks follow up: 6 weeks				Learning - 6 weeks was 11.5	
Preference Shifting trials to criterion - 6 weeks follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean preference Shifting trials to criterion - 6 weeks was 38.8	MD 5.8 lower (20.7 lower to 9.1 higher)
Preference Shifting reaction time - 6 weeks follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,	-	The mean preference Shifting reaction time - 6 weeks was 598.0	MD 40 higher (87.17 lower to 167.17 higher)
Response Shifting trials to criterion - 6 weeks follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean response Shifting trials to criterion - 6 weeks was 49.9	MD 0.6 lower (20.5 lower to 19.3 higher)
Response Shifting reaction time - 6 weeks follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean response Shifting reaction time - 6 weeks was 676.0	MD 20 lower (177.1 lower to 137.1 higher)
2-back commissions - 6 weeks follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back commissions - 6 weeks was 3.1	MD 1.1 higher (2.83 lower to 5.03 higher)
2-back omissions - 6 weeks follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 2-back omissions - 6 weeks was 1.4	MD 0.1 higher (0.65 lower to 0.85 higher)
2-back reaction time - 6 weeks follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean 2-back reaction time - 6 weeks was 680.0	MD 91 lower (243.91 lower to 61.91 higher)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

Table 40: Executive function-specific training vs. active control (responding quickly to visual stimuli), 12 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with active control (responding quickly to visual stimuli), 12 months	Risk difference with Executive function: executive function specific training
California Verbal Learning Test (CVLT) - Learning - 12 months follow up: 12 months	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test (CVLT) - Learning - 12 months was 11.5	MD 0 (1.85 lower to 1.85 higher)
Preference Shifting trials to criterion - 12 months follow up: 12 months	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean preference Shifting trials to criterion - 12 months was 45.7	MD 13.5 higher (9.26 lower to 36.26 higher)
Preference Shifting reaction time - 12 months follow up: 12 months	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean preference Shifting reaction time - 12 months was 734.0	MD 49 lower (226.08 lower to 128.08 higher)
Response Shifting trials to criterion - 12 months follow up: 12 months	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean response Shifting trials to criterion - 12 months was 49.9	MD 9.5 lower (40.23 lower to 21.23 higher)
Response Shifting reaction time - 12 months follow up: 12 months	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean response Shifting reaction time - 12 months was 747.0	MD 63 lower (306.46 lower to 180.46 higher)
2-back commissions - 12 months follow up: 12 months	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 2-back commissions - 12 months was 2.2	MD 2 higher (2.29 lower to 6.29 higher)
2-back omissions - 12 months follow up: 12 months	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 2-back omissions - 12 months was 3.5	MD 1.9 lower (3.26 lower to 0.54 lower)
2-back reaction time - 12 months follow up: 12 months	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean 2-back reaction time - 12 months was 587.0	MD 98 higher (105.15 lower to 301.15 higher)

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

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

c. MIDs used to assess imprecision were ±0.90

Executive function: goal management programme vs. psychoeducation

Table 41: Goal management programme vs. psychoeducation, 9 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation, 9 weeks	Risk difference with Executive function: Goal management programme
Sustained Attention to Response Task (SART) errors - 9 weeks - Commission errors (% of no-go trials) follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean sustained Attention to Response Task (SART) errors - 9 weeks - Commission errors (% of no-go trials) was 28.3	MD 11.1 higher (8.69 lower to 30.89 higher)
Sustained Attention to Response Task (SART) errors - 9 weeks - Omission errors (% of go trials) follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean sustained Attention to Response Task (SART) errors - 9 weeks - Omission errors (% of go trials) was 4.2	MD 0.2 higher (3.46 lower to 3.86 higher)
SART reaction time across go trials - 9 weeks follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SART reaction time across go trials - 9 weeks was 454.0	MD 6.2 lower (78.81 lower to 66.41 higher)
Test of Everyday Attention (TEA) - 9 weeks - Elevator counting with distraction follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean test of Everyday Attention (TEA) - 9 weeks - Elevator counting with distraction was 5.7	MD 1.9 higher (0.36 lower to 4.16 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation, 9 weeks	Risk difference with Executive function: Goal management programme
Test of Everyday Attention (TEA) - 9 weeks - Visual elevator follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean test of Everyday Attention (TEA) - 9 weeks - Visual elevator was 8.0	MD 0.4 higher (1.38 lower to 2.18 higher)
Test of Everyday Attention (TEA) - 9 weeks - Elevator counting with reversal follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean test of Everyday Attention (TEA) - 9 weeks - Elevator counting with reversal was 4.3	MD 1.2 higher (1.18 lower to 3.58 higher)
Delis-Kaplan Executive Function Scale (DKEFS) Tower Test achievement score - 9 weeks follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean delis-Kaplan Executive Function Scale (DKEFS) Tower Test achievement score - 9 weeks was 8.9	MD 1.3 lower (3.53 lower to 0.93 higher)
Hotel Test - tasks attempted - 9 weeks follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean hotel Test - tasks attempted - 9 weeks was 4.3	MD 0.4 higher (0.19 lower to 0.99 higher)
Hotel Test - deviation from optimal task time - 9 weeks follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean hotel Test - deviation from optimal task time - 9 weeks was 458.3	MD 54.7 lower (162.87 lower to 53.47 higher)
Cognitive Failures Questionnaire (CFQ; scale 0-100) - 9 weeks Scale from: 0 to 100 follow up: 9 weeks	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean cognitive Failures Questionnaire (CFQ; scale 0-100) - 9 weeks was 37.3	MD 5 higher (4.33 lower to 14.33 higher)
Dysexecutive Questionnaire (DEX; scale usually 0-80) -	27 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean dysexecutive Questionnaire (DEX; scale usually 0-80) -	MD 1.8 higher (5.18 lower to 8.78 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation, 9 weeks	Risk difference with Executive function: Goal management programme
9 weeks - Self-reported Scale from: 0 to 80 follow up: 9 weeks				9 weeks - Self-reported was 17.2	
Dysexecutive Questionnaire (DEX; scale usually 0-80) - 9 weeks - Informant-reported Scale from: 0 to 80 follow up: 9 weeks	27 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean dysexecutive Questionnaire (DEX; scale usually 0-80) - 9 weeks - Informant-reported was 22.5	MD 1 lower (14.8 lower to 12.8 higher)
Profile of Mood States (POMS) - Total Mood Disturbance (scale usually 0-200) - 9 weeks Scale from: 0 to 200 follow up: 9 weeks	27 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean profile of Mood States (POMS) - Total Mood Disturbance (scale usually 0-200) - 9 weeks was 23.0	MD 11.9 higher (3.64 lower to 27.44 higher)
Pittsburgh Sleep Quality Index (PSQI) - Global Sleep Disturbance (scale usually 0-21) - 9 weeks Scale from: 0 to 21 follow up: 9 weeks	27 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean Pittsburgh Sleep Quality Index (PSQI) - Global Sleep Disturbance (scale usually 0-21) - 9 weeks was 7.2	MD 0.2 lower (3.48 lower to 3.08 higher)
Goal attainment post-intervention - proportion achieving or exceeding target goal - 9 weeks follow up: 9 weeks	27 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	RR 2.51 (0.82 to 7.72)	Moderate	
				214 per 1,000	323 more per 1,000 (39 fewer to 1,438 more)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

Goal management programme vs. psychoeducation, 8 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation, 8 months	Risk difference with Executive function: Goal management programme
Sustained Attention to Response Task (SART) errors - 8 months - Commission errors (% of no-go trials) follow up: 8 months	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean sustained Attention to Response Task (SART) errors - 8 months - Commission errors (% of no-go trials) was 32.0	MD 10 higher (10.47 lower to 30.47 higher)
Sustained Attention to Response Task (SART) errors - 8 months - Omission errors (% of go trials) follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean sustained Attention to Response Task (SART) errors - 8 months - Omission errors (% of go trials) was 2.9	MD 0.7 higher (2 lower to 3.4 higher)
SART reaction time across go trials - 8 months follow up: 8 months	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SART reaction time across go trials - 8 months was 433.8	MD 10.8 lower (92.78 lower to 71.18 higher)
Test of Everyday Attention (TEA) - 8 months - Elevator counting with distraction follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean test of Everyday Attention (TEA) - 8 months - Elevator counting with distraction was 6.7	MD 0.9 higher (1.21 lower to 3.01 higher)
Test of Everyday Attention (TEA) - 8 months - Visual elevator	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean test of Everyday Attention (TEA) - 8 months - Visual elevator was 9.2	MD 0.5 lower (1.6 lower to 0.6 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation, 8 months	Risk difference with Executive function: Goal management programme
follow up: 8 months					
Test of Everyday Attention (TEA) - 8 months - Elevator counting with reversal follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean test of Everyday Attention (TEA) - 8 months - Elevator counting with reversal was 3.7	MD 1.1 higher (1.35 lower to 3.55 higher)
Delis-Kaplan Executive Function Scale (DKEFS) Tower Test achievement score - 8 months follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean delis-Kaplan Executive Function Scale (DKEFS) Tower Test achievement score - 8 months was 8.3	MD 0.5 lower (2.35 lower to 1.35 higher)
Hotel Test - tasks attempted - 8 months follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean hotel Test - tasks attempted - 8 months was 4.6	MD 0.3 higher (0.13 lower to 0.73 higher)
Hotel Test - deviation from optimal task time - 8 months follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean hotel Test - deviation from optimal task time - 8 months was 406.2	MD 60.9 lower (167.46 lower to 45.66 higher)
Cognitive Failures Questionnaire (CFQ; scale 0-100) - 8 months Scale from: 0 to 100 follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean cognitive Failures Questionnaire (CFQ; scale 0-100) - 8 months was 35.8	MD 5.9 higher (5.54 lower to 17.34 higher)
Dysexecutive Questionnaire (DEX; scale usually 0-80) - 8 months - Self-reported Scale from: 0 to 80	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean dysexecutive Questionnaire (DEX; scale usually 0-80) - 8 months - Self-reported was 16.9	MD 3.2 higher (4.91 lower to 11.31 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with psychoeducation, 8 months	Risk difference with Executive function: Goal management programme
follow up: 8 months					
Dysexecutive Questionnaire (DEX; scale usually 0-80) - 8 months - Informant-reported Scale from: 0 to 80 follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean dysexecutive Questionnaire (DEX; scale usually 0-80) - 8 months - Informant-reported was 18.4	MD 2.7 lower (12.37 lower to 6.97 higher)
Profile of Mood States (POMS) - Total Mood Disturbance (scale usually 0-200) - 8 months Scale from: 0 to 200 follow up: 8 months	23 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean profile of Mood States (POMS) - Total Mood Disturbance (scale usually 0-200) - 8 months was 20.5	MD 15.8 higher (2.6 lower to 34.2 higher)
Pittsburgh Sleep Quality Index (PSQI) - Global Sleep Disturbance (scale usually 0-21) - 8 months Scale from: 0 to 21 follow up: 8 months	23 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean Pittsburgh Sleep Quality Index (PSQI) - Global Sleep Disturbance (scale usually 0-21) - 8 months was 6.6	MD 1.3 higher (1.6 lower to 4.2 higher)

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

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

Improving language: RehaCom verbal fluency training vs. control (no intervention)

Table 42: Goal management programme vs. psychoeducation, 5-10 weeks

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention), 5-10 weeks	Risk difference with Improving language: RehaCom verbal fluency training
California Verbal Learning Test-II (CVLT-II) - 10 weeks follow up: 10 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean california Verbal Learning Test-II (CVLT-II) - 10 weeks was 46.62	MD 7.38 higher (0.77 higher to 13.99 higher)
Controlled Oral Word Association Test (COWAT) - 10 weeks follow up: 10 weeks	53 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean controlled Oral Word Association Test (COWAT) - 10 weeks was 46.62	MD 4.89 higher (0.65 higher to 9.13 higher)
Adherence - optional dropout of treatment - 5 weeks follow up: 5 weeks	60 (1 RCT)	⊕◯◯◯ VERY LOW a,c	RR 0.75 (0.18 to 3.07)	Moderate	
				133 per 1,000	33 fewer per 1,000 (109 fewer to 275 more)

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

b. Downgraded by 1 increment as the majority of the evidence came from studies reporting the outcome at a time-point <3-month minimum specified in the protocol

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

Data not suitable for GRADE analysis

A number of studies covering various comparisons reported either all or some of their results in a form that meant they could not be analysed using GRADE, for example where median values were reported instead of means and standard deviations or where outcomes were only reported for one of the two groups being compared, which was commonly the case for adherence outcomes where the control group was not an active control.

Multi-domain cognitive rehabilitation (computer tasks with no additional teaching strategies) vs. control (no rehabilitation)

Two papers covering a single study[45, 46] reported on the Rehacom intervention versus control at 3 and 9 months. One paper[46] (N=20) reported median and interquartile ranges at 3 months and found statistically significant differences in favour of rehabilitation for the Paced Auditory Serial Addition Test (PASAT) 2", PASAT 3 change score, Wisconsin Care Sorting Test total error, Montomery-Asberg Depression Rating Scale (MADRS). The second paper[45] (N=24) also reported medians and interquartile ranges at 9 months and found statistically significant

differences in favour of rehabilitation for PASAT 3", WCST perseverative errors, Controlled Oral Word Association – Semantic, MADRAS and MS Quality of Life.

Table 43: Mattioli 2010/2012 – outcomes reported as median (IQR) – 3 months

Outcome	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
PASAT 2 seconds – change score	22.00 (17.00 to 27.00)	10	0.00 (0.00 to 12.75)	10	P=0.04	Some concerns
PASAT 3 seconds – change score	36.00 (24.50 to 44.75)	10	7.00 (0.00 to 26.50)	10	P=0.023	Some concerns
Wisconsin Care Sorting Test total error – change score	20.00 (15.25 to 27.50)	10	45.00 (21.50 to 62.75)	10	P=0.037	Some concerns
Wisconsin Care Sorting Test perseverative responses – change score	17.50 (16.00 to 27.50)	10	37.90 (21.50 to 59.50)	10	P=0.08	Some concerns
Wisconsin Care Sorting Test perseverative errors – change score	14.50 (11.25 to 18.75)	10	28.50 (14.25 to 42.50)	10	P=0.051	Some concerns
Controlled Oral Word Association Phonemic (COWA/P) – change score	36.00 (27.50 to 44.50)	10	27.50 (17.75 to 39.75)	10	P=0.236	Some concerns
Controlled Oral Word Association Semantic (COWA/S) – change score	44.50 (27.25 to 47.00)	10	35.50 (29.00 to 42.00)	10	P=0.398	Some concerns
Test of Everyday Attention (TEA), auditory stimulus – change score	724.00 (596.50 to 848.75)	10	580.00 (551.75 to 670.75)	10	P=0.097	Some concerns
TEA, visual stimulus – change score	902.00 (857.25 to 1040.00)	10	1040.00 (829.75 to 1105.50)	10	P=0.771	Some concerns
TEA total omitted – change score	3.00 (2.00 to 4.75)	10	6.00 (3.00 to 6.75)	10	P=0.141	Some concerns
TEA total errors – change score	3.00 (2.00 to 4.75)	10	6.50 (4.00 to 8.00)	10	P=0.104	Some concerns
Selective Reminding Test (SRT), consistent long-term retrieval – change score	19.00 (14.00 to 29.50	10	16.00 (7.00 to 29.00)	10	P=0.559	Some concerns
SRT, delayed recall – change score	6.50 (4.50 to 8.75)	10	5.50 (4.25 to 7.75)	10	P=0.607	Some concerns
10/36 Spatial Recall Test (SPART), long-term retrieval – change score	17.50 (14.50 to 19.50)	10	14.00 (11.25 to 17.50)	10	P=0.204	Some concerns

