Cochrane Library

Cochrane Database of Systematic Reviews

Cognitive stimulation to improve cognitive functioning in people with dementia (Review)

Woods B, Rai HK, Elliott E, Aguirre E, Orrell M, Spector A

Woods B, Rai HK, Elliott E, Aguirre E, Orrell M, Spector A.
Cognitive stimulation to improve cognitive functioning in people with dementia.
Cochrane Database of Systematic Reviews 2023, Issue 1. Art. No.: CD005562.
DOI: 10.1002/14651858.CD005562.pub3.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Cognitive stimulation to improve cognitive functioning in people with dementia

Bob Woods[1], Harleen Kaur Rai[2], Emma Elliott[3], Elisa Aguirre[4], Martin Orrell[5], Aimee Spector[6]

[1]Dementia Services Development Centre Wales, Bangor University, Bangor, UK. [2]Department of Psychiatry and Applied Psychology, Institute of Mental Health, University of Nottingham, Nottingham, UK. [3]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [4]University College London, London, UK. [5]Institute of Mental Health, University of Nottingham, Nottingham, UK. [6]Research Department of Clinical, Educational and Health Psychology, University College London, London, UK

Contact: Bob Woods, b.woods@bangor.ac.uk.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.
Publication status and date: New search for studies and content updated (conclusions changed), published in Issue 1, 2023.

Citation: Woods B, Rai HK, Elliott E, Aguirre E, Orrell M, Spector A. Cognitive stimulation to improve cognitive functioning in people with dementia. *Cochrane Database of Systematic Reviews* 2023, Issue 1. Art. No.: CD005562. DOI: 10.1002/14651858.CD005562.pub3.

ABSTRACT

Background

Cognitive stimulation (CS) is an intervention for people with dementia offering a range of enjoyable activities providing general stimulation for thinking, concentration and memory, usually in a social setting, such as a small group. CS is distinguished from other approaches such as cognitive training and cognitive rehabilitation by its broad focus and social elements, aiming to improve domains such as quality of life (QoL) and mood as well as cognitive function.

Recommended in various guidelines and widely implemented internationally, questions remain regarding different modes of delivery and the clinical significance of any benefits. A systematic review of CS is important to clarify its effectiveness and place practice recommendations on a sound evidence base. This review was last updated in 2012.

Objectives

To evaluate the evidence for the effectiveness of CS for people with dementia, including any negative effects, on cognition and other relevant outcomes, accounting where possible for differences in its implementation.

Search methods

We identified trials from a search of the Cochrane Dementia and Cognitive Improvement Group Specialized Register, last searched on 3 March 2022. We used the search terms: cognitive stimulation, reality orientation, memory therapy, memory groups, memory support, memory stimulation, global stimulation, cognitive psychostimulation. We performed supplementary searches in a number of major healthcare databases and trial registers to ensure the search was up-to-date and comprehensive.

Selection criteria

We included all randomised controlled trials (RCTs) of CS for dementia published in peer review journals in the English language incorporating a measure of cognitive change.

Data collection and analysis

We used standard methodological procedures expected by Cochrane. As CS is a psychosocial intervention, we did not expect those receiving or delivering CS to be blinded to the nature of the intervention. Where necessary, we contacted study authors requesting data not provided in the papers. Where appropriate, we undertook subgroup analysis by modality (individual versus group), number of sessions

and frequency, setting (community versus care home), type of control condition and dementia severity. We used GRADE methods to assess the overall quality of evidence for each outcome.

Main results

We included 37 RCTs (with 2766 participants), 26 published since the previous update. Most evaluated CS groups; eight examined individual CS. Participants' median age was 79.7 years. Sixteen studies included participants resident in care homes or hospitals. Study quality showed indications of improvement since the previous review, with few areas of high risk of bias. Assessors were clearly blinded to treatment allocation in most studies (81%) and most studies (81%) reported use of a treatment manual by those delivering the intervention. However, in a substantial number of studies (59%), we could not find details on all aspects of the randomisation procedures, leading us to rate the risk of selection bias as unclear.

We entered data in the meta-analyses from 36 studies (2704 participants; CS: 1432, controls: 1272). The primary analysis was on changes evident immediately following the treatment period (median length 10 weeks; range 4 to 52 weeks). Only eight studies provided data allowing evaluation of whether effects were subsequently maintained (four at 6- to 12-week follow-up; four at 8- to 12-month follow-up). No negative effects were reported. Overall, we found moderate-quality evidence for a small benefit in cognition associated with CS (standardised mean difference (SMD) 0.40, 95% CI 0.25 to 0.55). In the 25 studies, with 1893 participants, reporting the widely used MMSE (Mini-Mental State Examination) test for cognitive function in dementia, there was moderate-quality evidence of a clinically important difference of 1.99 points between CS and controls (95% CI: 1.24, 2.74).

In secondary analyses, with smaller total sample sizes, again examining the difference between CS and controls on changes immediately following the intervention period, we found moderate-quality evidence of a slight improvement in self-reported QoL (18 studies, 1584 participants; SMD: 0.25 [95% CI: 0.07, 0.42]) as well as in QoL ratings made by proxies (staff or caregivers). We found high-quality evidence for clinically relevant improvements in staff/interviewer ratings of communication and social interaction (5 studies, 702 participants; SMD: 0.53 [95% CI: 0.36, 0.70]) and for slight benefits in instrumental Activities of Daily Living, self-reported depressed mood, staff/interviewer-rated anxiety and general behaviour rating scales. We found moderate-quality evidence for slight improvements in behaviour that challenges and in basic Activities of Daily Living and low-quality evidence for a slight improvement in staff/interviewer-rated depressed mood. A few studies reported a range of outcomes for family caregivers. We found moderate-quality evidence that overall CS made little or no difference to caregivers' mood or anxiety.

We found a high level of inconsistency between studies in relation to both cognitive outcomes and QoL. In exploratory subgroup analyses, we did not identify an effect of modality (group versus individual) or, for group studies, of setting (community versus care home), total number of group sessions or type of control condition (treatment-as-usual versus active controls). However, we did find improvements in cognition were larger where group sessions were more frequent (twice weekly or more versus once weekly) and where average severity of dementia among participants at the start of the intervention was 'mild' rather than 'moderate'. Imbalance in numbers of studies and participants between subgroups and residual inconsistency requires these exploratory findings to be interpreted cautiously.

Authors' conclusions

In this updated review, now with a much more extensive evidence base, we have again identified small, short-term cognitive benefits for people with mild to moderate dementia participating in CS programmes. From a smaller number of studies, we have also found clinically relevant improvements in communication and social interaction and slight benefits in a range of outcomes including QoL, mood and behaviour that challenges. There are relatively few studies of individual CS, and further research is needed to delineate the effectiveness of different delivery methods (including digital and remote, individual and group) and of multi-component programmes. We have identified that the frequency of group sessions and level of dementia severity may influence the outcomes of CS, and these aspects should be studied further. There remains an evidence gap in relation to the potential benefits of longer-term CS programmes and their clinical significance.

PLAIN LANGUAGE SUMMARY

Can cognitive stimulation benefit people with dementia?

Key messages

- For people with mild-to-moderate dementia, cognitive stimulation probably leads to small benefits in cognition (the general ability to think and remember).

- We found a range of other probable benefits, including improved well-being, mood and day-to-day abilities, but benefits were generally slight and, especially for cognition and well-being, varied greatly between studies.

- Most studies evaluated group cognitive stimulation. Future studies should try to clarify the effects of individual cognitive stimulation, assess how often group sessions should take place to have the best effect, and identify who benefits most from cognitive stimulation.

What is dementia?

Dementia is an umbrella term for numerous brain disorders. Alzheimer's disease is the most common of these. People of all ages can develop dementia, but most often it occurs in later life. People with dementia typically experience a decline in their cognitive abilities, which can impair memory, thinking, language and practical skills. These problems usually worsen over time and can lead to isolation, upset and distress for the person with dementia and those providing care and support.

Cognitive stimulation

Cognitive stimulation (CS) is a form of 'mental exercise' developed specifically to help people with dementia. It involves a wide range of activities aiming to stimulate thinking and memory generally, including discussion of past and present events and topics of interest, word games, puzzles, music and creative practical activities. Usually delivered by trained staff working with a small group of people with dementia for around 45 minutes twice-weekly, it can also be provided on a one-to-one basis. Some programmes have trained family carers to provide CS to their relative.

What did we want to find out?

We wanted to find out if CS was better for people living with dementia than usual care or unstructured social activities to improve:

- cognitive abilities (including memory, thinking and language skills)

- well-being and mood

- day-to-day abilities

- distress and upset for the person with dementia and/or carers

We also wanted to find out if family carers experienced any changes associated with the person with dementia receiving CS or if there were any unwanted effects.

What did we do?
We searched for studies that looked at group or individual CS compared with usual care or unstructured social activity in people living with dementia.

We compared and summarised the results of the studies and rated our confidence in the evidence, based on factors such as study methods and sizes.

What did we find?

We found 37 studies involving 2766 participants with mild or moderate dementia and an average age of 79 years. The biggest study involved 356 participants, the smallest 13. The studies were conducted in 17 countries from five continents, with most in Europe. Fewer than half (16) included participants living in care homes or hospitals. The length of the trials varied from four weeks to two years. Sessions per week varied from one to six. The overall number of sessions varied from eight to 520. Most studies lasted for around 10 weeks, with around 20 sessions. Most studies offered CS in groups, with just eight examining individual CS.

Main results

No negative effects were reported. We found that CS probably results in a small benefit to cognition at the end of the course of sessions compared with usual care/unstructured activities. This benefit equates roughly to a six-month delay in the cognitive decline usually expected in mild-to-moderate dementia. We found preliminary evidence suggesting that cognition benefited more when group sessions occurred twice weekly or more (rather than once weekly) and that benefits were greater in studies where participants' dementia at the outset was of mild severity.

We also found that participants improved on measures of communication and social interaction and showed slight benefits in day-to-day activities and in their own ratings of their mood. There is probably also a slight improvement in participants' well-being and in experiences that are upsetting and distressing for people with dementia and carers. We found CS probably made little or no difference to carers' mood or anxiety.

What are the limitations of the evidence?

Our confidence in the evidence is only moderate because of concerns about differences in results between studies. We cannot be certain of the exact reasons for these differences, but we noted that studies varied in:

• the way CS was delivered (individually, in groups, using an app) and the programme of activities included

• who delivered the programme (trained professionals, care workers, family carers)

• the frequency of sessions (1 per week to 5 per week)

• the duration of the programme (from 4 weeks to 1 or 2 years)

• the type(s) of dementia with which participants were diagnosed and the severity of the dementia

• whether participants lived in care homes and hospitals or in their own homes

We were unable to examine as many of these sources of potential difference as would have been desirable because of the relatively small number of studies reflecting each aspect.

How up-to-date is this evidence?

This review updates our previous review from 2012, with evidence up-to-date to March 2022.

SUMMARY OF FINDINGS

Summary of findings 1. Cognitive stimulation compared to no cognitive stimulation (post-treatment) in people with dementia

Cognitive stimulation compared to no cognitive stimulation (post-treatment) in people with dementia

Patient or population: people with dementia
Setting: care homes and long-term care facilities; community settings including daycare and outpatients
Intervention: cognitive stimulation
Comparison: no cognitive stimulation (post-treatment)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no cognitive stimulation (post-treatment)	Risk with cognitive stimulation				
Cognition Assessed with various brief cognitive tests including: ADAS-Cog, MMSE, Global Cognitive Score, Mattis Dementia Rating Scale, MoCA, ACE-III, CAM-COG DS, ENB2		SMD 0.4 SD higher (0.25 higher to 0.55 higher)	-	2340 (34 RCTs)	⊕⊕⊕⊖ Moderate [a]	Cognitive stimulation probably results in a small increase in cognition.
Quality of Life: self-report Assessed with: QoL-AD (17 studies) and EQ-5D (1 study)		SMD 0.25 SD higher (0.07 higher to 0.42 higher)	-	1584 (18 RCTs)	⊕⊕⊕⊖ Moderate [a]	Cognitive stimulation probably results in a slight increase in self-reported quality of life.
Communication and social interaction Assessed with: Holden Communication Scale; NOSGER Social Behaviour subscale; Narrative language - communicative abilities		SMD 0.53 SD higher (0.36 higher to 0.7 higher)	-	702 (7 RCTs)	⊕⊕⊕⊕ High	Cognitive stimulation results in an increase in communication and social interaction.
Mood: self-reported Assessed with: Geriatric Depression Scale (14; 15 and 30-item versions); HADS Depression Scale; CESD-R; Cornell Scale for Depression in Dementia (self-report)		SMD 0.11 SD higher (0.08 lower to 0.31 higher)	-	787 (10 RCTs)	⊕⊕⊕⊕ High	Cognitive stimulation results in a slight improvement in self-reported mood.
Mood: interviewer/staff-rated		SMD 0.35 SD higher	-	1011 (11 RCTs)	⊕⊕⊖⊖ Low [bc]	Cognitive stimulation may result in a slight improve-

Assessed with Cornell Scale for Depression in Dementia; NOSGER-Mood subscale; Montgomery-Asberg Depression Rating Scale		(0.09 higher to 0.61 higher)				ment in mood rated by an interviewer or by staff.
Instrumental ADL Assessed with: Lawton Brody IADL scale; Disability Assessment for Dementia; NOSGER IADL subscale; Bristol Activities of Daily Living Scale; ADCS-ADL scale; Rapid Disability Rating Scale		SMD 0.15 SD higher (0.04 higher to 0.26 higher)	-	1318 (13 RCTs)	⊕⊕⊕⊕ High	Cognitive stimulation results in a slight increase in Instrumental ADL.
Behaviour that challenges Assessed with: NPI; NPI-Agitation subscale; NOSGER-Challenging Behaviour subscale; BEHAVE-AD; Dementia Behaviour Disturbance Scale		SMD 0.18 SD higher (0.01 lower to 0.38 higher)	-	1340 (12 RCTs)	⊕⊕⊕⊖ Moderate [b]	Cognitive stimulation probably results in a slight improvement in behaviour that challenges.

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

ADL: activities of daily living; **CI:** confidence interval; **OR:** odds ratio; **RR:** risk ratio

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

[a] Downgraded one point for inconsistency as moderate heterogeneity was present.
[b] Downgraded one point for inconsistency as substantial heterogeneity was present.
[c] Downgraded one point for imprecision as 95% CIs included both a clinically important and a negligible benefit.

Summary of findings 2. Group cognitive stimulation compared to no cognitive stimulation (post-treatment) in people with dementia

Group cognitive stimulation compared to no cognitive stimulation (post-treatment) in people with dementia

Patient or population: people with dementia
Setting: care homes and long-term care facilities; community settings including daycare and outpatients
Intervention: group cognitive stimulation
Comparison: no cognitive stimulation (post-treatment)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no cognitive stimulation (post-treatment)	Risk with group cognitive stimulation				
Cognition Assessed with various brief cognitive tests including: ADAS-Cog, MMSE, Global Cognitive Score; Mattis Dementia Rating Scale, MoCA, ENB2		SMD 0.43 SD higher (0.26 higher to 0.59 higher)	-	1637 (27 RCTs)	⊕⊕⊕⊖ Moderate [a]	Group cognitive stimulation probably results in a small increase in cognition.
Quality of Life: self-report Assessed with: QoL-AD		SMD 0.28 SD higher (0.05 higher to 0.52 higher)	-	1058 (13 RCTs)	⊕⊕⊖⊖ Low [bc]	Group cognitive stimulation may result in a slight increase in self-reported quality of life.
Communication and social interaction Assessed with: Holden Communication Scale; NOSGER-Social Behaviour subscale; Narrative language - communicative abilities		SMD 0.53 SD higher (0.36 higher to 0.7 higher)	-	702 (7 RCTs)	⊕⊕⊕⊕ High	Group cognitive stimulation results in an increase in communication and social interaction.
Mood: self-reported assessed with: Geriatric Depression Scale (14 and 30-item versions); CESD-R		SMD 0.2 SD higher (0.06 lower to 0.45 higher)	-	299 (6 RCTs)	⊕⊕⊕⊖ Moderate [d]	Group cognitive stimulation probably results in a slight improvement in self-reported mood.
Mood: interviewer/staff-rated Assessed with: Cornell Scale for Depression in Dementia; NOSGER-Mood subscale; Montgomery-Asberg Depression Rating Scale		SMD 0.4 SD higher (0.14 higher to 0.67 higher)	-	959 (10 RCTs)	⊕⊕⊕⊖ Moderate [b]	Group cognitive stimulation probably results in a small improvement in interviewer/staff-rated mood.
Instrumental ADL Assessed with: Lawton-Brody IADL scale; Disability Assessment for Dementia; NOSGER IADL subscale; ADCS-ADL scale; Rapid Disability Rating Scale		SMD 0.2 SD higher (0.05 higher to 0.35 higher)	-	687 (8 RCTs)	⊕⊕⊕⊕ High	Group cognitive stimulation results in a slight increase in Instrumental ADL.
Behaviour that challenges assessed with: NPI; NPI-Agitation subscale; NOSGER-Challenging Behaviour		SMD 0.33 SD higher (0.11 higher to 0.54 higher)	-	754 (8 RCTs)	⊕⊕⊕⊖ Moderate [a]	Group cognitive stimulation probably results in a slight

subscale; BEHAVE-AD; Dementia Behaviour Disturbance subscale				improvement in behaviour that challenges.

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

ADL: activities of daily living; **CI:** confidence interval; **OR:** odds ratio; **RR:** risk ratio

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

[a] Downgraded one point for inconsistency as moderate heterogeneity was present.
[b] Downgraded one point for inconsistency as substantial heterogeneity was present.
[c] Downgraded one point for imprecision as 95% CIs included both a clinically important and a negligible effect.
[d] Downgraded one point for imprecision as fewer than 400 participants.

BACKGROUND

Description of the condition

Dementia is widely regarded as one of the greatest current challenges facing health and social care globally.

The dementias comprise a number of neurodegenerative disorders of the brain, which have in common that they:

- develop during life (i.e. they represent a change from a previous level of function or ability)
- lead to impairment in a number of areas of cognitive function, typically including memory and orientation but also language skills, reasoning, judgement, visuo-spatial skills, executive function and practical abilities may be affected
- have an impact on day-to-day life abilities
- may have an impact on personality and social relationships
- are usually progressive

The most common types of dementia are Alzheimer's disease and vascular dementia (Alzheimer's Research UK 2022). Mixed-types of dementia are common, with both Alzheimer and vascular changes evident at postmortem in the brains of people who have developed dementia in late life. The severity of dementia is often described as 'mild' in the early stages when, with support, the person is able to continue with many activities; 'moderate' when more support and personal care is required; and 'severe' when the person may need help with almost all aspects of day-to-day life. It is estimated that 55% of people with dementia will have a mild dementia, 32% moderate and 12% severe (Prince 2014).

It is estimated that there were over 55 million people worldwide living with dementia in 2020 and this number is projected to almost double every 20 years, reaching 78 million in 2030 and 139 million in 2050 (ADI 2022). The increase in numbers with dementia reflects the growth globally in the numbers of people living into later life, as the risk of developing dementia increases markedly with age. For example, whilst 2% of those age 65 to 69 years will have a dementia, this rises to 20% of those aged 85 to 89 years (Alzheimer's Research UK 2022).

Taking together, the direct costs of medical care and of social care and of costs attributed to the unpaid care provided by families and others, the annual global cost of dementia is now above US$1.3 trillion and is expected to rise to US$2.8 trillion by 2030. ADI 2022 point out that, if global dementia care were a country, it would be the 14th largest economy in the world.

In the UK, it is estimated that there are 944,000 people living with dementia, of whom 42,000 are aged under 65 years (Alzheimer's Research UK 2022). Of those aged 65 years and over, 39% live in care homes. People with dementia form the majority of care home residents in the UK, with 70% of care home residents having dementia (Alzheimer's Research UK 2022). However, most people with dementia live in their own homes, often with support from family members. This support carries a considerable human as well as economic cost. It is estimated that 1.1 billion hours are spent on unpaid care, from family and friends, for people with dementia each year and 36% of carers spend more than 100 hours per week providing care (Alzheimer's Research UK 2022). Nearly half (49%) of family carers report a significant sense of burden and nearly a third report depression (Collins 2020) and anxiety (Kaddour 2020).

Female carers are especially at risk of depression, and around two-thirds of carers for people with dementia are women (Alzheimer's Research UK 2022). The majority of carers (63.5%) report that they have had no or not enough support (Alzheimer's Research UK 2022).

Medications (notably acetylcholinesterase inhibitors) have been available for some years for Alzheimer's disease and these are seen as offering symptomatic help (Alzheimer's Research UK 2022). Non-pharmacological interventions are viewed as a key component of post-diagnostic support, with the potential for enhancing well-being and confidence, but Alzheimer's Society 2022 point out that, in the UK at least, provision of such support for people with dementia and their carers after diagnosis has often been lacking.

Description of the intervention

Interventions with a cognitive focus have a long history of development and application in dementia care (Woods 1977; Woods 2018a). Clare and Woods (Clare 2004) distinguished **'Cognitive stimulation'** from other therapeutic approaches with a cognitive focus, proposing the following definition:

'Cognitive stimulation is engagement in a range of activities and discussions (usually in a group) aimed at general enhancement of cognitive and social functioning.'

In contrast, **'Cognitive training'** is defined as:

'guided practice on a set of standard tasks designed to reflect particular cognitive functions; a range of difficulty levels may be available within the standard set of tasks to suit the individual's level of ability. It may be offered in individual or group sessions, with pencil and paper or computerised exercises.'

and **'Cognitive rehabilitation'** as:

'an individualised approach where personally relevant goals are identified and the therapist works with the person and his or her family to devise strategies to address these. The emphasis is on improving performance in everyday life rather than on cognitive tests, building on the person's strengths and developing ways of compensating for impairments.'

Cochrane reviews have been carried out or are in progress for each of these three distinct approaches to improving outcomes for people living with dementia (Bahar-Fuchs 2019; Kudlicka 2019; Woods 2012) using agreed definitions to provide consistency in a field where there has been a tendency for these terms to be used interchangeably.

Descriptions of group-based activities and discussions aimed at general improvements in cognitive function, communication and social well-being can be traced back for over 50 years. Reality orientation (RO) (Taulbee 1966) was developed in the late 1950s as a response to confusion and disorientation in older patients in hospital units in the USA, and was the precursor of the cognitive stimulation approach. 'RO Classes' were held for 30 minutes once or twice per day. Basic personal and current information was presented to the patient and a variety of materials used, such as individual calendars, word-letter games, building blocks and large-piece puzzles. A Reality Orientation board would be used in each session and would list the name of the unit and its location, the day, date, weather, current events etc. The approach emphasised the engagement of nursing assistants in a hopeful, therapeutic process.

Following the first controlled evaluation of RO groups being reported in the UK by Brook 1975, a number of other small-scale controlled evaluations of RO groups followed (Holden 1995), with outcome measures typically including assessments of orientation, other aspects of cognitive functioning and level of independent functioning. A Cochrane review specifically examining Reality Orientation (Spector 2000a; Spector 2000b) concluded that there was some evidence that RO had benefits for people with dementia on both cognition and behaviour. However, RO has been little practised or researched since 1990 and has attracted some criticism (Burton 1982; Dietch 1989), especially for being applied in a mechanical, inflexible, insensitive and confrontational manner, with the potential for a negative effect on the person. Doubts were also raised about the clinical significance of any improvements (e.g. Powell-Proctor 1982); the person with dementia might now know what day of the week it was but would this have any meaningful impact on the person's life?

Subsequently, there began to be increasing discussion of 'cognitive stimulation', in relation to both normal ageing and dementia (Breuil 1994; Small 2002), Recognising that RO fitted well with this concept of cognitive stimulation and that it had the beginnings of an evidence base, attempts were made to harness its positive aspects whilst ensuring that it was implemented in a properly sensitive and respectful manner (Spector 2001; Woods 2002), in keeping with best practice person-centred care, as had been influentially articulated by Kitwood 1997. At the same time, developments in measurement of outcomes for people with dementia (e.g. Logsdon 2002) meant that quality of life could now be envisaged as an outcome of psychosocial interventions, with cognitive function and orientation no longer the only indices of effectiveness. Given the concerns discussed above of a possible negative effect, measures of quality of life, well-being and mood are highly pertinent outcome measures, alongside any cognitive benefits.

The publication of a manual for cognitive stimulation groups (Spector 2006) setting out in some detail the approach used in a large randomised controlled trial in the UK (Spector 2003) facilitated the implementation of the approach in the UK and internationally, and has formed the basis of a number of further evaluations (Lobbia 2019). Implementation was also facilitated by the recommendation in the UK (NICE-SCIE 2006) Guidelines that ‘people with mild/moderate dementia of all types should be given the opportunity to participate in a structured group cognitive stimulation programme’. Protocols have been developed for the cultural adaptation of the approach, whilst maintaining the key principles and elements (Aguirre 2014), and the approach is used in over 20 countries. Other manualised cognitive stimulation approaches have also been developed and evaluated such as 'NEUROvitalis' (Middelstädt 2016) and 'MAKS' (Graessel 2011) but, to date, have been less widely implemented internationally.

How the intervention might work

Attempts to understand the mechanisms by which cognitive stimulation might lead to benefits for people living with dementia are at an early stage. The approach can be seen as bringing together three aspects: a component of *generalised cognitive exercise* and a component of *social interaction and support*, both underpinned by a *person-centred approach*, which upholds the dignity and value of the person living with dementia (Woods 2018a). In recent years, a number of qualitative studies have been published, reporting the experiences of people with dementia and their caregivers in relation to both group and individual cognitive stimulation (Gibbor 2020a). All three aspects emerge clearly from these qualitative studies on cognitive stimulation (Gibbor 2020a; Leung 2018; Orfanos 2020).

In relation to *generalised cognitive exercise*, Gibbor 2020a, for example, found that 'continued stimulation' was important and Leung 2018 reported on people with dementia emphasising the importance of 'being mentally active' and adopting the principle 'use it or lose it'. Although treating the brain as a muscle that can be made stronger through exercise may be a crude analogy, the notion that mentally stimulating activities may improve cognitive function or even prevent cognitive decline has been widely discussed in relation to normal ageing (e.g. Gallacher 2005; Salthouse 2006). Indeed, it has been also sometimes characterised in the literature as ‘use it or lose it’ (Hultsch 1999), perhaps reflecting a general view that lack of cognitive activity hastens cognitive decline. In similar vein, the influential Lancet commission recommends 'keeping cognitively, physically, and socially active in midlife and later life' (Livingston 2020). The argument then is that if cognitively engaging activities can have a role in slowing or even preventing decline in cognitive functioning in older people in general, could they have an effect on people experiencing a dementia?

Whilst feasibly cognitive training could also be seen as exercising the brain, cognitive stimulation places cognitive exercise within a social and interpersonal context. The *social interaction and support* aspects emerge in qualitative themes such as 'being with others' and 'relationships' (Gibbor 2020a), 'opportunities to communicate' (Leung 2018) and 'importance of companionship and getting to know others', and 'togetherness and shared identity' (Orfanos 2020). The underpinning values of the *person-centred approach* are evident in themes such as 'confidence' and 'relaxed environment' (Gibbor 2020a) and 'providing supportive/non-threatening group environment' (Leung 2018) and 'group support' (Orfanos 2020). Orfanos 2020 highlighted the contribution of the group format to the benefits of cognitive stimulation, which raises the question of how the social and interpersonal component can best be retained in individual cognitive stimulation. Neuropsychological studies have also provided some support for the importance of the interpersonal aspects, in that several studies have shown specific improvements in language skills and performance following cognitive stimulation (Hall 2013; Spector 2010).

A potential mechanism for improvements related to cognitive stimulation arising from the social and value-based elements relate to the concept of 'excess disability' (Sabat 1994). This suggests that the person with dementia may, for a variety of reasons, not be able to function at their optimal or potential level. The person may be socially withdrawn, feel a lack of confidence, feel unmotivated or anxious, for example. Cognitive stimulation may then address some of these barriers, enhancing social confidence, giving a greater sense of purpose, engaging the person with others, promoting relaxation and enjoyment. The mental exercise component may interact with these aspects, providing practice and experience of cognitive successes in a supportive environment, when previously fear of failure has inhibited the making of any response or contributing to conversation. The inter-connection of the cognitive exercise and social elements is supported by findings that improvements in quality of life are mediated by improvements in cognition (Woods 2006). This suggests that there is one process of

change, rather than two separate processes: one involving mental exercise, improving cognition, and one involving social enjoyment, improving quality of life.

Brain imaging has also been used to seek to explore potential mechanisms of action of cognitive stimulation (Liu 2018; Liu 2021). The preliminary results using structural and functional MRI (magnetic resonance imaging) scanning suggest that cognitive stimulation 'maintains/enhances brain reserve both structurally and functionally', with increased connectivity despite an overall decline in volume of grey matter (which also occurred in control participants not receiving cognitive stimulation). The connectivity increased in a network believed to be important for 'sense-making' in a social context and mental self-representation, and it is suggested may relate to the person-centred nature of the intervention. However, these studies involved a small number of participants and this work requires further development.

Why it is important to do this review

Since the previous version of this review (Woods 2012), cognitive stimulation continues to be recommended in guidelines for dementia care practice such as those published by NICE 2018 and Alzheimer's Disease International (Prince 2011). In a report on post-diagnostic support in the UK, the Alzheimer's Society recommend that people diagnosed with dementia should be offered 'equitable access to non-pharmacological interventions as per national guidance, such as cognitive stimulation therapy (CST), and ensure all memory services have access to CST by April 2024' (Alzheimer's Society 2022). Such guidelines and recommendations have increased the implementation of cognitive stimulation around the world and, in a survey of UK Memory Clinics, Holden 2020 found that 87% of services responding were offering cognitive stimulation interventions. It is important that both clinical guidelines and clinical practice are based on up-to-date evidence, and so an update of the review is essential.

As well as the approach being more widely used, more adaptations to delivery are being made. Notably, the initial emphasis was on a group approach, and thematic analysis of interviews with participants in cognitive stimulation groups reinforces the perceived importance of the group experience (Orfanos 2020; Spector 2011). However, interest from family caregivers in using the approach at home has led to the development of individual cognitive stimulation, delivered on a one-to-one basis (Yates 2014). This has the disadvantage of lacking some of the social elements inherent in a group context, but does allow greater individualisation of activities, in relation to the person's interests and preferences. It also allows the inclusion of people with dementia who are unable to attend or participate in group sessions, for reasons of logistics or due to sensory difficulties which can prevent engagement in group activities. Individual cognitive stimulation can be delivered by paid care staff and professionals or by volunteers, as well as by family members. The onus is on the person facilitating the session to ensure that it remains a social experience, albeit on a smaller scale than in a group context. Recommendations have largely been made relating to group cognitive stimulation, so it is timely to consider also the evidence-base for individual cognitive stimulation.

Other developments include the use of digital technology to support cognitive stimulation (e.g. Rai 2020). This could assist those offering the intervention, by readily providing a wider range of activities, games and materials for use in group or individual sessions. The development of a cognitive stimulation-based television programme also offers an alternative mode of delivery (Streater 2020). When using digital approaches, care would need to be taken to retain the social stimulation aspect of the approach, of course. As a means of maintaining services during the Covid pandemic, Cheung 2021 described the development and feasibility of virtual cognitive stimulation groups using Zoom video-conferencing, retaining the social and peer-interaction elements of the approach despite remote delivery.

There is also a trend to combine cognitive stimulation with other interventions and offer a multi-component approach. For example, physical exercise, which is widely recognised as beneficial for older people and for people with dementia, has been incorporated to some extent in many cognitive stimulation programmes. In order to make sense of the evidence base, it is important to be as clear as possible regarding the interventions included in any systematic review, and to adopt clear operational definitions for inclusion of studies. In the current review, we only included studies where the predominant intervention met our definition of cognitive stimulation.

Given the long-standing and typically progressive nature of difficulties associated with dementia, there is also interest as to the required duration and intensity of these approaches. If, say, cognitive stimulation is associated with benefits over a three-month period, will these benefits continue and/or will continued input be required to maintain them? How frequent do cognitive stimulation sessions need to be to have the most beneficial effects? Does effectiveness vary with the level of impairment or between care home and community settings?

In the review, we consider cognitive functioning as a primary outcome, as this would appear to be the minimum expectation of a general approach with this focus. However, weight must also be given to indices of quality of life and well-being, in view of the early criticism that a difference of, say, a few points on a test of orientation may, in itself, be of marginal benefit to the person living with dementia. The effects on the person's everyday life need to be considered in evaluating the meaning of any changes observed for the individual and his or her supporters, and have been highlighted by participants (Spector 2011). The impact on family caregivers and care workers is also important to consider as they are key partners in the process of care. Where family caregivers have the additional responsibility of delivering the intervention, as in some applications of individual cognitive stimulation, the impact on them is especially relevant.

This updated review (the third, including the original superseded Reality Orientation review) will then need to consider developments in research and practice including the following: the modality of delivery (individual or group); the duration and frequency of the intervention; the involvement of family caregivers in delivery of the intervention; the use of digital technology; and the classification of multi-component interventions, as well as the setting of the intervention (care home versus community), the severity of dementia-related impairment and the type of comparison groups employed.

OBJECTIVES

To evaluate the evidence for the effectiveness of CS for people with dementia, including any negative effects, on cognition and other relevant outcomes, accounting where possible for differences in its implementation.

METHODS

Criteria for considering studies for this review

Types of studies

We included studies that met the following criteria:
• Randomised controlled trials (RCTs) that used cognitive stimulation as an intervention for people living with dementia.
• Control activity was no treatment, treatment-as-usual or a passive treatment such as basic social contact.
• Study was written in English and published in a peer-reviewed journal article.

We included in the review trials published since the previous version of this review that did not publish (or later supply) adequate information about study design and results but we did not include these studies in the meta-analysis. Details are noted in Characteristics of included studies.

Types of participants

- Participants with a diagnosis of dementia, according to established diagnostic criteria. The main diagnostic categories that we included were Alzheimer's disease, vascular dementia or mixed Alzheimer's and vascular dementia. We considered these diagnostic categories together. We did not include participants with mild cognitive impairment, where the extent of cognitive impairment or its effects on day-to-day function were insufficient to justify a dementia diagnosis. In this revised review, we were also able to consider studies focusing on other forms of dementia, such as Lewy body dementia and Parkinson's Disease Dementia.
- We evaluated severity of dementia through group mean scores, range of scores, or individual scores on a standardised scale such as the Mini-Mental State Examination (MMSE) (Folstein 1975) or Clinical Dementia Rating (CDR) (Hughes 1982). We included all levels of severity.
- Qualifying participants received the intervention in a range of settings, including their own home, as outpatients and in daycare and residential settings.
- We did not apply any specific restrictions regarding age.
- We included data from family caregivers, where available.
- We documented whether participants were receiving concurrent treatment with acetylcholinesterase inhibitors, where possible.

Types of interventions

- We considered studies for this review if they described a cognitive stimulation intervention targeting cognitive and social functioning. These interventions may also have been described as RO groups, sessions or classes.
- We adopted the definition of cognitive stimulation as proposed by Clare 2004. This meant that we excluded some studies which described their intervention as 'cognitive stimulation'. Interventions needed to offer exposure to generalised cognitive activities rather than training in a specific modality.
- Where the intervention included multiple components e.g. cognitive stimulation and physical exercise, we included the study only if more than 50% of the intervention time was spent in activities meeting our definition of cognitive stimulation.
- We did not include studies where the predominant intervention involved reminiscence, defined as 'the discussion of memories and past experiences with other people using tangible prompts such as photographs or music to evoke memories and stimulate conversation', as this intervention is the subject of a separate Cochrane review (Woods 2018b).
- Interventions were typically conducted in a group to enhance social functioning, but could involve family or paid caregivers offering cognitive stimulation on an individual basis.
- We included studies if a comparison was made to 'no treatment', 'standard treatment' or 'placebo'. We defined standard treatment as the treatment that was normally provided to people with dementia in the study setting and could include provision of medication, clinic consultations, contact with a community mental health team, daycare, or support from voluntary organisations. Placebo conditions could consist, for example, of an equivalent number of sessions in which general support, but no structured intervention, was offered. We did not consider comparisons with other activities or therapies such as cognitive training in this review.
- For inclusion of a study, we required a minimum intervention duration of one month. We noted the number of treatment sessions, but we did not apply any restrictions on this.

Types of outcome measures

- We included only studies including at least one measure of cognitive function, as this was the focus of the review.
- We considered outcomes in relation to the impact of the intervention on the person with dementia and on the primary family caregiver. Studies could present data in both these categories.
- We considered short-term (immediately after the intervention) and medium-term (follow-up one month to one year after the intervention finished) outcomes.
- We considered outcomes for the person with dementia and the caregiver where these were assessed using scores on standardised tests, rating scales and questionnaires.
- We noted rates of attrition and reasons for participants dropping out from the study.

Primary outcomes

For the primary outcome measure we sought to identify whether short-term changes were observed following the intervention for the person with dementia on performance on at least one test of cognitive functioning (including tests of memory and orientation).

Secondary outcomes

Outcomes for the person with dementia

We considered the following variables as secondary outcome measures for the person with dementia.

- Self-reported, clinically-rated or carer-reported (proxy) measures for mood of the person with dementia.

- Self-reported or carer-reported (proxy) quality of life or well-being measures for the person with dementia.
- Observer or carer ratings of everyday functioning (activities of daily living) of the person with dementia.
- Carer ratings of the participant's behaviour.
- Clinician or carer ratings of 'behaviour that challenges' relating to the person with dementia.
- Clinician or carer ratings of the communication and social interaction of the person with dementia.
- Self-reported quality of relationship with the carer.

'Carer' in this context included care staff as well as family caregivers.

Outcomes for the family caregiver

We considered all outcomes for the family caregiver as secondary. We considered the following outcomes for the family caregiver.

- Self-reported quality of life.
- Self-reported depression and anxiety.
- Self-reported burden, stress and coping.
- Self-reported quality of relationship with the person with dementia.
- Self-reported resilience.

Adverse outcomes

There is a potential risk that participants may find the process of cognitive stimulation over- or under-challenging, if not targeted at the appropriate level. We monitored the potential for adverse outcomes by observing negative responses on the outcome measures.

Search methods for identification of studies

Electronic searches

We searched the Cochrane Dementia and Cognitive Improvement Group Specialised Register, on 3 March 2022. The search terms used were: cognitive stimulation, reality orientation, memory therapy, memory groups, memory support, memory stimulation, global stimulation, cognitive psychostimulation.

The Register is maintained by the Information Specialists of the Cochrane Dementia and Cognitive Improvement Group and contains studies in the areas of dementia prevention, dementia treatment and cognitive enhancement in healthy populations. The studies are identified from:

1. monthly searches of a number of major healthcare databases: MEDLINE, Embase, CINAHL, PsycINFO and LILACS;
2. monthly searches of a number of trial registers: meta Register of Controlled Trials; Umin Japan Trial Register; WHO portal (which covers ClinicalTrials.gov; ISRCTN; Chinese Clinical Trials Register; German Clinical Trials Register; Iranian Registry of Clinical Trials and the Netherlands National Trials Register, plus others);
3. quarterly search of the Cochrane Central Register of Controlled Trials (CENTRAL) (*The Cochrane Library*);
4. six-monthly searches of a number of grey literature sources: Web of Knowledge Conference Proceedings; Index to Theses; Australasian Digital Theses.

Details of the search strategies used for the retrieval of reports of trials from the healthcare databases, CENTRAL and conference proceedings can be viewed in the 'methods used in reviews' section within the editorial information about the Dementia and Cognitive Improvement Group (CDCIG).

We ran additional searches in each of the sources listed above to ensure that the search for the review was as up-to-date as possible. The search strategies used can be seen in Appendix 1

Searching other resources

We searched the reference lists of full-text papers, including those of relevant published systematic reviews, for further references, and review authors searched personal holdings of references to reports and trials. We sent emails to authors of included RCTs asking for essential information, where this was not available in the publication, such as relevant statistics.

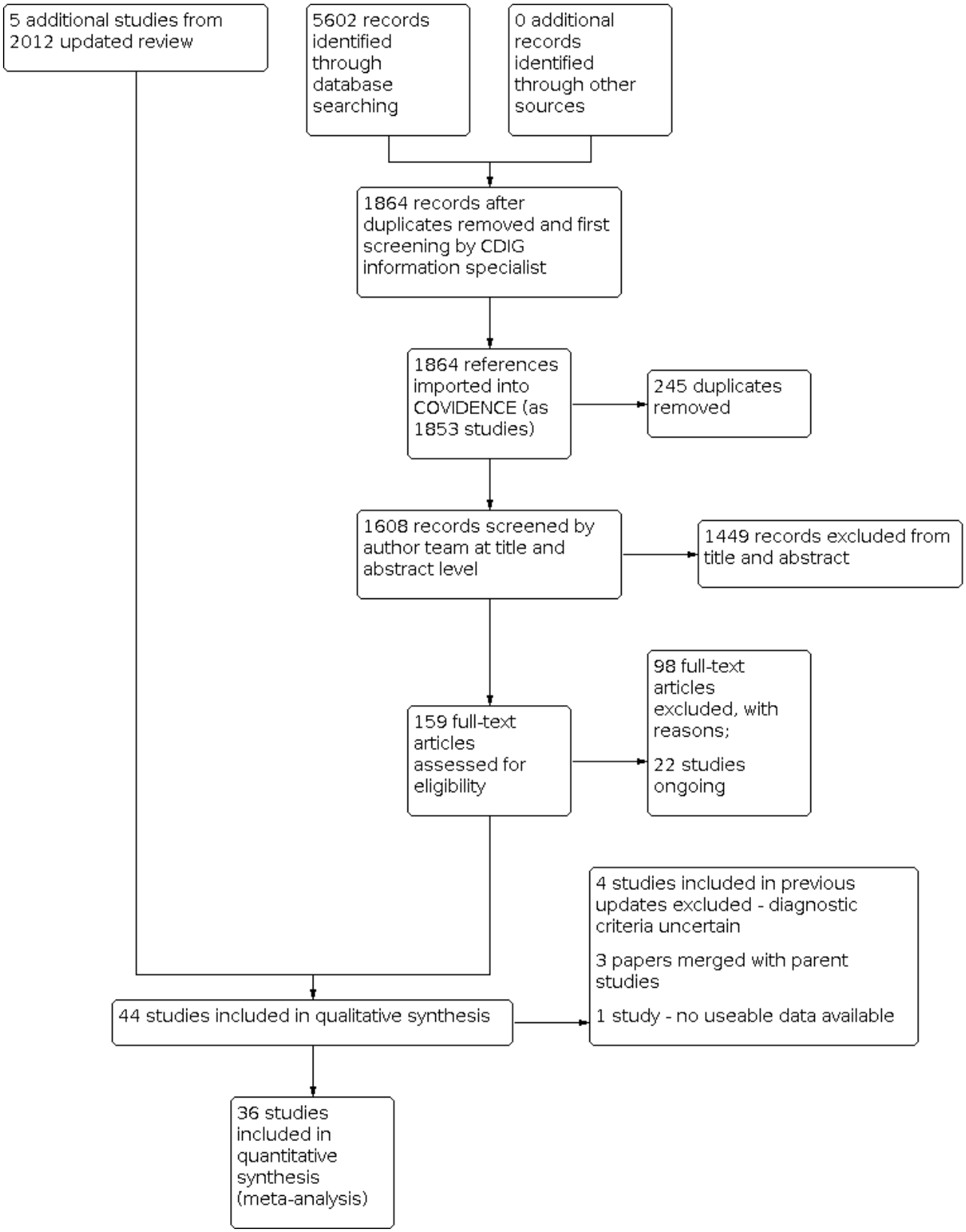
Data collection and analysis

Selection of studies

For this revised review, following deduplication, we imported into Covidence all references identified in the searches, their titles and abstracts. Three review authors (HR, EE and BW) undertook the selection of studies, with the title and abstract of each study being independently reviewed by two reviewers, using the inclusion and exclusion criteria detailed above. If one of the three reviewers had been involved in the study under consideration, s/he did not participate in the screening for that study. We automatically retained studies that had been included in previous versions of this review for further consideration. We discussed any disagreements and reached consensus by involving the third reviewer, where appropriate.