Outcome	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
10/36 SPART, delayed recall – change score	6.00 (4.25 to 6.75)	10	4.00 (3.25 to 5.75)	10	P=0.353	Some concerns
SDMT – change score	34.50 (31.00 to 44.75)	10	38.00 (28.50 to 45.75)	10	P=0.942	Some concerns
Montgomery-Asberg Depression Rating Scale – change score	4.50 (3.00 to 6.50)	10	14.00 (8.75 to 22.50)	10	P=0.01	Some concerns
MS Quality of Life (MSQoL) – change score	189.00 (165.75 to 208.75)	10	155.00 (142.50 to 184.50)	10	P=0.285	Some concerns

1 Note that 3-month outcomes were extracted from the 2010 paper.

2

3 **Table 44: Mattioli 2010/2012 – outcomes reported as median (IQR) – 9 months**

Outcome	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
PASAT 2 seconds – change score	11.0 (7.0 to 46.0)	13	0.0 (0.0 to 21.0)	11	Not significant	High
PASAT 3 seconds – change score	20.0 (14.0 to 30.0)	13	3.0 (0.0 to 21.0)	11	P<0.05	High
Wisconsin Care Sorting Test categories completed – change score	3.0 (0.0 to 6.0)	13	2.0 (0.0 to 4.0)	11	Not significant	High
Wisconsin Care Sorting Test total error – change score	-40.3 (-54.0 to 4.0)	13	-17.0 (-27.0 to 35.0)	11	Not significant	High
Wisconsin Care Sorting Test perseverative responses – change score	-31.5 (-45.0 to 8.0)	13	-14.0 (-30.0 to 30.0)	11	Not significant	High
Wisconsin Care Sorting Test perseverative errors – change score	-27.0 (-45.0 to 19.0)	13	-15.0 (-20.7 to 21.0)	11	P<0.05	High
Controlled Oral Word Association Phonemic (COWA/P) – change score	8.0 (1.0 to 12.0)	13	2.0 (0.5 to 9.0)	11	Not significant	High
Controlled Oral Word Association Semantic (COWA/S) – change score	8.0 (0.0 to 21.0)	13	0 (-3.5 to 7.0)	11	P<0.05	High
Test of Everyday Attention (TEA),	16.0 (-10.0 to 309.0)	13	-13.0 (-126.5 to 129.0)	11	Not significant	High

Outcome	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
auditory stimulus – change score						
TEA, visual stimulus – change score	98.0 (-119.0 to 395.0)	13	-55.0 (-136.0 to 148.0)	11	Not significant	High
TEA total omitted – change score	-1.0 (-5.0 to 2.0)	13	-1.0 (-4.0 to 3.0)	11	Not significant	High
TEA total errors – change score	-3.0 (-6.0 to 4.0)	13	-3.0 (-4.5 to 1.0)	11	Not significant	High
Selective Reminding Test (SRT), consistent long-term retrieval – change score	0.0 (-3.0 to 16.0)	13	2.0 (0.0 to 34.0)	11	Not significant	High
SRT, delayed recall – change score	2.0 (0.0 to 3.0)	13	1.0 (0.5 to 3.0)	11	Not significant	High
10/36 Spatial Recall Test (SPART), long-term retrieval – change score	0.0 (-1.0 to 4.0)	13	-1.0 (-3.0 to 7.0)	11	Not significant	High
10/36 SPART, delayed recall – change score	1.0 (0.0 to 5.0)	13	-1.0 (-1.5 to 4.0)	11	Not significant	High
SDMT – change score	3.0 (0.0 to 29.0)	13	2.0 (-3.0 to 11.0)	11	Not significant	High
Montgomery-Asberg Depression Rating Scale – change score	-8.0 (-15.0 to 6.0)	13	3.0 (-2.5 to 28.0)	11	P<0.05	High
MS Quality of Life (MSQoL) – change score	33.0 (-17.0 to 104.0)	13	-13.0 (-22.5 to 46.0)	11	P<0.05	High

Multi-domain cognitive rehabilitation tailored to individual (computer tasks with no additional teaching strategies) vs. control (psychoeducation with no cognitive training)

One study[44] reported data for this comparison in the form of median values. The results indicated significant differences between the two groups, with better scores in the intervention arm, for two cognitive tests, which were the 10/36 Spatial Recall Test and the Selective Reminding Test, with P-values for all other outcomes being >0.05.

Table 45: Mattioli 2014 – outcomes reported as median (IQR) – 12 months

Outcome	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
PASAT 3 seconds – change score	6 (2 to 10)	22	4 (0 to 9)	19	P=0.46	Some concerns
PASAT 2 seconds – change score	8 (0 to 10)	22	3 (0 to 8)	19	P=0.42	Some concerns

Outcome	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
10/36 Spatial Recall Test (SPART), long-term retrieval-change score	4 (1 to 7)	22	0 (-1 to 5)	19	P=0.0395	Some concerns
10/36 SPART, delayed recall – change score	1 (0 to 4)	22	0 (-1 to 3)	19	P=0.36	Some concerns
Selective Reminding Test (SRT), long-term storage – change score	10 (4 to 16)	22	6 (0 to 17)	19	P=0.34	Some concerns
SRT, consistent long-term retrieval – change score	7.5 (1 to 16)	22	4 (-4 to 12)	19	P=0.22	Some concerns
SRT, delayed recall – change score	1.5 (1 to 3)	22	0 (-1 to 1)	19	P=0.0076	Some concerns
SDMT – change score	3 (1 to 7)	22	1 (0 to 5)	19	P=0.24	Some concerns
Controlled Oral Words Association, Phonemic (COWA/P) – change score	3 (-1 to 8)	22	1 (-2 to 4)	19	P=0.36	Some concerns
Controlled Oral Words Association, Category (COWAC) – change score	3.5 (2 to 7)	22	2 (-2 to 6)	19	P=0.20	Some concerns
Stroop test – change score	2 (-1 to 7)	22	2 (-1 to 5)	19	P=0.96	Some concerns
MS Quality of Life-54 (0-100 scale)	0 (-12 to 9)	22	1 (-9 to 7)	19	P=0.98	High
Montgomery-Asberg Depression Rating Scale (0-60 sale)	-0.5 (-3 to 1)	22	0 (-4 to 1)	19	P=0.72	High
Modified Fatigue Impact Scale (0-84 scale)	-2.5 (-8 to 0)	22	-1 (-9 to 4)	19	P=0.52	High Indirectness: not specifically cognitive fatigue

General cognitive rehabilitation and psychotherapy vs. control

One study[34] (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)

External compensatory strategies (e.g., lists, diaries and visual mnemonics) vs. control

One study[39] (N=223 numbers varied per outcome and time-point) compared memory and problem solving rehabilitation with two comparators, including two control groups (patients received an assessment with feedback but no intervention or an assessment without feedback and intervention). The results were reported as median and inter-quartile ranges. No statistical significant differences were noted, apart from overall quality of life and satisfaction with quality of life at the 8-month time-point.

Table 46: Lincoln 2002 – external compensatory strategies vs. cognitive screening only with no feedback (comparator 1) and vs. cognitive assessment with feedback but no cognitive intervention (comparator 2) – outcomes reported as median (IQR) – 4 months

Outcome	Intervention results	Intervention group (n)	Comparator 1 results	Comparator group 1 (n)	Comparator 2 results	Comparator group 2 (n)	P-value (across three groups)	Risk of bias
General Health Questionnaire	22 (15-34)	74	21 (13 to 34)	77	21 (13 to 31)	72	0.73	Some concerns
SF-36 Physical health	31.4 (24 to 41)	74	25.6 (21 to 45)	77	27.1 (20 to 47)	72	0.45	Some concerns
SF-36 Mental health	46.9 (39 to 55)	74	44.7 (35 to 55)	77	44.7 (35 to 57)	72	0.55	Some concerns
Overall quality of life	6 (4-8)	74	7 (5-8)	77	6 (5 to 7)	72	0.15	Some concerns
Satisfaction with Quality of Life	4 (4-5)	74	4 (4-5)	77	4 (4-5)	72	0.32	Some concerns
Extended activities of daily living index	45 (25 to 56)	74	48 (37 to 60)	77	43 (37-60)	72	0.23	Some concerns
Everyday memory questionnaire	17 (7 to 35)	74	16.5 (7 to 42)	77	18.5 (5 to 31)	72	0.69	Some concerns
Dysexecutive syndrome questionnaire	20 (13 to 27)	74	17 (9 to 32)	77	16 (7 to 31)	72	0.77	Some concerns
Memory aids questionnaire	10 (5 to 14)	74	10 (7 to 14)	77	11 (7 to 14)	72	0.92	Some concerns
Carer outcome – General Health Questionnaire	22 (13 to 29)	74	22 (14 to 31)	77	24 (16 to 35)	72	0.35	Some concerns
Carer outcome – Everyday memory questionnaire	21 (5 to 34)	74	14 (3 to 35)	77	11.5 (4 to 28)	72	0.90	Some concerns

Outcome	Intervention results	Intervention group (n)	Comparator 1 results	Comparator group 1 (n)	Comparator 2 results	Comparator group 2 (n)	P-value (across three groups)	Risk of bias
Carer outcome – Dysexecutive questionnaire	11.5 (8 to 32)	74	17 (9 to 33)	77	11.5 (7 to 31)	72	0.80	Some concerns

Table 47: Lincoln 2002 – external compensatory strategies vs. cognitive screening only with no feedback (comparator 1) and vs. cognitive assessment with feedback but no cognitive intervention (comparator 2) – outcomes reported as median (IQR) – 8 months

Outcome	Intervention results	Intervention group (n)	Comparator 1 results	Comparator group 1 (n)	Comparator 2 results	Comparator group 2 (n)	P-value (across three groups)	Risk of bias
General Health Questionnaire	21 (15 to 36)	73	18 (13 to 35)	77	18.5 (13 to 35)	71	0.59	Some concerns
SF-36 Physical health	30.7 (24 to 38)	73	30.0 (25 to 38)	77	32.1 (25 to 42)	71	0.55	Some concerns
SF-36 Mental health	46.9 (36 to 54)	73	47.3 (36 to 57)	77	49.3 (33 to 58)	71	0.76	Some concerns
Overall quality of life	6 (4-8)	73	6.5 (5-8)	77	6 (4 to 7)	71	0.04	Some concerns
Satisfaction with Quality of Life	4 (3-5)	73	5 (4-8)	77	4 (3 to 5)	71	0.04	Some concerns
Extended activities of daily living index	42 (27-55)	73	47.5 (37 to 59)	77	44.5 (26 to 61)	71	0.21	Some concerns
Everyday memory questionnaire	15 (6 to 32)	73	14 (7 to 37)	77	15 (5 to 31)	71	0.76	Some concerns
Dysexecutive syndrome questionnaire	18 (10 to 29)	73	16.5 (9 to 32)	77	18 (7 to 31)	71	0.98	Some concerns
Memory aids questionnaire	10 (5 to 14)	73	10 (7 to 14)	77	9 (6 to 15)	71	0.80	Some concerns
Carer outcome – General Health Questionnaire	21 (12 to 32)	73	18 (13 to 30)	77	18.5 (13 to 32)	71	0.59	Some concerns
Carer outcome – Everyday memory questionnaire	13 (3 to 29)	73	10 (3 to 31)	77	10 (3 to 25)	71	0.88	Some concerns

Outcome	Intervention results	Intervention group (n)	Comparator 1 results	Comparator group 1 (n)	Comparator 2 results	Comparator group 2 (n)	P-value (across three groups)	Risk of bias
Carer outcome – Dysexecutive questionnaire	13 (8 to 31)	73	10 (9 to 32)	77	10 (7 to 28)	71	0.72	Some concerns

External compensatory training vs. restitution training (internal ability to code, organise and retrieve information) vs. self-help control group

One study[43] compared three different groups of external compensatory strategies, restitution training and a self-help control group in terms of improving memory. Outcomes were reported at 5 and 7 months. Results for all memory outcomes indicated no significant difference between the three groups, as P-values were all >0.05. A significant difference between the three groups was however observed for the Wimbledon Self-Report scale, which is a measure of mood.

Table 48: Martin 2014 – external compensatory strategies vs. restitution training (internal ability to code and retrieve information) vs self-help control group – outcomes reported as median (SD) – 5 months

Outcome	External comp. strategies results	External comp. Strategies group (n)	Restitution training results	Restitution training group (n)	Self-help control results	Self-help control group (n)	P-value (across three groups)	Risk of bias
Everyday Memory Questionnaire (scale usually 0-140)	43.0 (18.7)	11	36.0 (25.3)	16	38.0 (18.9)	10	0.99	High
Rivermead Behavioural Memory Questionnaire – Extended (scale unclear)	27.0 (7.7)	12	26.0 (7.6)	17	24.5 (9.8)	10	0.35	High
General Health Questionnaire (scale unclear)	2.0 (3.8)	12	4.0 (3.8)	17	3.0 (4.0)	10	0.96	High
Extended Activities of Daily Living (scale usually 0-66)	53.0 (11.9)	12	47.0 (12.9)	16	50.0 (14.1)	9	0.53	High
Wimbledon Self-Report Scale (scale usually 0-30), assesses mood	16.0 (4.1)	10	21.0 (7.6)	15	18.0 (7.9)	7	0.04	High

Table 49: Martin 2014 – external compensatory strategies vs. restitution training (internal ability to code and retrieve information) vs self-help control group – outcomes reported as median (SD) – 7 months

Outcome	External comp. strategies results	External comp. Strategies group (n)	Restitution training results	Restitution training group (n)	Self-help control results	Self-help control group (n)	P-value (across three groups)	Risk of bias
Everyday Memory Questionnaire (scale usually 0-140)	39.0 (19.2)	11	30.0 (25.2)	16	41.0 (20.6)	10	0.78	High
Rivermead Behavioural Memory Questionnaire – Extended (scale unclear)	26.5 (6.1)	12	29.0 (7.9)	17	22.5 (9.3)	10	0.26	High
General Health Questionnaire (scale unclear)	2.5 (3.6)	12	7.0 (4.4)	17	2.0 (3.8)	10	0.30	High
Extended Activities of Daily Living (scale usually 0-66)	54.0 (11.9)	12	48.5 (10.9)	16	55.0 (12.4)	9	0.62	High
Wimbledon Self-Report Scale (scale usually 0-30), assesses mood	16.5 (3.5)	10	22.0 (7.2)	15	20.0 (7.4)	7	0.05	High

Studies with some outcomes in a format that could not be analysed using GRADE but where other outcomes from the same study could be analysed using GRADE

Additional clinical outcomes

The table below contains additional clinical data for some studies, as follows:

- For the comparison of **general cognitive rehabilitation (multi-component and multi-domain) vs. control**, Mantynen 2014[42] reported the mean (SD) Goal Attainment Score (scale unclear) in the intervention group to be 56.2 (8.5), ranging between 41.0 and 75.0.
- For the comparison of **general cognitive rehabilitation – multi-component (different types of strategies combined, e.g. computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) vs. non-specific cognitive rehabilitation programme**, Lamargue 2020[36] reported the correct answers on various domains of the Test of Attentional Performances (Alertness – with and without warning, Visual Scanning – without a target and Divided Attention – Auditory attention, simple task

domains) to be very similar, with only one being significantly different (P<0.05) between the two groups (Divided Attention – Auditory attention domains, simple task version). This data could not be analysed with other outcomes given the SD of at least one of the groups was 0.

- For the comparison of **mindfulness vs. medical treatment and counselling**, Nazaribadie 2020[53] and Nazaribadie 2021[54] reported results for the Rey Complex Figure Test (copy) outcome to be identical in both groups, with the outcome not being analysed due to SDs being 0 in both groups.

Adherence, compliance and satisfaction outcomes

Remaining outcomes in the table below cover adherence, compliance and satisfaction outcomes. On the whole, adherence appeared to be good for many interventions, with many being >70%, though there were some where adherence was <40%. This was difficult to compare across interventions as the definition of adherence or compliance varied between studies and some studies even used more than one definition of adherence. These outcomes were also difficult to interpret as often there was no value in the control group to compare against, particularly in studies where a waitlist control or no intervention was the comparator.