We excluded obviously irrelevant studies. We then obtained the full text of remaining studies and excluded studies that did not meet the inclusion criteria with reasons outlined in the Characteristics of excluded studies table. We collated multiple reports of the same study, so that each study, rather than each report, was the unit of interest in the review. We recorded the selection process in sufficient detail to complete a PRISMA flow diagram (Figure 1).

Figure 1. Study flow diagram



Data extraction and management

Three review authors (HR, EE and BW) independently extracted descriptive characteristics, study methodology data and study results from the included studies and recorded them on a data collection form. Two reviewers extracted the data from each study, ensuring that these reviewers had not been involved in the study in question. We compared the data and resolved any disagreements through consensus. We transferred extracted data to RevMan.

For each outcome measure, the authors sought to obtain data on every participant randomised irrespective of whether the participant was excluded or dropped out of the intervention or research (i.e. data from an intention-to-treat (ITT) analysis). If these data were not available in the published studies, the review authors sought the data of those who completed the trials. Where necessary, we sent emails to trial authors requesting additional information.

Assessment of risk of bias in included studies

For each trial, two of three review authors (HR, EE and BW) independently assessed the risk of bias using the Cochrane risk of bias (version 1) tool (Higgins 2011). We resolved any initial disagreements with the third author. We attempted to obtain additional information from study authors when this was required. Based on the methods detailed in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011), we classified each category of bias as 'low risk of bias,' 'high risk of bias' or 'unclear risk of bias.' An outline of this can be seen in Table 1. We did not rate 'Blinding of participants and personnel', as it is a given in research on psychosocial interventions that the person receiving the intervention and the person delivering it will be aware of the nature of the intervention. We expanded 'Other sources of bias' to identify whether a structured treatment manual had been used, and whether those delivering the intervention had received training and/or supervision, two important aspects of ensuring a consistent intervention of good quality. For selection bias, the meta-analysis included only trials with a low or unclear risk of bias, in order to meet the study inclusion criteria as a randomised controlled trial. The ratings assigned with respect to each study's risk of bias are summarised in the risk of bias tables, Figure 2 and Figure 3.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies

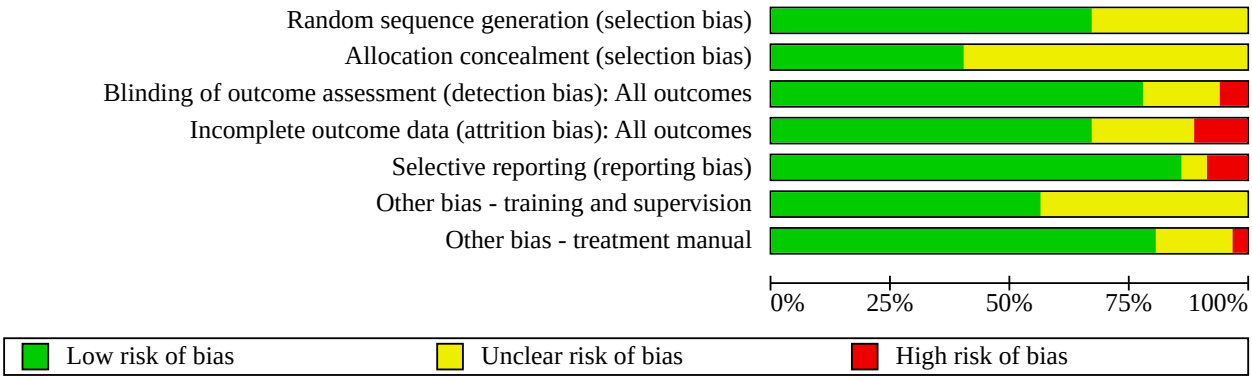


Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study

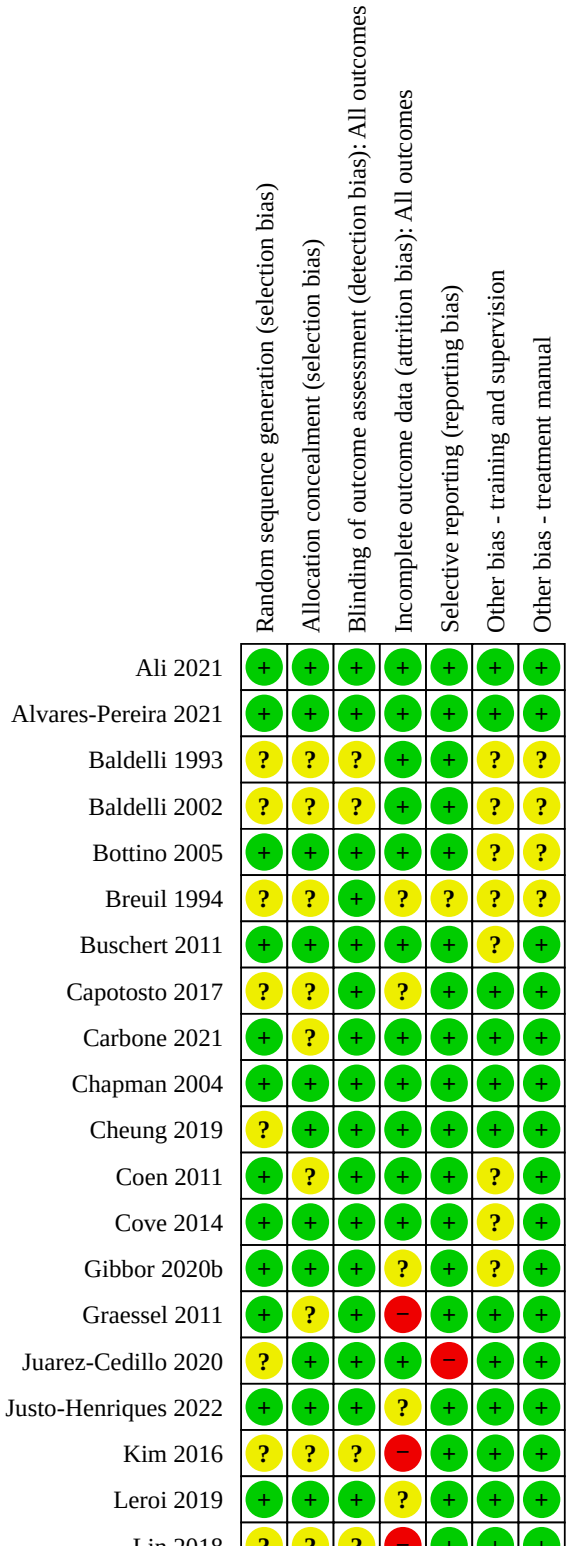
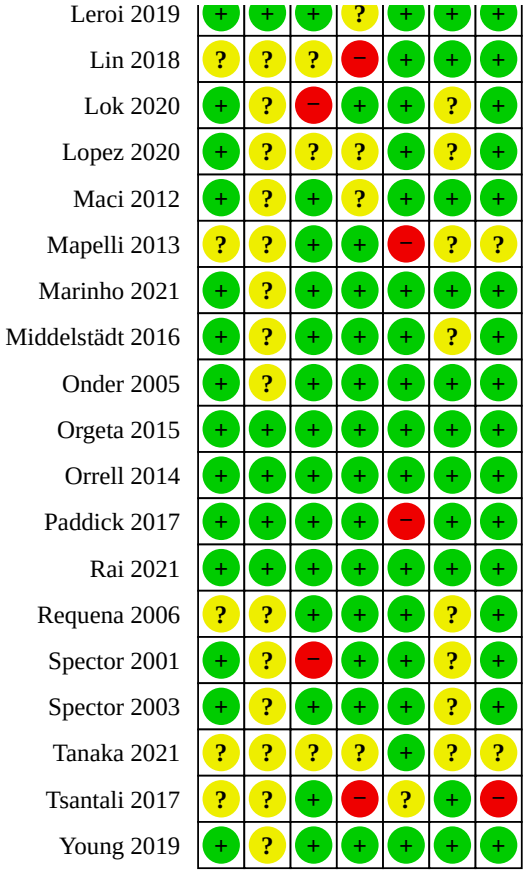


Figure 3. (Continued)



Measures of treatment effect

As the outcomes measured in clinical trials of dementia and cognitive impairment often arise from ordinal rating scales, where the rating scales had a reasonably large number of categories, we treated the data as continuous outcomes arising from a normal distribution. For meta-analysis of this type of data, the mean change scores from baseline, the standard deviation of the mean change and the number of participants for each treatment group at each assessment are required. The majority of study authors did not report change scores from baseline. We defined the baseline assessment as the latest available assessment prior to randomisation, but no longer than two months prior. Where change scores were not reported, we extracted the mean, standard deviation and number of participants for each treatment group at each time point and calculated the required summary statistics manually. In this case, we assumed a zero correlation between the measurements at baseline and assessment time. This method overestimates the standard deviation of the change from baseline, but this conservative approach is considered to be preferable in a meta-analysis. Some studies (e.g. Gibbor 2020b) reported mean differences and 95% confidence intervals, which enabled us to calculate the appropriate statistics in RevMan. Lok 2020 provided data as medians and interquartile ranges, from which we calculated the required summary statistics following the methods proposed by Wan 2014, as detailed in the Cochrane Handbook (sections 6.5.2.5 and 6.5.2.9).

The meta-analyses included the combination of data from trials that may not have used the same rating scale to measure a particular outcome. For example, cognition may have been measured by the MMSE in one study and the ADAS-Cog in another. In this situation we used the standardised mean difference (SMD; the absolute mean difference (MD) divided by the standard deviation) to measure the treatment difference. Where pooled trials used the same rating scale or test to measure an outcome, we used the MD.

To allow comparisons with other scales assessing similar outcomes, it was necessary to reverse the change scores on certain scales, for example, for the ADAS-Cog, where higher scores indicate worse cognitive performance.

Unit of analysis issues

In studies using a cross-over design, only data from the first treatment phase after randomisation were eligible for inclusion.

Three studies used cluster-randomisation. No analysable data were obtained from Lin 2018 and the studies reported by Cheung 2019 and Paddick 2017 were not large enough for planned adjustments to be made for clustering using estimated intraclass correlations.

Dealing with missing data

Where possible, review authors extracted data on all participants randomised. We preferred data from intention-to-treat analyses to per protocol or compliance analyses.

Assessment of heterogeneity

We performed assessments of heterogeneity using both the Chi^2 and I^2 statistic. Review authors followed guidance in the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2021),to interpret heterogeneity percentages (i.e. 0% to 40% might not be important; 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity; and 75% to 100% reflects considerable heterogeneity). Following these recommendations (Deeks 2021), we considered heterogeneity to be present when the Chi^2 statistic was significant at the P = 0.1 level, or when I^2 was greater than 40%. Where substantial heterogeneity was detected, we considered exploring the sources of heterogeneity by conducting subgroup analyses.

Assessment of reporting biases

If there were enough studies available, authors created a funnel plot to assess the risk of publication bias.

Data synthesis

The meta-analyses presented in this updated review provide overall estimates of the treatment effect using a random-effects model. In previous versions of the review, we have preferred a fixed-effects model where heterogeneity is low, but with a substantial number of analyses showing high heterogeneity, this protocol change leads to a more consistent approach. As a result, confidence intervals may be broader than would have been obtained from a fixed-effect model.

Subgroup analysis and investigation of heterogeneity

We performed planned subgroup analyses with respect to the modality of the cognitive stimulation intervention (individual versus group) and environmental context (care home versus community) and severity of dementia (mild versus moderate, considering MMSE scores of 20 and below as indicating 'moderate' and above 20 as 'mild' dementia, NICE-SCIE 2006). We undertook further subgroup analyses to investigate high levels of heterogeneity, considering feasible moderator variables such as frequency of sessions and duration of intervention. We planned a further subgroup analysis to compare studies where control participants took part in some form of alternate activity with those where the control condition was 'treatment-as-usual'. We only presented subgroup analyses where there were at least five studies per subgroup (Richardson 2019), to reduce concerns regarding covariate distributions. Where we presented subgroup analyses, we considered the difference between subgroups to be statistically significant where P < 0.10 (Richardson 2019). We considered all subgroup analyses as exploratory.

Sensitivity analysis

In planned sensitivity analyses we explored:

a) whether the trial of maintenance cognitive stimulation reported by Orrell 2014, where the control group had attended 14 sessions of cognitive stimulation before randomisation, might have an influence on the results of one session per week studies.

b) whether the findings on cognition, pooling different assessment measures, were reflected in the findings for the two most widely used assessment measures, the ADAS-Cog and the MMSE, for which figures for recognised minimum clinically important differences were available.

Summary of findings and assessment of the certainty of the evidence

We used GRADE methods to rate the quality of evidence (high, moderate, low or very low) behind each effect estimate in the review (Guyatt 2011). This rating referred to our level of confidence that the estimate reflected the true effect, taking account of the risk of bias in the included studies, inconsistency between studies, imprecision in the effect estimate, indirectness in addressing our review question and the risk of publication bias. We produced summary of findings tables for cognitive stimulation compared to no treatment to show the effect estimate and the quantity and quality of the supporting evidence for the following outcomes immediately post-treatment:

1. cognition
2. self-reported QoL
3. communication and social interaction
4. self-reported mood
5. interviewer/staff-rated mood
6. instrumental activities of daily living
7. behaviour that challenges

In view of the widespread use of group cognitive stimulation, and its specific recommendation in guidelines, we produced an additional table (with the same outcomes) to summarise the effects of group cognitive stimulation considered separately. We prepared the summary of findings tables using the GRADEpro GDT 2015 (gradepro.org).

RESULTS

Description of studies

Results of the search

From the searches of databases carried out since the previous version of this review was published (Woods 2012), we have identified a total of 5602 records (see Appendix 1; Figure 1). Following removal of duplicates and a first screening by the CDIG information specialist, 1864 records remained which we imported into COVIDENCE as 1853 studies. Further de-duplication resulted in 1608 studies, of which we excluded 1449 as it was clear from the title and/or abstract that the study did not meet the inclusion criteria.

We then obtained the full texts of the remaining 159 studies and the author team assessed their eligibility for inclusion. Twenty-two of these studies are ongoing (typically published on trial registries) and we excluded a further 98 studies at this point, leaving 39 studies to be transferred from COVIDENCE to RevMan. At this stage, it became clear that three of the 39 studies were in fact reports of specific aspects of larger studies (Graessel 2011; Orrell 2014), and so we merged these studies with their parent studies. We added five studies included in the previous version of this review which had not been re-included by the subsequent searches to these 36 studies (Baldelli 1993; Baldelli 2002; Bottino 2005; Spector 2001; Woods 1979). We recognised that some leeway regarding diagnostic criteria had been given in the previous version of the review, where the authors had stated, 'Older studies may be included where the review authors are satisfied

that the included population would now be described as having a dementia'. Given the growth of the evidence base, it now appeared timely to exclude the early, small-scale studies which had described participants as 'disorientated' and having 'significant memory impairment' (Woods 1979); 'demented/organic' (Wallis 1983); 'moderate to severe Impairment of cognitive functioning' (Baines 1987); and 'elderly patients with cognitive disturbances' (Ferrario 1991), predating current diagnostic criteria. Exclusion of these four studies resulted in inclusion of a total of 37 studies in this updated review, compared with 15 studies in the previous review (Woods 2012).

We present full details of the included studies and reasons for exclusion of selected excluded studies in the tables 'Characteristics of included studies' and 'Characteristics of excluded studies'. We present selected recent ongoing studies in 'Characteristics of ongoing studies'.

Included studies

We summarise key characteristics of the included studies in Table 2. Overall, the 37 included studies comprised 2766 participants, 1462 in the treatment groups and 1304 in the control groups. The 37 studies were carried out in 17 different countries, from five continents. The largest number of studies came from the UK (9) and Italy (7), with three from Germany and two each from China (Hong Kong), Spain, Portugal and Brazil. There was considerable variation in study sizes, with the mean intervention sample size being 39.5 (SD 39.1) and the mean control sample size being 35.2 (SD 35.3).The smallest study (Bottino 2005) had just 13 participants and the largest, (Orgeta 2015), a total of 356. Eight of the 37 studies had more than 100 participants in total (Alvares-Pereira 2021; Carbone 2021; Graessel 2011 (at 6 months); Onder 2005; Orgeta 2015; Orrell 2014; Spector 2003; and Young 2019). The included studies varied in many other aspects: (1) participant characteristics; (2) number and duration of cognitive stimulation sessions; (3) activities which defined cognitive stimulation; (4) the activity of the control group; and (5) outcome measures. We will consider these factors in turn.

1) Participant characteristics

The mean or median age of participants was over 80 years in 17 studies and over 70 years in all but one of the 36 studies reporting summary statistics for age. The exception was Ali 2021 where the mean age was 60.4 years, reflecting the more frequent younger onset of dementia in the people with an intellectual disability who participated in this study. The average mean age across the 36 studies was 79.4 years (median 79.7). Across the studies where the range of ages was reported, the lowest age was 50 and the highest 102 years, with most including participants aged 90 years and over.

Ten of the studies were carried out entirely in care homes, nursing homes or hospitals (Baldelli 1993; Baldelli 2002; Capotosto 2017; Coen 2011; Gibbor 2020b; Graessel 2011Lin 2018; Mapelli 2013; Middelstädt 2016; Tanaka 2021). The participants included in the Ali 2021; Alvares-Pereira 2021; Carbone 2021; Orrell 2014; Spector 2001 and Spector 2003 studies were recruited from both residential care homes and community settings, whilst the remaining 21 studies were recruited exclusively from community settings, including day-centres and outpatient populations. This is in contrast to the previous version of this review (Woods 2012), where most of the studies included participants from care homes and hospitals.

Eleven of the studies restricted participation to people with a diagnosis of Alzheimer's disease, with three citing NINCDS-ADRDA criteria and one DSM-V. One study (Leroi 2019) included people with Parkinson's disease dementia and dementia with Lewy Bodies, according to established diagnostic criteria. One study (Ali 2021) included adults with an intellectual disability with a confirmed clinical diagnosis of dementia. The remaining 24 studies did not specify subtypes of dementia, typically including Alzheimer's, vascular dementia and mixed dementia. Nine of these studies reported using DSM-IV criteria, six DSM-V, two specified ICD-10 and one used DSM-III.

In ten of the eleven studies where all participants had a diagnosis of Alzheimer's, all participants were receiving medication, most commonly being on a stable dose of an acetylcholinesterase inhibitor (ACHEI), such as donepezil or rivastigmine (Bottino 2005; Buschert 2011; Chapman 2004; Lok 2020; Lopez 2020; Maci 2012; Onder 2005; Requena 2006; Tsantali 2017). In the Buschert 2011 and Maci 2012 studies, other medications such as memantine were also used, and Kim 2016 did not detail the specific medications received. The one study (Baldelli 1993) where all participants had a diagnosis of Alzheimer's disease but were not receiving medication predated their wide availability. Several studies, where a mix of subtypes of dementia were included, reported the proportion receiving ACHEIs: Marinho 2021: 100%; Orgeta 2015: 76%; Rai 2021: 71%; Cove 2014: 62%; Ali 2021: 45%; Orrell 2014: 32%; Graessel 2011: 13.5%; and Tanaka 2021: 13%.

As an indication of the severity of cognitive impairment in participants in the included studies, the mean baseline MMSE score for the 30 studies reporting this information was 19.4 (SD 3.0), ranging from 13.1 (Spector 2001) to 24.9 (Buschert 2011), with all studies in the mild or moderate range of cognitive impairment. For 16 of these studies, the MMSE was over 20, suggesting a relatively mild degree of cognitive impairment (NICE-SCIE 2006). Of the remaining 14 studies, in eleven, the mean MMSE score at baseline was less than 18 and in five of these, all including participants from care home/inpatient settings (Graessel 2011; Lin 2018; Spector 2001; Spector 2003; Tanaka 2021), was less than 16. All of the 14 studies where the mean MMSE score was less than 20 at baseline offered cognitive stimulation on a group basis rather than individually.

2) Length, number and duration of sessions

The length of the intervention varied from four weeks, the minimum for inclusion in the review (Baldelli 1993), to 24 months (Requena 2006), with a median of 10 weeks. Requena 2006 presented data from both the 12-month and 24-month time point in their study. As there was less attrition at the 12-month time point, and this was more comparable (although still longer than most) in duration to the other studies, the 12-month data were used in combination with other studies in the meta-analyses, with the 24-month data reported separately. The stated length of sessions varied from 30 minutes to 120 minutes. The median session length across the studies was 45 minutes, and the median frequency was twice a week, ranging from once to six times a week. Seven studies offered sessions on one day per week, including Orrell 2014, where a six-month period of 'maintenance cognitive stimulation' followed on from a seven-week twice-weekly programme for those randomised to the intervention group, and Marinho 2021, where two 45-minute sessions took place on the same day, separated

by a short break, which we have viewed as, in effect, a 90-minute session.

The total possible exposure to the intervention varied dramatically, from eight sessions (Chapman 2004; Cheung 2019) to 300 sessions (Graessel 2011) and 520 sessions for the Requena 2006 24-month evaluation point. The median number of sessions was 20.

3) Activities during cognitive stimulation

Eight of the studies (Ali 2021; Gibbor 2020b; Justo-Henriques 2022; Leroi 2019; Onder 2005; Orgeta 2015; Rai 2021; Tsantali 2017) offered cognitive stimulation on an Individual basis, facilitated by a carer, rather than as a small group activity, with all, except Onder 2005, having been published since the previous version of this review. Although there is considerable overlap between the activities offered in individual and group studies, this difference of modality requires adaptations to certain group activities, removes the element of peer support and encouragement and places the onus on the carer to maintain the social and 'fun' elements often said to be essential aspects of the approach (Spector 2020). In four of the studies, individual sessions were predominantly delivered by a family caregiver, but four of the studies (Ali 2021; Gibbor 2020b; Justo-Henriques 2022; Tsantali 2017) have relied largely on paid carers or professionals.

Just three of the studies included in this review described the intervention used as reality orientation (Baldelli 1993; Baldelli 2002; Onder 2005), with four early studies using this term now excluded (Baines 1987: Ferrario 1991; Wallis 1983; Woods 1979). These studies described the use of an RO board and discussion of current orientating information through newspapers, photographs, calendars and clocks etc. with materials and activities selected to stimulate all five senses.

Sixteen of the included studies have used the 'Making a difference' treatment manuals, based initially on the development work reported by Spector 2001. These cover a 14-session group programme (Spector 2006; revised by Spector 2020), a six-month maintenance group cognitive stimulation programme (Aguirre 2011) and a manual designed for family caregivers to deliver individual cognitive stimulation (Yates 2014). These manuals provide detailed session plans and a range of suggested activities to suit different ability levels and interests. Activities in the sessions were designed with four themes: (1) the senses, (2) remembering the past, (3) people and objects, and (4) everyday practical issues. Activities included naming objects and people, association of words, remembering the past, discussion of hobbies, activities and current affairs, using money, knowing the way around and orientation topics. Sessions typically start with a period of introductions and orientation, followed by the main activity, closing with a summary, refreshments and farewells. Woods 2018a highlighted the considerable overlap between the suggested activities and those typically associated with RO, from which this approach developed. The 14-session programme - adapted as necessary - was used in ten studies (Alvares-Pereira 2021; Capotosto 2017; Carbone 2021; Coen 2011; Cove 2014; Marinho 2021; Paddick 2017; Spector 2001; Spector 2003; Young 2019), the maintenance programme was used by Orgeta 2015 and the individual programme by Ali 2021; Gibbor 2020b; Leroi 2019; Orrell 2014 and Rai 2021 (adapted as an app). In addition, Justo-Henriques 2022 reported that their individual cognitive stimulation protocol was largely based on Spector 2006 and two further studies, whilst not specifically referring to use of the treatment manual, appear to have been strongly influenced by it, using the session themes set out by Spector 2006 in fourteen-session (Lok 2020) and ten-session (Lin 2018) programmes, respectively.

The remaining studies have adopted a variety of protocols that were judged to meet our definition of 'cognitive stimulation'. For example, Tsantali 2017 stated that their individual cognitive stimulation programme comprised simple cognitive tasks, not targeted at specific cognitive impairments, such as drawing and painting, puzzles, looking at and naming images and listening to music and singing. In contrast, Lopez 2020 followed a similar session plan to 'Making a difference', with a main activity sandwiched between an orientation period and a summary discussion, but their activities were more specifically linked to cognitive domains such as memory, praxis, language and executive functions. The cognitive stimulation intervention described by Mapelli 2013 started with initial personal, spatial, and temporal orientation before proceeding to exercises offering structured stimulation of cognitive domains including memory, language, spatial and temporal orientation and attention with exercises adapted to the level of dementia. Similarly, Breuil 1994 introduced a number of more specific cognitive activities including drawing, associating words, object naming and categorising.and grouped into 3 levels of difficulty.

Chapman 2004 reported topics including current events, discussion of hobbies and activities, education regarding Alzheimer's disease, life story work, and links with daily life with groups of six to seven participants. Bottino 2005 described temporal and spatial orientation, discussion of interesting themes, reminiscence activities, naming people, planning of daily activities and use of calendars and clocks and other external memory aids. Requena 2006 described, for groups of five people with dementia, visual images being shown on a TV screen from a computer reflecting seven themes: orientation, bodily awareness, family and society, caring for oneself, reminiscing, household activities, animals, people and objects. These were accompanied by questions for discussion.

Buschert 2011 described their intervention as 'multi-component' and, for participants with mild Alzheimer's disease, this included exercises to stimulate social interaction, e.g. remote memory, reminiscence, language, imagination and creativity as well as exercises concerning global cognition and specific cognitive functions, such as memory, attention and executive functions and day-to-day activities. Kim 2016 also adopted a multi-domain approach to cognitive stimulation, including elements of music, art, reminiscence and horticultural activities. Middelstädt 2016 incorporated a period of relaxation alongside cognitive exercises and sensory stimulation in their 'NEUROvitalis senseful' programme, with groups of three to five people with dementia. Juarez-Cedillo 2020 evaluated the 'SADEM' programme, again described as 'multi-component', including some elements of daily activities as well as cognitive stimulation and orientation.

Several studies included physical activities as a (subsidiary) component of the overall approach. The participants in the study reported by Maci 2012 engaged in activities related to spatiotemporal orientation, memory, executive skills, and language, as well as a physical exercise session. Tanaka 2021 described a third of each 45-minute group session being devoted to seated exercises. Graessel 2011 described the MAKS programme,

including motor stimulation (M), practising activities of daily living (A), cognitive stimulation (K) and a short introductory phase (S) including a spiritual element ('for example, discussing topics such as happiness or singing a song, usually a hymn').

Finally, Cheung 2019 focused on the fun and enjoyment aspects of cognitive stimulation, developing 'CoS-Play', incorporating games and 'toys' providing visual, auditory and tactile stimulation. Activities included card games, balloon games, telling stories, making handicrafts, social interaction and playing percussion instruments. Participants were encouraged to 'exercise their creativity' in a non-judgemental, respectful and cheerful environment.

Apart from the use by Requena 2006 of digital images, only Rai 2021 of the included studies focused on cognitive stimulation using digital technology (an individual cognitive stimulation app 'Thinkability' based closely on the Yates 2014 iCST manual).

4) Control group(s) activities

'Treatment-as-usual' or no treatment was the control condition in 28 of the 37 studies (Alvares-Pereira 2021; Baldelli 1993; Bottino 2005; Breuil 1994; Chapman 2004; Coen 2011; Gibbor 2020b; Graessel 2011; Juarez-Cedillo 2020; Justo-Henriques 2022; Kim 2016; Leroi 2019; Lin 2018; Lok 2020; Lopez 2020; Maci 2012; Mapelli 2013; Marinho 2021; Middelstädt 2016; Onder 2005; Orgeta 2015; Orrell 2014; Rai 2021; Spector 2001; Spector 2003; Tanaka 2021; Tsantali 2017; Young 2019). Ali 2021, Cove 2014 and Paddick 2017 compared their cognitive stimulation groups with 'waiting-list controls', who effectively also received treatment-as-usual during the intervention period considered in this review. In those studies where all participants were also taking AChEIs or other medications, the control group was typically monitored in relation to the medication (e.g. Bottino 2005; Chapman 2004; Kim 2016; Lok 2020; Onder 2005; Requena 2006). Baldelli 2002 engaged both the control and cognitive stimulation participants in a physical therapy programme.

Five studies described more specific alternative activities for their control participants, meeting the inclusion criteria of being loosely structured and not comprising an alternative therapy. Requena 2006 reported that their control participants watched TV whilst the cognitive stimulation groups were in session and Buschert 2011 asked control participants to complete pencil and paper tasks at home, encouraged by monthly group meetings. Cheung 2019 described the control group participating in social activities (such as reading newspapers or watching TV) while Capotosto 2017 and Carbone 2021 reported an 'active control', involving reading, discussions and creative activities.

Three studies did include comparisons with other structured activities/therapies. These comparisons did not meet the inclusion criteria for this review. Lin 2018 had a reminiscence therapy comparison group and Tsantali 2017 a comparison with individual cognitive training. Mapelli 2013 reported (in addition to treatment-as-usual) a structured 'occupational therapy placebo' group.

5) Outcome measures

As a condition of inclusion, cognitive tests were used in all the studies. Twenty-seven studies used the MMSE (Folstein 1975) and 22 studies used the Alzheimer's Disease Assessment Scale - Cognitive (ADAS-Cog) (Rosen 1984). Unfortunately, only longer-term follow-up data on these and other measures could be utilised from Chapman 2004 (10 months follow-up, including both ADAS-Cog and MMSE) and Tsantali 2017 (8-month follow-up, MMSE) as it has not proved possible to obtain extractable data immediately after the end of the intervention period in these studies.

Nineteen studies used at least one self-report quality of life measure with participants with dementia, with all but one of these including the QoL-AD (Logsdon 2002); again (as with all the outcomes) the Chapman 2004 data were not in useable form for the immediate post-intervention analysis. The exception was Leroi 2019, using a widely-used generic health-related quality of life measure, the EQ-5D (Group EuroQol 1990). Ten studies (including Chapman 2004) additionally used the proxy version of the QoL-AD and Ali 2021 used only the QoL-AD rated by a proxy. Tanaka 2021 used a different proxy measure of quality of life, the short questionnaire for Quality of Life in Dementia (QoL-D) (Terada 2015).

Nineteen studies evaluated mood, although fewer (10) included a self-report measure (Baldelli 1993; Baldelli 2002; Coen 2011; Juarez-Cedillo 2020; Justo-Henriques 2022; Kim 2016; Leroi 2019; Orgeta 2015; Rai 2021; Requena 2006), with seven making use of a version of the Geriatric Depression Scale, which has a simple 'Yes/No' response format (Yesavage 1983). Ten studies used a depression scale completed from carer reports and/or interviews with the participants (Alvares-Pereira 2021; Buschert 2011; Capotosto 2017; Graessel 2011; Maci 2012; Marinho 2021; Orrell 2014; Rai 2021; Spector 2001; Spector 2003). The most frequently used (8 studies) was the Cornell Scale for Depression in Dementia (Alexopoulos 1988). Six studies used an anxiety measure completed in the same way (Alvares-Pereira 2021; Capotosto 2017; Coen 2011; Maci 2012; Orrell 2014; Spector 2001), with the Rating of Anxiety in Dementia scale (RAID) (Shankar 1999) being used in four studies.

In total, 18 studies evaluated activities of daily living. Seven of these included measures of basic Activities of Daily Living (ADLs) (Baldelli 1993; Baldelli 2002; Bottino 2005; Graessel 2011; Maci 2012; Onder 2005; Tanaka 2021), using four different scales. Fourteen studies included an assessment of more complex activities, Instrumental Activities of Daily Living (IADLs) (Ali 2021; Capotosto 2017; Carbone 2021; Chapman 2004; Graessel 2011; Juarez-Cedillo 2020; Justo-Henriques 2022; Maci 2012; Marinho 2021; Middelstädt 2016; Onder 2005; Orgeta 2015; Orrell 2014; Rai 2021). Again, there was little consistency in the scales used, with seven different measures included.

Thirteen studies included a measure of behaviour that challenges (Capotosto 2017; Carbone 2021; Chapman 2004; Graessel 2011; Juarez-Cedillo 2020; Leroi 2019; Mapelli 2013; Middelstädt 2016; Onder 2005; Orgeta 2015; Orrell 2014; Rai 2021; Tanaka 2021). Ten of these studies used the Neuropsychiatric Inventory (NPI - Cummings 1997). Six studies reported total scores on general behaviour rating scales which included both ADL items and items relating to behaviour that challenges (Alvares-Pereira 2021; Coen 2011; Graessel 2011; Juarez-Cedillo 2020; Spector 2001; Spector 2003). Four of these studies used the CAPE Behaviour Rating Scale (Pattie 1979).

Eight studies reported using assessments relevant to communication and social interaction (Alvares-Pereira 2021; Capotosto 2017; Carbone 2021; Chapman 2004; Graessel 2011; Spector 2001; Spector 2003; Tanaka 2021), using four different methods of assessment. Four studies evaluated the quality of the

relationship with the primary caregiver from the perspective of the person with dementia (Cove 2014; Leroi 2019; Orgeta 2015; Rai 2021), with three of these using the QCPR (Spruytte 2002).

Eleven studies reported evaluating outcomes for caregivers. One study, (Juarez-Cedillo 2020), which offered additional support to family caregivers, did not provide any data on caregiver outcomes despite reporting the inclusion of relevant outcome measures. In five of the remaining 10 studies (Ali 2021; Leroi 2019; Onder 2005; Orgeta 2015; Rai 2021), family caregivers received training and support to deliver individual cognitive stimulation to the person with dementia, so the effects on caregivers were especially pertinent for these studies (although it should be noted that the majority of carers in the Ali 2021 study were paid carers). A further study (Bottino 2005) offered additional carer support in addition to the group cognitive stimulation for the person with dementia and the remaining four studies (Maci 2012; Marinho 2021; Orrell 2014; Spector 2001) simply evaluated the effect on family caregivers of the person with dementia attending cognitive stimulation groups. As both Orrell 2014 and Spector 2001 included a mixed community/care home population, data from family caregivers were only collected in relation to a subsample of the participants.

The most common caregiver outcome evaluated was depressed mood, included in eight studies (Ali 2021; Bottino 2005; Leroi 2019; Maci 2012; Onder 2005; Orgeta 2015; Rai 2021; Spector 2001), with five different scales being used. Six studies evaluated anxiety in caregivers (Ali 2021; Bottino 2005; Leroi 2019; Onder 2005; Orgeta 2015; Rai 2021), two using the Hamilton Anxiety Scale (Hamilton 1959) and four the anxiety scale from the Hospital Anxiety and Depression Scales (HADS) (Zigmond 1983). However, Ali 2021 reported a total HADS score, combining anxiety and depression scores, so (as these scores are often correlated) results from this study have only been included in the meta-analysis of low mood. Caregiver stress was evaluated in seven studies (Ali 2021; Leroi 2019; Maci 2012; Marinho 2021; Onder 2005; Spector 2001), with two using the Caregiver Burden Inventory (Novak 1989), two the Relatives' Stress Scale (RSS) (Greene 1982) and two the Zarit Burden Interview (Zarit 1980) (including Leroi 2019 who also used the RSS). Ali 2021 used a different Caregiver Burden Scale (Macera 1993) as well as a measure of staff competence in dementia care (Schepers 2012). Generic health-related qua ility of life was assessed in four studies; Leroi 2019, Orgeta 2015 and Orrell 2014 all included both the EQ-5D (Group EuroQol 1990) and the SF-12 (Ware 1996) (broken down into physical and mental health components) whereas Onder 2005 used the SF-36 (Tarlov 1989) and Rai 2021 used only the EQ-5D. Finally, three studies (Leroi 2019; Orgeta 2015; Rai 2021) evaluated the quality of the relationship with the person with dementia from the caregiver's perspective, with the first two of these studies also including a measure of the caregiver's resilience, both using (different) brief Resilience Scales, developed by Smith 2008 and Wagnild 2009, respectively.

A full list of the outcome measures used in each of the included studies can be found in the table 'Characteristics of included studies'.

Excluded studies

The most frequent reasons for exclusion of studies at the full-text stage were that the study had been published only as a conference abstract (18 studies) or that the study did not allocate participants to intervention and control groups randomly (17 studies) or that, either the study did not include people with dementia or, if the study had included people with dementia, their data were not presented separately from those with mild cognitive impairment or no impairment (17 studies). We excluded sixteen studies as the intervention described did not meet the review definition of 'cognitive stimulation' - in four of these studies, the intervention appeared to meet the definition of 'cognitive training'. We detailed reasons for exclusion of selected specific studies in 'Characteristics of excluded studies'.

Risk of bias in included studies

We provide details for each study in the 'Characteristics of included studies' table. See also Figure 2 and Figure 3.

Allocation

Random sequence generation

We excluded studies from this review where allocation to intervention and control groups had clearly not been at random, or if an inadequate randomisation method had been used. For twelve studies, although it was stated that random allocation to treatment groups occurred, there was insufficient detail regarding the method of randomisation (e.g. no mention of the use of a computer programme) and so the risk of bias was rated unclear for these studies. For the remaining 25 studies, we rated selection bias related to random sequence generation as 'low'. Computerised randomisation was used in many of the more recent studies (e.g. Graessel 2011; Justo-Henriques 2022; Lok 2020; Marinho 2021; Orgeta 2015; Orrell 2014), with only a few earlier studies describing methods such as drawing names from a sealed container (Spector 2001; Spector 2003).

Allocation concealment

To reduce further the risk of selection bias, ideally the randomisation would be performed remotely by an independent person or, for example, by a clinical trials unit. For the majority of studies (22), it was unclear who had carried out the randomisation procedure, or it had been performed by a researcher involved in the day-to-day conduct of the study. We rated fifteen studies as having a low risk of bias, in that they stated an independent person or trials unit carried out the randomisation (e.g. Buschert 2011; Cheung 2019; Justo-Henriques 2022; Leroi 2019; Orgeta 2015; Orrell 2014; Paddick 2017) or that a centralised web-based system had been used (e.g. Ali 2021; Alvares-Pereira 2021; Rai 2021).

Blinding

Blinding of outcome assessment

The majority of studies took steps to ensure that the assessment of outcomes was carried out by assessors blind to treatment allocation. We rated two studies (Lok 2020; Spector 2001) as having 'high risk' of bias in this domain, in that the researcher conducting the treatment groups also carried out at least some of the assessments of outcomes. We rated six studies as having 'unclear risk' in that no details were provided of who carried out the assessments. We rated the remaining 29 studies as 'low risk', as they stated that assessments were carried out by a researcher blind to treatment allocation. Of course, even independent assessors may be given hints from participants regarding receiving the intervention during the assessments, and some studies took precautions to remind participants not to discuss treatment with

the assessor (e.g. Middelstädt 2016) and/or to check on the extent to which blinding was maintained by having assessors give their view on which group the person assessed had been allocated to, and their confidence in their judgement (e.g. Orgeta 2015).

Using independent assessors works well for evaluating changes in cognition or self-reported mood, well-being and quality of life, where the assessment is directly with the person with dementia. Ratings of day-to-day behaviour and function and proxy ratings of quality of life are usually completed by care staff or by family caregivers, who are typically not blinded to group allocation, often for reasons of logistics. There is likely then to be a higher risk of detection bias associated with outcome assessments of this nature.

Incomplete outcome data

We rated the majority of studies (25) as having a low risk of bias, reporting details of attrition from the study, with reasons, and/or reporting an intention-to-treat analysis, if necessary, using appropriate imputation methods so that all participants were accounted for in the analyses. We rated four studies as having a high risk of attrition bias. Although Graessel 2011 did carry out an intention-to-treat analysis, extractable data were reported only for per protocol analyses, for which as many as 38% of the participants were omitted. There was a large imbalance in attrition in the Kim 2016 study, with zero attrition in the intervention group and 34% in the control group, which we judged as likely to have an impact on the conclusions drawn. Lin 2018 and Tsantali 2017 both reported per protocol analyses, excluding 13% and 19% of cognitive stimulation participants, respectively. For the remaining eight studies, the information regarding attrition was unclear, or the potential effects of reported attrition were unclear.

Selective reporting

Most studies (31) reported results from all the outcome measures listed in the methods section of the report or in the protocol, where provided. For Orrell 2014, the results for several secondary measures were not included in the published report, but were obtained from the study authors. We rated three studies as having a high risk of reporting bias. Mapelli 2013 did not provide results for several measures included in the study methods (e.g. Geriatric Depression Scale and ADL Scale). Juarez-Cedillo 2020 did not provide results for caregiver outcomes detailed in the methods section of the report, although data for several outcomes were reported for people with dementia that were not listed. Paddick 2017 provided some results for all the measures included in their study, but, in view of the small sample size, only presented results for a measure of cognition for the comparison of interest for this review (cognitive stimulation versus treatment-as-usual).

Other potential sources of bias

Training and supervision

We rated all studies as having a low (21 studies) or unclear risk (16 studies) of bias in relation to the training and supervision of those delivering the cognitive stimulation intervention. Many studies rated as 'unclear' did not mention any training or supervision of those conducting the intervention, or simply reported their professional background, without detailing any specific training in cognitive stimulation. Those studies involving family caregivers in delivering individual cognitive stimulation (Ali 2021; Leroi 2019; Onder 2005; Orgeta 2015; Rai 2021) were amongst those detailing the provision of training and support. For example, in the Onder 2005 study, family caregivers were trained by a multi-disciplinary team including physicians, psychologists and therapists, and the training included a simulated therapy session. Further examples of good practice were provided by Graessel 2011, where therapists received three days of training and Chapman 2004, where the students assisting with the groups underwent a two-hour training session and were provided with written reference materials before commencing with the group. Graessel 2011 also reported compliance being checked three times at each participating nursing home, whereas Chapman 2004 described weekly meetings being held in order to ensure the proper implementation of the programme.