Table 50: Data that could not be analysed using GRADE

Outcome	Study	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
General cognitive rehabilitation (multi-component and multi-domain) vs. control							
Compliance/adherence							
Compliance with intervention	Mantynen 2014[42]	Reported to be 94.1%	58	NA	NA	NA	Some concerns
Adherence – Meeting or exceeding minimum times of practice sessions per week	Stuifbergen 2012[70]	79-82% each week	34	NA	NA	NA	High
Adherence - Meeting or exceeding minimum number of minutes of required practice per week		67-82% each week		NA	NA	NA	High
Goal attainment							
Achievement of personal goals (Goal attainment score, scale unclear) – mean (SD)	Mantynen 2014[42]	56.2 (8.5), range 41.0-75.0	58	NA	NA	NA	Some concerns
General cognitive rehabilitation – multi-component (different types of strategies combined, e.g. computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) vs. non-specific cognitive rehabilitation programme							

Outcome	Study	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
Cognitive measures where standard deviation values of 0 in one group did not allow analysis							
Alertness - Test of Attentional Performances subtest - correct answers, mean (SD)	Lamargue 2020[36]	Without warning, 4 months 40 (0) With warning, 4 months 40 (0) With warning, 4 months 40 (0) With warning, 4 months 40 (0)	18	Without warning, 4 months 40 (0) With warning, 4 months 40 (0) With warning, 4 months 40 (0) With warning, 4 months 40 (0)	17	NR – identical in both groups at both time-points	High
Visual scanning - Test of Attentional Performances subtest - correct answers, mean (SD)		Without a target 50 (0)	18	Without a target 49.8 (0.4)	17	Not significant	High
Divided Attention (auditory attention) - Test of Attentional Performances subtest - correct answers, mean (SD)		Simple task 16 (0)	18	Simple task 15.5 (1)	17	P<0.05	High
General cognitive rehabilitation + outpatient rehabilitation – multi-component and tailored to individual deficits (different types of strategies combined, e.g. computer training skills and teaching of other strategies such as internal/external learning strategies and/or psychoeducation components) vs. outpatient rehabilitation alone							
Satisfaction							
Satisfaction – overall rating of programme in terms of coping with existing cognitive impairments (rated on scale of 1-5 with 5 indicating very good in helping to cope with impairments and	Tesar 2005[72]	3/10 (30%) said programme was average and 7/10 (70%) said programme was above-average	10	NA	NA	NA	Some concerns

Outcome	Study	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
1 indicating not at all helpful)							
Multi-domain skills training (e.g. computer or pen/pencil tasks) without additional strategies (e.g. computer training for multiple domains such as attention, memory and information processing speed – does not include other techniques such as teaching learning techniques or psychoeducation) vs. control							
Compliance/adherence							
Adherence – Defined as attending at least 80% of hospital sessions and completed at least 80% of daily exercises.	Gich 2015[26]	8/22 (36.4%)	22	NA	NA	NA	Some concerns
Adherence – completing 10-week intervention	Messinis 2017[49]	32/32 (100%)	32	NA	NA	NA	Some concerns
Satisfaction – qualitatively reported benefits							
Benefits and recommending to someone else with MS	Messinis 2017[49]	n=30 reported large personal benefits gained, improvement in cognition and would recommend it	32	NA	NA	NA	High
Benefits in terms of everyday life activities		n=28 reported large benefits in terms of everyday life activities	32	NA	NA	NA	High
Multi-domain skills training tailored to individual (e.g., computer or pen/pencil tasks) without additional strategies (e.g., computer training for multiple domains such as attention, memory and information processing speed – does not include other techniques such as teaching learning techniques or psychoeducation) vs. control (no training)							
Compliance/adherence							
Adherence – unprompted adherence to the training in the intervention group (completed entire training	Shatil 2010[67]	22/59 (37.3%)	59	NA	NA	NA	High

Outcome	Study	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
regimen of 24 sessions)							
Brain training apps/games vs. control							
Compliance/adherence							
Treatment adherence % - Number of days in which the patient performed the training/total number of days required, mean (range)	de Giglio 2015[19]	96 (80-100)	18	NA	NA	NA	Some concerns
Compliance with protocol - Definition of compliance unclear	Vilou 2020[74]	12/23 (52.2%)	23	NA	NA	NA	Some concerns
Mindfulness vs. general cognitive rehabilitation (different types of strategies combined, e.g., computer training for skills + teaching other strategies)							
Compliance/adherence							
Adherence – completing all four weekly sessions	Manglani 2020[40]	13/20 (65.0%)	20	15/20 (75%)	20	P=0.48	Some concerns
Mindfulness vs. medical treatment and counselling							
Cognitive measures where standard deviation was 0 for at least one of the groups							
Cognition – Rey Complex Figure Test (Copy)	Nazaribadie 2020[53] and Nazaribadie 2021[54]	Mean (SD): 36 (0)	27	Mean (SD): 36 (0)	26	NR	High
Focus on information processing speed: Cognitive rehabilitation software focused on processing speed vs. control							
Compliance/adherence							
Average proportion of prescribed sessions played	Bove 2021[4]	0.84	20	1.06	20	NR	Some concerns
Focus on attention/working memory: Computer aided training for attention/working memory vs. control (no training or control intervention not related to cognitive training)							
Compliance/adherence							
Adherence - Completed at least 75% of prescribed sessions	Campbell 2016[6]	16/18 (88.9%)	18	NA	NA	NA	High
Adherence - Completed all		12/18 (66.7%)	18	NA	NA	NA	High

Outcome	Study	Intervention results	Intervention group (n)	Comparator results	Comparator group (n)	P-value	Risk of bias
prescribed sessions							
Focus on memory: Group memory programme (various learning techniques, including internal and external aids) vs. control							
Compliance/adherence							
Adherence – Attendance out of 10 sessions, mean (SD)	Carr 2014[7]	7.9 (0.23)	17	NA	NA	NA	Some concerns
Adherence - attended at least 3 sessions Defined as minimum that was considered likely to affect change.	Lincoln 2020[37]	208/245 (84.9%)	245	NA	NA	NA	High
Satisfaction							
Proportion reporting that attending had made a difference to how they coped with memory difficulties. Reported at end of final session	Carr 2014[7]	15/18 (83.3%)	18	NA	NA	NA	High
Focus on executive function: Goal management programme vs. control (psychoeducation)							
Adherence - attendance rate for group sessions	Richard 2013[60]	95.2%	13	94.4%	14	NR	High

1

1.1.7 Economic evidence

1.1.7.1 Included studies

One health economic study comparing cognitive rehabilitation for attention and memory problems plus usual care to usual care alone was included in this review[38]. This is summarised in the health economic evidence profile below (**Table 51**) and the health economic evidence table in the appendices.

No health economic studies were included comparing other comparators listed in the review protocol.

1.1.7.2 Excluded studies

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

See also the health economic study selection flow chart in the appendices.

1.1.8 Summary of included economic evidence

Table 51: Health economic evidence profile: Cognitive rehabilitation plus usual care versus usual care

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Lincoln 2020[38] (UK)	Partially applicable [(a)]	Minor limitations [(b)]	• Within-RCT analysis (CRAMMS RCT, same paper) • Cost-utility analysis (QALYs) • Population: Adults with MS who have cognitive problems • Comparators: 1. usual care 2. cognitive rehabilitation for attention and memory problems (10-week intervention, weekly 1.5-hour group session) + usual care Follow-up: 12 months	Saves £575 [(c)]	0.00 QALYs	Intervention 2 dominates	Probability intervention 2 cost effective (£20/£30K threshold): 84.8%/85.7% Cost per QALY using MSIS-8D to derive QALYs and cost per improvement in MSIS-Psy score were presented as sensitivity analyses. Intervention 2 remains dominant. Across all scenarios, the CIs for both incremental costs and incremental effects span zero, and for the costs, CIs are wide. Given this, caution should be applied in interpreting these results. One-way deterministic sensitivity analyses conducted. Using combinations of upper and lower bound for costs and effects resulted in four different conclusions highlighting uncertainty in

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
							the base-case analysis result.

Abbreviations: CRAMMS = Cognitive Rehabilitation for Attention and Memory in people with Multiple Sclerosis; ICER= incremental cost-effectiveness ratio; MSIS 8D= Multiple Sclerosis Impact Scale 8 dimensions; MSIS Psy= Multiple Sclerosis Impact Scale – psychological subscale; QALY= quality-adjusted life years; RCT= randomised controlled trial

(a) EQ5D-5L mapped to EQ5D-3L but mapping function used was not reported. Does not include all comparators in the review protocol.

(b) Based on a single RCT and so may not reflect full body of clinical evidence. RCT and HE analysis based on follow up of only 12 months and many not capture long term costs.

(c) 2017 UK pounds. Cost components incorporated: CRAMMS intervention (including training, implementation and delivery costs) (£209) and healthcare and personal social service resource use, and medication

1.1.9 Economic model

This area was not prioritised for new cost-effectiveness analysis.

1.1.10 Unit costs

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

Resource	Unit cost per working hour (a)
Hospital-based staff	
Consultant: Medical	£148
Consultant: psychiatric	£146
Clinical psychologist (band 8a)	£72
Hospital physiotherapist (band 7)	£62
Hospital occupational therapist (band 7)	£62
Hospital occupational therapist (band 6)	£52
Clinical Nurse specialist (band 7)	£62
Community-based staff	
Physiotherapy (band 7)	£60
Occupational therapy (band 7)	£60
Occupational therapy (band 6)	£50
Clinical psychologist, Counsellor (specialist) (band 7)	£60
Nurse (GP practice)	£41
Interventions	
Cognitive behavioural therapy (CBT) per session	£106 (b)
Mindfulness-based cognitive therapy – group-based intervention	£91 per hour of direct contact £181 per session, £16 per service user (c)

Source: PSSRU 2020[17]
(a) Qualification costs included (excluding individual and productivity costs)
(b) Taken from PSSRU (2017) and inflated to 2018/19 prices using OECD purchasing power parities (PPPs)[55]
(c) Taken from PSSRU (2013) and inflated to 2018/2019 prices using OECD purchasing power parities (PPPs)[55]

1.1.11 Evidence statements

Effectiveness/Qualitative

For results that could be assessed using GRADE, see summary of evidence in Tables 3-42.

A narrative summary of studies that did not report any outcomes suitable for GRADE is provided alongside tables in the first five sections of the results section entitled 'Data not suitable for GRADE analysis'.

For studies where some outcomes were suitable for GRADE analysis, other outcomes from these studies that were not suitable for GRADE analysis are presented in Table 50. A narrative summary of the data in this table is provided alongside the table. Much of the data in this table involves measure of adherence, compliance or satisfaction. Data for these types of outcomes could often not be analysed using GRADE due to the fact that the outcome only applied to the intervention group (for example, adherence or satisfaction could not be assessed in waitlist control groups as there was no intervention to adhere to or rate satisfaction for). Where it was possible to analyse using GRADE, this data is provided in the main GRADE summary tables for each comparison. This data was of limited use due to its non-comparative nature.

Economic

- One cost-utility analysis found that in people with MS who have cognitive problems, cognitive rehabilitation plus usual care dominated (less costly, equally effective) to usual care alone. This analysis was assessed as partially applicable with minor limitations.

1.1.12 The committee's discussion and interpretation of the evidence

1.1.12.1. The outcomes that matter most

All outcomes listed in the protocol were considered to be equally important for decision-making. The outcomes included in the protocol fell into five key areas: Objective measures of various cognitive functions, such as memory, attention and processing speed; subjective measures, including health-related quality of life scales, patient-reported cognitive outcomes, such as the Perceived Deficits Questionnaire, and self-efficacy outcomes; functional measures, which included medication management, mood, fatigue (preferably cognitive where reported) and activities of daily living; vocational measures, including employment and training, social engagement and relationship satisfaction or impact on carers; and engagement measures, including completion/adherence rates, acceptability and satisfaction.

Most outcomes were reported by at least one study, but the tests used, particularly for objective cognitive measures, varied across studies. This meant that, combined with the fact that interventions and comparisons varied across studies, limited pooling was possible for most comparisons. Of all outcomes, adherence/compliance and satisfaction data were least reported and where they were reported, this was of limited use as studies used a no training control or similar and rates could not be compared between the two groups.

The preferred format of continuous outcomes (as a continuous or dichotomous measure) was not specified in the protocol and any format these outcomes were reported in were therefore extracted. In the vast majority of cases studies reported outcomes in a continuous format. Only one study reported a continuous outcome in both a continuous and dichotomous format, and both versions of the same outcome were extracted (continuous value for Hopkins Verbal Learning Test and the same outcome but the proportion that improved on the baseline score at follow-up). Two studies[15, 25] only reported certain outcomes in a dichotomous format. Caution was noted when interpreting continuous outcomes that had been reported in a dichotomous format as there are various limitations associated with this. Although it can simplify interpretation, most often there is not a strong enough reason for selecting cut-points and dichotomisation of the data can lead to reduced statistical power, an increased risk of a false positive result, underestimation of the variation in outcome between groups and it reduces the data to two endpoints rather than representing the full spectrum of data when reported as a continuous measure. For example, when reported as the number achieving a 20% improvement in outcome compared to baseline, participants with improvements of 21% and 19% would be categorised into event and non-event groups, respectively, suggesting large differences between them when there is actually only a 2% difference between these two participants.

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

1.1.12.2 The quality of the evidence

A total of 63 RCTs were included in this review, all of which were parallel trials as crossover trials were not eligible to be included in the review. This included 15 studies that had already been included in the previous version of this review and an additional 48 studies identified as relevant during the update. Studies covered a wide range of cognitive interventions and different comparators. Pooling was performed where possible but often study characteristics were considered to be too different to pool leading to many comparisons with only a single study included. The majority of studies were very small, with the number of participants included ranging from 16 to 449. Despite the largest study having over 400 participants, very

few studies had a sample size >100 participants. The small size of included studies, combined with the fact that pooling was only possible for a small proportion of outcomes, even for comparisons that had multiple studies included, meant that the majority of reported outcomes across comparisons were based on data from very small populations, often <100 or <50 participants if only a single study reported the outcome, which was common as the cognitive measures reported differed widely across studies. This contributed to a lot of uncertainty in the size and direction of the effect, meaning the committee could not be confident in most of the results that were reported, based on confidence intervals for the absolute effect.

The frequency of the intervention, how they were delivered and who they were delivered by varied widely across studies. For frequency of sessions, some interventions were more intensive than other studies. For example, computerised, remote programmes were often shorter in duration but had a higher number of sessions weekly compared to interventions that involved training of specific techniques, such as specific techniques for memory or psychological techniques such as mindfulness and mental visual imagery, which were usually spread over a longer time-period. These types of intervention were also usually performed in person with the support of healthcare professionals, such as therapists or neuropsychologists. Most interventions lasted between one and four months.

The quality of the evidence as assessed by GRADE ranged from very low to moderate, with the majority being of low or very low quality. Across all outcomes, downgrading was primarily due to imprecision and/or risk of bias. Within risk of bias ratings, the most common reasons contributing to a rating of 'some concerns' or 'high' risk of bias for an outcome were a lack of information about allocation concealment or concerns about randomisation given baseline values for the outcome were different across groups, concerns about the degree of missing data and a lack of blinding for subjective outcome measures. Many outcomes were also downgraded for indirectness if the majority of the evidence for that outcome came from studies where the outcome was reported a time-point less than the 3-month minimum specified in the protocol (for example, at 6 weeks).

A number of outcomes were also downgraded for inconsistency as there was heterogeneity present in the meta-analyses that could not be explained by prespecified subgrouping strategies due to there being three or fewer studies included or most or all studies falling into the same subgroup categories and heterogeneity therefore not being explained by these subgrouping strategies. A random effects analysis was used for these outcomes and downgrading for inconsistency performed as part of the GRADE quality rating. As some studies differed at baseline for the outcome, this was also investigated as a possible reason for heterogeneity, and if this resolved heterogeneity those that were similar at baseline or adjusted for baseline values were separated from studies where there was a larger difference at baseline in the outcome between the two groups. In these cases, downgrading for heterogeneity was not performed but results presented separately. Outcomes where differences in baseline values appear to affect the result are as follows:

- General cognitive rehabilitation (multi-component and multi-domain) vs. control (Table 3):
    - SDMT

- Multi-domain cognitive rehabilitation (pen/paper tasks or computer tasks with no additional teaching strategies) vs. control (Table 11):
    - SDMT
    - Trail Making Test Part B
    - FAS verbal fluency test

- Brain training apps/games (targeting general cognitive function/multiple domains) vs. control (Table 14):
    - Trail Making Test Part A

- Trail Making Test Part B

Based on the limitations described above, including small study sizes and limited pooling, with uncertainty in the direction and/or size of the effect for most outcomes and low to very low quality for most reported outcomes, the committee were not able to make more specific recommendations about which interventions may be appropriate for memory and cognitive problems in MS. They noted the need for future research and made a research recommendation to involve larger trials and for similar outcome sets to be used across studies to enable better use and interpretation of data for future meta-analyses. Instead, the committee focused on edits to existing recommendations on assessment of cognitive and memory problems to ensure this is done where needed as this is important before deciding on intervention for those with these symptoms in MS. The committee also edited other recommendations based on current practice and clinical experience. This is discussed in the subsequent section in more detail.

1.1.12.3 Benefits and harms

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

When presented with the evidence, the committee concluded that across all interventions and comparisons included in the review, the evidence was too limited to be able to inform any recommendations on which interventions would be preferable in people with MS and memory and cognition impairments, which was based on multiple factors described in the subsequent paragraphs.

Although 63 RCTs were included, the ability to pool data was limited within the review because studies differed in the content of their interventions, the comparators used and the tests or scales used to assess outcomes, which was particularly the case for objective measures of cognitive functions. For example, across studies various different tests were used to assess verbal memory, including the California Verbal Learning Test in some studies and Hopkins Verbal Learning Test in others. This was also the case for measures of other cognitive functions, such as information processing and attention.

When considering clinically important benefits or harms across comparisons based on the absolute risk difference, the point estimate for many outcomes suggested a possible benefit of cognitive intervention compared to the control; however, for the vast majority of these, confidence intervals demonstrated uncertainty in this conclusion as the size and/or direction of the effect varied, meaning there could actually be no difference and/or a clinically important harm of the intervention. This meant that the committee were not confident in making conclusions based on these outcomes. There were some outcomes that were worse in the intervention group, but the committee noted that this did not represent a harm but rather a failure to benefit from the intervention with possible deterioration from time to assessment. There were some outcomes within particular comparisons where both the point estimates and confidence intervals were consistent in the conclusion (either a harm or benefit) but given there were very few of these relative to the vast number with uncertainty, the committee were not able to use these few outcomes to make recommendations on which interventions would be appropriate. There were also many outcomes, across the comparisons, where the point estimate suggested no clinically important difference between two groups, which further contributed to the uncertainty.

The committee noted that the variations in interventions across the included studies made it difficult to make generalised conclusions and agreed that the evidence was too limited to make recommendations about the types of interventions that should be given. The fact that current practice with regards to interventions for cognitive impairments in MS was variable meant that the committee could not make consensus-based recommendations on which

interventions would be most appropriate. Based on this and the limitations of the evidence already described, with the aim of improving the certainty in results for future meta-analyses, the committee agreed that a research recommendation for larger trials within this area should be made and that encouraging the use of particular tests or scales for measuring different cognitive functions would improve the ability to pool and interpret data in future meta-analyses. In addition to the research recommendation, the committee agreed edits to the existing recommendations to improve clarity, with any edits made being based on current practice and clinical experience or recommendations that were already included in the guideline but in a different section. The reasoning behind all edits is described in the subsequent paragraphs.

The committee agreed that a recommendation to assess cognition as part of the comprehensive review should be made. This was previously highlighted in another section of the guideline but not explicitly mentioned under the cognition section of the guideline. The committee explained that every person with MS and cognitive symptoms has a different cognitive profile (an individual's pattern of relative strengths and difficulties across several cognitive domains) that should be taken into account when deciding how to proceed. For this reason, access to a cognitive assessment for each person with MS and cognitive symptoms was considered to be important as the cognitive profile needs to be established before decisions about any interventions can be made, based on the person. It was highlighted that this would not consist of screening of every person with MS using a full formal assessment to identify any cognitive impairments, but that people with MS should be asked about cognitive symptoms as part of the comprehensive review. The committee agreed that if there were cognitive symptoms present before, intervention tests would need to be performed to confirm that there are impairments present and which cognitive functions are impaired. The type and complexity of testing required may differ for each person. For some, clinical interview with or without input from carers may be sufficient, while others may require a brief formal neuropsychological assessment (for example, Addenbrooke's or Montreal Cognitive Assessment; MoCA) or a full neuropsychological assessment; this would depend on the needs of each person – for example, a full neuropsychological assessment may be required if assessing the impact of other factors such as fatigue or other disorders on cognition. It was agreed that this was in line with current practice, as the experience of the committee was that cognitive assessment was usually available if the person had been referred for it, though there may be some regional differences. The committee emphasised the importance of assessing and offering interventions for memory and cognition so that people can perform activities of daily living and if applicable, maintain employment.

Edits were also made to the 2014 recommendation that highlights the possible role of anxiety, depression, sleeping and fatigue in cognition in MS. It was highlighted that medication that is being taken, for example drugs being taken for spasticity or anticholinergics for bladder symptoms, can also affect cognition in MS, so this was included as an additional factor in that recommendation. As well as being aware of these factors, the recommendation was edited to state that appropriate management of these issues should be offered if they are present, which incorporates assessment and treatment included in the previous wording of the recommendation.

The committee agreed that the recommendation to consider referring people with MS and persisting cognitive impairments for assessment and management of their cognitive impairments should be retained, but made minor edits based on current practice and clinical experience and to improve clarity. It was agreed that in current practice, many already have access to an occupational therapist who is skilled in cognitive assessment and interventions and that a proportion also have access to neuropsychologists as well. The wording of the original recommendation was changed from referring people to both an occupational therapist and neuropsychologist for this assessment and management to either one or both, as the committee noted that in current practice this depends on who is best suited to each person and that in some but not all cases it may involve a referral to both. It was also made clear in the recommendation that the assessment and management of cognitive impairments

should be tailored to the needs of the individual, as the cognitive profile of each person is likely to differ and therefore need to be managed slightly differently.

General cognitive rehabilitation (multi-component and multi-domain)

Compared to control

Up to 6 months

Depending on the outcome, up to five studies (up to 436 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most specific measures were only reported by one or two studies. Most outcomes were low to very low quality based on GRADE. The cognitive tests reported in studies were extremely varied meaning there were results for lots of different tests to consider and interpret. The results for most analyses of cognitive tests suggested no clinically important difference between the two groups based on point estimates.

Of the remaining analyses for cognitive tests, based on the point estimate, fifteen suggested a possible benefit of cognitive rehabilitation compared to control: Selective Reminding Test Learning Index (1 study; n=101; low quality); PASAT 3-second version (5 studies; n=436; very low quality); Stroop Test time (1 study; n=60; very low quality); Stroop Test 'interference' (1 study; n=42; very low quality); Wisconsin Card Sorting Test time (1 study; n=60; very low quality); four subdomains of the Test of Attentional Performances test, including Alertness-simple reaction time, Alertness-cued reaction time, Divided Attention-acoustic reaction time and Divided Attention-visual reaction time (1 study; n=40; very low quality for all four); t-scores of multiple tests combined for verbal learning, visuospatial memory, visuomotor speed and visual perception, as well as a t-score for the sum of all 11 tests, reported in one study (n=32; very low quality for all five); and a measure of information processing speed from one study with the test used unclear (n=60l very low quality). However, based on confidence intervals there was uncertainty in the size and/or direction of effect for all of these outcomes apart from Stroop Test time and Wisconsin Card Sorting Test time. Results for the following outcomes suggested a worse score in the intervention group compared to control: SDMT (specifically in a study where score was already worse in the intervention group at baseline; n=42; very low quality); Brief Test of Attention (1 study; n=42; very low quality); Calibrated Ideational Fluency Assessment (1 study; n=42; very low quality); MUSIC (unclear what this measures; 1 study; n=40; very low quality); DO80 assessing language (1 study; n=101; low quality); memory span composite based on multiple tests (1 study; n=32; very low quality); and t-scores for Similarities Test and Picture Arrangement test as part of the Wechsler Adult Intelligence Scale (1 study; n=32; very low quality for both). However, for all eight of these outcomes, there was uncertainty based on confidence intervals in terms of the size and or direction of effect.

Of patient-reported outcomes for cognition, four of the five analyses suggested no clinically important difference between groups based on point estimates. One study reporting the Perceived Deficits Questionnaire (n=98; low quality) did however suggest a clinically important benefit of the intervention compared to control, with no uncertainty present as confidence intervals were also consistent with this conclusion.

Similarly, of eight different analyses for quality-of-life scales/subscales, only one suggested a clinically important benefit in the intervention group compared to control (physical subscale of MSIS-29 scale; 1 study; n=98; low quality), with confidence intervals also consistent with this conclusion. Other analyses of quality of life suggested no clinically important difference between groups based on the point estimate, apart from one study of 34 people reporting the WHO Quality of Life and Satisfaction with Life composite in the form of a z-score (very low quality), with the point estimate suggesting a worse score in the intervention group but

uncertainty in the direction and size of the effect being present based on confidence intervals.

Only one study reported fatigue (n=98; low quality), with the results suggesting no clinically important difference between groups. Of six analyses covering psychological outcomes such as anxiety and depression, five suggested a clinically important benefit of the intervention based on the point estimate: Beck Depression Inventory (3 studies; n=164; very low quality); CES-D depression scale (1 study; n=183; very low quality); State and Trait sub scores of the State-Trait Anxiety Inventory (1 study; n=32; very low quality); and the Penn State worry Questionnaire (1 study; n=32; very low quality); however, in all five cases there was uncertainty in the size and direction of effect based on confidence intervals. Data for self-efficacy scales, Multifactorial Memory Questionnaire assessing the use of memory strategies and assessing activities of daily living demonstrated no clinically important difference between the two groups.

>6 months – 1 year

Fewer studies reported data for this later time-point and only one or two studies were included for all analyses (up to 243 people analysed). Most outcomes were low to very low quality based on GRADE. In terms of cognitive tests, the results for four of eleven analyses suggested a clinically important benefit of the intervention group compared to control based on point estimates: PASAT 3-second version (2 studies; n=243; low quality); Stroop Test time (1 study; n=60; very low quality); Wisconsin Card Sorting Test time (1 study; n=60; very low quality); and information processing speed measure (test name unclear; 1 study; n=60; very low quality). However, based on confidence intervals there was uncertainty in the size of effect for all of these outcomes apart from Stroop Test time and Wisconsin Card Sorting Test time. All remaining cognitive test analyses suggested no clinically important difference between groups based on point estimates.

Of patient-reported outcomes for cognition, three of four analyses suggested a clinically important benefit of the intervention compared to control: Perceived Deficits Questionnaire (1 study; n=78; low quality); patient-reported version of MS Neuropsychological Questionnaire (1 study; n=78; low quality); and PROMIS-Applied Cognition Abilities short form 8a (1 study; n=183; very low quality). However, confidence intervals for all indicated uncertainty in the size of the effect. The remaining analysis was the informant-reported version of the MS Neuropsychological Questionnaire and results suggested no clinically important difference between groups.

Of six different analyses for quality-of-life scales/subscales, including two subscales of the MSIS-29 scale and four subscales of the WHO-BREF Quality of Life scale, results for all suggested no clinically important difference between groups based on point estimates

Only one study reported fatigue (n=98; low quality), with the results suggesting no clinically important difference between groups. Of two analyses covering depression outcomes, one suggested a worse outcome in the intervention group based on the point estimate: Beck Depression Inventory (1 study; n=78; low quality); however, there was uncertainty in the size and direction of effect based on confidence intervals. Data for the other depression outcome (CES-D scale) suggested no clinically important difference between groups. Data for self-efficacy scales, Multifactorial Memory Questionnaire assessing the use of memory strategies and assessing activities of daily living demonstrated no clinically important difference between the two groups.

Compared to psychoeducation + information-sharing

A single study of only 30 people was included in this comparison, with data reported at 3 months and all outcomes assessed as low or very low quality based on GRADE. The only outcomes matching the protocol in this study were those from cognitive tests, with no data for other outcomes such as quality of life and fatigue, for example. Of nine analyses, based on point estimates all of them suggested a clinically important benefit of the intervention compared to the psychoeducation + information-sharing control group; however, the confidence intervals were only consistent with this conclusion in four cases: categories completed on the Wisconsin Card Sorting Test; perseverative errors on the Wisconsin Card Sorting Test; BRIEF-A Global Executive Function score; and Visual Memory subscore on the Weschler Memory Scale-Revised. For all other outcomes (Addenbrooke's cognitive examination; non-perseverative errors on the Wisconsin Card Sorting Test; time taken on the Wisconsin Card Sorting Test; General rating score on the Memory Functioning Questionnaire; and Verbal memory subscore on the Weschler Memory Scale-Revised), confidence intervals indicated uncertainty in the direction and/or size of the effect.

Compared to a non-specific cognitive rehabilitation programme

4 months

A single study of only 35 people was included in this comparison, with all outcomes assessed as low or very low quality based on GRADE. In terms of cognitive tests, a lot of data was reported, with multiple sub scores for each test reported. Of the 34 different analyses reported, only four suggested a clinically important benefit of the intervention compared to the non-specific cognitive rehabilitation programme. These were all sub scores on the Test of Attentional Performances test and were reaction time measures rather than correct answers (reaction time for Alertness-without warning, Divided Attention-visual attention simple task, Divided Attention-auditory attention simple task and N-back test). Even for these outcomes, there was uncertainty in the direction and size of effect based on confidence intervals. Results for a further six analyses, including three Test of Attentional Performances sub scores (Visual Scanning reaction time without a target, Visual Scanning reaction time with a target and Divided Attention-visual attention dual task reaction time), Stroop Test time, Trail Making Test Part A time and Rey Complex Figure time, suggested a worse score in the intervention group compared to the non-specific cognitive rehabilitation programme, but there was uncertainty in the size and direction of effect based on confidence intervals.

One patient-reported outcome for cognition was reported (Daily Cognitive Activities Questionnaire), with the point estimate suggesting a clinically important benefit of the intervention compared non-specific cognitive rehabilitation. There was however uncertainty in the size and direction of the effect based on confidence intervals. Quality of life was reported using the SF-36 scale and results indicated no clinically important difference between groups based on point estimates.

Results for cognitive fatigue indicated no clinically important difference between groups based on point estimates. However, results for two of the three depression and anxiety measures (Beck Depression Inventory and State subscale of the State-Trait Anxiety Inventory) suggested worse scores in the intervention group compared to non-specific cognitive rehabilitation, though there was uncertainty in the size and direction of the effect based on confidence intervals.

8 months

A single study of only 35 people was included in this comparison, with all outcomes assessed as low or very low quality based on GRADE. Of the 36 different analyses reported, only four suggested a clinically important benefit of the intervention compared to the non-

specific cognitive rehabilitation programme. These were all sub scores on the Test of Attentional Performances test and were reaction time measures rather than correct answers (reaction time for Alertness-without warning, Divided Attention-visual attention simple task, Divided Attention-visual attention dual task and N-back test). Even for these outcomes, there was uncertainty in the direction and size of effect based on confidence intervals. Results for a further five analyses, including three Test of Attentional Performances sub scores (Alertness reaction time with warning, Visual Scanning reaction time without a target and Visual Scanning reaction time with a target), Trail Making Test Part B time and Rey Complex Figure time, suggested a worse score in the intervention group compared to the non-specific cognitive rehabilitation programme, but there was uncertainty in the size and direction of effect based on confidence intervals.