Treatment manual

We rated the majority of studies (30) as 'low risk', in that there was evidence of a manual, protocol or structure outlining the content of each session. We rated the Tsantali 2017 study as 'high risk' in that there was no indication of a manual or a particular structure for the cognitive stimulation intervention. Notably, cognitive stimulation was not the main focus of this study, with much more detail provided regarding an individual cognitive training intervention. Four of the remaining studies, where we rated the risk as 'unclear' were among the earlier studies (Baldelli 1993; Baldelli 2002; Bottino 2005; Breuil 1994), conducted before the wide availability of cognitive stimulation therapy manuals (e.g. Spector 2006). We rated the remaining two studies as being 'unclear': (Mapelli 2013) outlined a clear structure, but the actual content of exercises was not described; Tanaka 2021 referred to the Japanese version of Spector 2006 but appeared to include additional components. As previously noted, 19 studies either used the 'Making a difference' manuals (Aguirre 2011; Spector 2006; Yates 2014) or adopted a very similar structure. The presence of detailed treatment protocols reduces the risk that the cognitive stimulation may not have been delivered as intended although, without checks on compliance and fidelity, this cannot be assured.

Effects of interventions

See: **Summary of findings 1** Cognitive stimulation compared to no cognitive stimulation (post-treatment) in people with dementia; **Summary of findings 2** Group cognitive stimulation compared to no cognitive stimulation (post-treatment) in people with dementia

Effect sizes

Evaluating the clinical meaningfulness of changes on the outcome measures used in studies of cognitive stimulation is challenging, as there are no internationally agreed standards to apply in this context. For SMDs, we have adopted the rule that an SMD of 0.5 or greater reflects an important difference. This is more conservative than the 0.40 SMD recommended by Howard 2011 for a minimum clinically important difference in dementia studies. We consider SMDs of 0.10 and less as being negligible. We describe effect sizes where the SMD falls between 0.10 and 0.40 as 'slight' and, in view of the threshold suggested by Howard 2011, 'small' if 0.40 or more, but less than 0.50. For analyses using the MMSE, we judged a difference of 1.5 points or more as clinically important. The rate of decline on this measure has been estimated, in mild-to-moderate dementia, to be between 2 and 4 points per annum (Mohs 2000) and so, 1.5 points is broadly equivalent to preventing six months of decline in cognition. Again, this is slightly more conservative than the 1.4 point difference suggested as the minimum clinically important

difference on this scale by Howard 2011. For the ADAS-Cog, 3 points has been suggested as a clinically important difference (Schrag 2012). For other measures, we did not have parallel criteria, so have applied the 0.5 of a standard deviation rule, taking the standard deviation from the baseline evaluations. Thus, for the QoL-AD, we have taken a difference of 3 points or more to be clinically meaningful, reflecting approximately half the typical standard deviation in samples of people with mild-to-moderate dementia (e.g. Woods 2016).

For meta-analyses, we used RevMan 5.4.1. See Summary of findings 1 for the main comparison, of cognitive stimulation versus controls, at the end of the intervention period, for the primary outcome, cognition, and for other important outcomes such as quality of life. Summary of findings 2 summarises findings for cognition and quality of life and other important outcomes for group cognitive stimulation separately.

Cognition

(*See* Figure 4).

Figure 4. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation, outcome: Cognition.

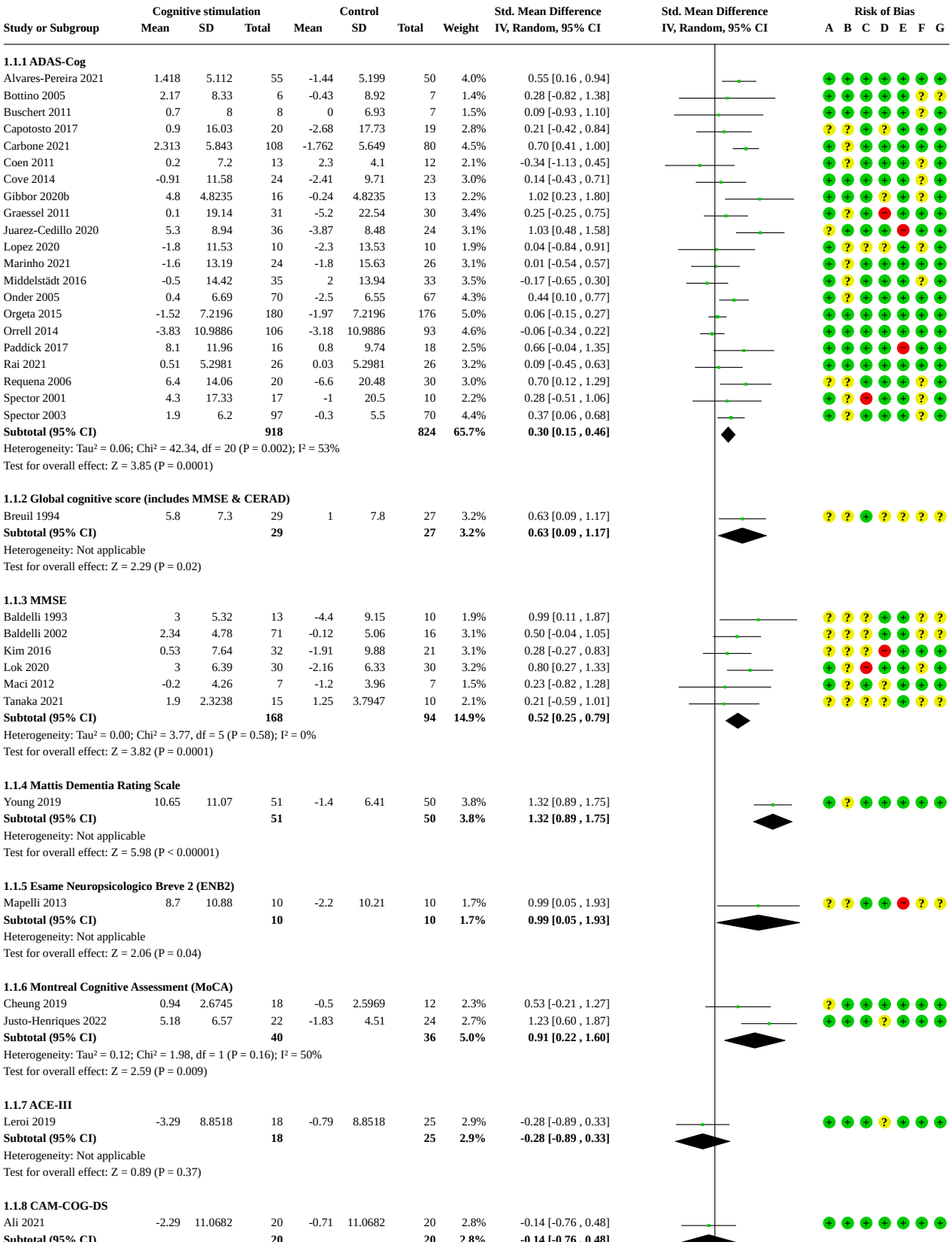
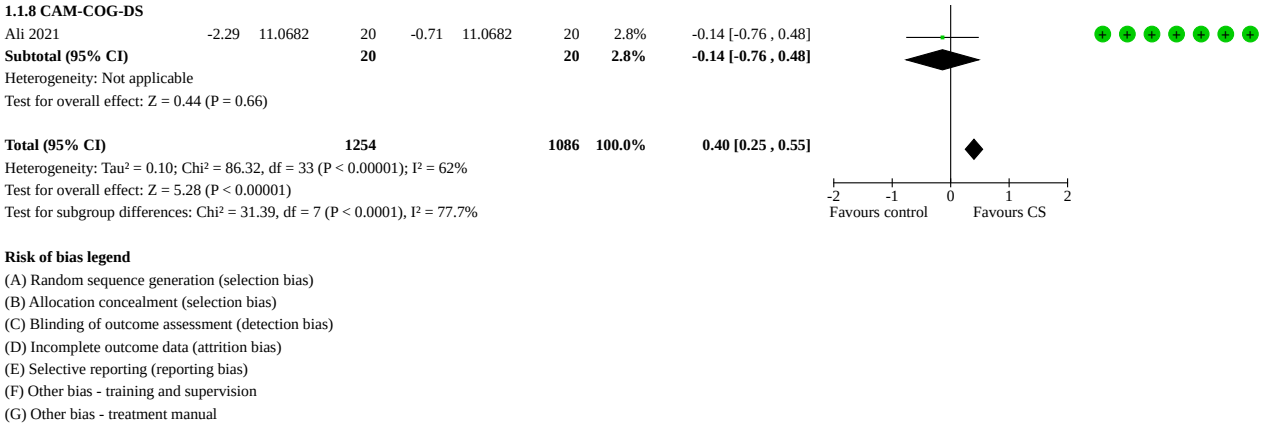


Figure 4. (Continued)



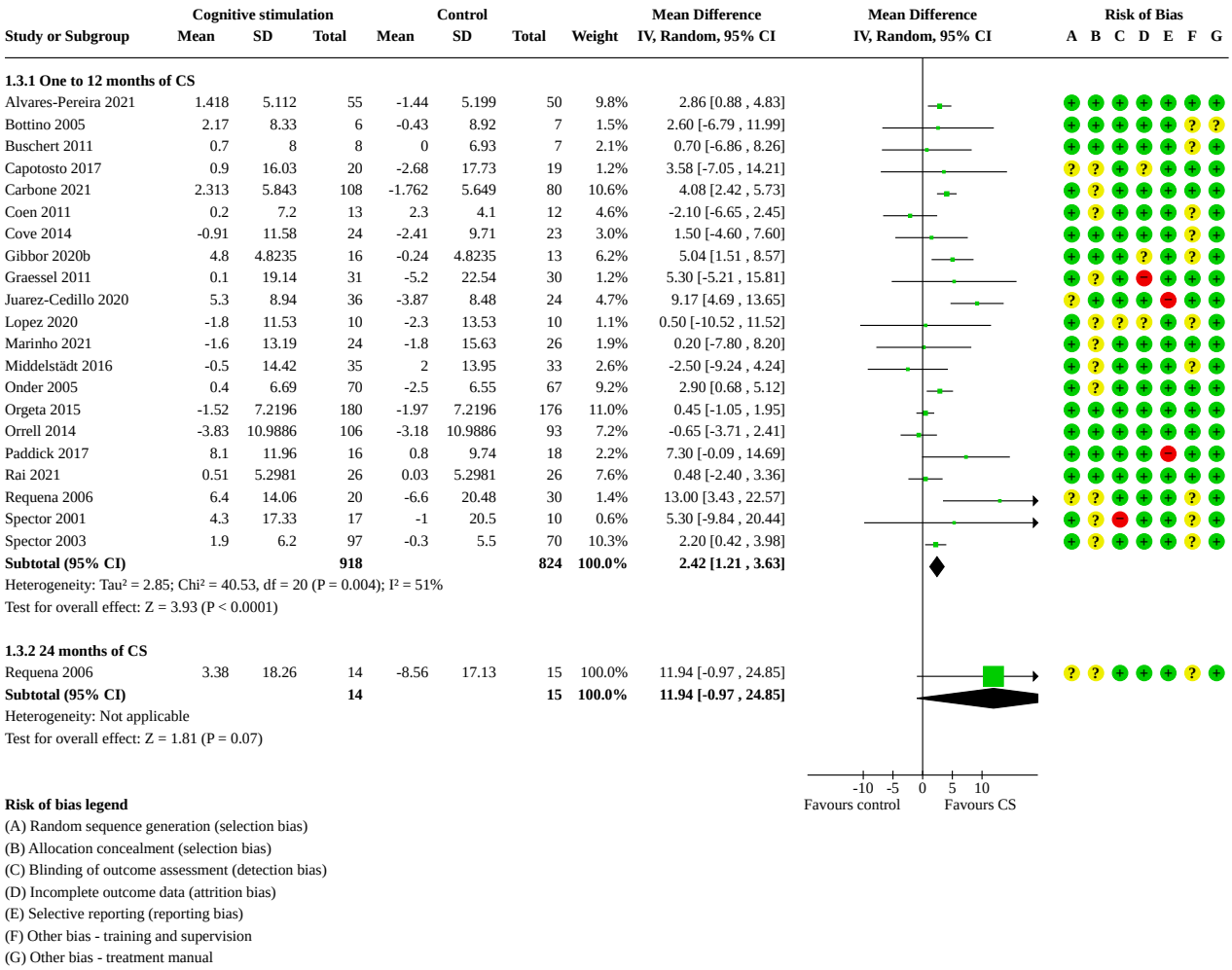
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of outcome assessment (detection bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias - training and supervision
(G) Other bias - treatment manual

For the overall evaluation of the effects of cognitive stimulation on cognitive function, all 34 RCTs having useable data immediately post-treatment were included. Only follow-up data were extractable from Chapman 2004 and Tsantali 2017 and data from Lin 2018 were not in an extractable form. These 34 studies involved a total of 2340 people with dementia, of whom 1254 received cognitive stimulation and 1086 received no treatment or a placebo treatment. As most studies included more than one measure of cognitive function, this analysis was conducted on the most extensive assessment included. For 21 studies, this was the ADAS-Cog, and for six it was the MMSE. Two studies used the Montreal Cognitive Assessment and the remaining five studies each used a different cognitive scale. Overall, cognitive stimulation probably leads to a small improvement in cognitive function (SMD 0.40, 95% CI 0.25 to 0.55; I^2 = 62%; moderate-quality evidence). The substantial degree of inconsistency evident in this analysis may result from a variety of factors, clinical and methodological, explored subsequently.

We wanted to explore whether the result was robust to the cognitive measure used in the analysis. Therefore, as a sensitivity analysis, we went on to analyse results separately for the two most frequently used cognitive measures, the ADAS-Cog and the MMSE, which differ in extent and focus. The MMSE was the most frequently used measure in the review, having been used in a total of 25 studies (in most cases alongside the ADAS-Cog).

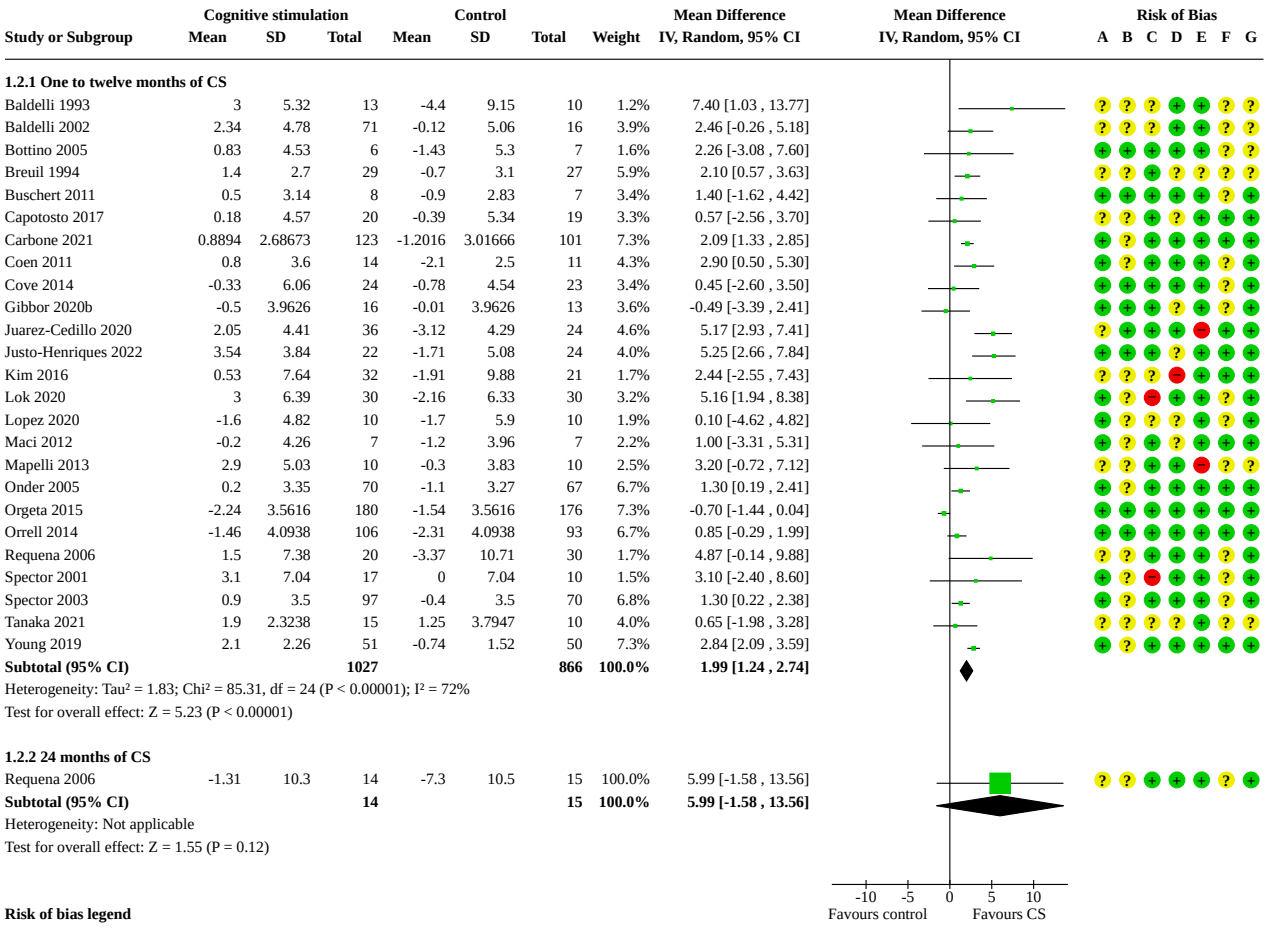
For the 21 studies, including 1742 participants, using the ADAS-Cog as an outcome measure (Figure 5), the results showed moderate inconsistency between studies. Cognitive stimulation probably leads to a small improvement in ADAS-Cog scores (mean difference 2.42 points, 95% CI 1.21 to 3.63; I^2 = 51%; moderate-quality evidence).

Figure 5. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation, outcome: ADAS-Cog



In total, 25 studies involving 1893 participants used the MMSE (Figure 6), indicating that cognitive stimulation is probably associated with a clinically relevant improvement in MMSE scores at post-treatment assessments. The overall mean difference was 1.99 points (95% CI 1.24 to 2.74; I^2 = 72%; moderate-quality evidence). Again, there is substantial heterogeneity between studies.

Figure 6. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation, outcome: MMSE



Risk of bias legend

(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of outcome assessment (detection bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias - training and supervision
(G) Other bias - treatment manual

The results of the various subgroup analyses carried out to explore heterogeneity and to address areas of clinical interest detailed in the following sections are summarised in Table 3.

Modality

A potential source of inconsistency between studies relates to whether cognitive stimulation was delivered individually or in a group context, with the essential social elements more readily apparent in the latter. A comparison of studies offering group cognitive stimulation with those where the cognitive stimulation was individual does not indicate a modality difference in cognitive outcomes (test for subgroup differences: $Chi^2 = 0.43$, df = 1, P = 0.51, $I^2 = 0\%$). However, while there were overall a relatively large number of trials and participants contributing to this comparison, only seven of the 34 studies related to individual cognitive stimulation. There were 27 studies (1637 participants) with cognitive data at post-treatment which offered cognitive stimulation primarily in a group format. For these studies, there remained a moderate degree of heterogeneity, with group CS probably leading to a small improvement in cognition (SMD 0.43, 95% CI 0.26 to 0.59; $I^2 = 56\%$; moderate-quality evidence). The remaining seven studies with cognitive data at post-treatment provided cognitive stimulation on an individual basis. In four studies (Leroi 2019; Onder 2005; Orgeta 2015; Rai 2021), family carers were trained to carry out the intervention; researchers (Gibbor 2020b), clinical psychologists (Justo-Henriques 2022) and a mix of paid and family carers (Ali 2021) delivered the sessions in the remaining studies. For cognition overall, data from 703 participants from the seven studies were available. The results indicate there may be a slight improvement in cognition, but the results are both inconsistent and imprecise (SMD 0.30, 95% CI -0.03 to 0.64; $I^2 = 72\%$; low-quality evidence).

Total exposure to and frequency of group intervention

For the studies of group CS, a wide range in the total number of group sessions was evident, from 8 to 300 within a 12-month period, with a median of 20 sessions. Twelve studies (301 participants) offered 20 sessions or more, and 15 studies (1336 participants) offered a smaller number. In exploring the number of group sessions as a potential source of heterogeneity, the overall number of studies and participants provided a reasonable basis

for comparison. Studies offering individual cognitive stimulation were not included in this analysis as documenting the number of sessions actually offered has proved challenging. There was no evidence that a greater number of sessions led to greater effect sizes (test for subgroup differences: $Chi^2 = 0.01$, df = 1, P = 0.94, I^2 = 0%) and moderate to substantial heterogeneity was evident for both subgroups.

However, there were also differences between studies in terms of the frequency of exposure, with six studies (354 participants) offering group sessions once per week, and the remaining studies having intensities varying from twice a week to five times a week. Eight studies (328 participants) offered group sessions three times a week or more and thirteen studies (955 participants) offered two sessions per week. These three subgroups differed significantly (test for subgroup differences: $Chi^2 = 10.82$, df = 2, P = 0.004, I^2 = 81.5%), with further analyses indicating there was a significant difference between once-a-week sessions and both three sessions or more (test for subgroup differences: $Chi^2 = 6.68$, df = 1, P = 0.010, I^2 = 85.0%) and two sessions per week (test for subgroup differences: $Chi^2 = 8.79$, df = 1, P = 0.003, I^2 = 88.6%), but no difference between two sessions per week and three or more sessions per week (test for subgroup differences: $Chi^2 = 0.14$, df = 1, P = 0.71, I^2 = 0%).

The 21 studies offering at least two sessions per week (1283 participants) are associated with a probable clinically relevant improvement in cognition (SMD 0.51, 95% CI 0.34, 0.69; I^2 = 51%; moderate-quality evidence), whereas the six studies where sessions took place once per week (354 participants) may have a negligible effect on cognition (SMD 0.04, 95% CI -0.17, 0.27; I^2 = 0%; low-quality evidence). A potential confound may arise from the Orrell 2014 study, which offered once-a-week sessions in the context of a maintenance study, where all participants had previously received twice-weekly sessions for seven weeks. However, a sensitivity analysis showed that the subgroup difference remained when this study was not included (test for subgroup differences: $Chi^2 = 3.33$, df = 1, P = 0.07, I^2 = 70.0%). The disparity in numbers of studies and participants between the twice-a-week or more studies and the once-a-week studies should also be noted, although the subgroups were more comparable for the three or more times per week and once per week comparison. Heterogeneity in these two subgroups was minimal, whereas it remained moderate-to-substantial for twice a week studies.

The 24-month data from Requena 2006, where group sessions continued for a further year, five times per week, indicated that effect sizes may be maintained through continued exposure at relatively high frequency (ADAS-Cog mean difference 11.94 points, 95% CI -0.97 to 24.85; MMSE mean difference 5.99 points, 95% CI -1.58 to 13.56; both low-quality evidence). However, these effects require replication as the confidence intervals were broad and included little or no effect.

Setting: care home v community

Of the 27 studies of cognitive stimulation groups with post-treatment data available on cognition, 22 were conducted wholly in either care home or community settings. Five (Alvares-Pereira 2021; Carbone 2021; Orrell 2014; Spector 2001; Spector 2003) included participants from both care homes and community settings and, of these, only one (Alvares-Pereira 2021) presented results separately for the two settings. A subgroup analysis with 15 community studies (642 participants) and nine care home studies (323 participants) indicates there is no evidence of a difference in effect size for cognition between the settings (test for subgroup differences: $Chi^2 = 0.03$, df = 1, P = 0.86, I^2 = 0%). Heterogeneity is substantial in the results from both settings, and it should be noted that community studies predominate in both numbers of studies and participants.

Type of control condition

Five studies (322 participants) of cognitive stimulation groups offered control participants an alternate activity, such as watching TV, social activities, reading and discussions (Buschert 2011; Capotosto 2017; Carbone 2021; Cheung 2019; Requena 2006). The remaining 22 studies (1315 participants) had 'treatment-as-usual' control conditions. A subgroup analysis did not indicate any difference in cognitive outcomes between the two types of control conditions (test for subgroup differences: $Chi^2 = 1.45$, df = 1, P = 0.23, I^2 = 30.9%). There was substantial inconsistency in the results from studies offering 'treatment-as-usual', and a large imbalance in numbers of studies and participants means the results should be interpreted cautiously.

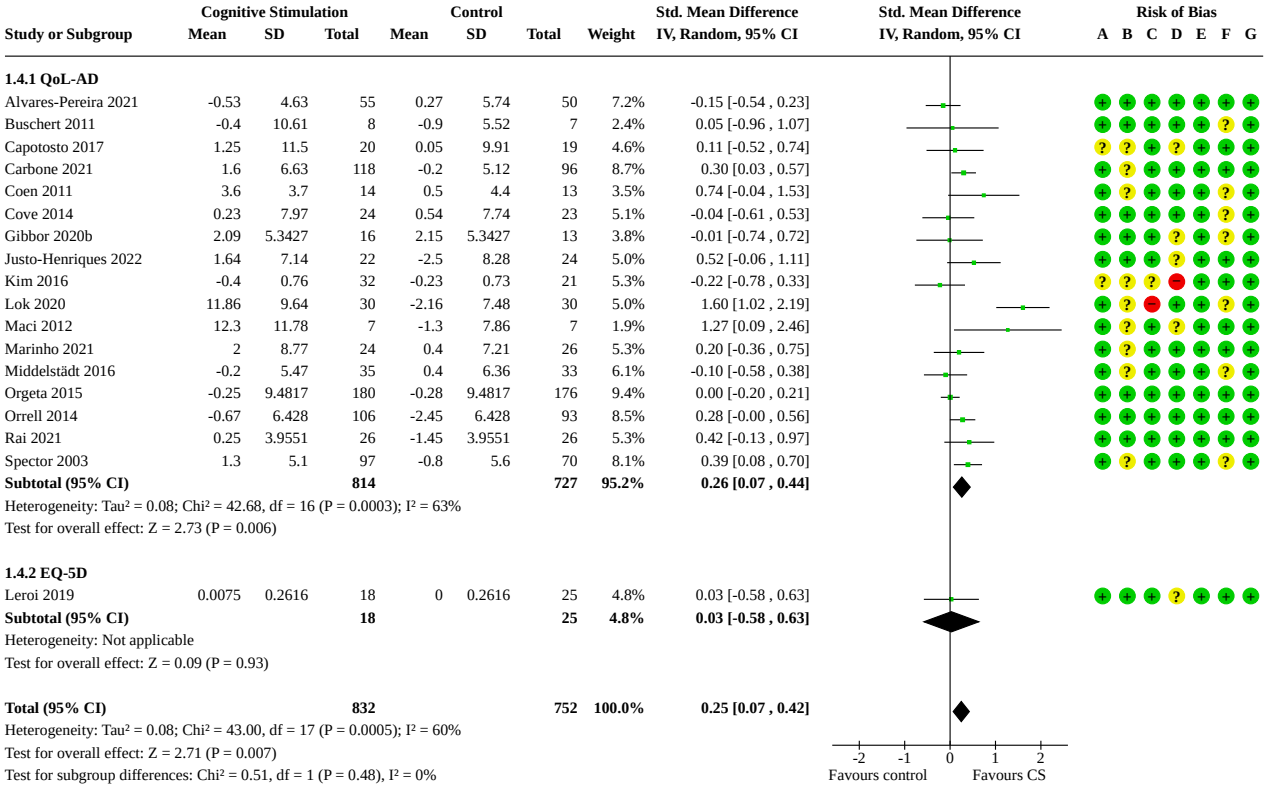
Severity of dementia

Considering studies of group cognitive stimulation with a post-treatment cognitive outcome, thirteen studies (778 participants) reported a mean baseline MMSE score in the moderate range (defined here as < 20), and ten studies (640 participants) reported their average level of dementia severity to be within the mild range (i.e. mean baseline MMSE > 20). This provided a reasonable covariate distribution for an analysis of any differences between studies based on initial dementia severity. The test for subgroup differences was statistically significant ($Chi^2 = 10.53$, df = 1, P = 0.001, I^2 = 90.5%), with greater improvement where dementia severity was initially mild. For moderate dementia studies, there is a slight improvement in overall cognition (SMD 0.21, 95% CI 0.03, 0.39; I^2 = 28%; high-quality evidence) whereas, for mild dementia studies, the improvement is probably of clinical importance (SMD 0.71, 95% CI 0.47, 0.95; I^2 = 43%; moderate-quality evidence). However, for the mild studies, a moderate degree of heterogeneity remains unexplained.

Quality of life (self-report)

(*See* Figure 7).

Figure 7. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation, outcome: self-report QoL



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of outcome assessment (detection bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias - training and supervision
(G) Other bias - treatment manual

Eighteen studies, involving 1584 participants, included relevant self-report measures. Seventeen used the QoL-AD whereas Leroi 2019 reported data from the EQ-5D. Where studies reported data from more than one quality of life scale, we have utilised the QoL-AD data where possible, as this is the most commonly used measure, and recommended in this field (Moniz-Cook 2008). The meta-analysis indicated that cognitive stimulation leads to a probable slight improvement in quality of life compared with no treatment (SMD 0.25, 95% CI 0.07, 0.42; I^2 = 60%; moderate-quality evidence). There appeared to be substantial inconsistency between studies, but we could only explore this further in relation to modality, as too few studies were available for subgroup analysis of number and frequency of sessions, setting and dementia severity.

Modality

Thirteen studies offered group cognitive stimulation (1058 participants), while five studies (526 participants) offered individual cognitive stimulation. The test for subgroup differences was not statistically significant (test for subgroup differences: Chi^2 = 1.27, df = 1, P = 0.26, I^2 = 21.5%). For group studies, there may be a slight improvement in self-reported quality of life (SMD 0.28, 95% CI 0.05, 0.52; I^2 = 67%; low-quality evidence), with substantial inconsistency still present. In contrast, the results for individual cognitive stimulation showed little heterogeneity, with again a slight effect on quality of life (SMD 0.11, 95% CI -0.09, 0.30; I^2 = 7%; high-quality evidence).

Quality of life (proxy-rated)

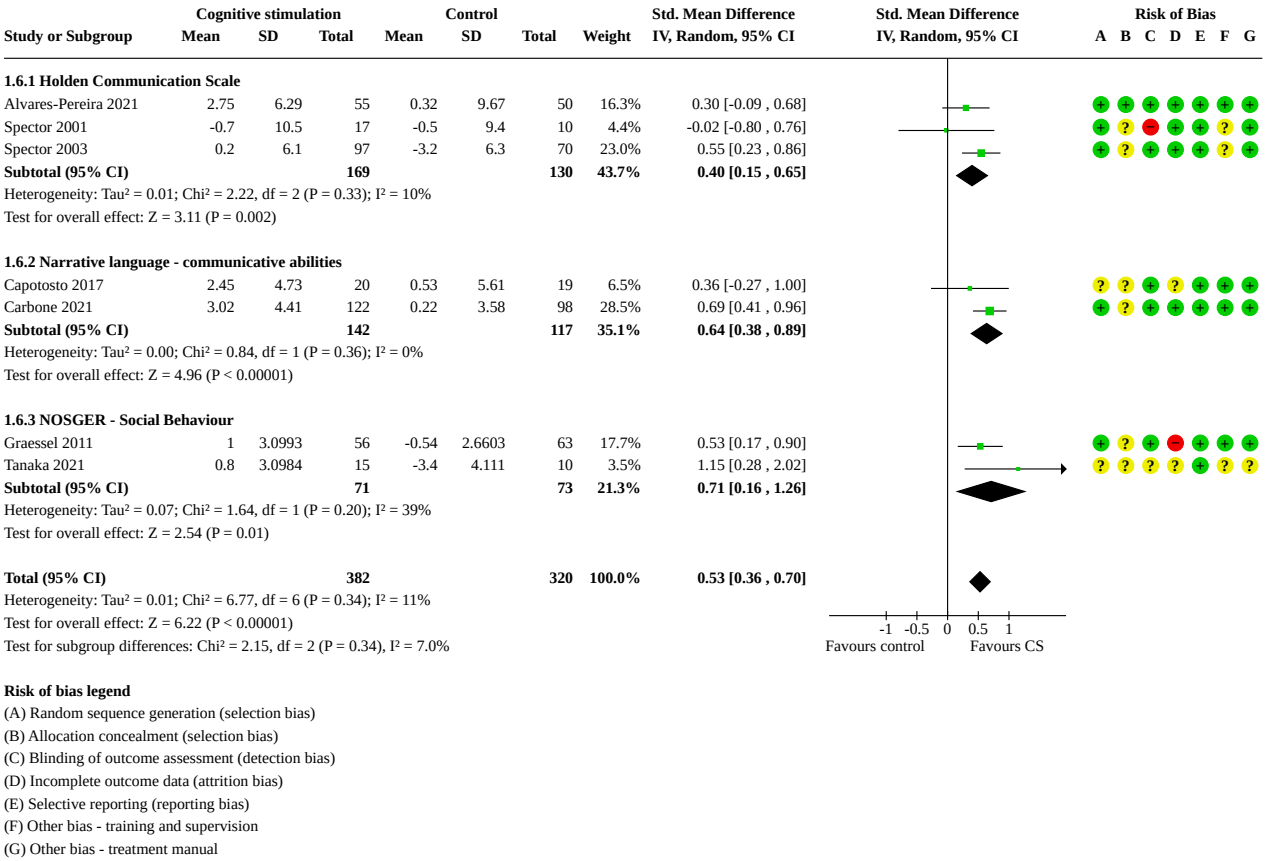
Eleven studies with 988 participants included quality of life rated by a proxy as an outcome measure. All studies, except Tanaka 2021, used the proxy form of the QoL-AD, in some cases also using the proxy version of the DEMQOL. Proxies were typically family carers, but could be care staff in studies conducted in care homes. Although results showed moderate inconsistency, there is probably a slight benefit (SMD 0.21, 95% CI 0.00, 0.42; I^2 =51%; moderate-quality evidence).

Investigation of the factors contributing to the heterogeneity was not possible, in view of the small number of studies making up the planned subgroups.

Communication and social interaction

(*See* Figure 8).

Figure 8. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation: post-treatment, outcome: Comunication and social interaction



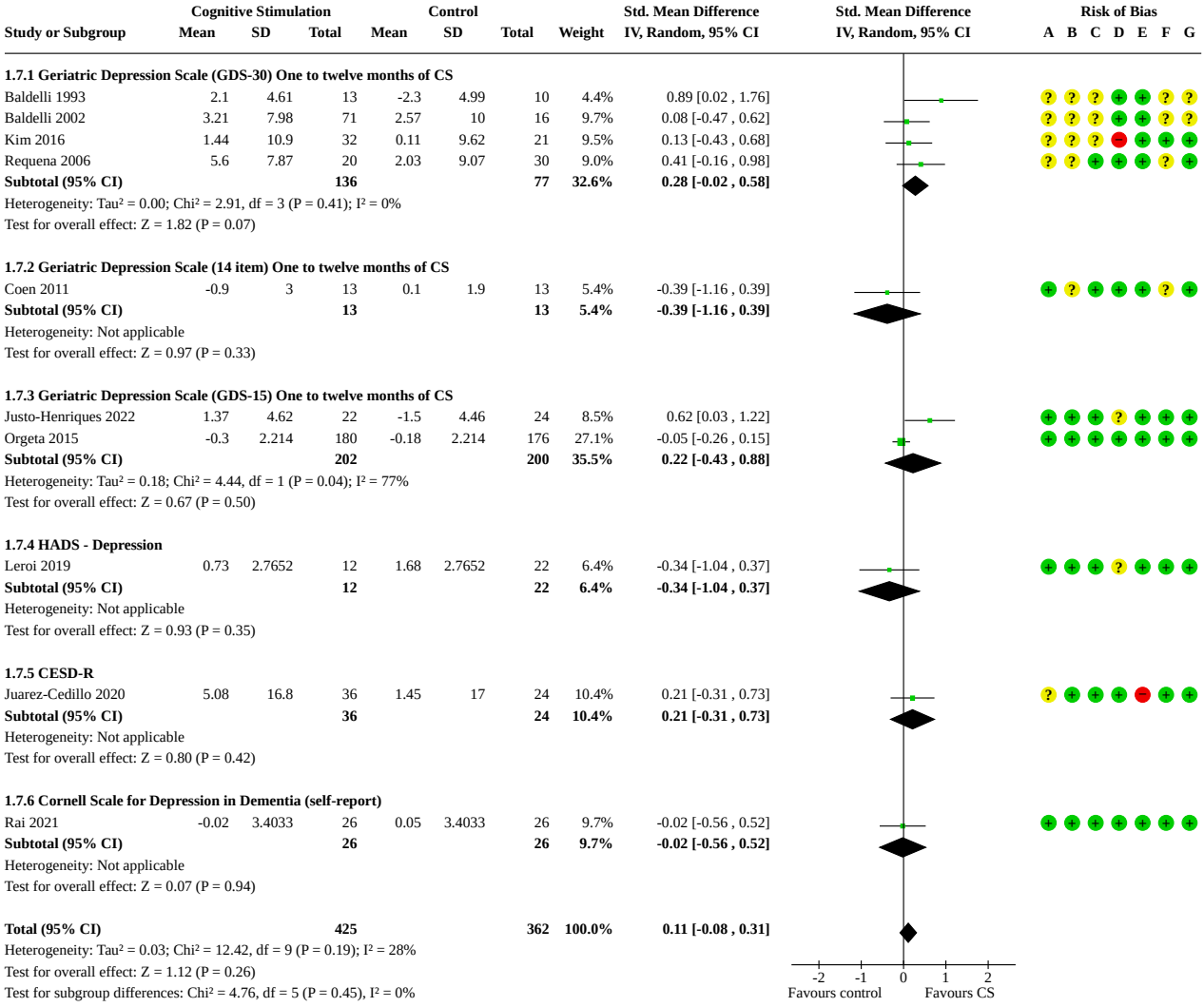
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of outcome assessment (detection bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias - training and supervision
(G) Other bias - treatment manual

Seven studies, involving 702 participants, included indices of the person's communication and social interaction. Five studies (Alvares-Pereira 2021; Graessel 2011, Spector 2001; Spector 2003; Tanaka 2021) used staff ratings (outside of the cognitive stimulation group), and the remaining two (Capotosto 2017; Carbone 2021) used ratings of the person's communication in a structured task. Notably all the seven studies were group studies conducted in care homes or included a mix of care home and community participants, with group sessions at least twice a week. The overall effect size (SMD) was 0.53 (95% CI 0.36, 0.70; I^2 =11%; high-quality evidence), indicating a clinically relevant benefit in this domain. A sensitivity analysis excluding the two studies with a high risk of bias in one domain (Graessel 2011; Spector 2001) did not change this conclusion (5 studies, 556 participants; SMD 0.56, 95% CI 0.36, 0.75; I^2 =15%; high-quality evidence).

Depressed mood

(*See* Figure 9; Figure 10).

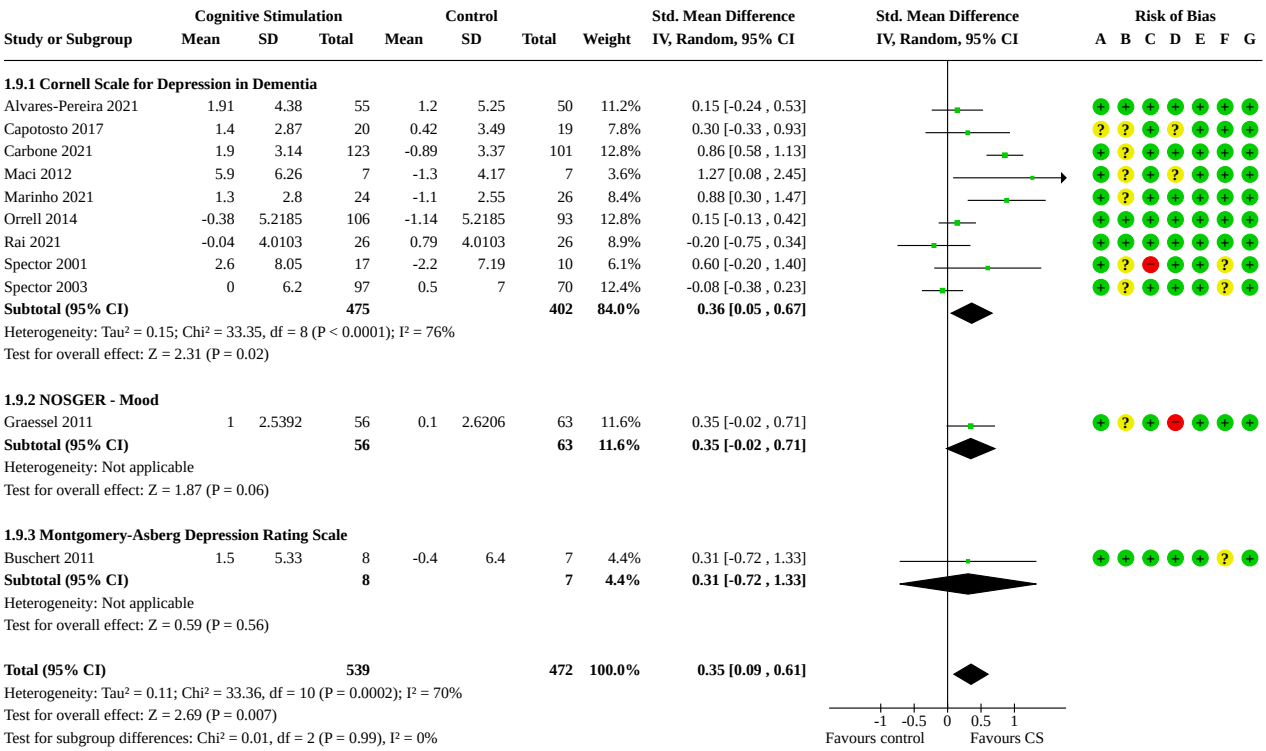
Figure 9. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation, outcome: self-reported depression



Risk of bias legend

(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of outcome assessment (detection bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias - training and supervision
(G) Other bias - treatment manual

Figure 10. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation: post-treatment, outcome: Mood: Staff-reported



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of outcome assessment (detection bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias - training and supervision
(G) Other bias - treatment manual

Ten studies, involving 787 participants, used a self-report measure of mood (seven used a version of the Geriatric Depression Scale and the remaining studies each used a different scale). Cognitive stimulation resulted in a slight improvement to self-reported mood across these studies (SMD 0.11, 95% CI -0.08, 0.31; I^2 =28%; high-quality evidence). We did not undertake further subgroup analyses for self-reported mood as inconsistency was low and, for modality, the number of individual cognitive stimulation studies was small. A sensitivity analysis indicated that there was probably a slight improvement for group cognitive stimulation (6 studies, 299 participants; SMD 0.20, 95% CI -0.06, 0.45; I^2 =7%; moderate-quality evidence).