Additional outcomes such as fatigue, depression and quality of life were not reported at the 8-month time-point.

Tailored to individual with outpatient rehabilitation, compared to outpatient rehabilitation only

A single study of only 19 people was included in this comparison, with data reported at 3 months and all outcomes assessed as low or very low quality based on GRADE. Of the nine different analyses reported for cognitive tests, four suggested a clinically important benefit of intervention compared to outpatient rehabilitation only: incorrect answers on computer-aided card sorting test (low quality); incorrect responses when assessing sustained attention (very low quality); reaction time on assessing sustained attention-variation (low quality); and score on verbal learning test (very low quality). However, for all four of these outcomes there was uncertainty in the size and direction of effect based on confidence intervals. For two analyses (correct answers on computer-aided card sorting test and correct answers when assessing sustained attention; low quality for both) where a worse outcome was identified in the intervention group, uncertainty in the size and direction of effect was also present based on confidence intervals. For the three remaining analyses of cognitive tests, point estimates suggested no clinically important difference between groups.

The study also reported data for depression (Beck Depression Inventory Scale; very low quality) and fatigue (Modified Fatigue Impact Scale; very low quality); based on point estimates, no clinically important difference was identified for depression, but scores were worse in the intervention group for the fatigue outcome, though there was uncertainty in this conclusion based on confidence intervals in terms of the size and direction of effect for fatigue.

Goal Attainment Scaling (GAS) goals (multi-component cognitive rehabilitation tailored to individual)

In addition to usual rehabilitation, compared to usual rehabilitation only

4 months

A single study of 102 people was included in this comparison, with data reported at 4 months and all outcomes assessed as low or very low quality based on GRADE. The study only reported four outcomes, with two being cognitive test measures (BRIEF-A General Executive Composite and BRIEF-A Metacognition index, both reported as T-scores; low quality for both), one measure of quality of life (psychological subscale of MSIS-29; very low quality) and one measure of psychological health (Hopkins Symptom Checklist-25; very low quality). For all four outcomes the point estimate suggested no clinically important difference between the intervention and the usual rehabilitation only group.

7 months

The same study of 102 people described in the previous paragraph reported data at 7 months, with all outcomes assessed as low or very low quality based on GRADE. Of the four outcomes reported, there was still no clinically important difference based on point estimates for three outcomes (BRIEF-A cognitive measures and Hopkins Symptom Checklist-25; low quality for all three measures). At 7 months the point estimate suggested a possible benefit of Goal Attainment Scaling for the psychological subscale of MSIS-29 (very low quality); however, confidence intervals indicated uncertainty in the size and direction of the effect for this outcome.

Multi-domain cognitive rehabilitation (pen/paper or computer tasks with no additional teaching strategies)

Compared to control

Up to 6 months

Depending on the outcome, up to four studies (up to 189 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most specific measures were only reported by one or two studies. All outcomes were low to very low quality based on GRADE. The cognitive tests reported in studies were extremely varied meaning there were results for lots of different tests to consider and interpret. The results for most analyses of cognitive tests suggested a clinically important benefit of the intervention compared to control based on point estimates; however, in only eight cases was there no uncertainty in this conclusion based on confidence intervals: SDMT (4 studies, those where change from baseline was reported or groups similar at baseline; n=189; very low quality); PASAT 2-second version (1 study; n=20; very low quality); Word List Generation Test (1 study; n=41; low quality); Brief Visuospatial Memory Test-Revised (2 studies; n=94; very low quality); Trail Making Test Part A (2 studies; n=77; very low quality); Letter-Number Sequencing as part of the Weschler Adult Intelligence Scale-III (1 study; n=41; very low quality); Boston Naming Test (1 study; n=41; low quality); and Greek Verbal Learning Test (1 study; n=36; low quality).

Other analyses where the point estimate suggested a possible benefit of the intervention but where uncertainty in the size and/or direction of effect was present included PASAT 3-second version, Phonemic cues component of Controlled Oral Word Association Test, various measures on the Wisconsin Card Sorting Test, verbal and non-verbal scores on the Delis-Kaplan Executive Function System card sorting test, 10/36 Spatial Recall Test total and delayed recall scores, total, long-term storage, delayed retrieval and long-term retrieval scores on the Selective Reminding Test, Trail Making Test Part B (where studies were similar at baseline or reported change from baseline scores), Stroop Neuropsychological Screening Test, auditory stimulus score and total errors on the Test of Everyday Attention, all eleven reported sub scores on the Integrated Auditory Visual-2 score, three remaining scores (Digit Span forward and backward, and Block Design) on the Weschler Intelligence Scale-III, Judgement of Line Orientation, FAS verbal fluency test (those similar at baseline or reporting change from baseline scores), and phonemic and semantic scores on a verbal fluency test.

Seven of the cognitive measures reported suggested no clinically important difference between groups based on point estimates (including SDMT, semantic cues and 'Animals' scores on Controlled Oral Word Association Test, long-term retrieval score on 10/36 Spatial Recall Test, Trail Making Test part B in those where the scores were higher at baseline in the intervention group, total omitted stimuli on the Test of Everyday Attention and FAS verbal fluency test results in those where scores were lower in the intervention group at baseline). For one analysis (visual stimulus on Test of Everyday Attention), scores were worse in the

intervention group compared to control; however, there was uncertainty in the size and direction of effect based on confidence intervals.

One study reported a patient-reported outcome for cognition (MS Neuropsychological questionnaire; n=62; very low quality), with results suggesting no clinically important difference between groups based on point estimates.

Three different quality of life scales (across four subscales) were reported in studies. Results from one study (n=62; very low quality) suggested worse scores on the physical and mental composite scores of the MSQoL-54 scale in the intervention group, though there was uncertainty in the size of effect based on confidence intervals. Results from two other studies for an MS quality of life scale (undefined; 1 study; n=20; very low quality) and EQ-5D visual analogue score suggested a possible benefit of intervention based on point estimates, though there was uncertainty in the size and/or direction of effect based on confidence intervals.

Of the depression scales reported, all three suggested a benefit of intervention compared to control, with the confidence intervals being consistent with this conclusion for Beck Depression Inventory-Fast Screen (1 study; n=36; very low quality) and Montgomery-Asberg Depression Scale (1 study; n=20; very low quality). For depression measured using the Hospital Anxiety and Depression scale (2 studies; n=103; very low quality), there was uncertainty in the size and direction of effect. Similarly, two studies reporting the anxiety score of the same scale (n=103; very low quality) suggested a benefit of intervention based on point estimates but there was uncertainty in the size of the effect. Fatigue was reported by two studies either on Modified Fatigue Impact Scale cognitive subscale (1 study; n=36; very low quality) or the Fatigue Severity Scale (1 study; n=62; very low quality). Results for the first measure suggested a clinically important benefit of intervention compared to control, with confidence intervals also consistent with this conclusion, but for Fatigue Severity Scale point estimates suggested no clinically important difference between groups.

9 months

A single study of only 18-19 people reported data for this comparison at 9 months. All outcomes were very low quality based on GRADE. Of the fifteen cognitive measures reported, the point estimates for eleven of these (PASAT 2- and 3-second versions, semantic cues on Controlled Oral Word Association Test, three scores on the Wisconsin Card Sorting Test, all three sub scores reported for the Selective Reminding Test and both sub scores reported for the Test of Everyday Attention) suggested a possible benefit of intervention compared to control; however, for all of these there was uncertainty in the size and direction of effect based on confidence intervals. For the remaining four measures, no clinically important difference was demonstrated based on confidence intervals (SDMT, phonemic cues on Controlled Oral Word Association Test, and both sub scores reported for the 10/36 Spatial Recall Test).

The study also reported one measure of quality of life (MS quality of life, unclear which scale used) and depression using the Montgomery-Asberg Depression Scale. For both of these outcomes the point estimate suggested a benefit of intervention compared to control, but there was uncertainty in the size and/or direction of effect based on confidence intervals.

Tailored to individual (CogniFit – computer tasks, with no additional teaching strategies), compared to control

A single study of only 46 people was included in this comparison, with data reported at 3 months and all outcomes assessed as very low quality based on GRADE. The study reported twelve outcomes, all of which were cognitive test measures. For five of these, the

point estimates suggested a benefit of intervention compared to control (general memory, naming, response time, time estimation and visual working memory), but for all of these there was uncertainty in direction and/or size of effect based on confidence intervals. There was also uncertainty in direction and size of effect for two outcomes where the point estimate suggested a worse score in the intervention group (hand-eye coordination and shifting attention scores). For the remaining outcomes, there was no clinically important difference based on point estimates (divided attention, avoiding distractions, spatial perception, visual scanning and verbal auditory working memory scores).

Brain training apps/games (targeting general cognitive function/multiple domains)

Compared to control

Depending on the outcome, up to four studies (up to 133 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most specific measures were only reported by one study. Most outcomes were low to very low quality based on GRADE. The cognitive tests reported in studies were extremely varied meaning there were results for lots of different tests to consider and interpret. The results for 26 analyses of cognitive tests (including one or more scores on Trail Making Test, Stroop Test, PASAT 2-second and 3-second versions, SDMT, Selective Reminding Test, 10/36 Spatial Recall Test, Brief Visuospatial Memory Test-Revised, Greek Verbal Learning Test, Repeatable Battery for Assessment of Neuropsychological Status, Visual span Corsi Block test and Delis-Kaplan Executive Function System) suggested a clinically important benefit of the intervention compared to control based on point estimates; however, in all cases there was uncertainty in this conclusion based on confidence intervals. For the remaining cognitive measures, no clinically important difference was suggested based on point estimates (PASAT 2 seconds reported in one study, three sub scores of the Selective Reminding Test, Word List Generation Test, two sub scores of the Repeatable Battery for Assessment of Neuropsychological Status, Letter-Number Sequencing as part of the Wechsler Adult Intelligence Scale-IV and Trail 5 score on Delis-Kaplan Executive Function System).

One study reported a patient-reported outcome for cognition (self-reported improvement in cognition; n=135; moderate quality), with results suggesting a clinically important benefit of intervention compared to control but there being uncertainty in the size of effect based on confidence intervals.

A measure of cognitive fatigue was reported by one study (n=34; very low quality), with the point estimate and confidence intervals being consisted with a clinically important benefit in the intervention group compared to control.

One study reported data for quality of life using MSQoL-54 scale (n=34; very low quality). The results for the mental health composite score suggested a benefit of intervention, though there was uncertainty based on confidence intervals, and the results for the physical composite score suggested no clinically important difference between groups.

Mental visual imagery

Compared to control

A single study of only 17 people was included in this comparison, with data reported at 6-8 weeks. The study only reported one outcome relevant to the review, which was assessed as very low quality based on GRADE. The study reported the number of details provided, which is a measure of mental visualisation ability), with the results suggesting no clinically important difference between mental visual imagery intervention and control.

Mindfulness

Compared to control

4 weeks

A single study of only 33 people was included in this comparison at 4 weeks, with all outcomes assessed as very low quality based on GRADE. The study reported nine cognitive measures. For three of these, the point estimates suggested a benefit of intervention compared to control (three sub scores on the Selective Reminding Test), but for all of these there was uncertainty in direction and size of effect based on confidence intervals. There was also uncertainty in direction and size of effect for two outcomes where the point estimate suggested a worse score in the intervention group (SDMT and one of two sub scores on the 10/36 Spatial Recall Test). For the remaining outcomes, there was no clinically important difference based on point estimates (PASAT 2-second and 3-second versions, World List Generation Test and one of two sub scores on the 10/36 Spatial Recall Test).

The study reported three measures of psychological outcome, with only one of these suggestive of a benefit of intervention based on point estimates (Beck Depression Inventory). Even for this outcome there was uncertainty based on confidence intervals and for the other two measures (Penn State Worry Questionnaire and Difficulties in Emotion Regulation scale) point estimates suggested no clinically important difference between groups.

The final outcome reported by the study was a measure of quality of life (WHO Quality of Life and Satisfaction with Life Scale composite, as a z-score). The point estimate suggested a benefit of intervention but as with other outcomes, confidence intervals indicated uncertainty in this conclusion.

12 months

A single study of only 60 people was included in this comparison at 12 months, with all outcomes assessed as low-very low quality based on GRADE. The study reported ten cognitive measures. For eight of these, the point estimates suggested a benefit of intervention compared to control (SDMT, PASAT 2-second and 3-second versions, two sub scores on the Controlled Oral Word Association verbal fluency test and three sub scores on the Wechsler Memory Scale-III), but for all of these there was uncertainty in direction and size of effect based on confidence intervals. For the remaining outcomes, there was no clinically important difference based on point estimates (two sub scores on the Wechsler Memory Scale-III).

The study reported a possible benefit of intervention compared to control for the Beck Depression Inventory Scale (very low quality); however, as with other outcomes there was uncertainty in this conclusion based on confidence intervals. For the remaining outcomes reported in the study (State-Trait Anxiety Inventory score and composite score for Functional Independence and Assessment), point estimates suggested no difference between intervention and control groups.

Compared to general cognitive rehabilitation (multi-component and multi-domain)

A single study of only 33 people was included in this comparison at 4 weeks, with all outcomes apart from adherence assessed as very low quality based on GRADE. The study reported nine cognitive measures. For three of these, the point estimates suggested a benefit of mindfulness compared to general cognitive rehabilitation (three sub scores on the Selective Reminding Test), but for all of these there was uncertainty in direction and size of

effect based on confidence intervals. There was also uncertainty in direction and size of effect for one outcome where the point estimate suggested a worse score in the mindfulness group (Word List Generation Test). For the remaining outcomes, there was no clinically important difference based on point estimates (SDMT, PASAT 2-second and 3-second versions, World List Generation Test and two sub scores on the 10/36 Spatial Recall Test).

The study reported three measures of psychological outcome, with only one of these suggestive of a benefit of mindfulness based on point estimates (Beck Depression Inventory). Even for this outcome there was uncertainty based on confidence intervals and for the other two measures (Penn State Worry Questionnaire and Difficulties in Emotion Regulation scale) point estimates suggested no clinically important difference between groups.

The final outcome reported by the study was a measure of quality of life (WHO Quality of Life and Satisfaction with Life Scale composite, as a z-score). The point estimate suggested a benefit of mindfulness but as with other outcomes, confidence intervals indicated uncertainty in this conclusion. Adherence results suggested that adherence was better in the mindfulness group compared to general cognitive rehabilitation, though there was also uncertainty based on confidence intervals.

Compared to medical treatment and counselling

A single study of only 53 people was included in this comparison at 16 weeks, with all outcomes as low-very low quality based on GRADE. The study reported twelve cognitive measures. For all but one of these, the point estimates suggested a benefit of mindfulness compared to medical treatment and counselling, but confidence intervals were only consistent with this conclusion in four cases: Digit Span Test as part of Wechsler Adult Intelligence Scale-Revised (very low quality); PASAT 3-second version (very low quality); and perseveration and total error scores on the Wisconsin Card Sorting Test (very low quality for both). Results for Symbol Coding test on Wechsler Adult Intelligence Scale-Revised, PASAT 2-second version and five other sub scores on the Wisconsin Card Sorting Test suggested a benefit based on point estimates but there as uncertainty in the conclusion. For the remaining outcome (Rey Complex Figure Test recall), no clinically important difference was indicated based on point estimates.

The only other outcome reported by the study was a measure of anxiety (Hamilton Anxiety Scale; low quality). The point estimate and confidence intervals were consistent with there being a clinically important benefit of mindfulness compared to medical treatment and counselling.

Information processing speed: cognitive rehabilitation software focused on processing speed

In addition to occupational therapy, compared to occupational therapy only

A single study of only 64 people was included in this comparison at 3 months. The study reported only two outcomes relevant to the protocol, both of which were cognitive measures and were assessed as very low quality based on GRADE. For both outcomes (SDMT and PASAT), point estimates suggested a possible benefit of cognitive rehabilitation software compared to occupational therapy only, though there was uncertainty in the direction and size of effect based on confidence intervals.

Compared to control (active game or no intervention)

A number of studies reported data for this comparison at 5-6 weeks (21 to 50 people analysed depending on outcome); however, due to no overlap in outcome reporting all analyses consisted of only one small study. All outcomes were assessed as very low quality based on GRADE. Outcomes reported included nine cognitive measures, with the results six of these suggesting a benefit of intervention compared to control based on point estimates (SDMT, Digit Symbol Coding test on Wechsler Adult Intelligence Scale-III, two sub scores on the California Verbal Learning Test-II, and results for perceptual speed assessed by letter comparison and pattern comparison tests). However, for all of these outcomes there was uncertainty in the direction and size of effect based on confidence intervals. Similar uncertainty was identified for two outcomes where point estimates indicated worse scores in the intervention compared to control group (Brief Visuospatial Memory Test-Revised and one subscore on the California Verbal Learning Test-II).

One study reported a measure of self-reported cognition (Perceived Deficits Questionnaire), with the point estimate suggesting a worse score in the intervention group but uncertainty being present in the direction and size of effect.

Of the remaining outcomes, point estimates suggested either a possible benefit of intervention (Timed Activities of Daily Living Test z-score State subscore of State-Trait Anxiety Index scale and Modified Fatigue Impact Scale) or worse scores in the intervention compared to control group (CES-D depression score and Trait subscore of State-Trait Anxiety Index scale), though for all of these outcomes there was uncertainty based on confidence intervals.

Information processing speed + working memory: n-back training focused on processing speed + working memory

Compared to sham training (training with no increasing difficulty)

A single study of only 64 people was included in this comparison at 6 weeks. All outcomes were assessed as very low quality based on GRADE. The study reported ten cognitive measures, with point estimates for four of these suggesting a possible benefit of cognitive n-back training but there being uncertainty in this conclusion based on confidence intervals (PASAT, Stroop Test, Controlled Oral Word Association Test and Trials 1-5 on Auditory Verbal Learning Task). For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, Letter-Number Sequencing, Digits-backwards, Raven's Advanced Progressive Matrices, Brief Visuospatial Memory Test and Conner's Continuous Performance Task).

The study reported quality of life using the MSQoL-54 scale, and the point estimate indicated no clinically important difference between groups. Of the three psychological outcomes reported, results for the State and Trait sub scores on the State-Trait Anxiety Inventory suggested no clinically important difference between groups, but for Beck Depression Inventory-Fast Screen the point estimate suggested a worse score in the intervention group, though confidence intervals meant there was uncertainty in this conclusion. Fatigue as measured by the Modified Fatigue Impact Scale suggested no clinically important difference between groups.

There was no difference in adherence between the two groups, but satisfaction may have been better in the sham training group compared to intervention, though there was uncertainty in this result.

Attention/working memory

Computer-aided RehaCom training (attention and information processing) compared to active control

A single study of only 23 people was included in this comparison at 6 weeks. All outcomes were assessed as very low quality based on GRADE. The study reported twelve cognitive measures, with point estimates for four of these suggesting a possible benefit of RehaCom training (all three sub scores on Selective Reminding Test reported and Stroop Test) and point estimates for two suggesting a worse score in the RehaCom group (both sub scores of the 10/36 Spatial Recall Test reported); however, for all of these there was uncertainty in this conclusion based on confidence intervals. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, PASAT 3-second version, Word List Generation Test and three sub scores on the Trail Making Test).

Of the three psychological outcomes reported, results for all three (State and Trait sub scores on the State-Trait Anxiety Inventory and Beck Depression Inventory-II) suggested a possible benefit in the intervention group, though confidence intervals meant there was uncertainty in this conclusion. A possible benefit in the intervention group for fatigue as measured by the Fatigue Severity Scale was also identified, however as with other outcomes there was uncertainty in this conclusion.

Computer-aided training for attention/working memory compared to control

Depending on the outcome, up to two studies (up to 53 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most specific measures were only reported by one study. All outcomes were low to very low quality based on GRADE. Fourteen analyses of cognitive measure were reported, with point estimates for three of these suggesting a possible benefit of intervention compared to control (Brief Visuospatial Memory Test-Revised, Digit Span-backward on the Wechsler Adult Intelligence Scale and 2-back errors on the N-back test) and point estimates for two suggesting a worse score in the intervention group (Spatial Span on the Wechsler Memory Scale-III and 1-back on the N-back test); however, for all of these there was uncertainty in this conclusion based on confidence intervals. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, PASAT, California Verbal Learning Test Total Immediate Recall, four sub scores on the Wechsler Adult Intelligence Scale-III, Color-Word Interference on the Delis-Kaplan Executive Function System and 0-back errors on the N-back test).

Four self-reported measures of cognitive function were reported. For the Cognitive Failure Questionnaire (1 study; n=22), Dysexecutive Questionnaire (1 study; n=22) and Perceived Deficits Questionnaire (1 study; n=22), point estimates suggested worse scores in the intervention group but there was uncertainty based on confidence intervals, while results for MS Neuropsychological Screening Questionnaire (2 studies; n=53) suggested no clinically important difference between groups.

Two quality of life outcomes were reported; results for SF-36 suggested a possible benefit of intervention, with uncertainty based on confidence intervals, and results for EQ-5D on a 0-1 scale suggested no clinically important difference between groups.

Data reported for Functional Assessment of MS scale (1 study; n=31), anxiety on the Hospital Anxiety and Depression Scale (2 studies; n=53) and Self-Efficacy Scale for MS (1 study; n=31) suggested a worse score in the intervention group, though uncertainty was again present in this conclusion.

Of the remaining outcomes reported, including Beck Depression Inventory-Fast Screen, depression on the Hospital Anxiety and Depression Scale, fatigue measured by Fatigue Severity Scale and Patient Activation Measure-13 measuring patient engagement in health, results suggested no clinically important difference between groups.

High-intensity working memory training, distributed working memory training and control (no training) compared to each other

High-intensity compared to distributed training

A single study of only 30 people was included in this comparison at 4-8 weeks. All outcomes were assessed as very low quality based on GRADE. The study reported ten cognitive measures, with point estimates for only one of these suggesting a possible benefit of high-intensity training over distributed training (Faces Symbol Test) and point estimates for three suggesting a worse score in the high-intensity group (SDMT, and reaction time and omissions on the 2-back test); however, for all of these there was uncertainty in this conclusion based on confidence intervals. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (PASAT, backward and forward scores on Corsi Blocks, forward and backward score on Digit Span test and 2-back number correct).

The study reported one measure of functional ability (Functional Assessment of MS) and one measure of depression (Allgemeine Depressionsskala), with point estimates for both suggesting worse score in the high-intensity group but there being uncertainty in this conclusion. For both of the fatigue measures reported (Motor and Cognitive Functions Scale and Modified Fatigue Impact Scale), the point estimates suggested no clinically important difference between groups.

High-intensity training compared to control

A single study of only 30 people was included in this comparison at 4 weeks. All outcomes were assessed as very low quality based on GRADE. The study reported ten cognitive measures, with point estimates for three of these suggesting a possible benefit of high-intensity training compared to control (forward and backward scores on Digit Span test and omissions on the 2-back test) and point estimates for one suggesting a worse score in the high-intensity group compared to control (forward score on Corsi Blocks); however, for all of these there was uncertainty in this conclusion based on confidence intervals. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, PASAT, backward score on Corsi Blocks, 2-back number correct and reaction time and Faces Symbol Test).

The study reported one measure of functional ability (Functional Assessment of MS), one measure of depression (Allgemeine Depressionsskala) and two fatigue measures (Motor and Cognitive Functions Scale and Modified Fatigue Impact Scale), with the point estimates for all four of these outcomes suggesting no clinically important difference between groups.

Distributed training compared to control

A single study of only 30 people was included in this comparison at 4-8 weeks. All outcomes were assessed as very low quality based on GRADE. The study reported ten cognitive measures, with point estimates for six of these suggesting a possible benefit of distributed training compared to control (PASAT, backward score on Corsi Blocks, forward and backward scores on Digit Span test, and reaction time and omissions on 2-back test) and point estimates for one suggesting a worse score in the distributed training group compared to control (Faces Symbol Test); however, for all of these there was uncertainty in this conclusion based on confidence intervals. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, forward score on Corsi Blocks and number correct on 2-back test).

The study reported one measure of functional ability (Functional Assessment of MS), one measure of depression (Allgemeine Depressionsskala) and two fatigue measures (Motor and Cognitive Functions Scale and Modified Fatigue Impact Scale), with the point estimates for all three of these (all apart from Modified Fatigue Impact Scale) suggesting a clinically important benefit of distributed training compared to control, with uncertainty still present based on confidence intervals. For Modified Fatigue Impact Scale, no clinically important difference between groups was indicated by point estimates.

Attention Processing Training (APT) + multidisciplinary rehabilitation compared to multidisciplinary rehabilitation only

A single study of only 34 people was included in this comparison with outcomes reported between 3 and 6 months depending on the outcome. All outcomes were assessed as low-very low quality based on GRADE. The study reported nine cognitive measures, with point estimates for only one of these suggesting a possible benefit of APT training compared to rehabilitation only (one of two sub scores on the Selective Reminding Test; low quality) and point estimates for two suggesting a worse score in the APT training group compared to rehabilitation only (Word List Generation Test and Stroop Test; very low and low quality, respectively); however, for all of these there was uncertainty in this conclusion based on confidence intervals. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, PASAT 2-second and 3-second versions, one of two sub scores reported for the Selective Reminding Test and both sub scores reported for the 10/36 Spatial Recall Test).

The other two outcomes reported by the study were a measure of depression (Montgomery-Asberg Depression scale; low quality) and a measure of activities of daily living (Barthel Index; low quality). For depression, point estimates suggested a possible benefit of intervention while for Barthel Index the score was worse in the intervention group; however, for both, there was uncertainty in the conclusion based on confidence intervals.

Reaction time tasks + usual rehabilitation compared to active control (cognitive software with no time component)

A single study of only 30 people was included in this comparison with outcomes reported at 2 weeks. The study only reported one cognitive measure and it was reported in a dichotomous format; results for the proportion with a T-value ≥40 for alertness (very low quality), indicating a normal result, suggested a possible benefit of reaction time tasks compared to control, though confidence intervals indicated uncertainty in the result.

The study also reported the proportion with fatigue according to WEIMuS (score ≥32) in the two groups, with results suggesting a benefit of intervention compared to control but again there was uncertainty in this conclusion. Results also suggested that adherence was better in the training group compared to those using an active control software with no time component, with uncertainty present as with other outcomes.

Memory

Computer-aided training for memory (with or without attention components) compared to control (no training)

A single study of 40 to 42 people depending on the outcome reported data for this comparison at 6-14 weeks. Of 21 cognitive measures reported, sixteen of these suggested a clinically important benefit of intervention based on point estimates; however, confidence

intervals were only consistent with this conclusion in seven cases: Spatial Span (Corsi; low quality); Paired Associates-hard (very low quality), Visual Reproduction (low quality), Luria-Nebraska Neuropsychological Battery Memory scale (low quality), Signal Detection reaction time (very low quality), Recognition Memory (very low quality) and Digit Span-forward (very low quality) scores. For the remaining nine measures, though the point estimate suggested a benefit of intervention, there was uncertainty based on confidence intervals (one of five sub scores for the California Verbal Learning Test, PASAT, Object Alternation errors, Alertness with and without cueing, Paired Associates-easy, Short Story Recall, Signal Detection hits and Digit Span-backward). For one measure (Object Alternation reaction time; very low quality) the score was worse in the intervention group compared to control, but there was uncertainty based on confidence intervals. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (four of five sub scores reported for the California Verbal Learning Test).

Additional outcomes reported for this comparison were quality of life using SF-12 mental and physical sub scores, Beck Depression Inventory and Fatigue Severity Scale (very low quality for all). For all of these outcomes, point estimates suggested no clinically important difference between groups.

Computer-aided RehaCom memory (and attention) training compared to active control

Across two studies, data for this comparison was reported for 40 or 77 people, depending on the study, with no pooling possible for any outcomes due to lack of overlap in outcome-reporting. Most outcomes were graded low-very low quality and results were reported at 14-16 weeks. Of 19 cognitive measures reported, eleven of these suggested a clinically important benefit of intervention based on point estimates; however, confidence intervals were only consistent with this conclusion in five cases: Word List Generation Test (n=77; low quality); Digit Span-forward (n=40; very low quality); Digit Span-backward (n=40; very low quality); Short Story Recall (n=40; low quality) and Luria-Nebraska Neuropsychological Battery Memory scale (n=40; very low quality. For the remaining six measures, though the point estimate suggested a benefit of intervention, there was uncertainty based on confidence intervals (one of two sub scores for the Selective Reminding Test, Spatial Span-Corsi, Paired Associates easy and hard scores, and reaction time and hits on Signal Detection). For five measures (PASAT 2-second version, one of two sub scores for the Selective Reminding Test, both sub scores reported for the 10/36 Spatial Recall Test and Recognition Memory) the score was worse in the intervention group compared to control, but there was uncertainty based on confidence intervals for all of these apart from the delayed score on 10/36 Spatial Recall Test (n=77; low quality). For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, Visual Reproduction and proportion achieving >20% improvement on Brief Repeatable Battery of Neuropsychological Tests).

Additional outcomes reported for this comparison were quality of life using MSQoL-54 scale (n=77; very low quality) and depression assessed using the Chicago Mood Depression Inventory (n=77; low quality). Results for quality-of-life suggested worse outcome in the intervention group based on point estimates but there was uncertainty in the direction and size of effect based on confidence intervals. For depression, the point estimate indicated no clinically important difference between groups.

Story Memory Technique compared to control

5-11 weeks

Depending on the outcome, up to two studies (up to 114 people analysed) reported data that could be pooled for this comparison between 5 and 11 weeks, though most specific

measures were only reported by one study. All outcomes were low to very low quality based on GRADE. Ten analyses of cognitive measure were reported, with point estimates for seven of these suggesting a possible benefit of intervention compared to control (Hopkins Verbal Learning Test as a continuous measure and when reporting the proportion with improvement on the test, two continuous measures of California Verbal Learning Test and also two analyses reporting the proportion achieving >10% improvement on this test, and Rivermead Behavioural Memory Test) and point estimates for one suggesting a worse score in the intervention group (SDMT as a z-score); however, for all of these there was uncertainty in this conclusion based on confidence intervals. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (Letter-Number Sequencing scaled score measuring working memory and Digit Span scaled score measuring attention).

Two self-reported measures of cognition were reported; results for the Awareness of Cognitive Deficits Questionnaire (1 study; n=28; very low quality) suggested a benefit of intervention compared to control based on point estimates, with uncertainty when considering confidence intervals, and no clinically important difference was indicated for Memory Functioning Questionnaire (1 study; n=20; very low quality) results.

Other outcomes reported for this comparison where a possible benefit of intervention was identified included Functional Assessment of MS-General Contentment (1 study; n=86; very low quality) and Satisfaction with Life Scale (1 study; n=20; very low quality); however, similar to other outcomes there was uncertainty in this conclusion. One study also suggested a harm for three sub scores on the Frontal Systems Behaviour Scale, though again there was uncertainty in this conclusion.

The remaining outcomes reported for this comparison were State and Trait sub scores on the State-Trait Anxiety Inventory (1 study; n=86; very low quality for both), the Chicago Multidimensional Depression Inventory (1 study; n=86; very low quality), and patient-reported and informant-reported versions of the Patient Competency Rating Scale (1 study; n=20 very low quality for both). For all of these outcomes no clinically important difference was demonstrated.

7 months

A single study of 78 people reported data for this comparison at 7 months, with all outcomes assessed as low-very low quality based on GRADE. Of seven cognitive measures reported, only two of these suggested a clinically important benefit of intervention based on point estimates, with there being uncertainty based on confidence intervals for both of these: total learning score on California Verbal Learning Test as a T-score (very low quality) and z-score for learning slope on the same. Results for one measure (z-score for SDMT; low quality), suggested a worse score in the intervention group compared to control and confidence intervals were also consistent with this conclusion. For the remaining four measures (two sub scores on the Rivermead Behavioural Memory Test, Letter-Number Sequencing scaled score for working memory and attention measured by Digit Span scaled score) point estimates suggested no clinically important difference between groups (four of five sub scores reported for the California Verbal Learning Test).