Eleven studies, involving 1011 participants, reported findings from Interviewer and staff ratings of mood. Nine of these studies used the interviewer rated Cornell Scale for Depression in Dementia, and one each used the interviewer rated Montgomery-Asberg Depression Rating Scale and the mood subscale of the staff-rated NOSGER. There may be a slight improvement in staff/interviewer rated mood, but there was substantial inconsistency between studies and the confidence intervals were wide including both a negligible effect and a clinically relevant improvement (SMD 0.35, 95% CI 0.09, 0.61, I^2 =70%, low-quality evidence). The small numbers of studies in each subgroup meant that exploration of the substantial inconsistency between studies in relation to staff/interviewer ratings of mood was not possible. It is notable that all but one of the eleven studies with interviewer or staff ratings of mood were group studies. Excluding this study (Rai 2021) in a sensitivity analysis did not change the level of inconsistency or the overall effect.

Anxiety

Six studies (410 participants) of group cognitive stimulation included a measure of anxiety, rated by an interviewer or staff member. Four used the Rating of Anxiety in Dementia (RAID), one study used the Hamilton Anxiety Rating Scale and, in one study, the anxiety subscale of the NPI was used. Overall, cognitive stimulation is associated with a slight improvement in anxiety (SMD 0.11, 95% CI -0.09, 0.30, I^2 =0%, high-quality evidence).

Quality of relationship with caregiver

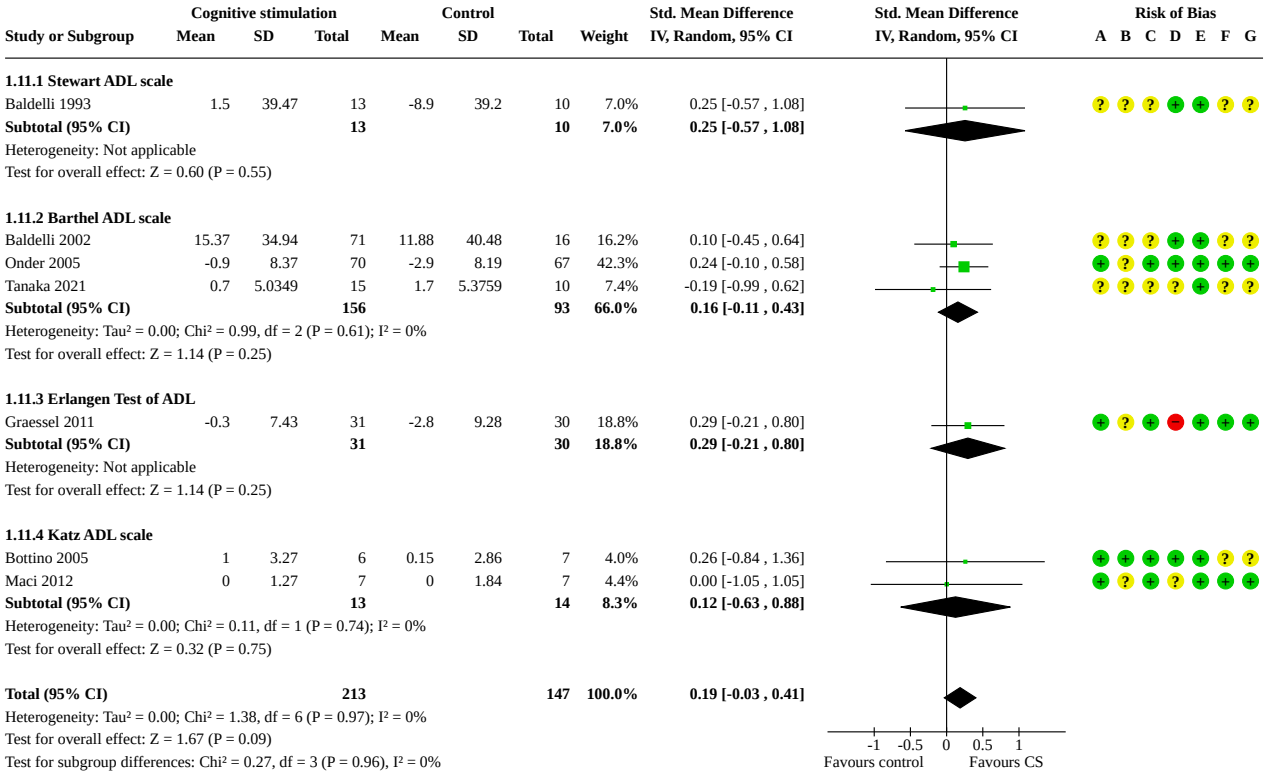
Four studies (Cove 2014; Leroi 2019; Orgeta 2015; Rai 2021) with 492 participants included a measure of the quality of the relationship with the person's caregiver, as rated by the person with dementia. Three studies (Cove 2014; Orgeta 2015; Rai 2021) used the Quality of Carer-Patient Relationship scale (QCPR), whilst Leroi 2019 used the Relationship Satisfaction Scale. The results were imprecise, with

the confidence intervals including both a slight positive and a slight negative effect (SMD -0.01, 95% CI -0.27, 0.25, I^2 = 30%, moderate-quality evidence). There is probably little or no difference in quality of relationship following a cognitive stimulation intervention. A sensitivity analysis excluding Cove 2014, the one study where the intervention had not been delivered by the family caregiver, did not influence the result, but increased inconsistency (445 participants: SMD 0.00, 95% CI -0.33, 0.33, I^2 =42%, low-quality evidence).

Activities of Daily Living

(See Figure 11).

Figure 11. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation, outcome: ADL



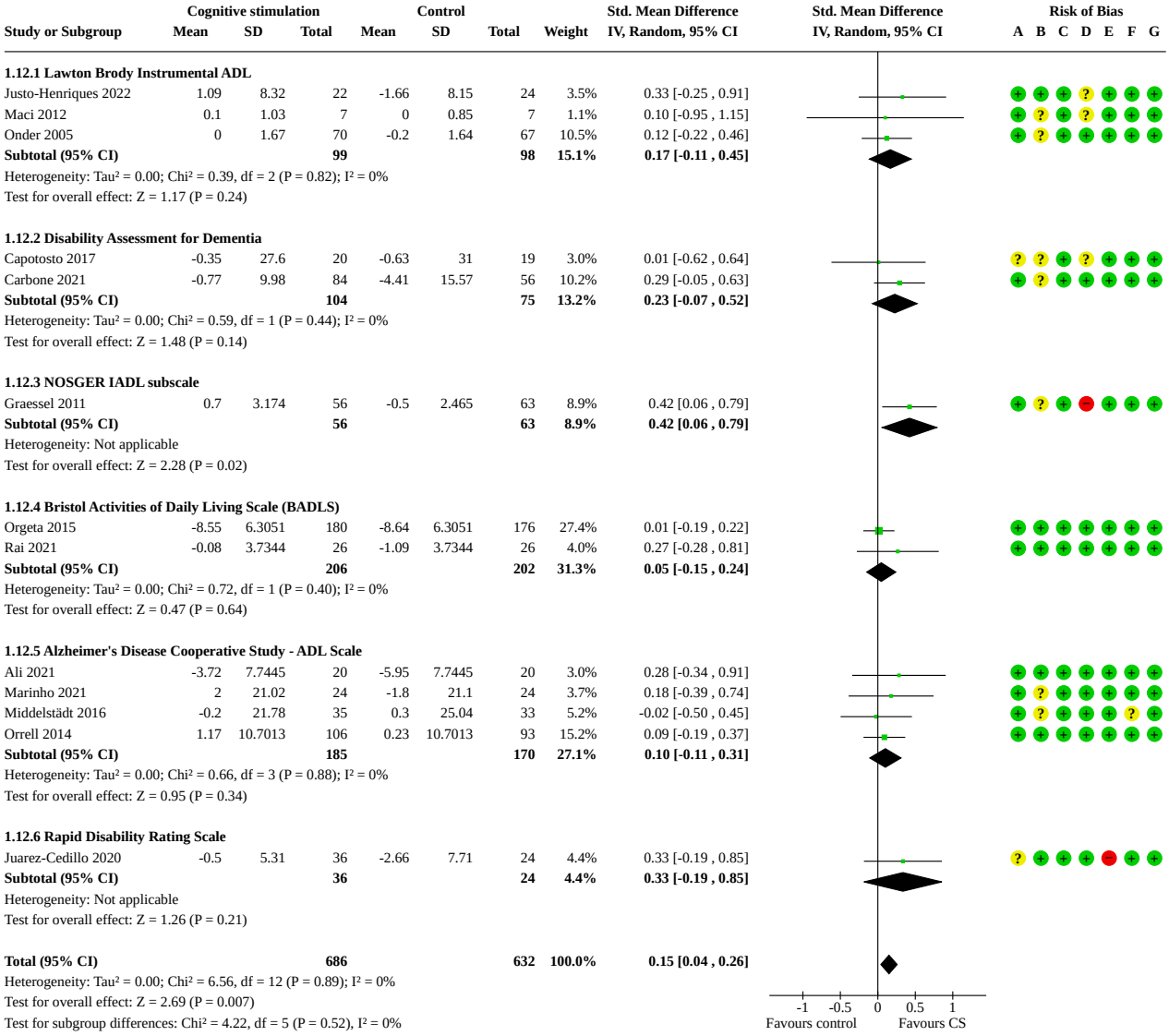
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of outcome assessment (detection bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias - training and supervision
(G) Other bias - treatment manual

Seven studies with 360 participants included a measure of basic activities of daily living such as washing and dressing. A variety of indices were used, with three studies using the Barthel Index and two the Katz ADL scale. There is probably a slight effect of cognitive stimulation on basic ADL function (SMD 0.19, 95% CI -0.03, 0.41, I^2 =0%, moderate-quality evidence). In view of the lack of heterogeneity and the small number of studies, we did not report subgroup analyses. A sensitivity analysis, excluding the one study of the seven that offered individual cognitive stimulation (Onder 2005), resulted in a comparable, though less precise, effect size (223 participants: SMD 0.15, 95% CI -0.14, 0.43, I^2 =0%, low-quality evidence).

Instrumental Activities of Daily Living

(See Figure 12).

Figure 12. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation, outcome: Instrumental ADL



Thirteen studies with a total of 1318 participants reported on a measure of Instrumental Activities of Daily Living, using six different indices, with the Alzheimer's Disease Cooperative Study - ADL Scale being used by four studies and the Lawton Brody scales by three. Overall, cognitive stimulation is associated with a slight improvement to this outcome (SMD 0.15, 95% CI 0.04, 0.26, I^2 =0%, high-quality evidence). As the results were consistent across studies, we only explored potential modality differences.

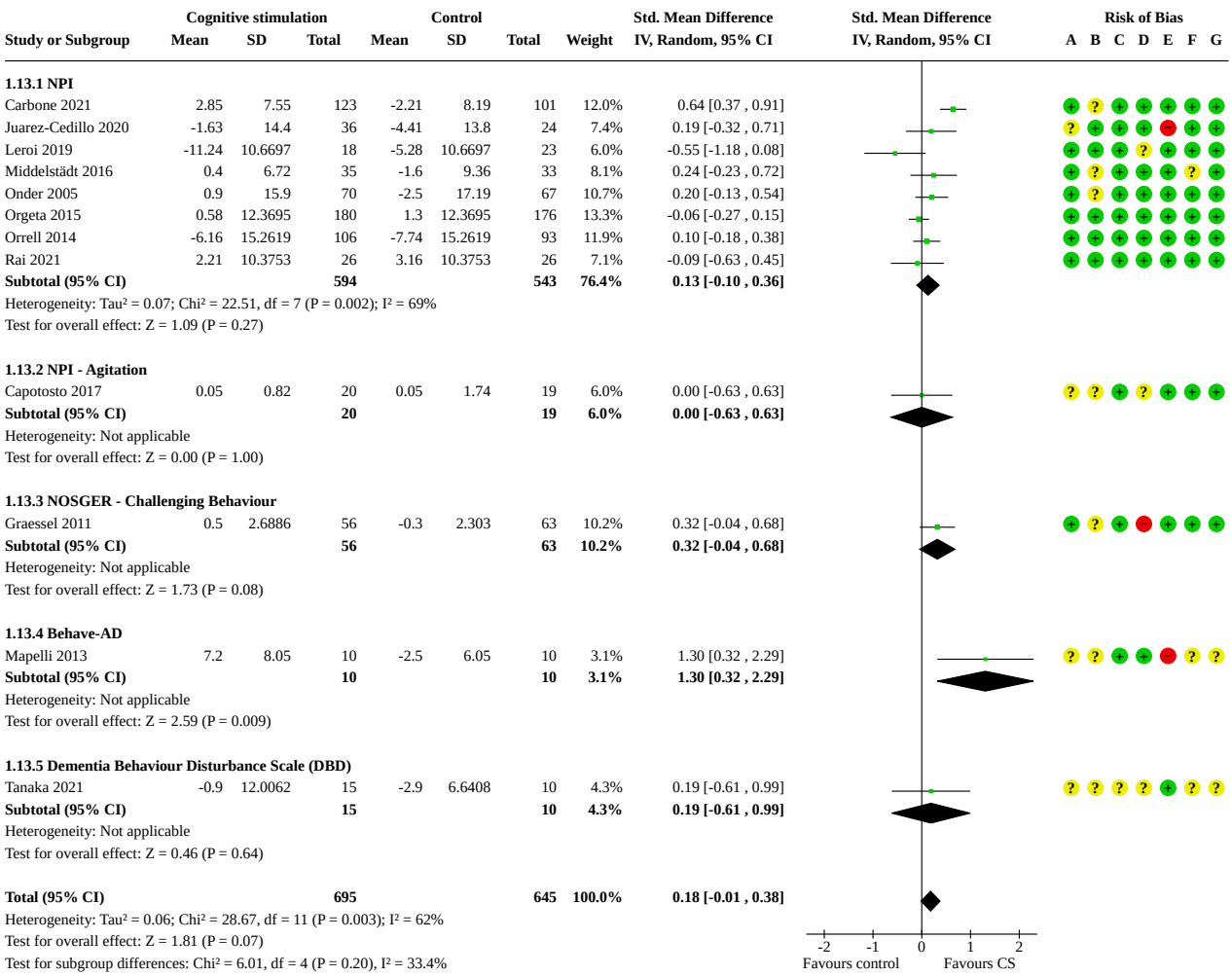
Modality

Eight studies (687 participants) offered cognitive stimulation groups and five studies (with a similar number of participants - 631) delivered individual cognitive stimulation. The test for subgroup differences did not show a significant effect (Chi^2 = 0.81, df = 1, P = 0.37, I^2 = 0%).

Behaviour that challenges

(See Figure 13).

Figure 13. Forest plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation: post-treatment, outcome: Behaviour that challenges



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of outcome assessment (detection bias)
(D) Incomplete outcome data (attrition bias)
(E) Selective reporting (reporting bias)
(F) Other bias - training and supervision
(G) Other bias - treatment manual

Many of the rating scales included in this domain focused solely on areas that have variously been described as 'Behavioural and Psychological Symptoms of Dementia (BPSD)', 'neuropsychiatric symptoms', 'behaviour that challenges' and 'agitation'. The appropriate terminology is the subject of much debate (Cunningham 2019; Wolverson 2019; Wolverson 2021). Here we use the term 'behaviour that challenges' reflecting that the difficulties are often as much for those providing care as for the person with dementia. Some scales include a mixture of items from this domain, together with items reflecting the person's skills and day-to-day function. We describe these as 'General Behaviour Rating Scales' (see below).

Twelve studies including a total of 1340 participants incorporated a measure of behaviour that challenges as an outcome measure. Eight of these used the Neuropsychiatric Interview (NPI). A further study (Capotosto 2017) also used the NPI, but gave subscale scores only - the agitation score was used for analysis as reflecting a key feature of this domain. Cognitive stimulation probably has a slight effect on behaviour that challenges (SMD 0.18, 95% CI -0.01, 0.38, I^2 =62%, moderate-quality evidence). The level of heterogeneity was substantial, but we could not explore this through subgroup analysis due to the low numbers of studies in each subgroup.

General Behaviour Rating Scales

Six studies (with 505 participants) used a General Behaviour Rating Scale. Four studies used the CAPE Behaviour Rating Scale, one study the NOSGER and one study the Blessed Dementia Rating Scale. All were studies of group cognitive stimulation, two based in care homes and three having a mix of care home and community

participants. Cognitive stimulation leads to improved scores on General Behaviour Rating Scales (SMD 0.35, 95% CI 0.13, 0.58, I^2 =30%, high-quality evidence).

Caregiver outcomes

A variety of outcome measures were utilised to evaluate the impact on family caregivers, with multiple measures used in all the studies. Domains evaluated included mood, caregiving stress and burden and health-related quality of life.

Caregiver stress/burden

Six of the studies, involving 288 participants, included a measure of caregiver burden or of stress related to caregiving. Onder 2005 and Maci 2012 used the Caregiver Burden Inventory and Marinho 2021 and Leroi 2019 used the Zarit Burden Inventory. The relatively small sample size added to the imprecision of the results, but it appears that there may be little or no effect on this domain (SMD 0.09, 95% CI -0.14, 0.32, I^2 =0%, low-quality evidence). Sensitivity analyses, including only the three studies (Ali 2021; Leroi 2019; Onder 2005) where caregivers delivered the intervention, did not change this conclusion. Leroi 2019 used the Relatives Stress Scale as well as the Zarit Burden Inventory; sensitivity analyses indicated that the results did not change if the alternate measure was included.

Depressed mood

Seven studies (Ali 2021; Bottino 2005; Leroi 2019; Maci 2012; Onder 2005; Orgeta 2015; Rai 2021) used a self-report or interviewer-rated depression scale, whilst Spector 2001 used the GHQ-12, which is regarded as an indicator of psychological distress. The analysis included 664 participants and showed that overall cognitive stimulation made little or no difference to caregivers' mood (SMD 0.05, 95% CI -0.10, 0.21, I^2 =0%, moderate-quality evidence). The five studies of individual cognitive stimulation where the caregiver delivered the intervention (Ali 2021; Leroi 2019; Onder 2005; Orgeta 2015; Rai 2021) showed a similar result (627 participants, SMD 0.02, 95% CI -0.14, 0.18, I^2 = 0%, moderate-quality evidence).

Anxiety

Five studies with a total of 600 participants provided information on the anxiety levels of family caregivers. In four studies (Leroi 2019; Onder 2005; Orgeta 2015; Rai 2021), the family caregivers delivered the intervention; in the fifth (Bottino 2005), family caregivers attended a support group. The intervention probably leads to little or no difference in anxiety for the family caregivers involved (SMD 0.00, 95% CI -0.19, 0.19, I^2 =14%, moderate-quality evidence). A sensitivity analysis, considering only the four studies (587 participants) where the family caregiver delivered the intervention, produced a similar result (SMD -0.02, 95% CI -0.18, 0.14, I^2 =0%, moderate-quality evidence).

Health-related quality of life

Two instruments were used to evaluate quality of life of family caregivers, both of which are generic health-related quality of life measures. Four studies ((Leroi 2019; Orgeta 2015; Orrell 2014; Rai 2021) used the EQ-5D, with all except Rai 2021 also reporting two indices from the SF-12: the physical component scale (PCS) and the mental component scale (MCS). Onder 2005 used the longer SF-36 (from which the short form SF-12 has been derived), reporting an overall score for this measure. Accordingly, in order to make best use of the available data, we have included the data from the SF-36 with that from the EQ-5D, to evaluate overall health-related quality of life (HRQoL), as each scale includes a combination of both mental and physical aspects. We have then analysed the data from the two components of the SF-12 separately.

For overall HRQoL, the four studies included had data from 651 participants. The analysis suggested that the intervention may have led to a slight improvement in caregiver HRQoL (SMD 0.17, 95% CI -0.14, 0.49, I^2 =66%, low-quality evidence). However, the results showed substantial inconsistency between studies and the confidence intervals were consistent with either a small positive or a slight negative effect. A sensitivity analysis, removing the one study where caregivers did not provide the intervention (Orrell 2014), again indicated there may be a slight improvement in HRQoL when family caregivers delivered cognitive stimulation (588 participants; SMD 0.28, 95% CI 0.01, 0.55, I^2 =49%, low-quality evidence), although there was moderate inconsistency and confidence intervals included a clinically relevant change and a negligible effect.

The three studies reporting data from the two components of the SF-12 ((Leroi 2019; Orgeta 2015; Orrell 2014) included 461 participants. For both the PCS and the MCS there was probably little or no difference in outcome between caregivers of people with dementia receiving cognitive stimulation and control participants (PCS: SMD 0.07, 95% CI -0.11, 0.25, I^2 =0%, moderate-quality evidence; MCS: SMD -0.05, 95% CI -0.23, 0.13, I^2 =0%, moderate-quality evidence). A sensitivity analysis, including only the two studies where caregivers delivered the intervention (Leroi 2019; Orgeta 2015), did not change this conclusion, although the slightly smaller number of participants (398) reduced the certainty associated with it.

Quality of relationship

Three studies provided data on the quality of relationship between the person with dementia and family caregiver as rated by the family caregiver. In all three studies, the family caregiver delivered the intervention, so this is a potentially important outcome. Orgeta 2015 and Rai 2021 used the Quality of Carer-Patient Rating Scale (QCPR) whilst Leroi 2019 used several scales to evaluate this domain. We have included data from the Relationship Satisfaction Scale in the meta-analysis, as this scale was also used with the participants with dementia in this study. Overall, there may be no effect of individual cognitive stimulation on quality of relationship (367 participants; SMD 0.06, 95% CI -0.26, 0.37, I^2 =36%, low-quality evidence). The confidence intervals were imprecise, including both a slight negative and a slight positive effect.

Resilience

Two studies, both of individual, caregiver-delivered, cognitive stimulation (Leroi 2019; Orgeta 2015) included (different) Brief Resilience Scales as a caregiver outcome. The intervention may have made little or no difference to this outcome (399 participants; SMD 0.06, 95% CI -0.13, 0.26, I^2 =0%, low-quality evidence).

Follow-up of outcomes

Cognition

Three studies (Baldelli 1993; Carbone 2021; Middelstädt 2016) reported a short-term follow-up in relation to cognitive outcomes, over periods of six to 12 weeks, with a total of 242 participants.

The results showed moderate inconsistency and a high level of imprecision, such that we are uncertain whether any improvement is maintained in the short-term (SMD 0.34, 95% CI -0.11, 0.80, I^2 =58%, very low-quality evidence).

Four studies, with 194 participants, provided data for a longer-term follow-up, from eight months (Tsantali 2017) to 10 months (Chapman 2004; Graessel 2011) or 12 months (Juarez-Cedillo 2020). To evaluate overall cognition, we used the most detailed cognitive assessment measure available, the ADAS-Cog for Chapman 2004, Graessel 2011 and Juarez-Cedillo 2020 and the CAM-COG for Tsantali 2017. For overall cognition, there may be a slight benefit from baseline compared with control participants (SMD 0.13, 95% CI -0.16, 0.42, I^2 =0%, low-quality evidence).

Quality of life

For self-report quality of life measures, two studies provided short-term (6-12 weeks) follow-up data (Carbone 2021; Middelstädt 2016) using the QoL-AD with 254 participants. The result was imprecise and there was substantial inconsistency, such that we are uncertain whether there is any improvement in self-reported quality of life at short-term follow-up (mean difference 0.36 points, 95% CI -2.47, 3.20, I^2 =65%, very low-quality evidence). The longer-term (10 months) follow-up from Chapman 2004, with 54 participants may show a slight improvement for the self-reported QoL-AD, but again was imprecise (mean difference 2.15 points, 95% CI -1.12, 5.42, low-quality evidence).

Two studies also reported results for the proxy QoL-AD: Middelstädt 2016 at six weeks and Chapman 2004 at 10 months. The results from these relatively small single studies indicate that cognitive stimulation may have little or no effect on proxy-rated quality of life at either six weeks or 10 months follow-up (six weeks follow-up: (Middelstädt 2016) 68 participants, mean difference -0.30 points, 95% CI -3.15, 2.55, low-quality evidence; ten months follow-up: (Chapman 2004) 54 participants, mean difference -0.28 points, 95% CI -3.14, 2.58, low-quality evidence).

Communication and interaction

One study reported on this outcome at each of short-term (12 weeks) and long-term (10 months) follow-up. At the short-term follow-up (Carbone 2021), there may be a slight improvement, but the result was imprecise (182 participants, SMD 0.33, 95% CI 0.04, 0.63, low-quality evidence). Chapman 2004 reported a similar finding at 10 months, but the result was again imprecise with broad confidence intervals (54 participants, SMD 0.15, 95% CI -0.38, 0.69, low-quality evidence).

Mood

One study (Carbone 2021)) reported a 12-week follow-up for staff/interviewer-rated mood, with a probable clinically relevant improvement (187 participants, SMD 0.54, 95% CI 0.24, 0.83, moderate-quality evidence). A further, relatively small, study (Juarez-Cedillo 2020) with 50 participants indicated there may be a slight improvement in self-reported mood 12 months after cognitive stimulation sessions have finished, although the results were imprecise (SMD 0.36, 95% CI -0.23, 0.94, low-quality evidence).

Activities of Daily Living/Instrumental Activities of Daily Living

Two studies, with 182 participants, reported short-term follow-up of Instrumental ADL outcomes (Carbone 2021; Middelstädt 2016). There may be a slight improvement, although the result was imprecise (SMD 0.12, 95% CI -0.19, 0.42, I^2 =0%, low-quality evidence). At long-term follow-up, data from three studies, with 156 participants, were available, each using a different scale to evaluate (instrumental) Activities of Daily Living (Chapman 2004, the Texas Functional Living Scale; Graessel 2011, the Erlangen Test of ADL; Juarez-Cedillo 2020, the Rapid Disability Rating Scale). Results were imprecise, with broad confidence intervals, but there may be a small benefit for measures of ADL/IADL 10 to 12 months after participation in cognitive stimulation groups (SMD 0.40, 95% CI 0.07, 0.72, I^2 =0%, low-quality evidence).

Behaviour that challenges

Short-term follow-up data for the NPI were reported by Middelstädt 2016 and Carbone 2021. There is probably a slight benefit from cognitive stimulation for this outcome (255 participants, SMD 0.19, 95% CI -0.06, 0.44, I^2 =0%, moderate-quality evidence).

Chapman 2004 and Juarez-Cedillo 2020 reported 10 to 12 months follow-up data for the Neuropsychiatric Inventory (NPI) severity score. There may be a small benefit at this time point, but the results were imprecise with broad confidence intervals (104 participants; SMD 0.43, 95% CI 0.03, 0.83, I^2 =0%, low-quality evidence). Chapman 2004 also reported the NPI caregiver distress score, with a broadly similar, and again imprecise, result (54 participants; SMD 0.41, 95% CI -0.13, 0.95, low-quality evidence).

DISCUSSION

Summary of main results

For this updated review, we included 36 RCTs with a total of 2704 participants (1432 receiving cognitive stimulation, 1272 in control groups) in the meta-analyses; we were unable to obtain useable data for the remaining study of the 37 we included in the review. Although the intervention in each study met our operational criteria for 'cognitive stimulation', there were considerable differences in implementation between studies. Notably, whilst the majority of interventions involved cognitive stimulation groups, eight offered individual cognitive stimulation, the frequency ranged from one to six sessions a week and the extent of exposure to cognitive stimulation ranged from eight sessions to 300. Only two of the included studies used digital technology in the delivery of cognitive stimulation.

Considering together the results from all modalities of, and all exposures to, cognitive stimulation, the overall finding for cognition is that cognitive stimulation interventions result in a probable small improvement in cognition immediately following the intervention, compared with control conditions. This finding includes data from 34 of the studies with 2340 participants, as two had only provided analysable data on cognitive tests at eight to 10 months follow-up. Sensitivity analyses indicated that findings were similar for the two cognitive tests most often used and recommended as core outcomes in dementia research (e.g. Webster 2017), the MMSE and the ADADS-Cog. The MMSE was used in 25 RCTs with 1893 participants. Overall, cognitive stimulation led to a probable improvement of 1.99 points, compared with control groups, a difference we judged to be clinically important. This can be taken to indicate a slowing down by six months or more of the rate of decline, which has been estimated, in mild to moderate dementia to be between 2 and 4 points on the MMSE per annum (Mohs 2000).

For the ADAS-Cog, used here in 21 studies with 1742 participants, cognitive stimulation may lead to a small improvement in ADAS-Cog scores, in comparison with control participants, with a mean difference of 2.42 points, compared with the 3-point difference we defined as clinically important.

However, as has been pointed out repeatedly over the years (Woods 2006), changes in cognition are not sufficient to justify an extensive programme of intervention, unless they are accompanied by other changes, for example, in quality of life, mood or day-to-day activities. Here there are several encouraging findings from the combined analyses of all forms and extents of cognitive stimulation. Firstly, from seven RCTs with 702 participants, results indicated a clinically relevant improvement in communication and social interaction evident in staff ratings outside the context of the cognitive stimulation group sessions either in the everyday environment or in a structured task. Secondly, slight improvements were identified in a number of domains. For Instrumental Activities of Daily Living, self-reported depressed mood, staff/interviewer-rated anxiety and general behaviour rating scales, the quality of evidence for these slight improvements was high; for quality of life (self-report and proxy), behaviour that challenges and basic Activities of Daily Living, the evidence was of moderate quality; and for staff/interviewer-rated depressed mood, it was low.

A small number of studies reported a range of outcomes for family caregivers; in most domains, little or no effect was identified. For depressed mood and anxiety and physical and mental health components of health-related quality of life, the quality of evidence was moderate; for caregiver stress and burden, resilience and the quality of the caregiving relationship (as perceived by either the person with dementia or the caregiver), the evidence was of low quality. However, there was low-quality evidence for a slight benefit to overall caregiver health-related quality of life. Sensitivity analyses indicated that our conclusions on family caregiver outcomes did not change if only those studies where the intervention was delivered by a carer were included.

These results all relate to an assessment immediately following the end of the course of cognitive stimulation sessions. Only seven studies reported follow-up evaluations some time after the end of the intervention. Three were short-term (6 to 12 weeks) and four were longer-term (8 to 12 months). However, partly due to the relatively low number of participants, the evidence was generally of low quality at best. In terms of cognition, for example, at eight to 12 months follow-up, data from four studies, with 194 participants, indicates there may be a slight benefit compared with control participants. For ADL/IADL outcomes, the results from three studies with 156 participants at a longer-term follow-up may show a slight benefit. For most of the other outcomes, data were available from just one or two studies at short- and longer-term follow-up. Here we identified a probable clinically relevant improvement in staff/interviewer-rated depressed mood at short-term follow-up (from one study with 187 participants), as well as slight improvements in behaviour that challenges at follow-up, both in the short-term (moderate-quality evidence) and longer-term (low-quality evidence).

These headline results are broadly comparable with those from the previous update of this review (Woods 2012), which highlighted improvements in cognition, communication and quality of life from the 15 studies included. The effect size for overall cognition is almost identical (0.41 in 2012, 0.40 here) although the mean differences on the two most commonly used cognitive tests are slightly higher (1.74 point mean difference on the MMSE compared with 1.99 points here; 2.27 points on the ADAS-Cog compared with 2.42 points here). The range of effects is now greater, with slight effects now emerging in additional domains such as Activities of Daily Living, mood and behaviour that challenges. However, the most striking difference is that the current results show a much greater level of heterogeneity, with much greater differences in effects between studies. For example, in the current review, I^2 was 62% for overall cognition and 60% for self-reported quality of life compared with 0% for both cognition and quality of life in the 2012 review. Examining factors involved in the substantial inconsistency between studies was then of particular importance for this updated review.

The modality of the intervention, individual versus group, was a particular issue, as the majority of the individual studies have been published since 2012, and it can be argued that the social aspects of the group setting are a core aspect of the cognitive stimulation approach, which may be difficult to replicate in the individual situation. We planned to explore whether modality did have an influence on outcomes but, in view of the relatively small number of studies using the individual approach, this was only possible for three outcomes: cognition, self-reported quality of life and instrumental activities of daily living. In each case, the subgroup analysis did not indicate that effect sizes for the individual modality were significantly different from those associated with the group approach. The imbalance in numbers of studies and participants between the two modalities means that caution is needed in drawing conclusions from these analyses and, for cognition, substantial inconsistency in both subgroups remains. Although there are important differences between studies using individual cognitive stimulation, for example, in who delivers the intervention (professional/paid worker versus family carer) and in participant population (typical dementia versus Parkinson's-related dementia versus dementia in adults with an intellectual disability), the small number of available studies means that we cannot explore these differences further at this time.

In view of group cognitive stimulation having been both widely recommended and used internationally, it may be helpful to summarise the findings for studies using this approach. For cognition overall, in 27 studies of group cognitive stimulation including 1637 participants, there is a probable small improvement compared with control participants at the evaluation following the end of the intervention. This is mirrored in a probable small improvement of 2.66 points on the ADAS-Cog (17 studies, 1168 participants) and a probable clinically important difference of 2.16 points on the MMSE (21 studies, 1325 participants). Slight improvements in quality of life, both self-reported (13 studies, 1058 participants) and proxy-rated (7 studies, 511 participants), may be associated with group cognitive stimulation, but the quality of the evidence is low, with inconsistency between studies evident as well as imprecision. There is high-quality evidence of a slight improvement in Instrumental Activities of Daily Living (8 studies, 687 participants) and moderate-quality evidence of slight improvements in self-reported mood (6 studies, 299 participants), staff/interviewer-rated mood (10 studies, 859 participants) and behaviour that challenges (8 studies, 754 participants). In addition, it should be noted that the clinically relevant improvement in communication and interaction was related entirely to studies of group cognitive stimulation.

Results for several domains, including overall cognition, self-reported and proxy-rated quality of life and staff/interviewer-rated mood continue to show a moderate or substantial degree of inconsistency between group cognitive stimulation studies, suggesting that further exploration of differences between studies would be helpful. Studies varied in a number of potentially important aspects: the total number of group sessions and their frequency, the location (care home or community), the type of control condition (alternate activity or 'treatment-as-usual') and the average level of dementia severity of participants. We only undertook subgroup analyses for overall cognition, as there were insufficient studies in each subgroup for the other outcome domains.

There was no evidence that the total number of sessions contributed to inconsistency between studies, with no difference in cognition effect sizes between those studies offering the median of 20 or more group sessions and those offering less, with both estimates associated with moderate or substantial heterogeneity. However, there did appear to be an effect of frequency of sessions, with studies offering two or more group sessions per week associated with a larger improvement in cognition (probably of clinical importance) than those offering sessions only once per week. However, moderate inconsistency remains in the estimated effect size for studies offering two or more sessions per week. .

The potential for results to be different in different settings was also explored, comparing studies of group cognitive stimulation recruiting participants from care homes with those recruiting people with dementia living at home in the community. These subgroup analyses did not include several large studies which recruited participants from both settings (Carbone 2021; Orrell 2014; Spector 2001; Spector 2003). There was no evidence of a difference in cognition effect sizes between studies carried out in the two settings, and inconsistency between studies was substantial in each setting.

The availability of a few studies offering an 'active' control condition (social activities, reading, watching TV) allowed a comparison with the majority of group studies where control participants received 'treatment-as-usual' There was no indication that cognition effect sizes were smaller where an alternate activity was offered to control participants. Inconsistency between studies was substantial in the 'treatment-as-usual' subgroup.

Finally, we explored possible differences between studies of cognitive stimulation groups based on the dementia severity level of the participants, as indicated by the average MMSE score at baseline, comparing those studies reporting an average MMSE score for their participants of above and below the median of 20 i.e. 'mild' and 'moderate' dementia, respectively. Studies where dementia severity was initially mild had a larger improvement in cognition (of probable clinical importance) than those where dementia severity was moderate. However, it should be noted that the 'mild dementia' studies showed a moderate degree of inconsistency, suggesting there remain other factors leading to differences between studies.

Overall completeness and applicability of evidence

A strength of this review is that it is based on a thorough, complete and current search of the relevant literature. However, by design, it only includes randomised controlled trials published in peer review journals. This has resulted in the exclusion of at least 35 other studies that may have met the other inclusion criteria for the review. A further exclusion that limits the completeness of the review is the restriction to people meeting diagnostic criteria for dementia, with a number of studies excluded as they additionally included individuals with Mild Cognitive Impairment (MCI), and results for people with dementia were not available separately. These exclusions lead to a more focused review, addressing directly our review question, and providing more certainty that effects observed are intervention effects than might be the case with studies lacking random allocation to treatment and control groups. Other reviews may be needed to identify the effects of cognitive stimulation therapy with people with MCI. A further limitation comes from the inclusion criteria requiring studies to have been published in the English language. This may have resulted in some studies being missed, but is in line with other reviews (e.g. Chan 2020; Kim 2017).

The review definition of 'cognitive stimulation' also led to the exclusion of a number of potentially relevant studies, including some that have used the term 'cognitive stimulation' to describe their intervention (e.g. Quayhagen 2000) and some that have been included in other reviews of 'cognitive stimulation'. Potentially a wide range of activities could be described as cognitively stimulating, including, for example, arts-based interventions, reminiscence work and so on, but we have adopted an agreed definition that underpins Cochrane reviews of other cognition-based interventions in dementia care, including those relating to cognitive training, cognitive rehabilitation and reminiscence therapy. Although clear in principle, the distinction between cognitive stimulation, with its general, wide-ranging approach and cognitive training, with its specific, domain-based cognitive exercises, can be difficult to make in practice, with some studies providing limited details of their intervention. With an evident trend towards multi-component interventions, recognising the broad range of relevant activities, we have included studies with other intervention components as long as more than half the intervention time was clearly engaged in cognitive stimulation.

The review can be considered broadly applicable to people with mild or moderate degrees of dementia. Although we could not analyse results separately for individuals with mild as compared with individuals with moderate dementia, it was possible to conduct an analysis of overall cognition based on the average severity of impairment reported for study participants, offering some indication of the effects at different (average) levels of dementia impairment. Cognitive stimulation interventions are unlikely to be appropriate for people living with advanced or severe dementia. Although this review has included one study where the participants were diagnosed as having Lewy body dementia and Parkinson's disease dementia, and another where all the participants had an intellectual disability as well as dementia, it has not been possible to examine the effects specifically on these or on other subtypes of dementia. More work may be needed to define if there are people with dementia, or subgroups, who are more or less likely to benefit from a cognitive stimulation intervention. The combined effects of cognitive stimulation and acetylcholinesterase inhibitors (AChEIs), commonly prescribed for people with Alzheimer's disease, have been of some interest (e.g. Orrell 2014). Given that AChEIs are typically 'treatment-as-usual' for people with an Alzheimer's diagnosis, it would not be considered ethical to include in trials a control group where such

medication was withheld. Separating effects of subtype (most commonly Alzheimer's and vascular dementia) then becomes confounded with medication effects. Given that the average age of participants across the review was 80 years, it is highly likely that neurodegenerative and vascular changes are co-occurring in the majority of participants, and so the distinction may be of limited pragmatic utility. Although some studies included participants as young as 54 years, none of the studies included focused specifically on younger people with dementia, or on subtypes of dementia more common in younger people (such as fronto-temporal dementia).

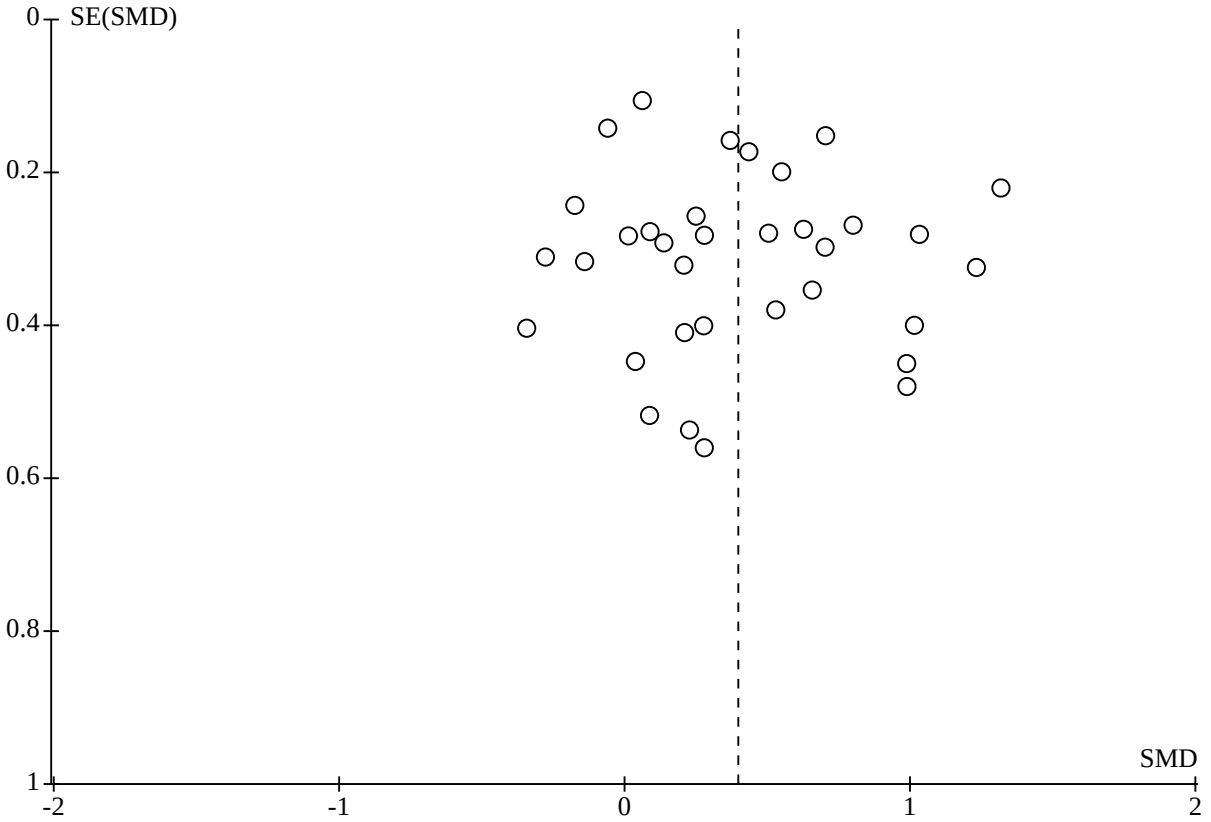
In relation to outcomes, whilst all included studies reported measures of cognitive function (as an inclusion criterion), only around half the studies reported measures of quality of life or evaluated changes in participants' mood or in aspects of day-to-day functioning. Less than a quarter reported on communication and social interaction. There is a risk that the evident emphasis on cognitive outcomes may have limited the potential to identify with certainty effects in other domains that arguably have at least as great, if not greater, impact on the experience of living with dementia.

We did not identify any reports of adverse events for study participants related to the intervention, but across the studies there has been an absence of evidence in this respect. Attrition was due to expected reasons in studies of this nature: illness, death, transfer to another facility and occasional refusal to complete follow-up assessments. The main, influential critiques identifying negative, depersonalising effects of some implementations of Reality Orientation were based on qualitative and observational reports (e.g. Dietch 1989), and so the systematic review of qualitative studies reported by Gibbor 2020a may be informative in this respect. Ten qualitative studies of cognitive stimulation were reviewed and analysed thematically; the only potentially negative aspect was that people with dementia and carers could 'sometimes experience activities as ‘childish’ or 'too easy''. Given the range of impairment of potential recipients of cognitive stimulation, these findings suggest that those providing the intervention need skills in achieving the appropriate level of difficulty and challenge for their specific group of participants. There is no indication from the current review of the amount or type of training required to deliver cognitive stimulation, with the included studies utilising therapists with a variety of backgrounds, experience and training. They included volunteers, family caregivers, speech and language therapists, occupational therapists, nurses, care workers and research staff. There appears to be broad agreement that whilst some training is needed, a professional qualification is not a prerequisite for delivery of cognitive stimulation.

We have undertaken a number of subgroup analyses, allowing some exploration of the high levels of inconsistency between studies, especially for cognitive outcomes. However, some caution is needed in interpretation of these exploratory findings. For example, whilst we may conclude that studies offering, say, group cognitive stimulation three times a week show higher effect sizes on a specific outcome than studies offering cognitive stimulation once per week, it is important to remember that these are not direct head-to-head comparisons, and may arise from a number of other differences between studies.

Finally, consideration should be given to the possibility of publication bias in this domain. By reviewing only the studies on cognitive stimulation that have been published in peer-reviewed journals, it must be acknowledged that these could represent a biased sample of the studies undertaken worldwide on this topic. In many fields of endeavour, trials that are not successful (in that they do not produce the expected positive findings) are less likely to be published. This may be especially the case with smaller trials. The welcome trend to preregistration of trials, and the publication of trial protocols, makes this less likely to occur in the future in relation to larger, well-funded trials. Funnel plots of the cognitive outcome (Figure 14) appear reasonably symmetrical, suggesting that publication bias is not a major issue in this case.

Figure 14. Funnel plot of comparison: 1 Cognitive stimulation versus no cognitive stimulation: post-treatment, outcome: 1.1 Cognition



Quality of the evidence

The quality of the included studies has improved from previous versions of this review, with 43% of studies (16/37) having at least six of the seven risk of bias items rated as 'low risk' and only 4% of risk of bias items rated as 'high risk'. There remains room for improvement, certainly in terms of reporting, and possibly in methodology, in that just over a quarter (28%) of risk of bias items were rated as 'unclear'. By design, risk of bias ratings with respect to selection bias were all rated as 'low risk' or 'unclear risk', in that only randomised controlled trials were included. However, in a third of studies, including some carried out more recently, details were sketchy regarding the procedure for random sequence generation and, in two-thirds of studies, the procedures for ensuring randomisation was carried out independently from researchers conducting evaluations were unclear. Attrition bias accounted for the highest number of 'high risk' ratings (in 4 studies), mainly due to reporting of 'per protocol' rather than 'intention-to-treat' analyses. Arrangements for training and supervision of those delivering the intervention were unclear in 43% of the studies although, encouragingly, over three-quarters reported use of some form of treatment manual or clear structure for their intervention. However, there appear to be few if any attempts to evaluate the extent to which the intervention as delivered followed the manual or protocol.

Although there were a few examples of large-scale, multicentre trials, with oversight and randomisation carried out by an independent clinical trials unit (e.g. Orgeta 2015; Orrell 2014), these are the exception rather than the rule, and the median sample size of 56 for included studies since the previous version of this review (i.e. post-2012) is only a little larger than the median of 50 for studies prior to this time.

The need for assessors to be blind to treatment allocation is widely recognised and attempted in almost all studies for outcomes such as cognition and self-reports of quality of life and mood, by these being assessed or recorded in interviews by researchers not involved at all in the intervention. Ratings by staff or family caregivers of functional abilities and behaviour that challenges or of proxy-rated quality of life are less readily blinded, depending on the role of the staff member or family member in either assisting the person with dementia in attending intervention sessions, or in the actual delivery of the intervention (as in most of the individual cognitive stimulation studies). Maintaining staff blinding to treatment allocation is more challenging than for an assessor who only visits the facility to carry out the assessments. Accordingly, findings for these domains may be more likely to be subject to bias in this respect.

In addition, other issues may arise with staff ratings. In a care home or hospital context, it is well known that achieving consistent staff ratings in research studies such as these is a major challenge. It can often be impossible to have the same staff member rate the person at each time point due to staff turnover and sickness. There can be a 'drift' in ratings over time, even with the same rater. The

observation period and opportunities for observation may vary over time. Assuring the quality of these ratings is essential if they are to be relied upon as useful outcome measures.

More attention may need to be given in future studies to demonstrating the extent to which the cognitive stimulation is delivered as planned. Well-developed treatment manuals, as used in the majority of studies now, will help with developing approaches to assuring the replicability of the intervention and adherence to its key principles. Assessment of fidelity to the treatment was not evident across the studies.

As was pointed out previously, the results of this updated review show a much greater level of inconsistency between studies than the previous review. Whilst the number of studies has more than doubled, our certainty in the evidence base has reduced in some respects, reflecting greater diversity in intervention modality, intervention methods, frequency and duration of the intervention and level of dementia severity. Examining the results for specific subgroups was only possible for cognition, but inconsistency was not fully attributable to any specific dimension of difference, suggesting there are other differences between studies that we have not been able to account for in this review, or that there is some interaction between these dimensions which could not be examined here.

Potential biases in the review process

We decided not to include in our risk of bias assessment for each study any judgement regarding the extent to which participants and interventionists had been blinded to group allocation. We consider that people with dementia and therapists cannot realistically be participating in a psychosocial intervention of this nature without awareness of the intervention being received. However, some commentators argue that an active placebo control is nonetheless needed, to ensure that any effects noted are not attributable to nonspecific effects such as meeting as a group, socialising, increased attention and so on or to motivational effects arising from not being allocated to the intervention group. For example, Huntley 2015 concluded that, in order to match the rigour of pharmacotherapy trials, "there is a need for theoretically based, well designed, blinded and adequate active control interventions, rather than relying on non-active 'treatment-as-usual' controls, which in themselves may differ widely from each other."

There are several issues to consider here. Firstly, it is suggested 'treatment-as-usual' may differ greatly, presumably between studies and for participants within studies. Within a study, random allocation and ensuring all other aspects of treatment and care remain the same for those receiving cognitive stimulation should allow results to be interpreted as the *additional, incremental* effect of cognitive stimulation. 'Treatment-as-usual' will inevitably differ greatly between studies, in relation to availability of services, resources, societal attitudes and policies. In comparison with pharmacotherapy, the implementation of psychosocial approaches is much more influenced by contextual and cultural aspects, and so results from different countries and contexts - as evident in this review - add greatly to an understanding of the effects of a specific intervention.

Secondly, Huntley 2015 suggested that 'active controls' are needed to account for nonspecific effects, although they recognise that a psychosocial equivalent of a placebo pill is difficult to achieve. Particularly for a rather general intervention such as cognitive stimulation is designed to be, a plausible active control group may well inadvertently include potential therapeutic 'ingredients'. An alternative view (Woods 2014) considered nonspecific effects an integral component of the overall intervention package, with a pragmatic randomised controlled trial focusing on what the intervention adds to usual treatment, rather than attempting to fractionate the contributions to efficacy of the various components. At a later stage, comparing different interventions of proven efficacy may be appropriate, but this is not the focus of the current review.

The great majority of the studies included in this review have included a 'treatment-as-usual' or 'waiting-list control' group (31 of the 37 studies) and one further study offered physical rehabilitation sessions to both groups (Baldelli 2002), in effect offering an augmented 'treatment-as-usual'. The remaining five studies offered a loosely structured 'active control' (watching TV, reading, social activities). Comparing the two types of control groups did not provide any support for the argument that the use of 'treatment-as-usual' controls overestimates the effects of cognitive stimulation. Three studies also reported comparisons with potentially effective comparison interventions but, for two of these, extractable data at post-treatment were not available and the third had a small sample size, and so we did not consider it appropriate to revisit the protocol and add consideration of such comparisons to this review.

A further area of potential bias arising from the review process is evident in that several members of the review team (BW, MO, AS, HR, EA) had been actively involved in one or more of the trials included in the review. In view of this, the remaining members of the review team carried out decisions regarding the inclusion of such studies, risk of bias judgements and data extraction.

Agreements and disagreements with other studies or reviews

We have identified seven relevant systematic reviews that have been published since the previous version of this review (Cafferata 2021; Chan 2020; Chen 2019; Huntley 2015; Kim 2017; Lobbia 2019; Watt 2021).

Four had a specific focus on cognitive stimulation; the others (Huntley 2015; Chan 2020; Watt 2021) included cognitive stimulation alongside other cognitive interventions or other non-pharmacological therapies, but reported results specifically for cognitive stimulation. Two of these broader reviews focused on outcomes relating to mood (Chan 2020; Watt 2021), whilst Huntley 2015 reported only on cognitive outcomes.

The number of studies included ranges greatly from 5 to 44, reflecting differences in inclusion criteria. Chen 2019 (5 studies, 315 participants) included only studies where participants had a diagnosis of Alzheimer's disease. Cafferata 2021 (44 studies, 2444 participants) and Huntley 2015 (21 studies, 1199 participants) had a broad definition of cognitive stimulation, including studies that are reviewed in the Cochrane review of reminiscence therapy (Woods 2018b). Lobbia 2019 included 12 studies that had used the treatment protocol developed by Spector 2006, and did not undertake a meta-analysis. Chan 2020 (14 studies, 1144 participants) included studies where participants met criteria for MCI and also had a broad definition of cognitive stimulation, including reminiscence interventions but, like Watt 2021 (13

studies, 805 participants), included fewer studies, as only those reporting mood outcomes were eligible. Kim 2017 (14 studies, 731 participants) drew most closely on the previous version of this review (Woods 2012), but predated a number of recent studies included in the current review.

From the five reviews reporting on cognitive outcomes, there is consensus that cognitive stimulation is associated with improved cognition compared with controls, with reviews reporting what they describe as a 'medium' or 'moderate' effect sizes. For example, Cafferata 2021 reported an overall effect size (Hedge's g) of 0.49 (95% CI 0.35 to 0.63). Several reviews cited mean improvements on specific cognitive measures from their meta-analyses. For the MMSE, the improvement ranged from 1.10 points (Chen 2019: 95% CI 0.62 to 1.58), to 1.78 points (Huntley 2015 versus non-active controls: 95% CI 1.23 to 2.33), with results from Kim 2017 (1.41 points, 95% CI 0.98 to 1.84) and Huntley 2015 (in comparison with active control groups: 1.45 points, 95% CI −0.11 to 3.02) falling in between. These results are broadly consistent with (although slightly lower than) our finding for cognitive stimulation overall of a mean difference of 1.99 points (95% CI 1.24 to 2.74). For the ADAS-Cog, an improvement of 2.41 points was reported by Cafferata 2021, reduced from 3.50 points after removing a positive outlier (Requena 2006). The other reviews reported slightly lower improvements on the ADAS-Cog ranging from 2.21 points (Kim 2017: 95% CI 0.93 to 3.49) to 1.92 points (Huntley 2015: 95% CI 0.40 to 3.43). Whilst these are all slightly lower than the estimate of 2.42 points (95% CI 1.21 to 3.63) from this review, they all fall within what we would consider to be the range of questionable clinical relevance.

The reviews provided mixed evidence in relation to mood, with Kim 2017 finding no benefit and Lobbia 2019 only weak evidence of an improvement, whereas Cafferata 2021 found strong evidence for an effect of cognitive stimulation on mood (Hedge's g = 0.46, 95% CI 0.15 to 0.78). The two reviews focusing specifically on mood showed positive effects, reporting effect sizes (standard mean differences) of 0.61 (Chan 2020: 95% CI 0.15 to 1.08) and 0.67 (Watt 2021: 95% CI 0.33 to 1.02). This lack of agreement between reviews is partially reflected in the current review, where the improvements on both self-report and staff/interviewer-rated measures are slight and, in the latter case, inconsistent.

Only three of the reviews examined the evidence on quality of life, with Lobbia 2019 reporting 'moderate' evidence of an effect and Kim 2017 citing an improvement on the QoL-AD of 2.05 points (95% CI 0.72, 3.38). In contrast, Cafferata 2021 reported no effect (Hedge's g = 0.16, 95% CI -0.16 to 0.48), but this is seen as an ambiguous finding, with heterogeneity between studies. Heterogeneity was also a feature of the current review findings on quality of life although, overall, a slight benefit was noted (SMD 0.25, 95% CI 0.07, 0.42; I^2 = 60%; moderate-quality evidence), consistent with the Kim 2017 review.

The same thee reviews reported on the effects on activities of daily living. For this domain, Kim 2017 and Lobbia 2019 did not find any evidence of an effect, whereas Cafferata 2021 identified a 'small' significant benefit (Hedge's g = 0.17, 95% CI 0.02 to 0.32), although the evidence was again seen as ambiguous in this case. The findings of the current review showed a slight effect both on basic daily living tasks, and on higher level instrumental ADLs, consistent with the findings of Cafferata 2021.

Finally, three reviews reported on outcomes relating to behaviour that challenges. Here, Kim 2017 and Cafferata 2021 reported no effect (although again for Cafferata 2021 this was an ambiguous finding), whereas Chen 2019, only looking at studies including people with Alzheimer's disease, reported a 2.14 point benefit on the NPI (95% CI 1.30 to 2.98 points). The current review showed a probable slight effect, with a substantial degree of inconsistency between studies (SMD 0.18, 95% CI -0.01, 0.38, I^2 =62%, moderate-quality evidence).

In summary, despite differences between reviews in study selection criteria and methods of analysis, there is emerging consensus that cognitive stimulation leads to improved cognition in people with dementia, with clinical relevance more likely to be apparent when the MMSE is used as an outcome measure, and with more doubtful clinical relevance when the ADAS-Cog is the outcome measure. The results for other outcomes are less clear-cut, with findings varying between reviews for mood, quality of life, activities of daily living and behaviour that challenges. Differences in measurement of outcomes and in study selection may contribute to these differences, with heterogeneity between studies evident in most domains.

Few moderating factors have been identified by the other reviews. A longer duration of cognitive stimulation was identified by Chen 2019 as having greater benefits in terms of MMSE scores and (tentatively) by Cafferata 2021 in terms of depression. Watt 2021 suggested that a combination of cognitive stimulation with acetylcholinesterase inhibitors was more effective in reducing depressed mood in people with dementia without major depression than some anti-depressants. Cafferata 2021 looked at subtypes of cognitive stimulation and found that, where Reality Orientation (RO) methods predominated, there was a greater effect on cognition than with other forms of cognitive stimulation. This finding may be in part due to a number of the RO studies having been carried out some years ago, with relatively small sample sizes - we have excluded several of those included by Cafferata 2021 in this category from the current version of this review, due to uncertainty regarding diagnostic criteria (e.g. Baines 1987; Ferrario 1991; Wallis 1983; Woods 1979). The only review to include modality (individual versus group) as a potential moderating variable was Huntley 2015, but this analysis was carried out on the whole range of cognitive interventions and not cognitive stimulation specifically, and did not identify an effect, as was the case in the current review.

AUTHORS' CONCLUSIONS

Implications for practice

This updated review adds to the evidence base for the effectiveness of cognitive stimulation for people with mild to moderate dementia in relation to cognitive function. Small benefits have been consistently demonstrated, reported in multiple trials on commonly used brief measures of cognitive function; adverse effects have not been reported. This is perhaps the most consistent finding in the literature on psychosocial interventions with people with dementia. The 37 studies included here come from 17 different countries, across five continents and a variety of contexts; from hospital, care home, nursing home, day-centre and outpatient settings; and administered in groups by staff or volunteers, or individually by family caregivers. They provide evidence of moderate quality that cognitive stimulation is associated with improved scores on cognitive tests, although there is substantial

inconsistency between studies. Exploratory subgroup analyses suggest that the frequency of group sessions and the severity of the person's initial cognitive impairment may influence the extent of cognitive benefits, and whether they are large enough to have an impact on the person's day-to-day life. Those studies offering two or more group sessions a week and those studies including people with, on average, a mild degree of dementia appeared to show the clearest benefits to cognition. In contrast to our Cochrane review of reminiscence therapy (Woods 2018a), care home studies did not show greater benefits to cognition than community studies. We cannot be certain whether the cognitive changes evident immediately after the intervention period are maintained in the following six to 12 weeks, but there may be a slight benefit eight to 12 months later.

We are generally less certain regarding the effects of cognitive stimulation on other important outcomes, such as the quality of life and mood of the person with dementia, with around half the studies having looked at these aspects. We conclude that slight improvements in quality of life, mood, activities of daily living and behaviour that challenges may be associated with cognitive stimulation, an important addition to any cognitive benefits. From a smaller number of studies, we have identified a clinically relevant improvement in communication and social interaction from studies of group cognitive stimulation.

Individual cognitive stimulation has become a research focus in recent years. Studies were diverse, with differences between studies in the participant population, the person delivering the intervention and the medium of delivery. We have not found evidence that the results are different from those obtained with group cognitive stimulation, but the small number of studies to date and their diversity, means that conclusions regarding the benefits of individual cognitive stimulation specifically would be premature.

The findings of this review are consistent with the reviews underpinning the recommendations from NICE-SCIE 2006 and NICE 2018 that group cognitive stimulation therapy should be offered to people living with mild to moderate dementia. The availability of treatment manuals setting out the cognitive stimulation approaches reviewed here, makes this intervention highly accessible to health and social care professionals, care workers and family caregivers.The World Alzheimer's Report makes a similar recommendation (Prince 2011). From a health economics viewpoint, Knapp 2022 provided data to suggest that offering cognitive stimulation groups to all those developing mild-to-moderate dementia is cost-effective and affordable, with maintenance cognitive stimulation sessions adding to the health-related quality of life of people with dementia, but at greater cost.

Implications for research

There are a number of areas, relating to both theory and practice, where further research is required. Now that its effects on cognition are well established, the theoretical basis of cognitive stimulation and the mechanisms by which change occurs would benefit from fuller investigation. This would involve studying cognitive changes both in relationship to neural processes and pathways and their linkage, if any, with outcomes such as mood, quality of life, day-to-day function and behaviour.

Research to explore further the differences that have emerged between subgroups of studies in this review would also be helpful. Direct, within-study comparisons of different intensities of group sessions and within-study analyses of the effects of dementia severity on outcomes would be informative, for example. Further research on individual cognitive stimulation is warranted, to explore more fully the benefits for people with dementia and carers, and to identify how to address the difficulties seen in some studies (e.g. Orgeta 2015) in achieving the planned level of engagement. Different modes of delivery, including digital and remote applications, for both individuals and groups require further research.

Little is known about the longer-term effects of cognitive stimulation, or how best to maintain any short-term improvements. Relatively few of the included studies reported a follow-up assessment some time after the end of the intervention, and few had an intervention period greater than six months. One study (Requena 2006) is exceptional in having a two-year intervention period, but the promising results from this study require replication. Although we were able to include a six-month once-weekly maintenance cognitive stimulation group study in our analyses (Orrell 2014), further work on the benefits of lower-level input following a period of more intensive cognitive stimulation is indicated.

Where studies have indicated the subtype dementia diagnosis of their participants, this has typically been Alzheimer's, vascular dementia, or a mixture of both. The inclusion in this review of single studies focusing on participants with Parkinson's disease dementia/Lewy body dementia and on dementia in adults with intellectual disability indicates that more research is required in relation to the effectiveness of cognitive stimulation with other dementia subtypes.

We have noted an increase in multi-component interventions, where cognitive stimulation forms one part of the overall intervention package. Here we have included studies where cognitive stimulation comprises at least half the intervention time. Whilst all studies met our definition of 'cognitive stimulation', there are doubtless differences between studies in the procedures, principles and activities followed. For example, some have included elements that might also feature in cognitive training approaches. Even where the same manual is used, there may be differences in approach and differences in fidelity to the manual that we have not been able to take account of in this review. It would be helpful to explore further in depth the content of the sessions and the additive and interactive benefits of different components such as physical exercise or more specific cognitive exercises.

In taking forward these areas of research, it is important that the wide range of outcome domains included in this review are incorporated in evaluations. As well as any effects on cognitive function, we need to understand changes in outcomes of particular significance to people with dementia and their supporters, which to date have been evaluated in much smaller numbers of studies. Further improvements in the quality of research, including adequate sample sizes, clear reporting of randomisation procedures and of attrition would also assist in developing the evidence base. A range of research designs may be considered. Qualitative studies have already been informative regarding potential mechanisms for change, and observational studies may assist in identifying components of cognitive stimulation that are

more or less engaging or help identify who benefits most from the intervention.

With the benefits of cognitive stimulation now well established, it is likely that it may be used more often in the future in comparative studies, where the challenge is for alternative interventions to provide equal or greater benefits. Whilst this would be of some interest, it will be important to recognise that interventions may benefit people living with dementia differently, and so the question should also be asked 'what works for whom?', rather than assuming that any intervention can have universal benefits.

ACKNOWLEDGEMENTS

The 2012 update of this review was supported by the Support at Home-Interventions to Enhance Life in Dementia (SHIELD) project (Application No RP-PG-0606-1083) which was funded by the NIHR Programme Grants for Applied Research funding scheme. The grant holders were Professors Orrell (UCL), Woods (Bangor), Challis (Manchester), Moniz-Cook (Hull), Russell (Swansea), Knapp (LSE) and Dr Charlesworth (UCL). The views and opinions expressed are those of the authors and do not necessarily reflect those of the Department of Health/NIHR.

For the 2022 update, we would like to thank peer reviewers Lee-fay Low, Michelle Kelly and Javier Olazaron and consumer reviewer Joanne Knowles for their comments and feedback and Anne Lethaby for skilful copy editing.

The authors wish to thank the Cochrane Dementia and Cognitive Improvement Group, especially Sue Marcus and Jenny McCleery, for support and guidance through the many versions of this review. Over a number of years the work of this group has been instrumental in the development of an evidence-base for psychosocial interventions for people living with dementia and their carers.

REFERENCES

References to studies included in this review

Ali 2021 *{published data only}*

Ali A, Brown E, Tsang W, Spector A, Aguirre E, Hoare S, et al. Individual cognitive stimulation therapy (iCST) for people with intellectual disability and dementia: a feasibility randomised controlled trial. *Aging & Mental Health* 2021;**26**:698-708. [DOI: 10.1080/13607863.2020.1869180]

Alvares-Pereira 2021 *{published data only}*

Alvares-Pereira G, Silva-Nunes MV, Spector A. Validation of the cognitive stimulation therapy (CST) program for people with dementia in Portugal. *Aging & Mental Health* 2021;**25**(6):1019-28. [DOI: 10.1080/13607863.2020.1836473]

Baldelli 1993 *{published data only}*

Baldelli MV, Pirani A, Motta M, Abati E, Mariani E, Manzi V. Effects of reality orientation therapy on elderly patients in the community. *Archives of Gerontology and Geriatrics* 1993a;**17**(3):211-8.

Baldelli 2002 *{published data only}*

Baldelli MV, Boiardi R, Fabbo A, Pradelli JM, Neri M. The role of reality orientation therapy in restorative care of elderly patients with dementia plus stroke in the subacute nursing home setting. *Archives of Gerontology and Geriatrics* 2002;**35 Suppl 8**:15-22.

Bottino 2005 *{published data only}*

Bottino CMC, Carvalho IAM, Alvarez AM, Avila R, Zukauskas PR, Bustamante SEZ, et al. Cognitive rehabilitation combined with drug treatment in Alzheimer's disease patients: a pilot study. *Clinical Rehabilitation* 2005;**19**:861-9.

Breuil 1994 *{published data only}*

Breuil V, De Rotrou J, Forette F, Tortrat D, Ganansia Ganem A, Frambourt A, et al. Cognitive stimulation of patients with dementia: preliminary results. *International Journal of Geriatric Psychiatry* 1994;**9**(3):211-7.

Buschert 2011 *{published data only}*

Buschert VC, Friese U, Teipel SJ, Schneider P, Merensky W, Rujescu D, et al. Effects of a newly developed cognitive intervention in amnestic mild cognitive impairment and mild Alzheimer's disease: a pilot study. *Journal of Alzheimer's Disease* 2011;**25**:679-94. [DOI: 10.3233/JAD-2011-100999]

Capotosto 2017 *{published data only}*

Capotosto E, Belacchi C, Gardini S, Faggian S, Piras F, Mantoan V, et al. Cognitive stimulation therapy in the Italian context: its efficacy in cognitive and non-cognitive measures in older adults with dementia. *International Journal of Geriatric Psychiatry* 2017;**32**(3):331-40. [DOI: 10.1002/gps.4521]

Carbone 2021 *{published and unpublished data}*

Carbone E, Gardini S, Pastore M, Piras F, Vincenzi, M, Borella E. Cognitive Stimulation Therapy (CST) for older adults with mild-to-moderate dementia in Italy: effects on cognitive functioning and on emotional and neuropsychiatric symptoms. *Journals of Gerontology Series B: Psychological Sciences & Social Sciences* 2021;**76**:1700-10. [DOI: 10.1093/geronb/gbab007]

Chapman 2004 *{published data only}*

Chapman SB, Weiner MF, Rackley A, Hynan LS, Zientz J. Effects of cognitive-communication stimulation for Alzheimer's disease patients treated with donepezil. *Journal of Speech, Language, and Hearing Research* 2004;**47**(5):1149-63.

Cheung 2019 *{published data only}*

Cheung DSK, Li B, Lai DWL, Leung AYM, Yu CTK, Tsang KT. Cognitive stimulating play intervention for dementia: a feasibility randomized controlled trial. *American Journal of Alzheimer's Disease and Other Dementias* 2019;**34**(1):63-71. [DOI: 10.1177/1533317518808036]

Coen 2011 *{published data only}*

Coen RF, Flynn B, Rigney E, O'Connor E, Fitzgerald L, Murray C, et al. Efficacy of a cognitive stimulation therapy programme for people with dementia. *Irish Journal of Psychological Medicine* 2011;**28**(3):145-7.

Cove 2014 *{published data only}*

Cove J, Jacobi N, Donovan H, Orrell M, Stott J, Spector A. Effectiveness of weekly cognitive stimulation therapy for people with dementia and the additional impact of enhancing cognitive stimulation therapy with a carer training program. *Clinical Interventions in Aging* 2014;**9**:2143-50. [DOI: http://dx.doi.org/10.2147/CIA.S66232]

Gibbor 2020b *{published data only}*

Gibbor L, Forde L, Yates L, Orfanos S, Komodromos C, Page H, et al. A feasibility randomised control trial of individual cognitive stimulation therapy for dementia: impact on cognition, quality of life and positive psychology. *Aging & Mental Health* 2020;**25**:999-1007. [DOI: 10.1080/13607863.2020.1747048]

Graessel 2011 *{published data only}*

* Graessel E, Stemmer R, Eichenseer B, Pickel S, Donath C, Kornhuber J, et al. Non-pharmacological, multicomponent group therapy in patients with degenerative dementia: a 12-month randomized, controlled trial. *BMC Medicine* 2011;**9**:129. [DOI: 10.1186/1741-7015-9-129]

Luttenberger K, Donath C, Uter W, Graessel E. Effects of multimodal nondrug therapy on dementia symptoms and need for care in nursing home residents with degenerative dementia: a randomized-controlled study with 6-month follow-up. *Journal of the American Geriatrics Society* 2012;**60**(5):830-40. [DOI: 10.1111/j.1532-5415.2012.03938.x]

Luttenberger K, Hofner B, Graessel E. Are the effects of a non-drug multimodal activation therapy of dementia sustainable? Follow-up study 10 months after completion of a randomised controlled trial. *BMC Neurology* 2012;**12**:151. [DOI: 10.1186/1471-2377-12-151]

Juarez-Cedillo 2020 *{published data only}*

Juarez-Cedillo T, Gutierrez-Gutierrez L, Sanchez-Hurtado LA, Martinez-Rodriguez N, Juarez-Cedillo E. Randomized controlled

trial of multi-component cognitive stimulation therapy (SADEM) in community-dwelling demented adults. *Journal of Alzheimer's Disease* 2020;**78**:1033-45. [DOI: 10.3233/JAD-200574]

Justo-Henriques 2022 *{published data only}*

Justo-Henriques SI, Perez-Saez E, Marques-Castro AE, Carvalho JO. Effectiveness of a year-long individual cognitive stimulation program in Portuguese older adults with cognitive impairment. *Aging, Neuropsychology & Cognition* 2022:1-15. [DOI: 10.1080/13825585.2021.2023458]

Kim 2016 *{published data only}*

Kim H-J, Yang Y, Oh J-G, Oh S, Choi H, Kim KH, et al. Effectiveness of a community-based multidomain cognitive intervention program in patients with Alzheimer's disease. *Geriatrics & Gerontology International* 2016;**16**(2):191-9. [DOI: 10.1111/ggi.12453]

Leroi 2019 *{published and unpublished data}*

Leroi I, Vatter S, Carter LA, Smith SJ, Orgeta V, Poliakoff E, et al. Parkinson's-adapted cognitive stimulation therapy: a pilot randomized controlled clinical trial.. *Therapeutic Advances in Neurological Disorders* 2019;**12**:1-20. [DOI: 10.1177/1756286419852217]

Lin 2018 *{published data only}*

Lin H-C, Yang Y-P, Cheng W-Y, Wang J-J. Distinctive effects between cognitive stimulation and reminiscence therapy on cognitive function and quality of life for different types of behavioural problems in dementia. *Scandinavian Journal of Caring Sciences* 2018;**32**(2):594-602. [DOI: 10.1111/scs.12484]

Lok 2020 *{published data only}*

Lok N, Buldukoglu K, Barcin E. Effects of the cognitive stimulation therapy based on Roy's adaptation model on Alzheimer's patients' cognitive functions, coping-adaptation skills, and quality of life: a randomized controlled trial. *Perspectives in Psychiatric Care* 2020;**56**:581-92. [DOI: 10.1111/ppc.12472]

Lopez 2020 *{published data only}*

López C, Sánchez JL, Martín J. The effect of cognitive stimulation on the progression of cognitive impairment in subjects with Alzheimer's disease. *Applied Neuropsychology: Adult* 2020;**29**:90-9. [DOI: 10.1080/23279095.2019.1710510]

Maci 2012 *{published data only}*

Maci T, Pira FL, Quattrocchi G, Nuovo SD, Perciavalle V, Zappia M. Physical and cognitive stimulation in Alzheimer Disease. The GAIA Project: a pilot study. *American Journal of Alzheimer's Disease and other Dementias* 2012;**27**(2):107-13. [DOI: 10.1177/1533317512440493]

Mapelli 2013 *{published data only}*

Mapelli D, Di Rosa E, Nocita R, Sava D. Cognitive stimulation in patients with dementia: randomized controlled trial. *Dementia & Geriatric Cognitive Disorders Extra* 2013;**3**:263-71. [DOI: 10.1159/000353457]

Marinho 2021 *{published and unpublished data}*

Marinho V, Bertrand E, Naylor R, Bomilcar I, Laks J, Spector A, et al. Cognitive stimulation therapy for people with dementia in Brazil (CST-Brasil): results from a single blind randomized controlled trial. *International Journal of Geriatric Psychiatry* 2021;**36**:286-93. [DOI: 10.1002/gps.5421]

Middelstädt 2016 *{published data only}*

Middelstaedt J, Folkerts A-K, Blawath S, Kalbe E. Cognitive stimulation for people with dementia in long-term care facilities: baseline cognitive level predicts cognitive gains, moderated by depression. *Journal of Alzheimer's Disease* 2016;**54**(1):253-68. [DOI: :10.3233/JAD-160181]

Onder 2005 *{published data only}*

Onder G, Zanetti O, Giacobini E, Frisoni GB, Bartorelli L, Carbone G, et al. Reality orientation therapy combined with cholinesterase inhibitors in Alzheimer's disease: randomised controlled trial. *British Journal of Psychiatry* 2005;**187**:450-5.

Orgeta 2015 *{published data only}*

* Orgeta V, Leung P, Yates L, Kang S, Hoare Z, Henderson C, et al. Individual cognitive stimulation therapy for dementia: a clinical effectiveness and cost-effectiveness pragmatic, multicentre, randomised controlled trial. *Health Technology Assessment* 2015;**19**(64):7-73. [DOI: 10.3310/hta19640]

Orrell M, Yates L, Leung P, Kang S, Hoare Z, Whitaker C, et al. The impact of individual Cognitive Stimulation Therapy (iCST) on cognition, quality of life, caregiver health, and family relationships in dementia: a randomised controlled trial. *PLOS Medicine* 2017;**14**(3):e1002269. [DOI: 10.1371/journal.pmed.1002269]

Orrell 2014 *{published and unpublished data}*

Aguirre E, Hoare Z, Spector A, Woods RT, Orrell M. The effects of a Cognitive Stimulation Therapy [CST] programme for people with dementia on family caregivers' health. *BMC Geriatrics* 2014;**14**:31. [DOI: 10.1186/1471-2318-14-31]

Aguirre E, Hoare Z, Streater A, Spector A, Woods B, Hoe J, et al. Cognitive Stimulation Therapy (CST) for people with dementia - who benefits most? *International Journal of Geriatric Psychiatry* 2013;**28**(3):284-90. [DOI: 10.1002/gps.3823]

Aguirre E, Spector A, Hoe J, Russell IT, Knapp M, Woods RT, et al. Maintenance Cognitive Stimulation Therapy (CST) for dementia: a single-blind, multi-centre, randomized controlled trial of maintenance CST vs. CST for dementia (study protocol). *Trials* 2010;**11**:46. [DOI: 10.1186/1745-6215-11-46]

D'Amico F, Rehill A, Knapp M, Aguirre E, Donovan H, Hoare Z, et al. Maintenance cognitive stimulation therapy: an economic evaluation within a randomized controlled trial. *Journal of the American Medical Directors Association* 2015;**16**(1):63-70. [DOI: 10.1016/j.jamda.2014.10.020]

* Orrell M, Aguirre E, Spector A, Hoare Z, Woods RT, Streater A, et al. Maintenance cognitive stimulation therapy for dementia: single-blind, multicentre, pragmatic randomised controlled trial. *British Journal of Psychiatry* 2014;**204**(6):454-61. [DOI: 10.1192/bjp.bp.113.137414]

Orrell M, Hoe J, Charlesworth G, Russell I, Challis D, Moniz-Cook E, et al. Support at Home: Interventions to Enhance Life in Dementia (SHIELD) – evidence, development and evaluation of complex interventions. *Programme Grants for Applied Research* 2017;**5**(5):1-184. [DOI: 10.3310/pgfar05050]

Paddick 2017 *{published data only}*

Paddick S-M, Mkenda S, Mbowe G, Kisoli A, Gray WK, Dotchin CL, et al. Cognitive stimulation therapy as a sustainable intervention for dementia in sub-Saharan Africa: feasibility and clinical efficacy using a stepped-wedge design. *International Psychogeriatrics* 2017;**29**:979-89. [DOI: 10.1017/S1041610217000163]

Rai 2021 *{published data only}*

Rai HK, Schneider J, Orrell M. An individual Cognitive Stimulation Therapy app for people with dementia and carers: results from a feasibility randomized controlled trial (RCT). *Clinical Interventions In Aging* 2021;**16**:2079-94. [DOI: 10.2147/CIA.S323994]

Requena 2006 *{published and unpublished data}*

Requena C, Lopez-Ibor MI, Maestu F, Campo P, Lopez-Ibor JJ, Ortiz T. Effects of cholinergic drugs and cognitive training on dementia. *Dementia and Geriatric Cognitive Disorders* 2004;**18**:50-4.

* Requena C, Maestu F, Fernandez A, Ortiz T. Effects of cholinergic drugs and cognitive training on dementia: 2-year follow-up. *Dementia and Geriatric Cognitive Disorders* 2006;**22**:339-45.

Spector 2001 *{published data only}*

Spector A, Orrell M, Davies S, Woods B. Can reality orientation be rehabilitated? Development and piloting of an evidence-based programme of cognition-based therapies for people with dementia. *Neuropsychological Rehabilitation* 2001;**11**(3-4):377-97.

Spector 2003 *{published data only}*

Knapp M, Thorgrimsen L, Patel A, Spector A, Hallam A, Woods B, et al. Cognitive stimulation therapy for people with dementia: cost-effectiveness analysis. *British Journal of Psychiatry* 2006;**188**:574-80.

* Spector A, Thorgrimsen L, Woods B, Royan L, Davies S, Butterworth M, et al. Efficacy of an evidence-based cognitive stimulation therapy programme for people with dementia: randomised controlled trial. *British Journal of Psychiatry* 2003;**183**:248-54.

Tanaka 2021 *{published data only}***10.1111/psyg.12627**

Tanaka S, Yamagami T, Yamaguchi H. Effects of a group-based physical and cognitive intervention on social activity and quality of life for elderly people with dementia in a geriatric health service facility: a quasi-randomised controlled trial. *Psychogeriatrics* 2021;**21**:71-9. [DOI: 10.1111/psyg.12627]

Tsantali 2017 *{published data only}*

* Tsantali E, Economidis D, Rigopoulou S. Testing the benefits of cognitive training vs. cognitive stimulation in mild Alzheimer's disease: a randomised controlled trial. *Brain Impairment* 2017;**18**(2):188-96. [DOI: https://doi.org/10.1017/BrImp.2017.6]

Young 2019 *{published data only}*

Young DKW, Ng PYN, Kwok T, Ho F, Cheng D, Mak V, et al. The effects of an expanded cognitive stimulation therapy model on the improvement of cognitive ability of elderly with mild stage dementia living in a community - a randomized waitlist controlled trial. *Aging & Mental Health* 2019;**23**(7):855-62. [DOI: 10.1080/13607863.2018.1471586]

References to studies excluded from this review

Alves 2014 *{published data only}*

Alves J, Alves-Costa F, Magalhaes R, Goncalves OF, Sampaio A. Cognitive stimulation for Portuguese older adults with cognitive impairment: a randomized controlled trial of efficacy, comparative duration, feasibility, and experiential relevance. *American Journal of Alzheimer's Disease and Other Dementias* 2014;**29**:503-12. [DOI: 10.1177/1533317514522541]

Apostolo 2013 *{published data only}*

Apostolo J, Gil I, Rosa A, Almeida J, Fernandes A. The effect of cognitive stimulation on cognition in the elderly. *Alzheimers & Dementia* 2013;**9**(4 Suppl. 1):P650-1. [DOI: 10.1016/j.jalz.2013.05.1331 MISC1 - 20140418]

Apostolo 2014 *{published data only}*

Apostolo JLA, Cardoso DFB, Rosa AI, Paul C. The effect of cognitive stimulation on nursing home elders: a randomized controlled trial. *Journal of Nursing Scholarship* 2014;**46**:157-66. [DOI: 10.1111/jnu.12072]

Arcoverde 2008 *{published data only}*

Arcoverde C, Deslandes A, Rangel A, Rangel A, Pavllo R, Nigri F, et al. Role of physical activity on the maintenance of cognition and activities of daily living in elderly with Alzheimer's disease. *Arquivos de Neuro-Psiquiatria* 2008;**66**(2b):323-7.

Baglio 2015 *{published data only}*

Baglio F, Griffanti L, Saibene FL, Ricci C, Alberoni M, Critelli R, et al. Multistimulation group therapy in Alzheimer's disease promotes changes in brain functioning. *NeuroRehabilitation and Neural Repair* 2015;**29**:13-24. [DOI: 10.1177/1545968314532833]

Baines 1987 *{published data only}*

Baines S, Saxby P, Ehlert K. Reality orientation and reminiscence therapy: a controlled cross-over study of elderly confused people. *British Journal of Psychiatry* 1987;**151**:222-31.

Basak 2008 *{published data only}*

Basak C, Boot WR, Voss MW, Kramer AF. Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging* 2008;**23**(4):765-77.

Brook 1975 *{published data only}*

Brook P, Degun G, Mather M. Reality orientation, a therapy for psychogeriatric patient: a controlled study. *British Journal of Psychiatry* 1975;**127**:42-5.

Buettner 2011 *{published data only}*

Buettner LL, Fitzsimmons S, Atav S, Sink K. Cognitive stimulation for apathy in probable early-stage Alzheimer's. *Journal of Aging Research* 2011;**2011**:no pagination. [DOI: 10.4061/2011/480890]

Calatayud 2022 *{published data only}*

Calatayud E, Jimenez-Sanchez C, Calvo S, Brandin De la Cruz N, Herrero P, Gomez-Soria I. Effectiveness of cognitive stimulation personalized by the preexisting cognitive level in older adults: a randomized clinical trial. *Topics in Geriatric Rehabilitation* 2022;**38**:73-80. [DOI: 10.1097/TGR.0000000000000345]

Camargo 2015 *{published data only}*

Camargo CHF, Justus FF, Retzlaff G. The effectiveness of reality orientation in the treatment of Alzheimer's Disease. *American Journal of Alzheimer's Disease and Other Dementias* 2015;**30**:527-32. [DOI: 10.1177/1533317514568004]

Camargo 2019 *{published data only}*

Camargo CHF, Ladeira MA, Serpa RA, Jobbins VA, Pucci CR, Welling LC, et al. The effectiveness of reality orientation therapy in the treatment of Parkinson disease dementia. *American Journal of Alzheimer's Disease and Other Dementias* 2019;**34**:124-30. [DOI: 10.1177/1533317518802461]

Carlson 2008 *{published data only}*

Carlson MC, Saczynski JS, Rebok GW, Seeman T, Glass TA, McGill S, et al. Exploring the effects of an "everyday" activity program on executive function and memory in older adults: Experience Corps. *Gerontologist* 2008;**48**(6):793-801.

Cassinello 2008 *{published data only}*

Cassinello DZ, Tarraga L, Fernandez-Ballesteros R. Cognitive plasticity in Alzheimer's disease patients receiving cognitive stimulation programmes. *Psychology in Spain* 2009;**13**(1):48-54.

* Cassinello DZ, Tarraga L, Fernandez-Ballesteros R. Cognitive plasticity in Alzheimer's disease patients receiving cognitive stimulation programs. *Psicothema* 2008;**20**:432-7.