One other outcome reported for this comparison where a possible benefit of intervention was identified was Functional Assessment of MS-General Contentment (very low quality); however, similar to other outcomes there was uncertainty in this conclusion.

The remaining outcomes reported for this comparison were two sub scores on the Frontal Systems Behaviour Scale (low or very low quality), State and Trait sub scores on the State-Trait Anxiety Inventory (very low quality for both) and the Chicago Multidimensional

Depression Inventory (very low quality). For all of these outcomes no clinically important difference was demonstrated.

Group memory programme (various learning techniques) compared to control

Up to 6 months

Depending on the outcome, up to three studies (up to 489 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most specific measures were only reported by one study. Most outcomes were low to very low quality based on GRADE. Twelve analyses of cognitive measures were reported, with point estimates only one of these suggesting a possible benefit of intervention compared to control (working memory assessed possible using Wechsler Memory Scale-III; 1 study; n=60; moderate quality), though confidence intervals indicated uncertainty in this conclusion. For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, two scores on the Selective Reminding Test, two scores on the 10/36 Spatial Recall Test, easy and hard versions of PASAT, Trail Making Test score B-A, Word Fluency, Doors and people score and Digit Span Test for attention).

Four self-reported measures of cognitive function were reported, including three versions of the Everyday Memory Questionnaire and the Prospective and Retrospective Memory Questionnaire. For the Everyday Memory Questionnaire, two analyses (self-reported and carer-reported on a 0-140 scale; 2-3 studies; 374 or 489 people analysed; low quality for both) suggested a benefit of intervention based on point estimates, with uncertainty in this conclusion based on confidence intervals; however, for the other analysis (0-175 scale; 1 study; n=60; moderate quality), results suggested no clinically important difference between groups. Results for the Prospective and Retrospective Memory Questionnaire (1 study; n=56; low quality) suggested a clinically important benefit of intervention, with confidence intervals also consistent with this conclusion.

Quality of life was reported using various scales. Results for MSIS-29 on a 0-100 scale (psychological and physical sub scores; 1 study; n=402 to 404 people; low quality) and EQ-5D (1 study; n=411; low quality) suggested no clinically important difference between groups. Results for MSQoL-54 from one study (n=53; low quality) suggested a possible benefit in the intervention group, with there being uncertainty for the physical health subscore but not for the mental health subscore. In addition, the final measure of quality of life (MSIS-29 overall on a 29-145 scale; 1 study; n=37; very low quality) suggested a worse score in the intervention group, though there was uncertainty based on confidence intervals.

Of the three different psychological measures reported for this comparison, only results for the Beck Depression Inventory (1 study; n=56; low quality) suggested a benefit of intervention compared to control and confidence intervals were consistent with this conclusion. For the other two measures reported (different versions of the General health Questionnaire differing in scale range), the results suggested no difference between groups.

Fatigue, carer burden and employment were each reported by one study. Results for fatigue (Fatigue Severity Scale; n=399; very low quality) and Carer Strain Index (n=327; low quality) suggested a benefit of intervention based on point estimates, with uncertainty based on confidence intervals. For employment (n=411; very low quality), the results suggested no difference between groups.

8-12 months

Depending on the outcome, up to two studies (up to 409 people analysed) reported data that could be pooled for this comparison between 8 and 12 months, though most specific

measures were only reported by one study. All outcomes were low to very low quality based on GRADE. Ten analyses of cognitive measures were reported, with point estimates for none of these suggesting a possible benefit of intervention compared to control and one suggested a worse score in the intervention compared to control group (PASAT hard version; n=374; low quality). For the remaining cognitive measures, point estimates suggested no clinically important difference between groups (SDMT, two scores on the Selective Reminding Test, two scores on the 10/36 Spatial Recall Test, easy version of PASAT, Trail Making Test score B-A, Word Fluency and Doors and people score).

Two self-reported measures of cognitive function were reported, which included self-reported and carer-reported versions of the Everyday Memory Questionnaire (2 studies; 336 or 409 people analysed; low quality for both). Results suggested a benefit of intervention based on point estimates, with uncertainty in this conclusion based on confidence intervals.

Quality of life was reported using various scales, including MSIS-29 on a 0-100 scale (psychological and physical sub scores; 1 study; n=387 people; low quality), MSIS-29 overall on a 29-145 scale (1 study; n=31; very low quality) and EQ-5D (1 study; n=382; low quality). Results for all of these scales suggested no clinically important difference between groups.

One psychological measure was reported for this comparison at this time-point (General Health Questionnaire-30 on a 0-90 scale; 1 study; n=376; low quality), with the results suggesting no clinically important difference between groups.

Fatigue, carer burden and employment were each reported by one study. Results for fatigue (Fatigue Severity Scale; n=378; very low quality) suggested a benefit of intervention based on point estimates, with uncertainty based on confidence intervals. For Carer Strain Index (n=300; low quality) and employment (n=382; very low quality), the results suggested no difference between groups.

<u>Behaviour intervention (self-generated learning) compared to control (memory tasks with no self-generated learning taught)</u>

A single study of 35 people reported data for this comparison at 3-4 weeks, with most outcomes assessed as low-very low quality based on GRADE. Of eight cognitive measures reported, seven of these suggested a clinically important benefit of intervention based on point estimates; however, confidence intervals were only consistent with this conclusion in two cases: Immediate and Delay scores for the Contextual Memory Test (moderate and low quality, respectively). For the remaining five measures, though the point estimate suggested a benefit of intervention, there was uncertainty based on confidence intervals (one of two sub scores reported for the California Verbal Learning Test, Memory For Intentions Test, Verbal Fluency Test, and errors and cognitive score on the Actual Reality™ test). For the remaining cognitive measure, the point estimate suggested no clinically important difference between groups (one of two sub scores reported for the California Verbal Learning Test).

One self-reported measures of cognitive function as reported (Memory Functioning Questionnaire; low quality), with the point estimate suggesting a benefit of intervention based but there being uncertainty in this conclusion based on confidence intervals.

The study also suggested a benefit for the following outcomes based on point estimates with confidence intervals indicating uncertainty in the conclusion for all of these: Functional Behavioural Profile (low quality); Self-awareness of Cognitive Deficits Questionnaire (low quality); and Chicago Multiscale Depression Inventory (low quality).

For the remaining outcomes reported in the study, results suggested no clinically important difference between groups: Functional Assessment of MS (low quality); Self-regulation skills interview (self-awareness and strategy use; low quality); Trait score on the State-Trait Anxiety Inventory (low quality); and Satisfaction with Life Scale (very low quality).

Executive function

Executive function-specific training compared to control (no training)

6 weeks

Across two studies reporting data for this comparison at 6 weeks, all outcomes were assessed as very low quality based on GRADE and no pooling was possible across studies. Ten analyses were reported across studies, all of which were cognitive measures. For six of these analyses, point estimates suggested a possible benefit of intervention compared to control, but in only two cases were confidence intervals also consistent with this conclusion: number of categories and total errors on the Wisconsin Card Sorting Test (n=20; very low quality for both). For the other four analyses (Preference shifting trials to criterion and reaction time, Response shifting reaction time and omissions on the 2-back test), confidence intervals indicated uncertainty in the conclusion of a benefit.

Of the remaining measures, a worse score in the intervention group was suggested based on point estimates in two cases (Response shifting trials to criterion and commissions on the 2-back test), with uncertainty based on confidence intervals, and in two cases the point estimates suggested no clinically important difference between groups (California Verbal Learning Test score and reaction time on the 2-back test).

12 months

A single study of only 12 people reported data for this comparison at 12 months, with all outcomes assessed as very low quality based on GRADE. Eight analyses were reported, all of which were cognitive measures. For two of these analyses, point estimates suggested a possible benefit of intervention compared to control, in both cases there was uncertainty in this conclusion-based confidence intervals: Response shifting trials to criterion and omissions on the 2-back test. For three analyses (Preference shifting trials to criterion and reaction time, and reaction time on the 2-back test), point estimates suggested a worse score in the intervention group, with confidence intervals also suggesting uncertainty in this result.

Of the remaining measures, all three suggested no difference between groups based on point estimates (California Verbal Learning Test score, reaction time for Response shifting and 2-back commissions).

Executive function-specific training compared to active control (responding quickly to visual stimuli)

6 weeks

A single study of only 25 people reported data for this comparison at 6 weeks, with all outcomes assessed as very low quality based on GRADE. Eight analyses were reported, all of which were cognitive measures. For three of these analyses, point estimates suggested a possible benefit of intervention compared to control, but in all cases, there was uncertainty in the conclusion-based confidence intervals: Preference shifting trials to criterion; Response shifting reaction time and 2-back test reaction time. For three analyses (Preference shifting reaction time, and commissions and omissions on the 2-back test), point estimates suggested a worse score in the intervention group, with confidence intervals also suggesting uncertainty in this result.

Of the remaining measures, both suggested no difference between groups based on point estimates (California Verbal Learning Test score and Response shifting trials to criterion).

12 months

A single study of only 14 people reported data for this comparison at 12 months, with all outcomes assessed as very low quality based on GRADE. Eight analyses were reported, all of which were cognitive measures. For four of these analyses, point estimates suggested a possible benefit of intervention compared to control, but in all cases, there was uncertainty in this conclusion based on confidence intervals: Preference shifting reaction time; Response shifting trials to criterion and reaction time; and 2-back test omissions. For three analyses (Preference shifting trials to criterion, and commission's reaction time on the 2-back test), point estimates suggested a worse score in the intervention group, with confidence intervals also suggesting uncertainty in this result.

The results for the remaining measure suggested no difference between groups based on point estimates (California Verbal Learning Test score).

Goal management programme compared to psychoeducation

9 weeks

A single study of only 27 people reported data for this comparison at 9 weeks, with all outcomes assessed as very low quality based on GRADE. Nine analyses for cognitive measures were reported. For four of these analyses, point estimates suggested a possible benefit of goal management compared to psychoeducation, but in all cases, there was uncertainty in this conclusion, based on confidence intervals: two of three reported elevator counting tests as part of the Test of Everyday Attention, and tasks attempted and deviation from optimal task time as part of the Hotel Test. For two analyses (commission errors on the Sustained Attention to Response Task and Tower Test as part of the Delis-Kaplan Function Scale), point estimates suggested a worse score in the intervention group, with confidence intervals also suggesting uncertainty in this result.

For the remaining three cognitive measures, results indicated no clinically important difference between groups: reaction time and omission errors as part of the Sustained Attention to Response Task; and one of three reported elevator counting tests as part of the Test of Everyday Attention.

One outcome assessing self-reported cognition (informant version of the Dysexecutive Questionnaire) suggested no difference between groups, while the self-reported version of the same questionnaire suggested a worse score in the goal management group compared to psychoeducation, with uncertainty based on confidence intervals. Similarly, results for the Cognitive Failures Questionnaire suggested worse scores in the goal management group, again with uncertainty in the conclusion present. Results for Global Sleep Disturbance on the Pittsburgh Sleep Quality Index demonstrated no clinically important difference between groups.

Point estimates for the two remaining outcomes (Total Mood Disturbance on Profile of Mood States scale and proportion achieving or exceeding their goal) suggested a worse score in the goal management group or a possible benefit of the goal management group, respectively, compared to psychoeducation; however, for both there was uncertainty based on confidence intervals.

8 months

A single study of only 23 people reported data for this comparison at 8 months, with all outcomes assessed as low-very low quality based on GRADE. Nine analyses for cognitive

measures were reported. For three of these analyses, point estimates suggested a possible benefit of goal management compared to psychoeducation, but in all cases, there was uncertainty in this conclusion, based on confidence intervals: two of three reported elevator counting tests as part of the Test of Everyday Attention; and deviation from optimal task time as part of the Hotel Test. For two analyses (omission and commission errors on the Sustained Attention to Response Task), point estimates suggested a worse score in the intervention group, with confidence intervals also suggesting uncertainty in this result.

For the remaining four cognitive measures, results indicated no clinically important difference between groups: reaction time on the Sustained Attention to Response Task; one of three reported elevator counting tests as part of the Test of Everyday Attention; Tower Test as part of the Delis-Kaplan Function Scale; and tasks attempted as part of the Hotel Test.

Two outcomes assessing self-reported cognition (self-reported version of the Dysexecutive Questionnaire and Cognitive Failures Questionnaire) suggested a worse score in the goal management group compared to psychoeducation, with uncertainty based on confidence intervals. However, the informant-reported version of the Dysexecutive Questionnaire suggested a benefit of goal management compared to psychoeducation, again with uncertainty in the conclusion present. Results for Global Sleep Disturbance on the Pittsburgh Sleep Quality Index demonstrated no clinically important difference between groups.

Point estimates for both of the remaining outcomes (Total Mood Disturbance on Profile of Mood States scale and Global Sleep Disturbance on the Pittsburgh Sleep Quality Index) suggested a worse score in the goal management group compared to psychoeducation; however, for both there was uncertainty based on confidence intervals.

Improving language

RehaCom verbal fluency training compared to control (no training)

A single study of only 53 to 60 people reported data for this comparison at 10 weeks depending on the outcome, with all outcomes assessed as very low quality based on GRADE. Only three outcomes were reported, with two cognitive measures reported and a data for adherence also available. For both of the cognitive measures (California Verbal Learning Test-II and Controlled Oral Word Association Test; n=53), the point estimates suggested a benefit of verbal fluency training but there was uncertainty in the size of the effect based on confidence intervals.

The results for adherence suggested there was no difference between groups in terms of optional dropout from treatment.

1.1.12.4 Cost effectiveness and resource use

One health economic analysis was identified for this review comparing cognitive rehabilitation plus usual care versus usual care. This was a within-trial cost utility analysis of the CRAMMS RCT which was included in the clinical review. The cognitive rehabilitation intervention lasted 10 weeks and involved once weekly 1.5-hour group sessions. This study was from a UK NHS perspective and had a 12-month follow-up. The base case found that group cognitive rehabilitation and usual care dominated usual care only (less costly and equally effective). The confidence intervals (CI) for both incremental costs and QALYs spanned zero and the CI for costs were wide. The sensitivity analyses further highlighted the uncertainty in the base case analysis. Given this, the committee were cautious in interpreting the results. The study was assessed as partially applicable (EQ5D-5L mapped to EQ5D-3L but mapping function used was not reported and does not include all comparators in the review protocol) with minor limitations (based on a single RCT and so may not reflect full

body of clinical evidence. RCT and HE analysis based on follow up of only 12 months and many not capture long term costs).

In addition to this study, relevant unit costs were presented to the committee to aid consideration of cost effectiveness.

The committee discussed the clinical and economic evidence and based on the limitations described above in the clinical evidence section and the uncertainty in the economic evidence, the committee were not able to make more specific recommendations about which interventions may be appropriate for memory and cognitive problems in MS. A research recommendation has been made.

The committee did agree to recommend assessment of cognition as part of the comprehensive review. They noted that some people with MS are not aware that cognitive problems can be a symptom of MS and highlighted the importance of assessing cognitive problems as it will help identify the most relevant intervention for the individual. The committee noted that cognitive assessments do happen for most but there may be regional variations. They noted that for some only a simple assessment (such as the Addenbrooke's or MoCA) is required which takes 10-15 mins and does not require specific expertise on behalf of the healthcare professional. This type of assessment may be a change in practice for some, but it is unlikely to have a significant resource impact. For others a longer neuropsychological assessment may be needed which will include checking mood and fatigue. The committee said that this longer assessment could take up to 10 hours over several appointments and therefore would be a more costly assessment. The committee noted that a very small proportion of people with MS are likely to require this longer neuropsychological assessment (fewer than 1%) and that once a baseline assessment has been done, future assessments should not be as resource intensive. Given the very small population for whom this more expensive assessment will apply to, the committee did not think it would have a significant resource impact.