Cheng 2006 *{published data only}*

Cheng ST, Chan ACM, Yu ECS. An exploratory study of the effect of Mahjong on the cognitive functioning of persons with dementia. *International Journal of Geriatric Psychiatry* 2006;**21**(7):611-7.

Constantinidou 2008 *{published data only}*

Constantinidou F, Thomas RD, Robinson L. Benefits of categorization training in patients with traumatic brain injury during post-acute rehabilitation: additional evidence from a randomized controlled trial. *Journal of Head Trauma Rehabilitation* 2008;**23**(5):312-28.

Croisile 2006 *{published data only}*

Croisile B. Memory stimulation. What rational? What exercises? *Revue de Geriatrie* 2006;**31**(6):421-33.

Davis 2001 *{published data only}*

Davis RN, Massman PJ, Doody RS. Cognitive intervention in Alzheimer disease: a randomized placebo-controlled study. *Alzheimer Disease and Associated Disorders* 2001;**15**(1):1-9.

Devita 2021 *{published data only}*

Devita M, Masina F, Mapelli D, Anselmi P, Sergi G, Coin A. Acetylcholinesterase inhibitors and cognitive stimulation, combined and alone, in treating individuals with mild Alzheimer's disease. *Aging: Clinical & Experimental Research* 2021;**33**:3039-45. [DOI: 10.1007/s40520-021-01837-8]

Eckroth-Bucher 2009 *{published data only}*

Eckroth-Bucher M, Siberski J. Preserving cognition through an integrated cognitive stimulation and training program. *American Journal of Alzheimer's Disease and Other Dementias* 2009;**24**(3):234-45.

Eggermont 2009a *{published data only}*

Eggermont LH, Knol DL, Hol EM, Swaab DF, Scherder EJ. Hand motor activity, cognition, mood, and the rest-activity rhythm in dementia: a clustered RCT. *Behavioural Brain Research* 2009;**196**(2):271-8.

Eggermont 2009b *{published data only}*

Eggermont LH, Swaab DF, Hol EM, Scherder EJ. Walking the line: a randomised trial on the effects of a short term walking programme on cognition in dementia. *Journal of Neurology, Neurosurgery and Psychiatry* 2009;**80**(7):802-4.

Evans 2009 *{published data only}*

Evans JJ, Greenfield E, Wilson BA, Bateman A. Walking and talking therapy: improving cognitive-motor dual-tasking in neurological illness. *Journal of the International Neuropsychological Society* 2009;**15**(1):112-20.

Faggian 2007 *{published data only}*

Faggian S. An intervention protocol on cognitive abilities for subjects with severe cognitive impairment. *Giornale di Gerontologia* 2007;**55**(3):134-43.

Fanto 2002 *{published data only}*

Fanto F, Tisci C, Molaschi M, Ostacoli L, Furlan PM, Riondato A, et al. Long-term reality orientation therapy for subjects with dementia in an Italian suburban setting. In: 8th International Conference on Alzheimer's Disease and Related Disorders; 2002 July 20-25, Stockholm, Sweden. 2002:Abstract No 1966.

Farina 2006a *{published data only}*

Farina E, Mantovani F, Fioravanti R, Pignatti R, Chiavari L, Imbornone E, et al. Evaluating two group programmes of cognitive training in mild-to-moderate AD: is there any difference between a 'global' stimulation and a 'cognitive-specific' one? *Aging & Mental Health* 2006;**10**(3):211-8.

Farina 2006b *{published data only}*

Farina E, Mantovani F, Fioravanti R, Rotella G, Villanelli F, Imbornone E, et al. Efficacy of recreational and occupational activities associated to psychologic support in mild to moderate Alzheimer disease - a multicenter controlled study. *Alzheimer Disease & Associated Disorders* 2006;**20**(4):275-82.

Fernandez-Calvo 2010 *{published data only}*

Fernandez-Calvo B, Contador I, Serna A, De Lucena VM, Ramos F. The effect of an individual or group intervention format in cognitive stimulation of patients with Alzheimer's disease [El efecto del formato de intervencion individual o grupal en la estimulacion cognitiva de pacientes con enfermedad de Alzheimer]. *Revista de Psicopatología y Psicología Clínica* 2010;**15**:115-23.

Ferrario 1991 *{published data only}*

Ferrario E, Cappa G, Molaschi M, Rocco M, Fabris F. Reality orientation therapy in institutionalized elderly patients: preliminary results. *Archives of Gerontology and Geriatrics* 1991;**12 Suppl 2**:139-42.

Folkerts 2018 *{published data only}*

Folkerts AK, Dorn ME, Roheger M, Maassen M, Koerts J, Tucha O, et al. Cognitive stimulation for individuals with Parkinson's disease dementia living in long-term care: preliminary data from a randomized crossover pilot study. *Parkinson's Disease* 2018;**8104673**:no pagination. [DOI: 10.1155/2018/8104673]

Gerber 1991 *{published data only}*

Gerber GJ, Prince PN, Snider HG, Atchison K, Dubois L, Kilgour JA. Group activity and cognitive improvement among patients with Alzheimer's disease. *Hospital and Community Psychiatry* 1991;**42**(8):843-5.

Goldstein 1982 *{published data only}*

Goldstein G. An evaluation of reality orientation therapy. *Journal of Behavioral Assessment* 1982;**4**(2):165-78.

Gonzalez-Abraldes 2010 *{published data only}*

Gonzalez-Abraldes I, Millan-Calenti JC, Balo-Garcia A, Tubio J, Lorenzo T, Maseda A. Accessibility and usability of computer-based cognitive stimulation: telecognitio. *Revista Espanola de Geriatria y Gerontologia* 2010;**45**(1):26-9.

Green 2009 *{published data only}*

Green HA, Patterson K. Jigsaws - a preserved ability in semantic dementia. *Neuropsychologia* 2009;**47**(2):569-76.

Greenaway 2008 *{published data only}*

Greenaway MC, Hanna SM, Lepore SW, Smith GE. A behavioral rehabilitation intervention for amnestic mild cognitive impairment. *American Journal of Alzheimer's Disease and other Dementias* 2008;**23**(5):451-61.

Han 2017 *{published data only}*

Han JW, Lee H, Hong JW, Kim K, Kim T, Byun HJ, et al. Multimodal cognitive enhancement therapy for patients with mild cognitive impairment and mild dementia: a multi-center, randomized, controlled, double-blind, crossover trial. *Journal of Alzheimer's Disease* 2017;**55**:787-96. [DOI: 10.3233/JAD-160619]

Hanley 1981 *{published data only}*

Hanley IG, McGuire RJ, Boyd WD. Reality orientation and dementia: a controlled trial of two approaches. *British Journal of Psychiatry* 1981;**138**:10-4.

Hill 2014 *{published data only}***10.3928/19404921-20140311-01**

* Hill NL, Kolanowski AM, Fick D, Chinchilli VM, Jablonski RA. Personality as a moderator of cognitive stimulation outcomes in older adults at high risk for cognitive decline. *Research in Gerontological Nursing* 2014;**7**(4):159-70. [DOI: 10.3928/19404921-20140311-01]

Holden 1978 *{published data only}*

Holden UP, Sinebruchow A. Reality orientation therapy: a study investigating the value of this therapy in the rehabilitation of elderly people. *Age and Ageing* 1978;**7**(2):83-90.

Johnson 1981 *{published data only}*

Johnson CH, McLaren SM, McPherson FM. The comparative effectiveness of three versions of 'classroom' reality orientation. *Age and Ageing* 1981;**10**(1):33-5.

Justo-Henriques 2019 *{published data only}*

Justo-Henriques SI, Marques-Castro AE, Otero P, Vazquez FL, Torres AJ. Long-term individual cognitive stimulation program in patients with mild neurocognitive disorder: a pilot study. *Revista de Neurologia* 2019;**7**:281-9. [DOI: 10.33588/rn.6807.2018321]

Kim 2015 *{published data only}*

Kim KW, Han JW, Yoon JC, Ryu S-H, Lee N-J, Hong JW, et al. Effects of multimodal cognitive enhancement therapy (MCET) for people with mild cognitive impairment and early stage dementia: a randomized, controlled, double-blind, cross-over trial. *Alzheimers & Dementia* 2015;**11**(7 Suppl. 1):P465.

Kim 2020 *{published data only}*

Kim D. The effects of a recollection-based occupational therapy program of Alzheimer's disease: a randomized controlled trial. *Occupational Therapy International* 2020;**2020**:6305727. [DOI: 10.1155/2020/6305727]

Kolanowski 2016 *{published data only}*

Kolanowski A, Fick D, Litaker M, Mulhall P, Clare L, Hill N, et al. Effect of cognitively stimulating activities on symptom management of delirium superimposed on dementia: a randomized controlled trial. *Journal of American Geriatrics Society* 2016;**64**:2424-32. [DOI: 10.1111/jgs.14511]

Liu 2021 *{published data only}*

Liu T, Spector A, Mograbi DC, Cheung G, Wong GH. Changes in default mode network connectivity in resting-state fMRI in people with mild dementia receiving cognitive stimulation therapy. *Brain Sciences* 2021;**11**(9):1137.

Luttenberger 2019 *{published data only}*

Luttenberger K, Graessel E, Behrndt EM, Ozbe D, Donath C, Scheel J. Responder analysis of a multicomponent non-pharmacological intervention (MAKS) for people with cognitive impairment in the German day-care study (DeTaMAKS). *Frontiers in Psychiatry* 2019;**10**:587. [DOI: 10.3389/fpsyt.2019.00587]

Matsuda 2007 *{published data only}*

Matsuda O. Cognitive stimulation therapy for Alzheimer's disease: the effect of cognitive stimulation therapy on the

progression of mild Alzheimer's disease in patients treated with donepezil. *International Psychogeriatrics* 2007;**19**(2):241-52.

McCormick 2019 *{published data only}*

McCormick SA, Vatter S, Carter LA, Smith SJ, Orgeta V, Poliakoff E, et al. Parkinson's-adapted cognitive stimulation therapy: feasibility and acceptability in Lewy body spectrum disorders. *Journal of Neurology* 2019;**266**:1756-70.

Menna 2016 *{published data only}*

Menna LF, Santaniello A, Gerardi F, Di Maggio A, Milan G. Evaluation of the efficacy of animal-assisted therapy based on the reality orientation therapy protocol in Alzheimer's disease patients: a pilot study. *Psychogeriatrics* 2016;**16**:240-6. [DOI: 10.1111/psyg.12145]

Meza-Kubo 2009 *{published data only}*

Meza-Kubo V, Gonzalez-Fraga A, Moran AL, Tentori M. Augmenting cognitive stimulation activities in a nursing home through pervasive computing. In: La-Web: 2009 Latin American Web Congress. 2009:8-15.

Milev 2008 *{published data only}*

Milev RV, Kellar T, McLean M, Mileva V, Luthra V, Thompson S, et al. Multisensory stimulation for elderly with dementia: a 24-week single-blind randomized controlled pilot study. *American Journal of Alzheimer's Disease and Other Dementias* 2008;**23**(4):372-6.

Mudge 2008 *{published data only}*

Mudge AM, Giebel AJ, Cutler AJ. Exercising body and mind: an integrated approach to functional independence in hospitalized older people. *Journal of the American Geriatrics Society* 2008;**56**(4):630-5.

Muniz 2015 *{published data only}*

Muniz R, Serra CM, Reisberg B, Rojo JM, Del Ser T, Pena Casanova J, et al. Cognitive-motor intervention in Alzheimer's disease: long-term results from the Maria Wolff trial. *Journal of Alzheimer's Disease* 2015;**45**:295-304. [DOI: 10.3233/JAD-142364]

Newson 2006 *{published data only}*

Newson R, Kemps E. The influence of physical and cognitive activities on simple and complex cognitive tasks in older adults. *Experimental Aging Research* 2006;**32**(3):341-62.

Niu 2010 *{published data only}*

Niu YX, Tan JP, Guan JQ, Zhang ZQ, Wang LN. Cognitive stimulation therapy in the treatment of neuropsychiatric symptoms in Alzheimer's disease: a randomized controlled trial. *Clinical Rehabilitation* 2010;**24**:1002-11. [DOI: 10.1177/0269215510376004]

Okamura 2018 *{published data only}*

Okamura H, Otani M, Shimoyama N, Fujii T. Combined exercise and cognitive training system for dementia patients: a randomized controlled trial. *Dementia* 2018;**45**:318-25. [DOI: 10.1159/000490613]

Olazaran 2004 *{published data only}*

Olazaran J, Muniz R, Reisberg B, Pena-Casanova J, Del Ser T, Cruz-Jentoft AJ, et al. Benefits of cognitive-motor intervention in MCI and mild to moderate Alzheimer disease. *Neurology* 2004;**63**:2348-53.

Oliveira 2021 *{published data only}*

Oliveira J, Gamito P, Souto T, Conde R, Ferreira M, Corotnean T, et al. Virtual reality-based cognitive stimulation on people with mild to moderate dementia due to Alzheimer's disease: a pilot randomized controlled trial. *International Journal of Environmental Research & Public Health* 2021;**18**:16. [DOI: 10.3390/ijerph18105290]

Onieva-Zafra 2018 *{published data only}*

Onieva-Zafra MD, Hernandez-Garcia L, Gonzalez-del-Valle MT, Parra-Fernandez ML, Fernandez-Martinez E. Music intervention with Reminiscence Therapy and Reality Orientation for elderly people with Alzheimer disease living in a nursing home: a pilot study. *Holistic Nursing Practice* 2018;**32**:43-50. [DOI: 10.1097/HNP.0000000000000247]

Orrell 2005 *{published data only}*

Orrell M, Spector A, Thorgrimsen L, Woods B. A pilot study examining the effectiveness of Maintenance Cognitive Stimulation Therapy (MCST) for people with dementia. *International Journal of Geriatric Psychiatry* 2005;**20**:446-51.

Perkins 2022 *{published data only}*

Perkins L, Fisher E, Felstead C, Rooney C, Wong GHY, Dai R, et al. Delivering Cognitive Stimulation Therapy (CST) virtually: developing and field-testing a new framework. *Clinical Interventions in Aging* 2022;**17**:97-116. [DOI: 10.2147/CIA.S348906]

Piras 2017 *{published data only}*

Piras F, Carbone E, Faggian S, Salvalaio E, Gardini S, Borella E. Efficacy of cognitive stimulation therapy for older adults with vascular dementia. *Dementia & Neuropsychologia* 2017;**11**:434-41. [DOI: 10.1590/1980-57642016dn11-040014]

Quayhagen 1995 *{published data only}*

Corbeil RR, Quayhagen MP, Quayhagen M. Intervention effects on dementia caregiving interaction: a stress-adaptation modeling approach. *Journal of Aging and Health* 1999;**11**(1):79-95.

* Quayhagen MP, Quayhagen M, Corbeil RR, Roth PA, Rodgers JA. A dyadic remediation program for care recipients with dementia. *Nursing Research* 1995;**44**(3):153-9.

Quayhagen 2000 *{published data only}*

Quayhagen MP, Quayhagen M, Corbeil RR, Hendrix RC, Jackson JE, Snyder L, et al. Coping with dementia: evaluation of four non pharmacologic interventions. *International Psychogeriatrics* 2000;**12**(2):249-65.

Raggi 2007 *{published data only}*

Raggi A, Iannaccone S, Marcone A, Ginex V, Ortelli P, Nonis A, et al. The effects of a comprehensive rehabilitation program of

Alzheimer's disease in a hospital setting. *Behavioural Neurology* 2007;**18**(1):1-6.

Reeve 1985 *{published data only}*

Reeve W, Ivison D. Use of environmental manipulation and classroom and modified informal reality orientation with institutionalized, confused elderly patients. *Age and Ageing* 1985;**14**(2):119-21.

Riegler 1980 *{published data only}*

Riegler J. Comparison of a reality orientation program for geriatric patients with and without music. *Journal of Music Therapy* 1980;**17**(1):26-33.

Rueda 2021 *{published data only}*

Rueda AV, Cabaco AS, Mejia-Ramirez M, Justo-Henriques SI, Carvalho JO. Improvement of the quality of life in aging by stimulating autobiographical memory. *Journal of Clinical Medicine* 2021;**10**:3168. [DOI: 10.3390/jcm10143168]

Ruiz Sanchez de Leon 2007 *{published data only}*

Ruiz Sanchez De Leon JM, Llanero Luque M. Computer-based cognitive training and donepezil: combined therapy effects in cognitive impairment. *Mapfre Medicina* 2007;**18**(1):25-33.

Schecker 2013 *{published data only}*

Schecker M, Pirnay-Dummer P, Schmidtke K, Hentrich-Hesse T, Borchardt D. Cognitive interventions in mild Alzheimer's disease: a therapy-evaluation study on the interaction of medication and cognitive treatment. *Dementia and Geriatric Cognitive Disorders Extra* 2013;**3**(1):301-11.

Schecker M. Cognitive interventions in mild Alzheimer's disease. In: Europe Challenges the Burden of Mental Disorders. 21st European Congress of Psychiatry, European Psychiatric Association; 2013 April 6-9; Nice (France). 2013.

Schmitter-Edgecombe 2008 *{published data only}*

Schmitter-Edgecombe M, Howard JT, Pavawalla SP, Howell L, Rueda A. Multidyad memory notebook intervention for very mild dementia: a pilot study. *American Journal of Alzheimer's Disease & Other Dementias* 2008;**23**(5):477-87.

Scott 2003 *{published data only}*

Scott J, Clare L. Do people with dementia benefit from psychological interventions offered on a group basis? *Clinical Psychology and Psychotherapy* 2003;**10**(3):186-96.

Silva 2017 *{published data only}*

Silva AR, Pinho MS, Macedo L, Moulin CJA. The cognitive effects of wearable cameras in mild Alzheimer Disease - an experimental study. *Current Alzheimer Research* 2017;**14**:1270-82. [DOI: 10.2174/1567205014666170531083015]

Silva 2021 *{published data only}*

Silva R, Bobrowicz-Campos E, Santos-Costa P, Cruz AR, Apostolo J. A home-based individual cognitive stimulation program for older adults with cognitive impairment: a randomized controlled trial. *Frontiers in Psychology* 2021;**12**:5416. [DOI: 10.3389/fpsyg.2021.741955]

Skov 2022 *{published data only}*

Skov SS, Nielsen MBD, Krolner RF, Oksnebjerg L, Ronbol Lauridsen SM. A multicomponent psychosocial intervention among people with early-stage dementia involving physical exercise, cognitive stimulation therapy, psychoeducation and counselling: results from a mixed-methods study. *Dementia* 2022;**21**:316-34. [DOI: 10.1177/14713012211040683]

Smith 2009 *{published data only}*

Smith GE, Housen P, Yaffe K, Ruff R, Kennison RF, Mahncke HW, et al. A cognitive training program based on principles of brain plasticity: results from the improvement in memory with plasticity-based adaptive cognitive training (IMPACT) study. *Journal of the American Geriatrics Society* 2009;**57**(4):594-603.

Tadaka 2004 *{published data only}*

Tadaka E, Kanagawa K. A randomized controlled trial of a group care program for community-dwelling elderly people with dementia. *Japan Journal of Nursing Science* 2004;**1**:19-25.

Tanaka 2017 *{published data only}*

Tanaka S, Honda S, Nakano H, Sato Y, Araya K, Yamaguchi H. Comparison between group and personal rehabilitation for dementia in a geriatric health service facility: single-blinded randomized controlled study. *Psychogeriatrics* 2017;**17**:177-85. [DOI: 10.1111/psyg.12212]

Tarraga 2006 *{published data only}*

Tarraga L, Badenas S, Modinos G, Espinosa A, Diego S, Balcells J, et al. Randomized pilot study to evaluate the efficacy of an interactive multimedia internet-based tool for the cognitive stimulation of patients with Alzheimer's disease. In: 57th Annual Meeting of the American Academy of Neurology; 2005 April; Miami Beach. 2005:P01.138.

* Tarraga L, Boada M, Modinos G, Espinosa A, Diego S, Morera A, et al. A randomised pilot study to assess the efficacy of an interactive, multimedia tool of cognitive stimulation in Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 2006;**77**:1116-21.

Tarraga L, Boada M, Modinos G, Espinosa A, Diego S, Morera A, et al. Interactive cognitive stimulation with a computer and Alzheimer's disease. *Neuroscientist* 2007;**13**(2):97.

Tarraga L. Randomised pilot study to evaluate the efficacy of an interactive multimedia system for the cognitive stimulation of Alzheimer's disease patients. In: 1st International Consensus Conference: Non pharmacological Therapies for Alzheimer's, 2005 May 12-13; Madrid. 2005:35-6.

Thickpenny-Davis 2007 *{published data only}*

Thickpenny-Davis KL, Barker-Collo SL. Evaluation of a structured group format memory rehabilitation program for adults following brain injury. *Journal of Head Trauma Rehabilitation* 2007;**22**:303-13.

Tsai 2008 *{published data only}*

Tsai AY, Yang M, Lan C, Chen C. Evaluation of effect of cognitive intervention programs for the community-dwelling elderly with subjective memory complaints. *International Journal of Geriatric Psychiatry* 2008;**23**(11):1172-4.

Tsai 2019 *{published data only}*

Tsai AY, Lee MC, Lai CC, Chou YC, Su CY. The outcomes of Cognitive Stimulation Therapy (CST) for community-dwelling older adults with cognitive decline in Taiwan. *Topics in Geriatric Rehabilitation* 2019;**35**:306-12. [DOI: 10.1097/TGR.0000000000000248]

Van Zon 2016 *{published data only}*

Van Zon L, Kirby JR, Anderson N. The efficacy of a volunteer-administered cognitive stimulation program in long-term care homes. *International Psychogeriatrics* 2016;**28**:995-1004. [DOI: 10.1017/S1041610215002392]

Wallis 1983 *{published data only}*

Wallis GG, Baldwin M, Higginbotham P. Reality orientation therapy - a controlled trial. *British Journal of Medical Psychology* 1983;**56**(3):271-7.

Wenisch 2007 *{published data only}*

Wenisch E, Cantegreil-Kallen I, De Rotrou J, Garrigue P, Moulin F, Batouche F, et al. Cognitive stimulation intervention for elders with mild cognitive impairment compared with normal aged subjects: preliminary results. *Aging Clinical & Experimental Research* 2007;**19**(4):316-22.

Wettstein 2004 *{published data only}*

Wettstein A, Schmid R, Konig M. Who participates in psychosocial interventions for caregivers of patients with dementia? *Dementia and Geriatric Cognitive Disorders* 2004;**18**(1):80-6.

Williams 1987 *{published data only}*

Williams R, Reeve W, Ivison D, Kavanagh D. Use of environmental manipulation and modified informal reality orientation with institutionalized, confused elderly subjects: a replication. *Age and Ageing* 1987;**16**(5):315-8.

Woods 1979 *{published data only}*

Woods RT. Reality orientation and staff attention: a controlled study. *British Journal of Psychiatry* 1979;**134**:502-7.

Yamanaka 2013 *{published data only}*

Yamanaka K, Kawano Y, Noguchi D, Nakaaki S, Watanabe N, Amano, et al. Effects of cognitive stimulation therapy Japanese version (CST-J) for people with dementia: a single-blind, controlled clinical trial.. *Aging & Mental Health* 2013;**17**:579-86. [DOI: 10.1080/13607863.2013.795742]

Young 2020 *{published data only}*

Young DK. Multicomponent intervention combining a cognitive stimulation group and tai chi to reduce cognitive decline among community-dwelling older adults with probable dementia: a multi-center, randomized controlled trial. *Dementia* 2020;**19**:2073-89. [DOI: 10.1177/1471301218814637]

Zanetti 1995 *{published data only}*

Zanetti O, Frisoni GB, De Leo D, Dello Bueno M, Bianchetti A, Trabucchi M. Reality orientation therapy in Alzheimer disease: useful or not? A controlled study. *Alzheimer Disease and Associated Disorders* 1995;**9**(3):132-8.

Zepelin 1981 *{published data only}*

Zepelin H, Wolfe CS, Kleinplatz F. Evaluation of a year long reality orientation program. *Journal of Gerontology* 1981;**36**(1):70-7.

Zientz 2007 *{published data only}*

Zientz J, Rackley A, Chapman SB, Hopper T, Mahendra N, Cleary S. Evidence-based practice recommendations: caregiver-administered active cognitive stimulation for individuals with Alzheimer's disease. *Journal of Medical Speech-Language Pathology* 2007;**15**(3):27-34.

References to ongoing studies

ChiCTR1800018600 *{published data only}*

ChiCTR1800018600. Evaluation on the effect of maintenance cognitive stimulation therapy for dementia patients. www.chictr.org.cn/showprojen.aspx?proj=26814 (first received 29 September 2018).

NCT04550975 *{published data only}*

NCT04550975. Advanced Cognitive Stimulation Therapy (ACST). clinicaltrials.gov/ct2/show/NCT04550975 (first received 16 September 2020).

NCT04828434 *{published data only}*

NCT04828434. Virtual individual Cognitive Stimulation Therapy: a proof of concept study (V-iCST). clinicaltrials.gov/ct2/show/NCT04828434 (first received 2 April 2021).

Additional references

ADI 2022

Alzheimer's Disease International. Dementia statistics. www.alzint.org/about/dementia-facts-figures/dementia-statistics/ (accessed 7 December 2022).

Aguirre 2011

Aguirre E, Spector A, Streater A, Hoe J, Woods B, Orrell M. Making a Difference 2: An evidence-based group programme to offer maintenance Cognitive Stimulation Therapy (CST) to people with dementia. London: Hawker Publications, 2011.

Aguirre 2014

Aguirre E, Spector A, Orrell M. Guidelines for adapting cognitive stimulation therapy to other cultures. *Clinical Interventions in Aging* 2014;**9**:1003-7.

Alexopoulos 1988

Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell Scale for depression in dementia. *Biological Psychiatry* 2008;**23**:271-84.

Alzheimer's Research UK 2022

Alzheimer's Research UK. Dementia Statistics Hub. www.alzint.org/about/dementia-facts-figures/dementia-statistics/ (accessed 7 December 2022).

Alzheimer's Society 2022

Alzheimer's Society. Left to Cope Alone: the unmet support needs after a dementia diagnosis. London: Alzheimer's Society, 2022. [https://www.alzheimers.org.uk/sites/default/files/2022-07/left-to-cope-alone-after-diagnosis-report.pdf]

Bahar-Fuchs 2019

Bahar-Fuchs A, Martyr A, Goh AM, Sabates J, Clare L. Cognitive training for people with mild to moderate dementia. *Cochrane Database of Systematic Reviews* 2019, Issue 3. Art. No: CD013069. [DOI: 10.1002/14651858.CD013069.pub2]

Burton 1982

Burton M. Reality orientation for the elderly: a critique. *Journal of Advanced Nursing* 1982;**7**:427-33.

Cafferata 2021

Cafferata RM, Hicks B, Von Bastian CC. Effectiveness of cognitive stimulation for dementia: a systematic review and meta-analysis. *Psychological Bulletin* 2021;**147**(5):455-76. [DOI: 10.1037/bul0000325]

Chan 2020

Chan JY, Chan TK, Kwok TC, Wong SY, Lee AT, Tsoi KK. Cognitive training interventions and depression in mild cognitive impairment and dementia: a systematic review and meta-analysis of randomized controlled trials.. *Age & Ageing* 2020;**49**(5):738-47. [DOI: 10.1093/ageing/afaa063]

Chen 2019

Chen J, Duan Y, Li H, Lu L, Liu J, Tang C. Different durations of cognitive stimulation therapy for Alzheimer's disease: a systematic review and meta-analysis.. *Clinical Interventions in Aging* 2019;**14**:1243–54. [DOI: 10.2147/CIA.S210062]

Cheung 2021

Cheung G, Peri K. Challenges to dementia care during COVID-19: innovations in remote delivery of group Cognitive Stimulation Therapy. *Aging & Mental Health* 2021;**25**(6):977-9. [DOI: 10.1080/13607863.2020.1789945]

Clare 2004

Clare L, Woods RT. Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: a review. *Neuropsychological Rehabilitation* 2004;**14**:385-401.

Collins 2020

Collins RN, Kishita N. Prevalence of depression and burden among informal care-givers of people with dementia: a meta-analysis. *Ageing & Society* 2020;**40**:2355-92. [DOI: 10.1017/S0144686X19000527]

Cummings 1997

Cummings JL. The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. *Neurology* 1997;**48**(5 (Supplement 6)):10S-16S. [DOI: 10.1212/WNL.48.5_Suppl_6.10S]

Cunningham 2019

Cunningham C, Macfarlane S, Brodaty H. Language paradigms when behaviour changes with dementia: # BanBPSD. *International Journal of Geriatric Psychiatry* 2019;**34**:1109-13.

Deeks 2021

Deeks JJ, Higgins JPT, Altman DG (editors), Cochrane Statistical Methods Group. Chapter 10: Analysing data and undertaking meta-analysis. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al (editors). Cochrane Handbook for Systematic Reviews of Interventions, version 6.2, Cochrane 2021. Available from training.cochrane.org/handbook. [www.training.cochrane.org/handbook]

Dietch 1989

Dietch JT, Hewett LJ, Jones S. Adverse effects of reality orientation. *Journal of American Geriatrics Society* 1989;**37**:974-6.

Folstein 1975

Folstein MF, Folstein SE, McHugh PR. 'Mini Mental State': a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;**12**:189-98.

Gallacher 2005

Gallacher J, Bayer A, Ben-Shlomo Y. Commentary: Activity each day keeps dementia away — does social interaction really preserve cognitive function? *International Journal of Epidemiology* 2005;**34**:872-3.

Gibbor 2020a

Gibbor L, Yates L, Volkmer A, Spector A. Cognitive stimulation therapy (CST) for dementia: a systematic review of qualitative research. *Aging & Mental Health* 2020;**25**:980-90. [DOI: 10.1080/13607863.2020.1746741]

Greene 1982

Greene JG, Smith R, Gardiner M, Timbury GC. Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: a factor analytic study. *Age & Ageing* 1982;**11**:121-6.

Group EuroQol 1990

Group EuroQol. EuroQoL: a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199-208.

Guyatt 2011

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 2011;**64**(4):383-94.

Hall 2013

Hall L, Orrell M, Stott J, Spector A. Cognitive stimulation therapy (CST): neuropsychological mechanisms of change. *International Psychogeriatrics* 2013;**25**:479-89.

Hamilton 1959

Hamilton M. The assessment of anxiety states by rating. *British Journal of Medical Psychology* 1959;**32**:50-5.

Higgins 2011

Higgins JPT, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. (updated March 2011). The Cochrane Collaboration, 2011. Available from: training.cochrane.org/handbook/archive/v5.1.

Holden 1995

Holden UP, Woods RT. Positive Approaches to Dementia Care. 3rd edition. Edinburgh: Churchill Livingstone, 1995.

Holden 2020

Holden E, Stoner CR, Spector A. Cognitive stimulation therapy for dementia: provision in National Health Service settings in England, Scotland and Wales. *Dementia* 2020;**20**:1553-64. [DOI: 10.1177/1471301220954611]

Howard 2011

Howard R, Phillips P, Johnson T, O'Brien J, Sheehan B, Lindesay J, et al. Determining the minimum clinically important differences for outcomes in the DOMINO trial. *International Journal of Geriatric Psychiatry* 2011;**26**:812-7. [DOI: 10.1002/gps.2607]

Hughes 1982

Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *British Journal of Psychiatry* 1982;**140**:566-72.

Hultsch 1999

Hultsch DF, Hertzog C, Small BJ, Dixon RA. Use it or lose it: engaged lifestyle as a buffer of cognitive decline in aging? *Psychology & Aging* 1999;**14**(2):245-63.

Huntley 2015

Huntley JD, Gould RL, Liu K, Smith M, Howard RJ. Do cognitive interventions improve general cognition in dementia? A meta-analysis and meta-regression. *BMJ Open* 2015;**5**:e005247. [DOI: 10.1136/bmjopen-2014-005247]

Kaddour 2020

Kaddour L, Kishita N. Anxiety in informal dementia carers: a meta-analysis of prevalence. *Journal of Geriatric Psychiatry and Neurology* 2020;**33**:161-72. [DOI: 10.1177/0891988719868313]

Kim 2017

Kim K, Han JW, So Y, Seo J, Kim YJ, Park JH, et al. Cognitive stimulation as a therapeutic modality for dementia: a meta-analysis. *Psychiatry Investigation* 2017;**14**:626-39. [DOI: 10.4306/pi.2017.14.5.626]

Kitwood 1997

Kitwood T. Dementia Reconsidered: The Person Comes First. Buckingham: Open University Press, 1997.

Knapp 2022

Knapp M, Bauer A, Wittenberg R, Comas-Herrera A, Cyhlarova E, Hu B, et al. What are the current and projected future cost and health-related quality of life implications of scaling up cognitive stimulation therapy? *International Journal of Geriatric Psychiatry* 2022;**37**:1-10. [DOI: 10.1002/gps.5633]

Kudlicka 2019

Kudlicka A, Martyr A, Bahar-Fuchs A, Woods B, Clare L. Cognitive rehabilitation for people with mild to moderate dementia. *Cochrane Database of Systematic Reviews* 2019, Issue 8. Art. No: CD013388. [DOI: 10.1002/14651858.CD013388]

Leung 2018

Leung P. People's experiences of cognitive stimulation therapy: a qualitative understanding. In: Yates LA, Yates J, Orrell M, Spector A, Woods B, editors(s). Cognitive Stimulation Therapy for Dementia: History, Evolution and Internationalism. London: Routledge, 2018:131-51.

Liu 2018

Liu BTY, Au ACL, Wong GHY. Neuropsychological aspects of cognitive stimulation therapy. In: Yates, LA, Yates J, Orrell M, Spector A, Woods B, editors(s). Cognitive Stimulation Therapy for Dementia: History, Evolution and Internationalism. London: Routledge, 2018:153-73.

Livingston 2020

Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 2020;**396**(10248):413-46.

Lobbia 2019

Lobbia A, Carbone E, Faggian S, Gardini S, Piras F, Spector A, et al. The efficacy of cognitive stimulation therapy (CST) for people with mild-to-moderate dementia. *European Psychologist* 2019;**24**:257-77. [DOI: 10.1027/1016-9040/a000342]

Logsdon 2002

Logsdon RG, Gibbons LE, McCurry SM, Teri L. Assessing quality of life in older adults with cognitive impairment. *Psychosomatic Medicine* 2002;**64**:510-9.

Macera 1993

Macera CA, Eaker ED, Jannarone RJ, Davis DR, Stoskopf CH. A measure of perceived burden among caregivers. *Evaluation & the Health Professions* 1993;**16**:204-11. [DOI: 10.1177/016327879301600205]

Mohs 2000

Mohs R. Neuropsychological assessment of patients with Alzheimer's disease. In: Bloom FE, Kupfer DJ, editors(s). Psychopharmacology: 4th Generation of Progress. American College of Neuropsychopharmacology, 2000.

Moniz-Cook 2008

Moniz-Cook E, Vernooij-Dassen M, Woods R, Verhey F, Chattat R, De Vugt M, et al, INTERDEM group. A European consensus on outcome measures for psychosocial intervention research in dementia care. *Aging & Mental Health* 2008;**12**:14-29.

NICE 2018

NICE (National Institute of Health and Care Excellence). Dementia: Assessment, Management and Support for People Living with Dementia and Their Carers. NICE guideline [NG97]. London: National Institute of Health and Care Excellence, 2018. [https://www.nice.org.uk/guidance/ng97/evidence]

NICE-SCIE 2006

NICE-SCIE. Dementia: Supporting People with Dementia and their Carers in Health and Social Care: Clinical Guideline 42. London: NICE-SCIE, 2006.

Novak 1989

Novak M, Guest C. Application of a multidimensional caregiver burden inventory. *Gerontologist* 1989;**29**:798-803.

Orfanos 2020

Orfanos S, Gibbor L, Carr C, Spector A. Group-based cognitive stimulation therapy for dementia: a qualitative study on experiences of group interactions. *Aging & Mental Health* 2020;**25**:991-8. [DOI: 10.1080/13607863.2020.1746740]

Pattie 1979

Pattie A, Gilleard C. Clifton Assessment Procedures for the Elderly (CAPE). Sevenoaks: Hodder & Stoughton, 1979.

Powell-Proctor 1982

Powell-Proctor L, Miller E. Reality orientation: a critical appraisal. *British Journal of Psychiatry* 1982;**140**:457-63.

Prince 2011

Prince M, Bryce R, Ferri C. World Alzheimer Report 2011: the Benefits of Early Diagnosis and Intervention. London: Alzheimer's Disease International, 2011. [http://www.alz.co.uk/research/WorldAlzheimerReport2011.pdf]

Prince 2014

Prince M, Knapp M, Guerchet M, McCrone P, Prina M, Comas-Herrera A, et al. Dementia UK; Update. London: Alzheimer's Society, 2014.

Rai 2020

Rai HK, Griffiths R, Yates L, Schneider J, Orrell M. Field-testing an iCST touch-screen application with people with dementia and carers: a mixed method study. *Aging & Mental Health* 2020;**25**:1008-18. [DOI: 10.1080/13607863.2020.1783515]

Richardson 2019

Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. *Clinical Epidemiology & Global Health* 2019;**7**:192-8. [DOI: 10.1016/j.cegh.2018.05.005]

Rosen 1984

Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *American Journal of Psychiatry* 1984;**141**:1356-64.

Sabat 1994

Sabat SR. Excess disability and malignant social psychology: a case study of Alzheimer's disease. *Journal of Community & Applied Social Psychology* 1994;**4**:157-66.

Salthouse 2006

Salthouse TA. Mental exercise and mental aging: evaluating the validity of the "use it or lose it" hypothesis. *Perspectives on Psychological Science* 2006;**1**:68-87.

Schepers 2012

Schepers AK, Orrell M, Shanahan N, Spector A. Sense of competence in Dementia Care Staff (SCIDS) scale: development, reliability, and validity. *International Psychogeriatrics* 2012;**24**:1153-62. [DOI: 1017/S104161021100247X]

Schrag 2012

Schrag A, Schott JM, Alzheimer's Disease Neuroimaging Initiative. What is the clinically relevant change on the ADAS-Cog? *Journal of Neurology, Neurosurgery & Psychiatry* 2012;**83**(2):171-3.

Shankar 1999

Shankar KK, Walker M, Frost D, Orrell MW. The development of a valid and reliable scale for rating anxiety in dementia (RAID). *Aging & Mental Health* 1999;**3**:39-49.

Small 2002

Small GW. What we need to know about age related memory loss. *BMJ* 2002;**324**:1502-5.

Smith 2008

Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. *International Journal of Behavioral Medicine* 2008;**15**:194-200.

Spector 2000b

Spector A, Orrell M, Davies S, Woods RT. Reality orientation for dementia. *Cochrane Database of Systematic Reviews* 2000, Issue 3. Art. No: CD001119. [DOI: 10.1002/14651858.CD001119]

Spector 2006

Spector A, Thorgrimsen L, Woods B, Orrell M. Making a Difference: an Evidence-Based Group Programme to Offer Cognitive Stimulation Therapy (CST) to People with Dementia. London: Hawker Publications, 2006.

Spector 2010

Spector A, Orrell M, Woods B. Cognitive Stimulation Therapy (CST): effects on different areas of cognitive function for people with dementia. *International Journal of Geriatric Psychiatry* 2010;**25**:1253-8.

Spector 2011

Spector A, Gardner C, Orrell M. The impact of Cognitive Stimulation Therapy groups on people with dementia: views from participants, their carers and group facilitators. *Aging & Mental Health* 2011;**15**(8):945-9.

Spector 2020

Spector A, Woods B, Stoner CR, Orrell M. Making a Difference - 1. 2nd revised edition. London: Hawker Publications, 2020. [ISBN: 9781874790891]

Spruytte 2002

Spruytte N, Van Audenhove C, Lammertyn F, Storms G. The quality of the caregiving relationship in informal care for older adults with dementia and chronic psychiatric patients. *Psychology and Psychotherapy: Theory, Research and Practice* 2002;**75**(3):295-311.

Streater 2020

Streater A, Yates L, Orrell M, Rosen J, Taylor Smith A, Schneider J. Interacting with television in one's own home: the development of a cognitive stimulation television pilot episode for older people with dementia (Innovative Practice). *Dementia* 2020;**19**(8):2881-8. [DOI: 10.1177/1471301219836001]

Tarlov 1989

Tarlov AR, Ware JE, Greenfield S. The Medical Outcomes Study: an application of methods for monitoring the results of medical care. *Journal of the American Medical Association* 1989;**262**:925-30.

Taulbee 1966

Taulbee LR, Folsom JC. Reality orientation for geriatric patients. *Hospital & Community Psychiatry* 1966;**17**:133-5.

Terada 2015

Terada S, Oshima E, Ikeda C, Hayashi S, Yokota O, Uchitomi Y. Development and evaluation of a short version of the quality of life questionnaire for dementia. *International Psychogeriatrics* 2015;**27**:103-10.

Wagnild 2009

Wagnild GM. The Resilience Scale: User's Guide for the US English Version of the Resilience Scale and the 14-Item Resilience Scale (RS-14). Worden, MT: Resilience Center, 2009.

Wan 2014

Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology* 2014;**14**:135. [DOI: 10.1186/1471-2288-14-135]

Ware 1996

Ware JE Jr, Kosinski M, Keller SD. A 12 item short form health survey: construction of scales and preliminary tests of reliability and validity. *Medical Care* 1996;**34**:220-3.

Watt 2021

Watt JA, Goodarzi Z, Veroniki AA, Nincic V, Khan PA, Ghassemi M, et al. Comparative efficacy of interventions for reducing symptoms of depression in people with dementia: systematic review and network meta-analysis. *British Medical Journal* 2021;**372**:n352. [DOI: 10.1136/bmj.n532]

Webster 2017

Webster L, Groskreutz D, Grinbergs-Saull A, Howard R, O'Brien JT, Mountain G, et al. Core outcome measures for interventions to prevent or slow the progress of dementia for people living with mild to moderate dementia: systematic review and consensus recommendations. *PLOS One* 2017;**12**(6):e0179521. [DOI: 10.1371/journal.pone.0179521]

Wolverson 2019

Wolverson E, Birtles H, Moniz-Cook E, James I, Brooker D, Duffy F. Naming and framing the behavioural and psychological symptoms of dementia (BPSD) paradigm: professional stakeholder perspectives. *OBM Geriatrics* 2019;**3**:1-19. [DOI: 10.21926/obm.geriatr.1904080]

Wolverson 2021

Wolverson E, Dunn R, Moniz-Cook E, Gove D, Diaz-Ponce A. The language of behaviour changes in dementia: a mixed methods survey exploring the perspectives of people with dementia. *Journal of Advanced Nursing* 2021;**77**:1992-2001. [DOI: 10.1111/jan.14787]

Woods 1977

Woods RT, Britton PG. Psychological approaches to the treatment of the elderly. *Age & Ageing* 1977;**6**:104-12.

Woods 2002

Woods B. Editorial: Reality Orientation: a welcome return? *Age & Ageing* 2002;**31**:155-6.

Woods 2006

Woods B, Thorgrimsen L, Spector A, Royan L, Orrell M. Improved quality of life and cognitive stimulation therapy in dementia. *Aging & Mental Health* 2006;**10**(3):219-26.

Woods 2014

Woods B, Russell I. Randomisation and Chance-Based Designs in Social Care Research. NIHR School for Social Care Research Methods Review 17. London: NIHR SSCR, 2014. [http://sscr.nihr.ac.uk/PDF/MR/MR17.pdf]

Woods 2016

Woods RT, Orrell M, Bruce E, Edwards RT, Hoare Z, Hounsome B, et al. REMCARE: Pragmatic multi-centre randomised trial of reminiscence groups for people with dementia and their family carers: effectiveness and economic analysis. *PLOS One* 2016;**11**:e0152843. [DOI: 10.1371/journal.pone.0152843]

Woods 2018a

Woods B. CST: development process. In: Yates L, Yates J, Orrell M, Spector A, Woods B, editors(s). Cognitive Stimulation Therapy for Dementia: History, Evolution and Internationalism. Abingdon: Routledge, 2018:33-47.

Woods 2018b

Woods B, O'Philbin L, Farrell EM, Spector AE, Orrell M. Reminiscence therapy for dementia. *Cochrane Database of Systematic Reviews* 2018, Issue 3. Art. No: CD001120. [DOI: 10.1002/14651858.CD001120.pub3]

Yates 2014

Yates L, Orrell M, Leung P, Spector A, Woods B, Orgeta V. Making a Difference 3. Individual Cognitive Stimulation Therapy: a Manual for Carers. London: Hawker, 2014.

Yesavage 1983

Yesavage JA, Brink TL, Rose TL. Development and validation of a geriatric depression scale: a preliminary report. *Journal of Psychiatric Research* 1983;**17**:37-49.

Zarit 1980

Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. *Gerontologist* 1980;**20**:649-55.

Zigmond 1983

Zigmond AS, Snaith RP. The Hospital Anxiety and Depression scale. *Acta Psychiatrica Scandinavica* 1983;**67**:361-70.

References to other published versions of this review

Aguirre 2013

Aguirre E, Woods RT, Spector A, Orrell M. Cognitive Stimulation for dementia: a systematic review of the evidence of effectiveness from randomized controlled trials. *Ageing Research Reviews* 2013;**12**(1):253-62. [DOI: 10.1016/j.arr.2012.07.001]

Spector 2000a

Spector A, Davies S, Woods B, Orrell M. Reality orientation for dementia: a systematic review of the evidence for its effectiveness. *Gerontologist* 2000;**40**(2):206-12.

Woods 2012

Woods B, Aguirre E, Spector A, Orrell M. Cognitive stimulation to improve cognitive functioning in people with dementia. *Cochrane Database of Systematic Reviews* 2012, Issue 2. Art. No: CD005562. [DOI: 10.1002/14651858.CD005562.pub2]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Ali 2021

Study characteristics	
Methods	RCT
Participants	N = 40 (23F/17M) Confirmed clinical diagnosis of mild or moderate dementia in adults (aged 40 and over) with premorbid mild or moderate intellectual disability Age 60.4 (SD 8.2) Most (72.5%) living in supported housing/residential care 45% receiving anti-dementia medication
Interventions	Individual cognitive stimulation delivered by carers (N = 20) (most carers (82.5%) were paid carers; remainder are relatives or friends) Waiting-list controls receiving usual treatment (N = 20)
Outcomes	Cognition: Cambridge Cognitive Examination for Older Adults with Down Syndrome (CAM-COG-DS); Modified Memory for Objects tests from Neuropsychological Assessment of Dementia in Intellectual Disabilities Battery; Cognitive Scale for Down Syndrome (CSDS) (proxy-completed) Quality of life: QoL-AD (proxy-completed) ADL: Alzheimer's Dementia Cooperative Study-Activities of Daily Living Inventory (ADCS-ADL) Caregiver outcomes: Care Giving Burden Scale; Sense of Competence in Dementia Care Staff Scale (SCIDS); Hospital Anxiety and Depression Scale (HADS)
Notes	Recommended 30 minutes, twice a week, for 20 weeks

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"The randomisation process was performed centrally using a web-based system".

Ali 2021 *(Continued)*

Allocation concealment (selection bias)	Low risk	Randomisation carried out by external administrator and unblinded researcher informed participants of allocation.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinded evaluator
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition at end-point and intention-to-treat analyses reported
Selective reporting (reporting bias)	Low risk	All outcomes cited in protocol paper appeared to be reported.
Other bias - training and supervision	Low risk	One-to-one training and ongoing support provided to carers implementing the intervention
Other bias - treatment manual	Low risk	Used adapted version of iCST Manual (Yates 2014)

Alvares-Pereira 2021

Study characteristics	
Methods	RCT
Participants	N = 105 (91F/14M) Met the DSM-5 criteria for neurocognitive disorder (dementia) (no information regarding medication) Age 83.6 (SD 7.6; range 59-98) Eight centres participated (2 day-centres, 2 nursing homes, 2 psychogeriatric centres, 1 hospital, 1 rehabilitation centre)
Interventions	Cognitive stimulation group sessions (N = 55) Treatment-as-usual (N = 50)
Outcomes	Cognition: ADAS-Cog; Clinical Dementia Rating (CDR) Cognitive Reserve Questionnaire Quality of life: QoL-AD (self-report & proxy scores combined in paper - separate scores provided by authors) Communication: Holden Communication Scale Mood: Cornell Scale for Depression in Dementia (CSDD); RAID (Rating of Anxiety in Dementia) Behaviour: Behaviour Rating Scale (CAPE)
Notes	45 minutes, twice a week, for 7 weeks
Risk of bias	

Alvares-Pereira 2021 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomisation process performed centrally using a web-based system
Allocation concealment (selection bias)	Low risk	Randomisation carried out by external administrator; participants informed of group allocation by an unblinded researcher
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Outcome assessors blinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition in intervention group - reasons given for larger dropout in control group
Selective reporting (reporting bias)	Low risk	All outcomes cited in Methods section were reported.
Other bias - training and supervision	Low risk	Facilitators trained by researchers who had trained at International CST centre
Other bias - treatment manual	Low risk	Used Portugese adaptation of CST Manual (Spector 2006)

Baldelli 1993

Study characteristics	
Methods	RCT
Participants	N = 23 (23F/0M) Alzheimer's (SDAT) Mean MMSE 20.6 (SD 4.9) Mean age 84.5 (SD 6.4; range 75-94) All resident in institution
Interventions	RO group sessions (N = 13) Treatment-as-usual (N = 10)
Outcomes	Cognition: MMSE; Berg Orientation Scale Mood: GDS-30 ADL: Stewart ADL scale
Notes	60 minutes, 3 times a week for 3 months; 3-month follow-up data on cognitive measures

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**

Baldelli 1993 *(Continued)*

Random sequence generation (selection bias)	Unclear risk	Stated by email that their trials were randomised (with no detail of the methods used)
Allocation concealment (selection bias)	Unclear risk	Stated by email that their trials were randomised (with no detail of the methods used)
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	No details of who assessors were
Incomplete outcome data (attrition bias) All outcomes	Low risk	Zero attrition at 3-month post-treatment assessment; no attrition reported at follow-up 3 months later
Selective reporting (reporting bias)	Low risk	Scores for all measures reported
Other bias - training and supervision	Unclear risk	No details of who carried out the groups
Other bias - treatment manual	Unclear risk	Described as 'formal ROT'

Baldelli 2002

Study characteristics	
Methods	RCT
Participants	N = 87 (61F/26M) 'Degenerative senile dementia of the Alzheimer's type (SDAT)' (N = 46) and "vascular multi-infarct dementia" (N = 41) Mean MMSE 20.7 (SD 3.0) Mean age 80.0 (range 65-97) Resident in subacute care nursing home All had at least elementary schooling. "All had comorbid conditions consisting of vascular accidents with acute motor deficits of recent onset".
Interventions	RO + physical therapy programme (N = 71) Physical therapy programme only (N = 16)
Outcomes	Cognition: MMSE Mood: Geriatric Depression Scale (GDS-30) ADL: Barthel
Notes	60 minutes, 5 days per week for one month

Baldelli 2002 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Stated by email that their trials were randomised (with no detail of the methods used)
Allocation concealment (selection bias)	Unclear risk	Stated by email that their trials were randomised (with no detail of the methods used)
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	No details of assessors given
Incomplete outcome data (attrition bias) All outcomes	Low risk	Zero attrition reported
Selective reporting (reporting bias)	Low risk	Scores for all measures reported
Other bias - training and supervision	Unclear risk	No information given on therapists
Other bias - treatment manual	Unclear risk	No information given - 'ROT sessions'

Bottino 2005

Study characteristics	
Methods	RCT
Participants	N = 13 (9F/4M) 'Mildly impaired probable Alzheimer's diagnosis' All participants taking rivastigmine 6-12 mg/day for 2 months Mean MMSE 22.3 (SD 3.6; range 16-28) Age 73.7 (range 62-83) Outpatients
Interventions	'cognitive rehabilitation' plus rivastigmine; carers attended a support group at same time (N = 6) Treatment-as-usual: rivastigmine plus 30 minute monthly consultation with doctor in relation to medication (N = 7)
Outcomes	Participants: Cognition: MMSE; ADAS-Cog, plus battery of neuropsychological tests ADL (rated by carer) Carers' mood: Hamilton Anxiety and Montgomery-Asberg Depression Rating Scales

Bottino 2005 *(Continued)*

Notes	90 minutes, once a week, for 5 months

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomised blocks design
Allocation concealment (selection bias)	Low risk	Random allocation to either group by telephone made by an assessor blind to the patient group
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessment made by assessors blinded to group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Zero attrition reported
Selective reporting (reporting bias)	Low risk	Scores for all measures reported
Other bias - training and supervision	Unclear risk	No information provided on therapists
Other bias - treatment manual	Unclear risk	Some detail given of 'cognitive rehabilitation training sessions', meeting criteria for cognitive stimulation, but no mention of manual or protocol

Breuil 1994

Study characteristics

Methods	RCT
Participants	N = 61 (37F/24M) Diagnosis of dementia (DSM-III) (90% have Alzheimer's disease) Age 77.1 (range 61-93) Mean MMSE 21.5 (range 9-29) Outpatients
Interventions	Cognitive stimulation (N = 29) Treatment-as-usual (N = 27)
Outcomes	Cognition: MMSE, CERAD ADL: ECA scale rated by family members
Notes	60 minutes, 2 times a week, for 5 weeks

Risk of bias

Breuil 1994 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Described as randomised, but no details of randomisation reported
Allocation concealment (selection bias)	Unclear risk	No details of randomisation reported
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Cognitive assessments made by an assessor blind to group allocation; ADL assessment open
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Five patients excluded as did not attend all training and evaluation sessions (3 from treatment group, 2 from controls) - reasons for non-attendance not provided
Selective reporting (reporting bias)	Unclear risk	Data on all measures reported, except for several scales which were deemed unsuitable due to ceiling effects
Other bias - training and supervision	Unclear risk	Two therapists - psychologist and physician - training in cognitive stimulation techniques not specified
Other bias - treatment manual	Unclear risk	Described as a 'cognitive stimulation programme' and some examples given, but no information regarding a manual or protocol

Buschert 2011

Study characteristics	
Methods	RCT
Participants	N = 39 24 amnestic MCI; 15 mild Alzheimer's disease (only data on Alzheimer's patients reported in this review) 8F/7M Mean MMSE 24.9 (SD 1.6; range 22-27) All on stable doses of AChEIs or memantine Age 75.9 (SD 8.1) Outpatients
Interventions	Multi-component cognitive group intervention - for AD group (N = 8) emphasis on cognitive stimulation (for MCI group more emphasis on cognitive training); Control group (N = 7) had pencil and paper exercises for self-study and monthly meetings.
Outcomes	Cognition: MMSE; ADAS-Cog, Trail Making Test, RBANS story memory & recall Quality of life: QoL-AD Mood: Montgomery Asberg Depression Rating Scale
Notes	2 hours, once a week for 6 months (20 sessions)
Risk of bias	

Buschert 2011 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Blocked randomisation procedure; participants pooled in pairs with respect to age, gender, education and ApoE genotype, then randomly assigned pairs to intervention or control using a computerised random number generator
Allocation concealment (selection bias)	Low risk	Blocked randomisation procedure - "a study-independent person then randomly assigned pairs to the intervention or control arm".
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors blind to group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Zero attrition in AD group
Selective reporting (reporting bias)	Low risk	Data on all measures reported
Other bias - training and supervision	Unclear risk	No information provided regarding training or supervision of the group leader, who remained constant throughout (first author)
Other bias - treatment manual	Low risk	"Manual with reproducible detailed protocols" developed before start of programme

Capotosto 2017

Study characteristics	
Methods	RCT
Participants	N = 39 (27F/12M) Mild/moderate dementia (Alzheimer's, vascular or mixed); MMSE > 13; Clinical Dementia Rating (CDR) 1 or 2 Not receiving AChEI medication Mean MMSE 18.2 (SD 3.4) Age 87.4 (SD 5.4) Two residential homes for older people
Interventions	Cognitive stimulation groups (N = 20) 'Active control' (Reading, discussions, creative activities) (N = 19)
Outcomes	Cognition: MMSE; ADAS-Cog; Digit Span backwards Communication: Narrative Language test Quality of Life: QoL-AD Mood: Cornell Scale for Depression in Dementia (CSDD); De Jong social and emotional loneliness scale Behaviour Problems: NPI

Capotosto 2017 *(Continued)*

	ADL/IADL: Disability Assessment for Dementia (DAD)
Notes	45 minutes, twice a week, for 7 weeks

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Very little detail given regarding randomisation except that random allocation took place
Allocation concealment (selection bias)	Unclear risk	No information given regarding who carried out randomisation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Stated as 'single-blind'. Not clear whether care home staff rating NPI and DAD would have been blind to treatment allocation
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	No information given regarding attrition after beginning of treatment; 5/44 had dropped out by that point (?pre-randomisation).
Selective reporting (reporting bias)	Low risk	All outcome measures mentioned in methods were reported.
Other bias - training and supervision	Low risk	"Two trained operators acted as facilitators".
Other bias - treatment manual	Low risk	Adapted for Italy from Spector 2006 manual

Carbone 2021

Study characteristics

Methods	RCT
Participants	N = 225 (149F/76M) 'Major neurocognitive disorder' DSM-V mild-to-moderate range (no information regarding medication) Mean MMSE 20.1 (SD 4.0) Age 83.6 (SD 8.1; range 50-99) 16 residential homes and day-centres
Interventions	Cognitive stimulation groups (N = 123) Usual treatment group sessions (N = 102)
Outcomes	Cognition: ADAS-Cog; MMSE Language: Narrative Language Test Quality of life: QoL-AD (self-report) Mood: Cornell Scale for Depression in Dementia

Carbone 2021 *(Continued)*

	ADL: Disability Assessment for Dementia (DAD)
	Behaviour problems: NPI
Notes	45 minutes, twice a week, for 7 weeks; 3-month follow-up

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Covariate adaptive randomization was used at each participating centre".
Allocation concealment (selection bias)	Unclear risk	No information provided as to who carried out the randomisation or the process used for concealment of allocation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	"Conducted by trained psychologists who did not participate in the treatment program, and they had no information on participants' group allocation".
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition at post-treatment; reasons given for those dropping out; data sought for all participants irrespective of compliance
Selective reporting (reporting bias)	Low risk	All outcomes reported
Other bias - training and supervision	Low risk	One-day training for at least one facilitator of each group, plus experience of dementia care and group facilitation skills
Other bias - treatment manual	Low risk	Italian version of 'Making a difference' manual used (Spector 2006)

Chapman 2004

Study characteristics

Methods	RCT
Participants	N = 54 (29F/25M)
	Probable AD, on stable dose of donepezil for at least 3 months
	Mean MMSE 20.87 (SD 3.55, range 12-28)
	Age 76.4 (SD 7.9; range 54-91)
	Living at home initially
Interventions	Cognitive stimulation + donepezil
	Donepezil only
Outcomes	Cognition: MMSE; ADAS-Cog
	ADL: Texas Functional Living Scale
	Behavioural problems: NPI - Irritability and Apathy

Chapman 2004 *(Continued)*

	Quality of Life: QoL-AD Global functioning: CBIC Verbalisation: Composite discourse score Carer distress - derived from the NPI 10-month follow-up data available
Notes	90 minutes, once a week, for 8 weeks

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Remote telephone randomisation, using a SAS procedure
Allocation concealment (selection bias)	Low risk	Independent randomisation procedure
Blinding of outcome assessment (detection bias) All outcomes	Low risk	All raters underwent extensive training; assessors blinded to group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis used. 24% attrition rate at end of study
Selective reporting (reporting bias)	Low risk	Data on all measures reported
Other bias - training and supervision	Low risk	Programme led by trained speech therapist, assisted by three Master's level speech language pathology students, who underwent a 2-hr training session before beginning treatment of each group and were provided with written reference materials. Weekly meetings were held in order to ensure the programme was implemented as designed.
Other bias - treatment manual	Low risk	No indication of a manual, but a clear structure to follow was evident.

Cheung 2019

Study characteristics

Methods	Cluster-RCT
Participants	N = 30 (22F/8M) Early-to-moderate dementia (any type) Global Deterioration Scale stages 4 to 6 Mean MoCA 7.9 (SD 4.4) Age 83.2 (SD 7.2) 73.3% had received either only a primary education or no education (73.3%).

Cheung 2019 *(Continued)*

	Two daycare centres
Interventions	Cognitive stimulating play intervention Control - social activities (reading newspapers, watching TV)
Outcomes	Cognitive: Montreal Cognitive Assessment (MoCA); Fuld Object Memory Evaluation; Modified Verbal Fluency Test
Notes	45-60 minutes, once a week, for 8 weeks

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Cluster-randomised, but only 2 clusters, so chance of comparable groups was less, and there was a significant baseline difference on the MoCA cognitive assessment.
Allocation concealment (selection bias)	Low risk	Randomisation carried out by independent researcher
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Data collection and entry undertaken by a blinded research assistant
Incomplete outcome data (attrition bias) All outcomes	Low risk	Some attrition due to sickness, 3/18 from cognitive stimulation and 1/12 from control, but used ITT analysis
Selective reporting (reporting bias)	Low risk	All outcomes reported
Other bias - training and supervision	Low risk	Training provided by co-ordinator
Other bias - treatment manual	Low risk	Clear structure for the approach and example session outline provided

Coen 2011

Study characteristics

Methods	RCT
Participants	N = 27 (14F/13M) Dementia - MMSE 10-23 MMSE: 16.9 (SD 5.0) Age: 79.8 (SD 5.6) Groups ran in 2 long-term care facilities and a private nursing home
Interventions	Cognitive stimulation (N = 14)

Coen 2011 *(Continued)*

	No treatment (N = 13)
Outcomes	Cognition: MMSE; ADAS-Cog Quality of life: QoL-AD Mood: Geriatric Depression Scale (14-item); RAID (Rating of Anxiety in Dementia) Behaviour: Behaviour Rating Scale (CAPE) Clinical Dementia Rating (sum of boxes)
Notes	45 minutes, 2 times a week for 7 weeks

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Stated that participants were randomly assigned. Author confirmed computerised randomisation and random number tables were used.
Allocation concealment (selection bias)	Unclear risk	Unclear who carried out randomisation and whether this was independent
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Tests administered by staff blind to group membership
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition reported
Selective reporting (reporting bias)	Low risk	Data for all measures reported
Other bias - training and supervision	Unclear risk	Sessions led by occupational therapists and activity coordinator - training/supervision unclear
Other bias - treatment manual	Low risk	Used Spector 2006 manual

Cove 2014

Study characteristics

Methods	RCT
Participants	N = 47 (22F/25M) DSM-IV criteria for dementia of any type (62% Alzheimer's or mixed) 62% receiving dementia medications Mean MMSE 22.8 (SD 3.4) Age 77.3 (SD 7.0) Living in the community

Cove 2014 *(Continued)*

Interventions	CST groups + carer training (N = 21) CST groups (N = 24) Waiting-list controls (N = 23)
Outcomes	Cognitive: MMSE, ADAS-Cog Quality of Life: QoL-AD Quality of Caregiver-Patient Relationship (QCPR)
Notes	Comparison of interest was CST groups versus controls; 45 minutes, once a week for 14 weeks.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Participants were randomized using the block method to achieve equal group sizes using Random Allocation Software".
Allocation concealment (selection bias)	Low risk	Although randomisation appeared to have preceded baseline assessment, it seemed to be independent of assessment.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors clearly blinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	3/24 withdrew from CST and 2/24 from controls, but ITT analysis used (LOCF). 1 person withdrew from controls before assessment, and was not included, but probably minimal effect.
Selective reporting (reporting bias)	Low risk	All outcomes reported
Other bias - training and supervision	Unclear risk	Groups run by clinicians - no mention of training or supervision
Other bias - treatment manual	Low risk	The study followed the standardised CST manual (Spector 2006).

Gibbor 2020b

Study characteristics

Methods	RCT
Participants	N = 33 (16F/17M) DSM-V criteria for dementia Mean MMSE 21.7 (SD 3.5; range 14-27) Age 81.9 (SD 10.3; range 56-98) Care homes
Interventions	Individual CST (delivered by researchers) (N = 17)

Gibbor 2020b *(Continued)*

	Treatment-as-usual (N = 16)
Outcomes	Cognition: SMMSE, ADAS-Cog Well-being and Quality of Life: QoL-AD (self-report), QoL-AD (proxy), Positive Psychology Outcome Measure (PPOM), Engagement and Independence in Dementia Questionnaire (EID-Q)
Notes	45 minutes, 2 times a week for 7 weeks

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Participants were randomly allocated by an independent web-based randomiser to receive either iCST or TAU within the care home, with a 1:1 ratio.
Allocation concealment (selection bias)	Low risk	Randomisation conducted independently
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Researchers conducting follow-up assessments were blinded to group allocation. Some risk for proxy QoL-AD
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	The treatment of missing data was unclear, with some imputation, some items omitted and some participants excluded for some measures.
Selective reporting (reporting bias)	Low risk	All outcome measures mentioned in methods were reported.
Other bias - training and supervision	Unclear risk	All sessions conducted by members of the research team, also described as 'professionals', but no specific mention of their training or supervision
Other bias - treatment manual	Low risk	Adapted from CST (Spector 2006) and iCST manuals (Yates 2014)

Graessel 2011

Study characteristics

Methods	RCT
Participants	N = 96 (80F/16M) (For 6-month evaluation, N = 139 (115F/24M)) ICD-10 criteria for primary degenerative dementia (vascular and secondary dementia excluded) 13.5% receiving anti-dementia medication Mean MMSE 14.6 (SD 5.4) Age 85.1 (SD 5.1) Nursing homes
Interventions	MAKS groups (motor stimulation, practice in activities of daily living, and cognitive stimulation) (N = 50; 6 months N = 71)

Graessel 2011 *(Continued)*

	Treatment-as-usual (N = 46; 6 months N = 68)
Outcomes	Cognition: ADAS-Cog* ADL: Erlangen Test of Activities of Daily Living (E-ADL)*; Barthel Index**; NOSGER ADL** and IADL subscales** Behaviour: Nurses' Observation Scale for Geriatric Patients (NOSGER)** Mood: NOSGER Mood subscale** Problem Behaviour: NOSGER Challenging behaviour subscale** Social Interaction: NOSGER Social Behaviour subscale** Care requirements: Resource Utilization in Dementia—Formal Care (RUD-FOCA)** (*10-month follow-up data provided; **Data after 6 months of intervention only)
Notes	2 hours, 6 times a week for 12 months

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated randomisation used
Allocation concealment (selection bias)	Unclear risk	Not clear who carried out the randomisation, or how independent they were. The randomisation appears to be carried out before baseline assessment, but evaluators were blinded, so this should not introduce bias.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Evaluators were independent and blinded to treatment allocation.
Incomplete outcome data (attrition bias) All outcomes	High risk	Per protocol analysis only has extractable data (although ITT analysis was also done). 38% lost for per protocol analysis
Selective reporting (reporting bias)	Low risk	All outcomes reported (time points varied)
Other bias - training and supervision	Low risk	Three days of training provided and compliance checked three times at each nursing home throughout study.
Other bias - treatment manual	Low risk	"Therapists and aides received a standardized handbook from the central study site describing in detail the steps to be taken on each day of therapy." Handbook developed specifically for the study

Juarez-Cedillo 2020

Study characteristics

Methods	iRCT
Participants	N = 67 (46F/21M)

Juarez-Cedillo 2020 *(Continued)*

	Diagnosis of mild dementia, using DSM-5 criteria (MMSE 19-24) Mean MMSE 22.6 (SD 0.9) Age 77.7 (SD 8.2) Outpatients
Interventions	'SADEM' cognitive stimulation groups (N = 39) Treatment-as-usual (N = 28)
Outcomes	Cognition: ADAS-Cog; MMSE; Syndrom-Kurztest (SKT); verbal fluency (semantic and phonological) Mood: CESD-R, Center for Epidemiologic Studies Depression Scale-Revised ADL: Rapid Disabilty Rating Scale (RDRS) Behaviour: Blessed Dementia Rating Scale Behaviour problems: NPI Caregiver outcomes (N.B. No data available for these): Zarit Burden Interview; Beck Depression Inventory; Beck Anxiety Inventory
Notes	90 minutes, 2 times a week, for 48 weeks; 12-month follow-up

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The exact randomisation method was unclear.
Allocation concealment (selection bias)	Low risk	Randomisation performed by independent researcher
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Evaluators blind to treatment allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	At end of intervention period attrition was low, and intention-to-treat analysis was used.
Selective reporting (reporting bias)	High risk	The results for caregiver outcomes do not appear to be reported. The results for people with dementia appear to include scales not mentioned in the Methods section.
Other bias - training and supervision	Low risk	Some indication training was provided.
Other bias - treatment manual	Low risk	Paper reported a detailed manual was used for the intervention.

Justo-Henriques 2022

Study characteristics

Justo-Henriques 2022 *(Continued)*

Methods	RCT
Participants	N = 59 (36F/23M) Formal diagnosis of a neurocognitive disorder according to DSM-5 criteria (51% Alzheimer's disease) No details available regarding AChEI medication Mean MMSE: 23.2 (SD 3.2) Age 78.9 (SD; 7.5; range 65-98) Community
Interventions	Home-based Individual cognitive stimulation delivered by clinical psychologist (N = 30) Treatment-as-usual (N = 29)
Outcomes	Cognition: MoCA, MMSE Quality of life: QoL-AD (self-report) Mood: GDS-15 ADL: adapted Lawton-Brody Index
Notes	45 minutes, once a week, for 47 weeks

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Non-stratified permuted block randomization process" carried out by blinded investigator using computer software
Allocation concealment (selection bias)	Low risk	"Allocation was unknown to participants and the therapist until the intervention started".
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Evaluator blinded to participant allocation
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Relatively high level of attrition (8/30 in intervention group), but lengthy treatment period (47 weeks) and reasons given. Similar attrition in control group
Selective reporting (reporting bias)	Low risk	Results for all outcomes cited in Methods section were reported.
Other bias - training and supervision	Low risk	Experienced therapist received additional training.
Other bias - treatment manual	Low risk	The study protocol provided a detailed account of the activities for each session.

Kim 2016

Study characteristics	
Methods	RCT
Participants	N = 53 (37F/16M) Probable Alzheimer's Disease (NINCDS and ADRDA criteria) All participants receiving 'pharmacotherapy' for dementia Mean MMSE 18.0 (SD 5.8) Age 78.5 (SD 1.5; range 61-94) Community residents attending dementia centre
Interventions	Multi-domain cognitive stimulation (N = 32) Control group (N = 21)
Outcomes	Cognition: CERAD (Korean version), including word fluency; short Boston naming test; tests of praxis, recall and recognition; MMSE Mood: Geriatric Depression Scale Quality of Life: QoL-AD (Korean version) (self-rated and proxy) Clinical Dementia Rating Scale
Notes	1 hour, 5 times a week for 6 months

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Description of randomisation unclear ('using random numbering')
Allocation concealment (selection bias)	Unclear risk	No mention of randomisation being carried out independently
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	No information about blinding
Incomplete outcome data (attrition bias) All outcomes	High risk	No attrition from intervention group; 34% attrition from control group
Selective reporting (reporting bias)	Low risk	All outcome measures mentioned in methods were reported.
Other bias - training and supervision	Low risk	"Implementation and management of the cognitive intervention program were carried out by two skilled and professionally educated occupational therapists."
Other bias - treatment manual	Low risk	Clear structure for the intervention

Leroi 2019

Study characteristics	
Methods	RCT
Participants	N = 61 Parkinson's disease dementia (PDD); dementia with Lewy bodies (DLB) according to standard clinical diagnostic criteria Median age (whole sample) 75 (range 55-90) Community
Interventions	Individual cognitive stimulation adapted for Parkinson's disease (delivered by informal carers) (CST-PD) (N = 31) Treatment-as-usual controls (N = 30)
Outcomes	Cognition: ACE-III; Dementia Cognitive Fluctuation Scale Quality of Life: EQ-5D; Parkinson's Disease Questionnaire-39 Quality of relationship: Relationship Satisfaction Scale Mood: Hospital Anxiety & Depression Scale; Lille Apathy Rating Scale; Brief Resilience Scale ADL: The Pill Questionnaire (specific ADL task) Interpersonal Reactivity Index Behaviour problems: NPI Caregiver outcomes: Hospital Anxiety & Depression Scales; EQ-5D; Short Form-12 Health Survey (SF-12); Relationship Satisfaction Scale; Dyadic Relationship Scale; Zarit Burden Interview; Relatives Stress Scale; Family Caregiving Role Scale; Brief Resilience Scale
Notes	30 minutes, 2-3 times a week for 12 weeks Study also included mild cognitive impairment in Parkinson's disease (PD-MCI); study authors have provided data for PDD and DLB separately for this review.

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"an independent arbiter, applied a single-stratum, blocked randomization to CST-PD or TAU at a 1:1 level by participant–dyad".
Allocation concealment (selection bias)	Low risk	Independent process: "A tamper proof process of single-strata, blocked randomisation will be applied and communicated via telephone and confirmatory email by an independent arbiter" (trial protocol).
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Clearly blinded:"'Procedures were in place to conceal the allocation from the independent, blinded outcome raters".
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	In the additional data provided by the study authors for e.g. the main cognition outcome (ACE-III), 10/28 were missing at post-treatment (cf. 3/28 for the

Leroi 2019 *(Continued)*

		control group). Analyses were for complete case only, so unclear what effect this might have
Selective reporting (reporting bias)	Low risk	Results for all outcomes were reported.
Other bias - training and supervision	Low risk	Care partners received training and support to deliver the intervention.
Other bias - treatment manual	Low risk	Manual developed (from iCST manual, Yates 2014) for the study

Lin 2018

Study characteristics	
Methods	Cluster-RCT
Participants	N = 105 (66F/39M) People with dementia with symptoms of agitation or depression Mean MMSE 14.9 (SD 3.7) Age 79.5 (SD 7.7) Long-term care institutions
Interventions	Cognitive stimulation (N = 30) Reminiscence therapy (N = 43) Treatment-as-usual control (N = 32)
Outcomes	Cognition: MMSE Quality of Life: QoL-AD Three-month follow-up
Notes	50 minutes, once a week for 10 weeks; cognitive stimulation versus control was relevant comparison for this review. Analysable data requested - not received

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Long-term care institutions randomly assigned into the different groups through a randomised block technique - but no information about block size and sequence
Allocation concealment (selection bias)	Unclear risk	Insufficient information regarding concealment
Blinding of outcome assessment (detection bias)	Unclear risk	No mention of blinding

Lin 2018 *(Continued)*

All outcomes		
Incomplete outcome data (attrition bias) All outcomes	High risk	Participants excluded if attended < 70% of sessions i.e. per protocol analysis - 13% of cognitive stimulation group excluded at post-test; 15% excluded at follow-up
Selective reporting (reporting bias)	Low risk	All outcome measures mentioned in Methods were reported.
Other bias - training and supervision	Low risk	Researcher attended a training course in CST.
Other bias - treatment manual	Low risk	Clear structure - followed CST manual, but fewer sessions and adapted content to make more culturally relevant

Lok 2020

Study characteristics	
Methods	RCT
Participants	N = 60 (30F/30M) Typical Alzheimer's disease in accordance with International Working Group-2 diagnostic criteria All receiving AChEI medication Mean MMSE 16.9 (SD 4.3) Age - no information provided Community residents
Interventions	CST groups based on Roy's adaptation model (RAM) (N = 30) Treatment-as-usual controls (N = 30)
Outcomes	Cognition: Standardised MMSE Quality of Life: QoL-AD Coping and Adaptation Processing Scale (CAPS)
Notes	45 minutes, twice a week for 7 weeks

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Electronic randomisation programme used
Allocation concealment (selection bias)	Unclear risk	No information about allocation concealment and independence of randomisation
Blinding of outcome assessment (detection bias)	High risk	"Pre-test and post-test data were collected and RAM-based CST was applied by the same person".

Lok 2020 *(Continued)*

All outcomes		
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition reported as zero and no cases were excluded from analysis
Selective reporting (reporting bias)	Low risk	All outcome measures mentioned in methods were reported.
Other bias - training and supervision	Unclear risk	Intervention delivered by first author of the paper but little information about receiving training in CST
Other bias - treatment manual	Low risk	Based on the manualised CST programme (Spector 2006)

Lopez 2020

Study characteristics	
Methods	RCT
Participants	N = 20 (15F/5M) Alzheimer's dementia according to DSM-5 criteria All receiving AChEI medication Mean MMSE 17.9 (SD 3.9) Age 81.9 (SD 5.5; range 66-89) Community - daycare centre
Interventions	Cognitive stimulation groups (N = 10) Treatment-as-usual controls (N = 10)
Outcomes	Cognition: ADAS-Cog; MMSE; WAIS-III, Wechsler Memory Scale III and Neuropsychological test battery (32 tests in total), including verbal fluency, digit span, Boston Naming Test
Notes	60 minutes, three times a week for 6 months

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Sequence of block randomisation was described.
Allocation concealment (selection bias)	Unclear risk	No information about blinding at point of randomisation
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	No information about blinding of outcome assessments
Incomplete outcome data (attrition bias)	Unclear risk	There was no indication of attrition, despite being a 6-month intervention study.

Lopez 2020 *(Continued)*

All outcomes

Selective reporting (reporting bias)	Low risk	All outcome measures mentioned in methods were reported.
Other bias - training and supervision	Unclear risk	Little to no information about who delivered the intervention
Other bias - treatment manual	Low risk	Intervention was extracted from a manualised, Spanish cognitive psychostimulation therapy (cuadernos de repaso).

Maci 2012

Study characteristics	
Methods	RCT
Participants	N = 14 (8F/6M) Alzheimer's Disease, according to National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria All receiving a stable dose regimen of memantine and/or cholinesterase inhibitors and/or antidepressants for at least 6 months prior to the beginning of the study Mean MMSE 17.8 (SD 2.8) Age 72.6 (SD 9.5) Community (gymnasium)
Interventions	Cognitive stimulation/physical activity groups (N = 7) Treatment-as-usual (N = 7)
Outcomes	Cognition: MMSE; Frontal Assessment Battery Clinical Dementia Rating ADL: Katz ADL scale; Lawton & Brody Instrumental Activities of Daily Living scale Mood: Cornell Scale for Depression in Dementia; Hamilton Anxiety Rating Scale Quality of Life: Cornell-Brown Scale for Quality of life in Dementia; Apathy Evaluation Scale; Quality of Life–Alzheimer's Disease (QoL-AD; self-report and caregiver proxy) Family caregiver outcomes: Caregiver Burden Inventory; Beck Depression Inventory
Notes	120 minutes, five times a week for 3 months

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Randomly assigned to the active treatment group or to the usual care as control group according to a list of randomly generated sequence of numbers."
Allocation concealment (selection bias)	Unclear risk	No indication of who carried out randomisation or their independence

Maci 2012 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	Examiners blinded
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	No mention of attrition - none reported over 3-month period?
Selective reporting (reporting bias)	Low risk	All outcomes appear to be reported.
Other bias - training and supervision	Low risk	Training and supervision provided
Other bias - treatment manual	Low risk	There was a clear structure, though no indication of a manual as such.

Mapelli 2013

Study characteristics	
Methods	RCT
Participants	N = 30 (gender not specified) Alzheimer's disease (N = 16); vascular dementia (N = 13); mixed dementia (N = 1) according to DSM-IV-TR criteria Mean MMSE 19.5 (SD 3.4) Age 83.7 (SD 4.6) Nursing home
Interventions	Cognitive stimulation groups (N = 10) Occupational therapy 'Placebo' groups (N = 10) Treatment-as-usual control (N = 10)
Outcomes	Cognition: MMSE; Esame Neuropsicologico Breve 2 (ENB2) (comprises 14 subtests including digit span, verbal fluency and Trail Making Test); Clinical Dementia Rating Mood: Geriatric Depression Scale (GDS-30) ADL: ADL scale Behaviour problems: Behavioral Pathology in Alzheimer's Disease Rating Scale (Behave-AD scale)
Notes	60 minutes, five times a week for 8 weeks. Relevant comparison for this review was cognitive stimulation versus treatment-as-usual.

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**

Mapelli 2013 *(Continued)*

Random sequence generation (selection bias)	Unclear risk	Allocation appears to be simple randomisation using 'a simple computerized randomization technique', but no further details given
Allocation concealment (selection bias)	Unclear risk	No mention of who carried out randomisation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinded rater
Incomplete outcome data (attrition bias) All outcomes	Low risk	Full information provided - no participants lost to follow-up, no participants excluded from analyses
Selective reporting (reporting bias)	High risk	No information was reported on some outcome measures despite being included in the methods (e.g. ADL, GDS-30).
Other bias - training and supervision	Unclear risk	No mention of training
Other bias - treatment manual	Unclear risk	Although structure was outlined, exact content of exercises was unclear - did not appear to be a treatment manual as such

Marinho 2021

Study characteristics	
Methods	RCT
Participants	N = 47 (29F/18M) Clinical diagnosis of dementia according to DSM-IV criteria (MMSE 10-24) All participants receiving AChEI medication Clinical Dementia Rating: 23 mild; 24 moderate Age 77.8 (SD; 8.4; range 60-91) Outpatients
Interventions	Cognitive stimulation groups (N = 23) Treatment-as-usual (N = 24)
Outcomes	Cognition: ADAS-Cog Quality of life: QoL-AD (self-report and proxy) Mood: Cornell Scale for Depression in Dementia ADL: ADCS-ADL scale Caregiver outcomes: Zarit Burden Interview
Notes	45 minutes, 2 times a week, for 7 weeks (both weekly sessions on same day, separated by short break)
Risk of bias	

Marinho 2021 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Used a random list generated by a computer program; stratified by dementia severity
Allocation concealment (selection bias)	Unclear risk	Not detailed who carried out the randomisation, and whether external to the study
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Outcome assessments carried out by researchers blind to the intervention
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data provided for intention-to-treat analysis; reasons for dropouts described
Selective reporting (reporting bias)	Low risk	Results for all outcomes reported
Other bias - training and supervision	Low risk	Facilitators trained by International CST centre
Other bias - treatment manual	Low risk	Used Brazilian CST manual, adapted from Spector 2006

Middelstädt 2016

Study characteristics	
Methods	RCT
Participants	N = 71 (60F/11M) Mild-to-moderate dementia according to ICD-10 criteria Mean MMSE 16.9 (SD 4.5; range 10-24) Age 86.4 (SD 4.5; range 74-102) Nursing homes
Interventions	Cognitive stimulation groups 'NEUROvitalis senseful' (N = 36) Treatment-as-usual control group (N = 35)
Outcomes	Cognition: ADAS-Cog Quality of Life: QoL-AD (self-report and proxy) ADL: ADCS-ADL Behaviour problems: NPI-NH Six-week follow-up data provided
Notes	60 minutes, twice a week for 8 weeks

Middelstädt 2016 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Used computer generated random numbers
Allocation concealment (selection bias)	Unclear risk	"The randomization process was then realized by a member of the research group who was blinded to the participants". Not clear how independent this process was?
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors for ADAS-Cog and QoL-AD blind to group assignment and not involved in any other part of the study. Group secrecy emphasised
Incomplete outcome data (attrition bias) All outcomes	Low risk	Details of attrition provided and followed ITT principle (also per protocol analysis)
Selective reporting (reporting bias)	Low risk	All outcome measures mentioned in methods were reported.
Other bias - training and supervision	Unclear risk	Facilitator was researcher experienced in conducting non-pharmacological interventions, but no details of training or supervision
Other bias - treatment manual	Low risk	"Every session was described in detail in a manual, including advice for instructions."

Onder 2005

Study characteristics	
Methods	RCT
Participants	N = 156 (113F/43M) Probable Alzheimer's Disease, on donepezil for at least 3 months MMSE 20.1 (SD 3.1) Age 75.8 (SD 7.1) Living at home
Interventions	Individual RO (delivered by family caregiver) + donepezil (N = 79) Donepezil only (N = 77)
Outcomes	Cognition: MMSE; ADAS-Cog ADL: Barthel; Lawton & Brody IADL scale Behaviour problems: NPI Family caregiver outcomes: Hamilton anxiety and depression scales; Caregiver Burden Inventory; SF-36
Notes	30 minutes, 3 times a week, for 25 weeks; plus informal contacts 2 or 3 times a day

Onder 2005 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computerised block randomisation process
Allocation concealment (selection bias)	Unclear risk	No information regarding the independence of the randomisation process
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessments made by blind assessors
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition data reported with reasons: 9 from RO group, 10 from control group i.e. 12%. Intention-to-treat analysis
Selective reporting (reporting bias)	Low risk	Data for all outcome measures reported
Other bias - training and supervision	Low risk	Family caregivers trained by a multidisciplinary team including physicians, psychologists and therapists. Training included a simulated therapy session.
Other bias - treatment manual	Low risk	The family caregivers were also provided with a manual of instruction on the therapy and specific schedules for each session.