1.1.12.5 Other factors the committee took into account

The committee noted that MS is predominantly a disease involving white matter damage, meaning the most common cognitive impairments are related to the subcortical area, such as information processing, executive function, working memory and multitasking, rather than focal-based impairments such as language. However, it was highlighted that there may also be some more specific lesions in individuals that cause impairments in other less common areas.

1.1.13 Recommendations supported by this evidence review

This evidence review supports recommendations 1.5.38 to 1.5.41.

1.1.14 References

1. Arian Darestani A, Naeeni Davarani M, Hassani-Abharian P, Zarrindast MR, Nasehi M. The therapeutic effect of treatment with RehaCom software on verbal performance in patients with multiple sclerosis. Journal of Clinical Neuroscience. 2020; 72:93-97

2. Azimian M, Yaghoubi Z, Ahmadi Kahjoogh M, Akbarfahimi N, Haghgoo HA, Vahedi M. The effect of cognitive rehabilitation on balance skills of individuals with multiple sclerosis. Occupational Therapy in Health Care. 2021; 35(1):93-104

3. Blair M, Goveas D, Safi A, Marshall C, Rosehart H, Orenczuk S et al. Does cognitive training improve attention/working memory in persons with MS? A pilot study using the Cogmed Working Memory Training program. Multiple Sclerosis and Related Disorders. 2021; 49:102770

4. Bove R, Rowles W, Zhao C, Anderson A, Friedman S, Langdon D et al. A novel in-home digital treatment to improve processing speed in people with multiple sclerosis: A pilot study. Multiple Sclerosis Journal. 2021; 27(5):778-789

5. Brissart H, Omorou AY, Forthoffer N, Berger E, Moreau T, De Seze J et al. Memory improvement in multiple sclerosis after an extensive cognitive rehabilitation program in groups with a multicenter double-blind randomized trial. Clinical Rehabilitation. 2020; 34(6):754-763

6. Campbell J, Langdon D, Cercignani M, Rashid W. A randomised controlled trial of efficacy of cognitive rehabilitation in multiple sclerosis: A cognitive, behavioural, and MRI study. Neural Plasticity. 2016; 2016:4292585

7. Carr SE, das Nair R, Schwartz AF, Lincoln NB. Group memory rehabilitation for people with multiple sclerosis: a feasibility randomized controlled trial. Clinical Rehabilitation. 2014; 28(6):552-561

8. 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

9. Charvet LE, Shaw MT, Haider L, Melville P, Krupp LB. Remotely-delivered cognitive remediation in multiple sclerosis (MS): protocol and results from a pilot study. Multiple Sclerosis Journal Experimental Translational & Clinical. 2015; 1:2055217315609629

10. Charvet LE, Yang J, Shaw MT, Sherman K, Haider L, Xu J et al. Cognitive function in multiple sclerosis improves with telerehabilitation: Results from a randomized controlled trial. PloS One. 2017; 12(5):e0177177

11. 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

12. Chiaravalloti ND, Goverover Y, Costa SL, DeLuca J. A pilot study examining speed of processing training (SPT) to improve processing speed in persons with multiple sclerosis. Frontiers in Neurology. 2018; 9:685

13. Chiaravalloti ND, Moore NB, DeLuca J. The efficacy of the modified Story Memory Technique in progressive MS. Multiple Sclerosis. 2020; 26(3):354-362

14. 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

15. Chiaravalloti ND, Wylie G, Leavitt V, Deluca J. Increased cerebral activation after behavioral treatment for memory deficits in MS. Journal of Neurology. 2012; 259(7):1337-1346

16. Chmelarova D, Fiala L, Dostal M, Lenz J. Intensive computer-assisted cognitive rehabilitation in persons with multiple sclerosis - results of a 12-week randomized study. Ceska a Slovenska Neurologie a Neurochirurgie. 2020; 83(4):408-415

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

18. das Nair R, Martin KJ, Lincoln NB. Memory rehabilitation for people with multiple sclerosis. Cochrane Database of Systematic Reviews 2016, Issue 3. Art. No.: CD008754. DOI: 10.1002/14651858.cd008754.pub3.

19. De Giglio L, De Luca F, Prosperini L, Borriello G, Bianchi V, Pantano P et al. A low-cost cognitive rehabilitation with a commercial video game improves sustained attention and executive functions in multiple sclerosis: a pilot study. Neurorehabilitation and Neural Repair. 2015; 29(5):453-461

20. De Giglio L, Tona F, De Luca F, Petsas N, Prosperini L, Bianchi V et al. Multiple sclerosis: Changes in thalamic resting-state functional connectivity induced by a home-based cognitive rehabilitation program. Radiology. 2016; 280(1):202-211

21. De la Torre GG, Mato I, Doval S, Espinosa R, Moya M, Cantero R et al. Neurocognitive and emotional status after one-year of mindfulness-based intervention in patients with relapsing-remitting multiple sclerosis. Applied Neuropsychology Adult. 2020:1-10

22. Ernst A, Sourty M, Roquet D, Noblet V, Gounot D, Blanc F et al. Functional and structural cerebral changes in key brain regions after a facilitation programme for episodic future thought in relapsing-remitting multiple sclerosis patients. Brain and Cognition. 2016; 105:34-45

23. Filippi M, Riccitelli G, Mattioli F, Capra R, Stampatori C, Pagani E et al. Multiple sclerosis: effects of cognitive rehabilitation on structural and functional MR imaging measures - an explorative study. Radiology. 2012; 262(3):932-940

24. 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

25. Flachenecker P, Meissner H, Frey R, Guldin W. Neuropsychological training of attention improves MS-related fatigue: Results of a randomized, placebo-controlled, double-blind pilot study. European Neurology. 2017; 78(56):312-317

26. Gich J, Freixanet J, Garcia R, Vilanova JC, Genis D, Silva Y et al. A randomized, controlled, single-blind, 6-month pilot study to evaluate the efficacy of MS-Line!: a cognitive rehabilitation programme for patients with multiple sclerosis. Multiple Sclerosis. 2015; 21(10):1332-1343

27. Goverover Y, Chiaravalloti N, Genova H, DeLuca J. A randomized controlled trial to treat impaired learning and memory in multiple sclerosis: The self-GEN trial. Multiple Sclerosis. 2018; 24(8):1096-1104

28. Grasso MG, Broccoli M, Casillo P, Catani S, Pace L, Pompa A et al. Evaluation of the impact of cognitive training on quality of life in patients with multiple sclerosis. European Neurology. 2017; 78(12):111-117

29. Hancock LM. Processing speed and working memory training in multiple sclerosis: A blinded randomized controlled trial. Kansas City, Missouri. University of Missouri. 2014. Ph.D.

30. Hancock LM, Bruce JM, Bruce AS, Lynch SG. Processing speed and working memory training in multiple sclerosis: a double-blind randomized controlled pilot study. Journal of Clinical and Experimental Neuropsychology. 2015; 37(2):113-127

31. Hanssen KT, Beiske AG, Landro NI, Hofoss D, Hessen E. Cognitive rehabilitation in multiple sclerosis: a randomized controlled trial. Acta Neurologica Scandinavica. 2016; 133(1):30-40

32. 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

33. Janssen A, Boster A, Lee H, Patterson B, Prakash RS. The effects of video-game training on broad cognitive transfer in multiple sclerosis: A pilot randomized controlled trial. Journal of Clinical and Experimental Neuropsychology. 2015; 37(3):285-302

34. 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

35. Krch D, Lequerica A, Aguayo Arelis A, Rabago Barajas BV, Arango-Lasprilla JC, Chiaravalloti ND. Efficacy of the Spanish modified Story Memory Technique in Mexicans with multiple sclerosis: A pilot randomized controlled trial. NeuroRehabilitation. 2019; 45(3):349-358

36. Lamargue D, Koubiyr I, Deloire M, Saubusse A, Charre-Morin J, Moroso A et al. Effect of cognitive rehabilitation on neuropsychological and semiecological testing and on daily cognitive functioning in multiple sclerosis: The REACTIV randomized controlled study. Journal of the Neurological Sciences. 2020; 415:116929

37. Lincoln NB, Bradshaw LE, Constantinescu CS, Day F, Drummond AE, Fitzsimmons D et al. Cognitive rehabilitation for attention and memory in people with multiple sclerosis: a randomized controlled trial (CRAMMS). Clinical Rehabilitation. 2020; 34(2):229-241

38. Lincoln NB, Bradshaw LE, Constantinescu CS, Day F, Drummond AE, Fitzsimmons D et al. Group cognitive rehabilitation to reduce the psychological impact of multiple sclerosis on quality of life: The CRAMMS RCT. Health Technology Assessment. 2020; 24(4)

39. 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

40. Manglani HR, Samimy S, Schirda B, Nicholas JA, Prakash RS. Effects of 4-week mindfulness training versus adaptive cognitive training on processing speed and working memory in multiple sclerosis. Neuropsychology. 2020; 34(5):591-604

41. Mani A, Chohedri E, Ravanfar P, Mowla A, Nikseresht A. Efficacy of group cognitive rehabilitation therapy in multiple sclerosis. Acta Neurologica Scandinavica. 2018; 137(6):589-597

42. 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

43. Martin K, Lincoln N, das Nair R. Group-based memory rehabilitation for people with multiple sclerosis: subgroup analysis of the ReMiND trial. International Journal of Therapy & Rehabilitation. 2014; 21(12):590-596

44. Mattioli F, Stampatori C, Bellomi F, Danni M, Compagnucci L, Uccelli A et al. A RCT comparing specific intensive cognitive training to aspecific psychological intervention in RRMS: The SMICT study. Frontiers in Neurology. 2015; 13(5):278

45. 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

46. 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(12):101-105

47. Mendozzi, Pugnetti L, Motta A, Barbieri E, Gambini A, Cazzullo CL. Computer-assisted memory retraining of patients with multiple sclerosis. Italian Journal of Neurological Sciences. 1998; 19:S431-S438

48. Messinis L, Kosmidis MH, Nasios G, Konitsiotis S, Ntoskou A, Bakirtzis C et al. Do Secondary Progressive Multiple Sclerosis patients benefit from computer- based cognitive neurorehabilitation? A randomized sham controlled trial. Multiple Sclerosis and Related Disorders. 2020; 39:101932

49. Messinis L, Nasios G, Kosmidis MH, Zampakis P, Malefaki S, Ntoskou K et al. Efficacy of a computer-assisted cognitive rehabilitation intervention in relapsing-remitting multiple sclerosis patients: A multicenter randomized controlled trial. Behavioural Neurology. 2017; 2017:5919841

50. Mousavi S, Zare H, Etemadifar M. Evaluating the effectiveness of cognitive rehabilitation on everyday memory in multiple sclerosis patients. Neuropsychological Rehabilitation. 2020; 30(6):1013-1023

51. Mousavi S, Zare H, Etemadifar M, Taher Neshatdoost H. Memory rehabilitation for the working memory of patients with multiple sclerosis (MS). Journal of Clinical and Experimental Neuropsychology. 2018; 40(4):405-410

52. Naeeni Davarani M, Arian Darestani A, Hassani-Abharian P, Vaseghi S, Zarrindast MR, Nasehi M. RehaCom rehabilitation training improves a wide-range of cognitive functions in multiple sclerosis patients. Applied Neuropsychology Adult. 2020; May 5:DOI: 10.1080/23279095.23272020.21747070

53. Nazaribadie M, Ghaleiha A, Ahmadpanah M, Mazdeh M, Matinnia N, Zarabian MK. Effectiveness of detached mindfulness intervention on cognitive functions in multiple sclerosis patients, results from a randomized controlled study. Pakistan Journal of Medical and Health Sciences. 2020; 14(4):2022-2029

54. Nazaribadie M, Ghaleiha A, Ahmadpanah M, Mazdeh M, Matinnia N, Zarabian MK. Metacognitive model of mindfulness can improve executive function in multiple sclerosis patients. Pakistan Journal of Medical and Health Sciences. 2021; 15(1):590-597

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

56. Parisi L, Rocca MA, Mattioli F, Copetti M, Capra R, Valsasina P et al. Changes of brain resting state functional connectivity predict the persistence of cognitive rehabilitation effects in patients with multiple sclerosis. Multiple Sclerosis. 2014; 20(6):686-694

57. Perez-Martin MY, Gonzalez-Platas M, Eguia-Del Rio P, Croissier-Elias C, Jimenez Sosa A. Efficacy of a short cognitive training program in patients with multiple sclerosis. Neuropsychiatric Disease and Treatment. 2017; 13:245-252

58. Pusswald G, Mildner C, Zebenholzer K, Auff E, Lehrner J. A neuropsychological rehabilitation program for patients with Multiple Sclerosis based on the model of the ICF. NeuroRehabilitation. 2014; 35(3):519-527

59. Rahmani M, Rahimian Boogar I, Talepasand S, Nokani M. Comparing the effectiveness of computer-based, manual-based, and combined cognitive rehabilitation on cognitive functions in relapsing-remitting multiple sclerosis patients. Iranian Journal of Neuroscience. 2020; 11(1):99-110

60. Richard NM. Rehabilitation of executive dysfunction in multiple sclerosis: Cognitive, behavioural and neurophysiological effects of goal management training. Toronto, Canada. University of Toronto. 2013. Ph.D.

61. Rilo O, Pena J, Ojeda N, Rodriguez-Antiguedad A, Mendibe-Bilbao M, Gomez-Gastiasoro A et al. Integrative group-based cognitive rehabilitation efficacy in multiple sclerosis: a randomized clinical trial. Disability and Rehabilitation. 2018; 40(2):208-216

62. Rosti-Otajärvi E, Mantynen A, Koivisto K, Huhtala H, Hamalainen P. Patient-related factors may affect the outcome of neuropsychological rehabilitation in multiple sclerosis. Journal of the Neurological Sciences. 2013; 334(12):106-111

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

64. Schirda B, Duraney E, Lee HK, Manglani HR, Andridge RR, Plate A et al. Mindfulness training for emotion dysregulation in multiple sclerosis: A pilot randomized controlled trial. Rehabilitation Psychology. 2020; 65(3):206-218

65. Shahpouri MM, Barekatain M, Tavakoli M, Sanaei S, Shaygannejad V. Evaluation of cognitive rehabilitation on the cognitive performance in multiple sclerosis: A randomized controlled trial. Journal of Research in Medical Sciences. 2019; 24(110):DOI:10.4103/jrms.JRMS_4124_4119

66. Sharifi A, Yazdanbakhsh K, Momeni K. The effectiveness of computer-based cognitive rehabilitation in executive functions in patients with multiple sclerosis. Journal of Kermanshah University of Medical Sciences. 2019; 23(1):e83092

67. 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

68. 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(12):99-104

69. Steultjens E, Dekker JJ, Bouter LM, Cardol MM, Van den Ende E, van de Nes J. Occupational therapy for multiple sclerosis. Cochrane Database of Systematic Reviews 2003, Issue 3. Art. No.: CD003608. DOI: 10.1002/14651858.cd003608.

70. 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

71. Stuifbergen AK, Becker H, Perez F, Morrison J, Brown A, Kullberg V et al. Computer-assisted cognitive rehabilitation in persons with multiple sclerosis: Results of a multi-site randomized controlled trial with six month follow-up. Disability & Health Journal. 2018; 11(3):427-434

72. 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(2122):747-754

73. Thomas PW, Thomas S, Hillier C, Galvin K, Baker R. Psychological interventions for multiple sclerosis. Cochrane Database of Systematic Reviews 2006, Issue 1. Art. No.: CD004431. DOI: 10.1002/14651858.cd004431.pub2.

74. Vilou I, Bakirtzis C, Artemiadis A, Ioannidis P, Papadimitriou M, Konstantinopoulou E et al. Computerized cognitive rehabilitation for treatment of cognitive impairment in multiple sclerosis: an explorative study. Journal of Integrative Neuroscience. 2020; 19(2):341-347

75. 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