Orgeta 2015

Study characteristics	
Methods	RCT
Participants	N = 356 (165 F/191 M) Mild-to-moderate dementia according to DSM-IV criteria 76% receiving AChEI medication Mean MMSE 21.2 (SD 4.3) Age 78.2 (SD 7.5) Community
Interventions	Individual cognitive stimulation (delivered by informal caregivers) (N = 180) Treatment-as-usual controls (N = 176)
Outcomes	Cognition: ADAS-Cog; MMSE Quality of Life: QoL-AD (self-report and proxy); DEMQOL (self-report and proxy) Quality of relationship: Quality of the Carer-Patient Relationship Scale (QCPR) ADL: Bristol ADL Scale Mood: GDS-15

Orgeta 2015 *(Continued)*

	Behaviour problems: NPI Caregiver outcomes: Short Form-12 Health Survey (physical (PCS) and mental (MCS) components); Hospital Anxiety & Depression Scale (HADS); EQ-5D-3L; Resilience Scale-14; NPI carer distress; QCPR
Notes	30 minutes, 3 times a week, for 25 weeks Imputed adjusted means used in meta-analysis

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomisation was completed using a dynamic adapative allocation method, with an overall allocation ratio of 1: 1. Random allocation was stratified by site and receipt of acetylcholinesterase inhibitors (AChEIs)."
Allocation concealment (selection bias)	Low risk	Randomisation database held at trials unit; only unblinded researchers informed of allocation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors blinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Full details given on attrition. ITT analyses used
Selective reporting (reporting bias)	Low risk	Results given for all planned outcomes
Other bias - training and supervision	Low risk	Carers all received training in delivering iCST and support throughout study.
Other bias - treatment manual	Low risk	Carers all received standardised training package and manual (Yates 2014).

Orrell 2014

Study characteristics

Methods	RCT
Participants	N = 236 (150 F/86 M) Dementia according to DSM-IV criteria (31% Alzheimer's disease) 32% receiving AChEI medication Mean MMSE 17.8 (SD 5.5) Age 83.1 (SD; 7.6) Nine care homes and nine community settings
Interventions	Maintenance cognitive stimulation groups (following 7 weeks of twice-weekly cognitive stimulation groups) (N = 123)

Orrell 2014 *(Continued)*

	Treatment-as-usual controls (following 7 weeks of twice-weekly cognitive stimulation groups) (N = 113)
Outcomes	Cognition: ADAS-Cog; MMSE Quality of life: QoL-AD (self-report and proxy); DEMQOL Mood: Cornell Scale for Depression in Dementia; RAID ADL: ADCS-ADL scale Behaviour problems: NPI Caregiver outcomes: SF-12
Notes	45 minutes, once a week, for 24 weeks No data were imputed for those cases in which all assessments were missing (see page 30, SHIELD report). Adjusted means used for meta-analysis

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The random allocation sequence was computer-generated and in the ratio of 1:1." Carried out by clinical trials unit
Allocation concealment (selection bias)	Low risk	Only intervention researcher informed of allocation by clinical trials unit
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinded researchers completed interviews and questionnaires. Proxy measures completed by carers/care staff not blinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrittion rates and reasons given, analysis by treatment allocated
Selective reporting (reporting bias)	Low risk	Results given for all planned outcomes
Other bias - training and supervision	Low risk	"All facilitators had at least 1 year of experience in dementia care, and had attended the 1-day CST training course."
Other bias - treatment manual	Low risk	Published treatment manual

Paddick 2017

Study characteristics

Methods	Stepped wedge design
Participants	N = 34 (29F/5M) Dementia according to DSM-IV criteria None receiving AChEI medication

Paddick 2017 *(Continued)*

	Mean Clinical Dementia Rating 1.65 (range 1-2) Median age 80 (IQR 76.5-85.3) Community
Interventions	Cognitive stimulation groups (N = 16) Wait-list controls (N = 18)
Outcomes	Cognition: ADAS-Cog Quality of life: World Health Organization Quality of Life assessment (WHOQOL) Mood: Hospital Anxiety & Depression Scales ADL: World Health Organization Disability Assessment Schedule (WHODAS 2.0) Behaviour problems: NPI Caregiver outcomes: WHOQOL; Zarit Burden Inventory; Hospital Anxiety & Depression Scales; NPI Caregiver distress
Notes	45 minutes, twice a week, for 7 weeks Relevant comparison was at assessment 8 weeks after baseline, where those who received cognition stimulation could be compared with the wait-list controls, who had received treatment-as-usual. Only ADAS-Cog data were available for this comparison.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Cluster-randomised. Simple randomisation with random number generator
Allocation concealment (selection bias)	Low risk	Randomisation carried out by independent statistician, blinded to participant allocation
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Outcome assessors were blind to group allocation and did not deliver the CST sessions; participants reminded not to disclose
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat protocol - no attrition in relation to comparison of immediate and delayed start groups
Selective reporting (reporting bias)	High risk	Although all outcome measures were reported for the comparison of interest (between immediate and delayed start CST groups), only data on ADAS-Cog provided
Other bias - training and supervision	Low risk	Facilitators received training on CST in the UK and had been involved in the adaptation and pilot of adapted CST for Sub-Saharan Africa.
Other bias - treatment manual	Low risk	Published manual - adapted from CST manual

Rai 2021

Study characteristics	
Methods	RCT
Participants	N = 61 (19F/42M) Dementia of any type according to DSM-IV criteria 71% receiving AChEI medication Age 73.0 (SD 7.7; range: 50–89) Community
Interventions	Individual cognitive stimulation therapy (iCST) app, delivered by family caregiver (N = 31) Treatment-as-usual controls (N = 30)
Outcomes	Cognition: ADAS-Cog Quality of Life: QoL-AD (self-report and proxy); EQ-5D Quality of relationship: Quality of the Carer-Patient Relationship Scale (QCPR) ADL: Bristol ADL Scale Mood: Cornell Scale for Depression in Dementia (CSDD) (self- and proxy-rated) Behaviour problems: NPI Caregiver outcomes: Hospital Anxiety & Depression Scale (HADS); EQ-5D; QCPR
Notes	Recommended 30 minutes, 2-3 times a week, for 11 weeks

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	On-line central randomisation service used for block randomisation
Allocation concealment (selection bias)	Low risk	Steps taken to ensure allocation concealment appropriate
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinded researcher carried out all post-baseline outcome assessments.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analyses and reasons for attrition explained
Selective reporting (reporting bias)	Low risk	All outcomes cited in Methods section were reported.
Other bias - training and supervision	Low risk	Training and support provided for carers implementing the intervention
Other bias - treatment manual	Low risk	App based closely on conventional iCST manual (Yates 2014)

Requena 2006

Study characteristics	
Methods	RCT
Participants	N = 86 (61F/25M) Alzheimer-type dementia (severe dementia excluded) MMSE 21.9 (6.3) Age 77.0 (SD 7.5) Attending daycare centre
Interventions	1) Cognitive stimulation + donepezil (N = 20) 2) Donepezil only (N = 30) 3) Cognitive stimulation only 4) No treatment
Outcomes	Cognition: MMSE, ADAS-Cog Mood: GDS-30 12-month and 24-month data reported
Notes	45 minutes, 5 times a week for 24 months 'No treatment' group were not part of the randomisation process. Comparison of interest to this review was cognitive stimulation + donepezil versus donepezil alone.

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Randomisation by registration order: "subjects were randomly distributed in groups at the time they arrived at the Centre".
Allocation concealment (selection bias)	Unclear risk	Not clear if randomisation was independent
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Spanish paper stated "Evaluator was blind to treatment allocation".
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition reported: 6/20 in CS + donepezil group; 10/30 in donepezil alone group i.e. 32% over 2-year period
Selective reporting (reporting bias)	Low risk	Data for all measures reported
Other bias - training and supervision	Unclear risk	Spanish paper stated that groups were led by an independent member of the research team - training and supervision unclear

Requena 2006 *(Continued)*

Other bias - treatment manual	Low risk	Clear structure for the cognitive stimulation approach, seven areas of stimulation; two levels of difficulty in relation to questions asked concerning images on TV screen

Spector 2001

Study characteristics		
Methods	RCT	
Participants	N = 35 Diagnosis of dementia according to DSM-IV criteria MMSE 13.1 (SD 4.4) Age 85.7 (SD 6.7) Living at home: 12; living in residential home: 23	
Interventions	Cognitive stimulation (N = 21) Treatment-as-usual (N = 14)	
Outcomes	Cognition: MMSE; ADAS-Cog Communication: Holden Communication Scale Mood: Cornell Scale for Depression in Dementia; RAID Behaviour: Behaviour Rating Scale (CAPE). Family caregivers: Relatives Stress Scale; GHQ	
Notes	45 minutes, 2 times a week, for 7 weeks	
Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Randomly allocated to either group by drawing names from a sealed container
Allocation concealment (selection bias)	Unclear risk	Randomly allocated to either group by drawing names from a sealed container - would have been preferable for randomisation to have been carried out independently
Blinding of outcome assessment (detection bias) All outcomes	High risk	At least some of the assessments were carried out by staff aware of group allocation.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition reported (with reasons): 4 in CS group, 4 in control group i.e. 23%
Selective reporting (reporting bias)	Low risk	Data from all measures reported

Spector 2001 *(Continued)*

Other bias - training and supervision	Unclear risk	Groups led by a member of the research team, with a member of staff as co-facilitator - training and supervision unclear
Other bias - treatment manual	Low risk	Manual (Spector 2006) developed through this study

Spector 2003

Study characteristics	
Methods	RCT
Participants	N = 201 (158F/43M) Dementia (DSM-IV criteria) - MMSE 10-24 MMSE: 14.4 (SD 3.8) Age: 85.3 (SD 7.0) Groups ran in 18 residential homes; 5 day-centres
Interventions	Cognitive stimulation (N = 115) Treatment-as-usual (N = 86)
Outcomes	Cognition: MMSE; ADAS-Cog Quality of life: QoL-AD Communication: Holden Communication Scale Mood: Cornell Scale for Depression in Dementia Behaviour: Behaviour Rating Scale (CAPE)
Notes	45 minutes, 2 times a week, for 7 weeks

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Randomly allocated to either group by drawing names from a sealed container
Allocation concealment (selection bias)	Unclear risk	Randomly allocated to either group by drawing names from a sealed container - would have been preferable for randomisation to have been carried out independently
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Cognitive assessments and quality of life interview conducted by a blind assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	34/201 did not complete study (18 CS/16 controls); 17% attrition; 'intention to treat analysis'

Spector 2003 *(Continued)*

Selective reporting (reporting bias)	Low risk	Data from all measures reported
Other bias - training and supervision	Unclear risk	Groups led by a member of the research team with a member of staff as co-facilitator; training and supervision unclear
Other bias - treatment manual	Low risk	Manual developed for this study (Spector 2006)

Tanaka 2021

Study characteristics	
Methods	RCT
Participants	N = 31 (18F/13M) 'Dementia' - MMSE 5-25 4/31 receiving donepezil MMSE: 15.5 (SD 5.8) Age: 86.2 (SD 7.8) Geriatric health service facility (inpatient)
Interventions	Group exercise and cognitive stimulation (N = 16) Treatment-as-usual (N = 15)
Outcomes	Cognition: MMSE Quality of life: QoL-D (short version, proxy completed) Social interaction: nurses' observation scale of geriatric patients (NOSGER) sub-item 'social behaviour' Apathy scale ADL: Barthel Index Behaviour problems: Dementia Behavior Disturbance Scale
Notes	45 minutes, 2 times a week, for 8 weeks

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	No detail of how randomisation was carried out
Allocation concealment (selection bias)	Unclear risk	No information as to who carried out the randomisation or whether an external person was involved
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Although observational ratings made by staff not involved in the intervention, not clear whether assessors were blinded or not

Tanaka 2021 *(Continued)*

Incomplete outcome data (attrition bias) All outcomes	Unclear risk	33% of participants in control group dropped out (compared with 6% in intervention group) - unclear what effect this may have had on results
Selective reporting (reporting bias)	Low risk	Results reported for all outcomes mentioned in Methods section
Other bias - training and supervision	Unclear risk	Although staff had at least 5 years experience of group interventions in this context, arrangements for specific training/supervision in the intervention used here not described
Other bias - treatment manual	Unclear risk	Although the Japanese version of the CST manual was referenced, the current study appears to include additional components.

Tsantali 2017

Study characteristics	
Methods	RCT
Participants	N = 55 Mild Alzheimer's disease according to NINCDS-ADRDA and DSM-IV-TR criteria All receiving AChEI medication for at least two years Mean MMSE 23.0 (SD 1.3) Age 73.7 (SD 5.3) Community
Interventions	Individual cognitive stimulation delivered by psychologists (N = 17) Individual cognitive training delivered by psychologists (N = 17) Treatment-as-usual controls (N = 21)
Outcomes	Cognition: MMSE; CAM-COG; Boston Naming Test; Pyramids and Palm Trees Test; Rivermead Behavioural Memory Test
Notes	90 minutes, 3 times a week, for 4 months Data for post-treatment assessment requested - not received; only 8-month follow-up data available

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"Participants were...randomly allocated by lot into the three conditions".
Allocation concealment (selection bias)	Unclear risk	No detail provided on the randomisation process
Blinding of outcome assessment (detection bias)	Low risk	Clinicians conducting assessments were blinded.

Tsantali 2017 *(Continued)*

All outcomes		
Incomplete outcome data (attrition bias) All outcomes	High risk	Per protocol analysis - 4/21 dropped out from cognitive stimulation group; 0/21 from control group
Selective reporting (reporting bias)	Unclear risk	Very little reported of results at post-treatment - nearly all data seems to be from 12-month follow-up.
Other bias - training and supervision	Low risk	"Four licensed psychologists, who had sufficient clinical experience in providing cognitive remediation programmes to an aging population administered the programmes." Activities according to specialisation
Other bias - treatment manual	High risk	No indication of a manual or a particular structure for the cognitive stimulation

Young 2019

Study characteristics	
Methods	RCT
Participants	N = 101 (81F/20M) Mild dementia according to DSM-V criteria Mean MMSE 20.7 (SD 2.3) Age 80.2 (SD 6.4) Community
Interventions	Cognitive stimulation plus Tai Chi groups (N = 51) Treatment-as-usual controls (N = 50)
Outcomes	Cognition: MMSE; Mattis Dementia Rating Scale
Notes	60 minutes, twice a week, for 7 weeks

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"The randomization and group assignment process was conducted using computer software."
Allocation concealment (selection bias)	Unclear risk	Insufficient information about concealment
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	"Intention- to-treat principle using last observation carried forward (LOCF) analysis for any missing data." Attrition rates given

Young 2019 *(Continued)*

Selective reporting (reporting bias)	Low risk	All outcome measures mentioned in methods were reported.
Other bias - training and supervision	Low risk	Training and supervision given and checks on adherence
Other bias - treatment manual	Low risk	Standardised manual produced and structure described.

ACE-III: Addenbrooke's Cognitive Examination -III
AChEI: Acetylcholinesterase inhibitor
AD: Alzheimer's disease
ADAS-COG: Alzheimer's Disease Assessment Scale - Cognitive Subscale
ADCS-ADL: Alzheimer's Dementia Cooperative Study - Activities of Daily Living Inventory
ADL: Activities of Daily Living
ADRDA: Alzheimer's Disease and Related Disorders Association
ApoE: apolipoprotein E
Behave-AD: Behavioral Pathology in Alzheimer's Disease Rating Scale
CAM-COG(-DS): Cambridge Cognitive Examination adapted for individuals with Down Syndrome
CAPE: Clifton Assessment Procedures for the Elderly
CBIC: Clinician's Interview-Based Impression of Change
CDR: Clinical Dementia Rating
CERAD: Consortium to Establish a Registry for Alzheimer's Disease test
CESD-R: Center for Epidemiologic Studies Depression Scale-Revised
CS: cognitive stimulation
CSDD: Cornell Scale for Depression in Dementia
CSDS: Cognitive Scale for Down Syndrome
CST: cognitive stimulation therapy
CST-PD: cognitive stimulation therapy – Parkinson's disease
DAD: Disability Assessment for Dementia
DEMQOL: Dementia Quality of Life Instrument
DLB: dementia with Lewy bodies
DSM(-III/-IV(-TR)/-5/-V): Diagnostic and Statistical Manual of Mental Disorders (-III/-IV(-TR)/-5/-V)
E-ADL: Erlangen Test of Activities of Daily Living
ECA: 'Echelle comportementale adaptive' (behaviour adaption scale)
EID-Q: Engagement and Independence in Dementia Questionnaire
ENB2: Esame Neuropsicologico Breve 2
EQ-5D(-3L): EuroQol 5 dimension scale (-3 level)
GDS(-15/-30): Geriatric Depression Scale (-15/-30 item)
GHQ: General Health Questionnaire
HADS: Hospital Anxiety and Depression Scale
IADL: Instrumental Activities of Daily Living
ICD-10: International Classification of Diseases -10
IQR: interquartile range
ITT: intention to treat
LOCF: last observation carried forward
MAKS: motor stimulation; activities of daily living; cognitive stimulation; spiritual element
MCI: mild cognitive impairment
MCS: mental component score
MoCA: Montreal Cognitive Assessment
MMSE: Mini-Mental State Examination
NINCDS: National Institute of Neurological and Communicative Diseases and Stroke
NOSGER: Nurses' Observation Scale for Geriatric Patients
NPI(-NH): Neuropsychiatric Inventory (-Nursing Home)
PCS: physical component score
PDD: Parkinson's disease dementia
PPOM: Positive Psychology Outcome Measure
QCPR: Quality of Caregiver-Patient Relationship
QOL-AD): Quality of Life - Alzheimer's disease

QOL-D: Quality of Life in dementia
RAID: Rating of Anxiety in Dementia
RAM: Roy's adaptation model
RBANS: Repeatable Battery for the Assessment of Neuropsychological Status
RCT: randomised controlled trial
RDRS: Rapid Disability Rating Scale
RO: reality orientation
ROT: reality orientation therapy
RUD-FOCA: Resource Utilization in Dementia—Formal Care
SADEM: study on ageing and dementia in Mexico
SAS: Statistical Analysis System
SCIDS: Sense of Competence In Dementia Care Staff Scale
SDAT: senile dementia of the Alzheimer's type
SF(-12/-36): Short Form-12/36 Health Survey
SKT: Syndrom-Kurztest
SMMSE: Standardised Mini-Mental State Examination
TAU: treatment as usual
WAIS-III: Wechsler Adult Intelligence Scale - III
WHODAS: World Health Organization Disability Assessment Schedule
WHOQOL: World Health Organization Quality of Life assessment

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Alves 2014	No separate data for people with dementia (MCI)
Apostolo 2013	No data for people with dementia
Apostolo 2014	No data for people with dementia
Arcoverde 2008	Intervention did not meet the inclusion criteria for cognitive stimulation. Diagnoses varied, not purely dementia
Baglio 2015	Multi-component intervention
Baines 1987	Unclear diagnostic criteria: "moderate-to-severe impairment of cognitive function"
Basak 2008	Intervention described did not meet the inclusion criteria for cognitive stimulation; better fit for cognitive training
Brook 1975	Did not not include a measure of cognitive function
Buettner 2011	Included both people with dementia and with MCI - results not given separately
Calatayud 2022	No data provided for people with dementia (MCI)
Camargo 2015	Intervention allocation not randomised
Camargo 2019	Intervention allocation not randomised
Carlson 2008	Intervention described did not meet the inclusion criteria for cognitive stimulation but for cognitive training. Diagnoses varied, not purely dementia
Cassinello 2008	Did not report an RCT
Cheng 2006	Intervention described did not meet the inclusion criteria for cognitive stimulation

Study	Reason for exclusion
Constantinidou 2008	Intervention described did not meet the inclusion criteria for cognitive stimulation
Croisile 2006	Did not report results from an RCT
Davis 2001	Cognitive stimulation (delivered for 30 minutes a day, 6 days a week by family caregivers) confounded with cognitive training-spaced retrieval and face name associations
Devita 2021	Intervention allocation not randomised
Eckroth-Bucher 2009	Intervention did not meet the inclusion criteria for cognitive stimulation. Diagnoses varied, not purely dementia
Eggermont 2009a	Intervention did not meet the inclusion criteria for cognitive stimulation
Eggermont 2009b	Intervention did not meet the inclusion criteria for cognitive stimulation
Evans 2009	Intervention did not meet the inclusion criteria for cognitive stimulation
Faggian 2007	Intervention did not meet the inclusion criteria for cognitive stimulation
Fanto 2002	No mention of randomisation; available only as a conference abstract
Farina 2006a	Non-randomised allocation; comparison was with an active treatment group.
Farina 2006b	Non-randomised allocation; comparison was with an active treatment group.
Fernandez-Calvo 2010	In Spanish - only the abstract was in English
Ferrario 1991	Unclear diagnostic criteria: "Elderly patients with cognitive disturbances"
Folkerts 2018	No data for people with a dementia diagnosis
Gerber 1991	Eligible study, but no extractable data provided. Only data available were for a composite cognitive and behavioural scale.
Goldstein 1982	Around 25% of participants appeared not to have dementia; other diagnoses included schizophrenia, epilepsy and ruptured aneurysm.
Gonzalez-Abraldes 2010	Intervention did not meet the inclusion criteria for cognitive stimulation
Green 2009	Intervention did not meet the inclusion criteria for cognitive stimulation
Greenaway 2008	Intervention did not meet the inclusion criteria for cognitive stimulation
Han 2017	No separate data for people with dementia (MCI)
Hanley 1981	Eligible study, but no extractable data available
Hill 2014	Study sample was people with dementia AND superimposed delirium.
Holden 1978	Diagnoses varied, not purely dementia. Not clear that participants were randomised to the intervention and control groups
Johnson 1981	Allocation of patients to treatment was not random for various practical reasons.

Study	Reason for exclusion
Justo-Henriques 2019	No data for people with dementia (MCI)
Kim 2015	No data for people with dementia (MCI)
Kim 2020	Intervention did not not meet definition of cognitive stimulation.
Kolanowski 2016	Delirium superimposed on dementia
Liu 2021	Allocation to intervention and control not random
Luttenberger 2019	Data not available for people with dementia - included MCI
Matsuda 2007	Non-randomised study
McCormick 2019	No cognitive outcome measure
Menna 2016	Intervention allocation not randomised
Meza-Kubo 2009	Not a randomised control trial
Milev 2008	Intervention did not meet the inclusion criteria for cognitive stimulation - related to 'snoezelen'.
Mudge 2008	Intervention did not meet the inclusion criteria for cognitive stimulation.
Muniz 2015	No data for people with dementia - combined with MCI
Newson 2006	Intervention did not meet the inclusion criteria for cognitive stimulation. Diagnoses varied, not dementia
Niu 2010	Intervention met criteria for cognitive training rather than cognitive stimulation.
Okamura 2018	Not cognitive stimulation - described cognitive training
Olazaran 2004	12/84 participants with diagnoses of MCI; results not presented separately for those with Alzheimer's disease. Interventions included additional elements as physical exercise.
Oliveira 2021	Did not meet definition of cognitive stimulation
Onieva-Zafra 2018	Intervention allocation not randomised
Orrell 2005	Allocation to intervention and control groups not random for maintenance study
Perkins 2022	Intervention allocation not randomised
Piras 2017	Comparison was with alternate treatment.
Quayhagen 1995	Although intervention was described as cognitive stimulation, it appeared to focus on specific cognitive modalities, and so fits better with cognitive training definition.
Quayhagen 2000	Although intervention was described as cognitive stimulation, it appeared to focus on specific cognitive modalities, and so fits better with cognitive training definition.
Raggi 2007	Not an RCT
Reeve 1985	No indication of random allocation to groups

Study	Reason for exclusion
Riegler 1980	Comparison of RO plus music versus RO. No control groups without RO
Rueda 2021	Comparison was with alternative treatment.
Ruiz Sanchez de Leon 2007	Not an RCT and intervention did not meet the criteria for inclusion under cognitive stimulation.
Schecker 2013	Intervention did not meet inclusion criteria for cognitive stimulation.
Schmitter-Edgecombe 2008	Intervention did not meet the inclusion criteria for cognitive stimulation.
Scott 2003	Did not report on a study intervention. No RCT
Silva 2017	Not cognitive stimulation
Silva 2021	No data for people with dementia (MCI)
Skov 2022	No control group
Smith 2009	Intervention did not meet the inclusion criteria for cognitive stimulation.
Tadaka 2004	The intervention combined elements of Reality Orientation (RO) and reminiscence. The RO element appeared to be only an orientation board, used to reinforce orientation for time, place and person. The reminiscence element appears to be predominant, with a variety of reminiscence-based triggers, and so the study would be a better fit for a review of reminiscence work with people with dementia.
Tanaka 2017	Multi-component intervention
Tarraga 2006	Allocation to groups was not entirely random. For the comparison of interest, integrated psychostimulation programme versus medication only control, allocation was clearly non-random.
Thickpenny-Davis 2007	Intervention did not meet the inclusion criteria for cognitive stimulation; participants included with other diagnosis than dementia
Tsai 2008	Not RCT
Tsai 2019	Intervention allocation not randomised
Van Zon 2016	Intervention allocation not randomised
Wallis 1983	Unclear diagnostic criteria: "Demented/organic"
Wenisch 2007	Participants included with MCI, not dementia
Wettstein 2004	Did not report an intervention study
Williams 1987	Not an RCT; compared two wards; not cognitive stimulation; involved environmental modification and informal RO
Woods 1979	Unclear diagnostic criteria: "Significant memory impairment"
Yamanaka 2013	Intervention allocation not randomised
Young 2020	No data for people diagnosed with dementia

Study	Reason for exclusion
Zanetti 1995	Allocation non-randomised
Zepelin 1981	Not an RCT; compared residents at one home with those in another
Zientz 2007	Did not report an intervention study

MCI: mild cognitive impairment
RCT: randomised controlled trial
RO: reality orientation

Characteristics of ongoing studies *[ordered by study ID]*

ChiCTR1800018600

Study name	Evaluation on the effect of maintenance cognitive stimulation therapy for dementia patients
Methods	Random allocation to groups
Participants	Participants "should meet the diagnostic criteria for 10/66 dementia and be evaluated as mild dementia or moderate dementia by using the Clinical Dementia Rating (CDR)"'
Interventions	Maintenance cognitive stimulation (planned sample size 55) Treatment-as-usual (planned sample size 55)
Outcomes	Cognitive function Quality of life of person with dementia Caregiver burden Quality of life of caregiver
Starting date	01/03/2018
Contact information	Peking University Sixth Hospital; Dr H Chen chenhg@bjmu.edu.cn
Notes	

NCT04550975

Study name	Feasibility Randomised Controlled Trial (RCT) of Advanced Cognitive Stimulation Therapy (ACST) for people with moderate to severe dementia
Methods	RCT
Participants	People with severe dementia (SMMSE 5-12); DSM-IV criteria
Interventions	ACST (planned sample size 16) Treatment-as-usual (planned sample size 16)
Outcomes	Cognition: SMMSE Quality of life: QoL-AD

NCT04550975 *(Continued)*

	Communication: Holden Communication Scale Behaviour that challenges: NPI Engagement: Group Observational Measurement of Engagement Tool Overall well-being: Adapted Greater Cincinnati Chapter Well-Being Observation Tool
Starting date	01/03/2021
Contact information	University College London: Dr E Hui esther.hui.19@ucl.ac.uk
Notes	14 group sessions, twice weekly for 7 weeks

NCT04828434

Study name	Virtual individual Cognitive Stimulation Therapy: a proof of concept study (V-iCST)
Methods	RCT
Participants	DSM-IV dementia diagnosis
Interventions	Virtual iCST (planned sample size 17) Treatment-as-usual (planned sample size 17)
Outcomes	Cognition: MoCA BLIND; ADAS-Cog Quality of life: QoL-AD Mood: GDS-15 Communication: Holden Communication Scale Engagement: Adapted Greater Cincinnati Chapter Well-Being Observation Tool
Starting date	01/04/2021
Contact information	University College London: Prof. A. Spector a.spector@ucl.ac.uk
Notes	Two 45-minute sessions per week for 7 weeks

ACST: advanced cognitive stimulation therapy
ADAS-Cog: Alzheimer's Disease Assessment Scale - Cognitive Subscale
CDR: Clinical Dementia Rating
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders IV
GDS-15: Geriatric Depression Scale – 15 items
MoCA BLIND: Montreal Cognitive Assessment Blind version
NPI: Neuropsychiatric Inventory
QoL-AD: Quality of Life – Alzheimer's Disease
SMMSE: Standardised Mini-Mental State Examination
V-iCST: virtual individual cognitive stimulation therapy

DATA AND ANALYSES

Comparison 1. Cognitive stimulation versus no cognitive stimulation: post-treatment

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Cognition	34	2340	Std. Mean Difference (IV, Random, 95% CI)	0.40 [0.25, 0.55]
1.1.1 ADAS-Cog	21	1742	Std. Mean Difference (IV, Random, 95% CI)	0.30 [0.15, 0.46]
1.1.2 Global cognitive score (includes MMSE & CERAD)	1	56	Std. Mean Difference (IV, Random, 95% CI)	0.63 [0.09, 1.17]
1.1.3 MMSE	6	262	Std. Mean Difference (IV, Random, 95% CI)	0.52 [0.25, 0.79]
1.1.4 Mattis Dementia Rating Scale	1	101	Std. Mean Difference (IV, Random, 95% CI)	1.32 [0.89, 1.75]
1.1.5 Esame Neuropsicologico Breve 2 (ENB2)	1	20	Std. Mean Difference (IV, Random, 95% CI)	0.99 [0.05, 1.93]
1.1.6 Montreal Cognitive Assessment (MoCA)	2	76	Std. Mean Difference (IV, Random, 95% CI)	0.91 [0.22, 1.60]
1.1.7 ACE-III	1	43	Std. Mean Difference (IV, Random, 95% CI)	-0.28 [-0.89, 0.33]
1.1.8 CAM-COG-DS	1	40	Std. Mean Difference (IV, Random, 95% CI)	-0.14 [-0.76, 0.48]
1.2 MMSE	25		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.2.1 One to twelve months of CS	25	1893	Mean Difference (IV, Random, 95% CI)	1.99 [1.24, 2.74]
1.2.2 24 months of CS	1	29	Mean Difference (IV, Random, 95% CI)	5.99 [-1.58, 13.56]
1.3 ADAS-Cog	21		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.3.1 One to 12 months of CS	21	1742	Mean Difference (IV, Random, 95% CI)	2.42 [1.21, 3.63]
1.3.2 24 months of CS	1	29	Mean Difference (IV, Random, 95% CI)	11.94 [-0.97, 24.85]
1.4 Quality of Life: self-report	18	1584	Std. Mean Difference (IV, Random, 95% CI)	0.25 [0.07, 0.42]
1.4.1 QoL-AD	17	1541	Std. Mean Difference (IV, Random, 95% CI)	0.26 [0.07, 0.44]
1.4.2 EQ-5D	1	43	Std. Mean Difference (IV, Random, 95% CI)	0.03 [-0.58, 0.63]
1.5 Quality of Life: proxy-rated	11	988	Std. Mean Difference (IV, Random, 95% CI)	0.21 [0.00, 0.42]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.5.1 QoL-AD proxy	10	963	Std. Mean Difference (IV, Random, 95% CI)	0.20 [-0.02, 0.41]
1.5.2 QoL-D	1	25	Std. Mean Difference (IV, Random, 95% CI)	0.52 [-0.29, 1.34]
1.6 Communication and social interaction	7	702	Std. Mean Difference (IV, Random, 95% CI)	0.53 [0.36, 0.70]
1.6.1 Holden Communication Scale	3	299	Std. Mean Difference (IV, Random, 95% CI)	0.40 [0.15, 0.65]
1.6.2 Narrative language - communicative abilities	2	259	Std. Mean Difference (IV, Random, 95% CI)	0.64 [0.38, 0.89]
1.6.3 NOSGER - Social Behaviour	2	144	Std. Mean Difference (IV, Random, 95% CI)	0.71 [0.16, 1.26]
1.7 Mood: Self-reported	10	787	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.08, 0.31]
1.7.1 Geriatric Depression Scale (GDS-30) One to twelve months of CS	4	213	Std. Mean Difference (IV, Random, 95% CI)	0.28 [-0.02, 0.58]
1.7.2 Geriatric Depression Scale (14 item) One to twelve months of CS	1	26	Std. Mean Difference (IV, Random, 95% CI)	-0.39 [-1.16, 0.39]
1.7.3 Geriatric Depression Scale (GDS-15) One to twelve months of CS	2	402	Std. Mean Difference (IV, Random, 95% CI)	0.22 [-0.43, 0.88]
1.7.4 HADS - Depression	1	34	Std. Mean Difference (IV, Random, 95% CI)	-0.34 [-1.04, 0.37]
1.7.5 CESD-R	1	60	Std. Mean Difference (IV, Random, 95% CI)	0.21 [-0.31, 0.73]
1.7.6 Cornell Scale for Depression in Dementia (self-report)	1	52	Std. Mean Difference (IV, Random, 95% CI)	-0.02 [-0.56, 0.52]
1.8 Anxiety: Interviewer/staff-rated	6	410	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.09, 0.30]
1.8.1 Hamilton Anxiety Rating Scale	1	14	Std. Mean Difference (IV, Random, 95% CI)	0.84 [-0.27, 1.95]
1.8.2 NPI - anxiety subscale	1	39	Std. Mean Difference (IV, Random, 95% CI)	-0.07 [-0.69, 0.56]
1.8.3 Rating of Anxiety in Dementia (RAID)	4	357	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.11, 0.31]

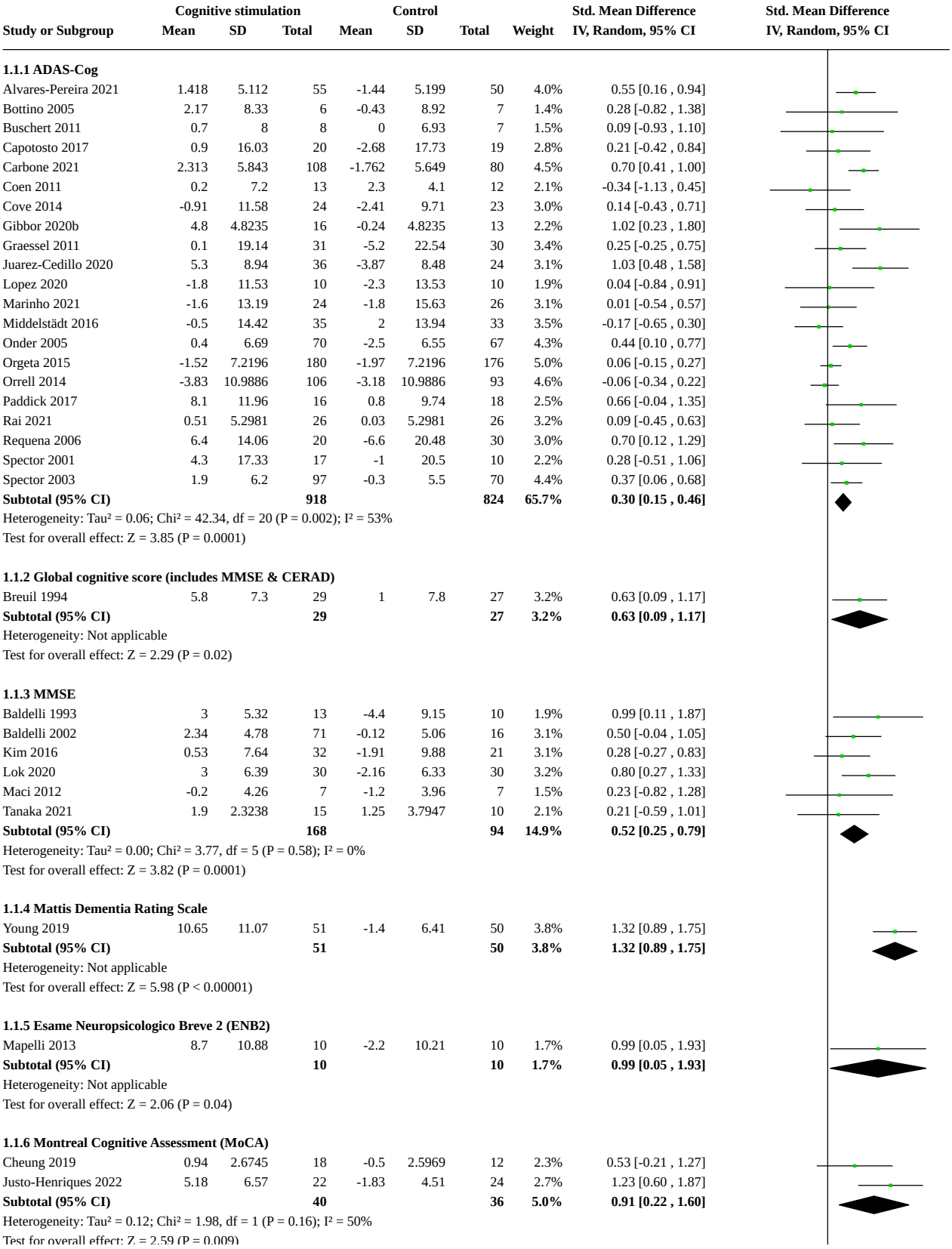
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.9 Mood: Interviewer/staff-rated	11	1011	Std. Mean Difference (IV, Random, 95% CI)	0.35 [0.09, 0.61]
1.9.1 Cornell Scale for Depression in Dementia	9	877	Std. Mean Difference (IV, Random, 95% CI)	0.36 [0.05, 0.67]
1.9.2 NOSGER - Mood	1	119	Std. Mean Difference (IV, Random, 95% CI)	0.35 [-0.02, 0.71]
1.9.3 Montgomery-Asberg Depression Rating Scale	1	15	Std. Mean Difference (IV, Random, 95% CI)	0.31 [-0.72, 1.33]
1.10 Quality of Relationship: self-report	4	492	Std. Mean Difference (IV, Random, 95% CI)	-0.01 [-0.27, 0.25]
1.10.1 QCPR	3	455	Std. Mean Difference (IV, Random, 95% CI)	0.08 [-0.15, 0.31]
1.10.2 Relationship Satisfaction Scale	1	37	Std. Mean Difference (IV, Random, 95% CI)	-0.37 [-1.03, 0.29]
1.11 ADL scales	7	360	Std. Mean Difference (IV, Random, 95% CI)	0.19 [-0.03, 0.41]
1.11.1 Stewart ADL scale	1	23	Std. Mean Difference (IV, Random, 95% CI)	0.25 [-0.57, 1.08]
1.11.2 Barthel ADL scale	3	249	Std. Mean Difference (IV, Random, 95% CI)	0.16 [-0.11, 0.43]
1.11.3 Erlangen Test of ADL	1	61	Std. Mean Difference (IV, Random, 95% CI)	0.29 [-0.21, 0.80]
1.11.4 Katz ADL scale	2	27	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.63, 0.88]
1.12 Instrumental ADL	13	1318	Std. Mean Difference (IV, Random, 95% CI)	0.15 [0.04, 0.26]
1.12.1 Lawton Brody Instrumental ADL	3	197	Std. Mean Difference (IV, Random, 95% CI)	0.17 [-0.11, 0.45]
1.12.2 Disability Assessment for Dementia	2	179	Std. Mean Difference (IV, Random, 95% CI)	0.23 [-0.07, 0.52]
1.12.3 NOSGER IADL subscale	1	119	Std. Mean Difference (IV, Random, 95% CI)	0.42 [0.06, 0.79]
1.12.4 Bristol Activities of Daily Living Scale (BADLS)	2	408	Std. Mean Difference (IV, Random, 95% CI)	0.05 [-0.15, 0.24]
1.12.5 Alzheimer's Disease Cooperative Study - ADL Scale	4	355	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.11, 0.31]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.12.6 Rapid Disability Rating Scale	1	60	Std. Mean Difference (IV, Random, 95% CI)	0.33 [-0.19, 0.85]
1.13 Behaviour that challenges	12	1340	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.01, 0.38]
1.13.1 NPI	8	1137	Std. Mean Difference (IV, Random, 95% CI)	0.13 [-0.10, 0.36]
1.13.2 NPI - Agitation	1	39	Std. Mean Difference (IV, Random, 95% CI)	0.00 [-0.63, 0.63]
1.13.3 NOSGER - Challenging Behaviour	1	119	Std. Mean Difference (IV, Random, 95% CI)	0.32 [-0.04, 0.68]
1.13.4 Behave-AD	1	20	Std. Mean Difference (IV, Random, 95% CI)	1.30 [0.32, 2.29]
1.13.5 Dementia Behaviour Disturbance Scale (DBD)	1	25	Std. Mean Difference (IV, Random, 95% CI)	0.19 [-0.61, 0.99]
1.14 Behaviour, General Rating Scales	6	505	Std. Mean Difference (IV, Random, 95% CI)	0.35 [0.13, 0.58]
1.14.1 CAPE Behaviour Rating Scale	4	326	Std. Mean Difference (IV, Random, 95% CI)	0.21 [-0.01, 0.43]
1.14.2 NOSGER	1	119	Std. Mean Difference (IV, Random, 95% CI)	0.65 [0.28, 1.02]
1.14.3 Blessed Dementia Rating Scale	1	60	Std. Mean Difference (IV, Random, 95% CI)	0.51 [-0.02, 1.03]
1.15 Caregiver outcome - anxiety	5	600	Std. Mean Difference (IV, Random, 95% CI)	-0.00 [-0.19, 0.19]
1.15.1 Hamilton Anxiety Scale	2	150	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.33, 0.68]
1.15.2 Hospital Anxiety & Depression Scale - Anxiety	3	450	Std. Mean Difference (IV, Random, 95% CI)	-0.06 [-0.33, 0.21]
1.16 Caregiver outcome - depressed mood	8	664	Std. Mean Difference (IV, Random, 95% CI)	0.05 [-0.10, 0.21]
1.16.1 Hospital Anxiety & Depression Scales - depression	4	490	Std. Mean Difference (IV, Random, 95% CI)	0.02 [-0.16, 0.20]
1.16.2 Hamilton Depression Scale	1	137	Std. Mean Difference (IV, Random, 95% CI)	0.03 [-0.30, 0.37]
1.16.3 Beck Depression Inventory	1	14	Std. Mean Difference (IV, Random, 95% CI)	1.04 [-0.10, 2.18]

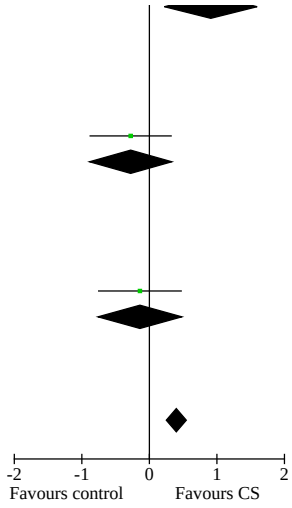
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.16.4 Montgomery-Asberg Depression Rating Scale	1	13	Std. Mean Difference (IV, Random, 95% CI)	0.19 [-0.91, 1.28]
1.16.5 General Health Questionnaire (GHQ-12)	1	10	Std. Mean Difference (IV, Random, 95% CI)	0.94 [-0.41, 2.29]
1.17 Caregiver outcome - caregiving stress/burden (2)	6	288	Std. Mean Difference (IV, Random, 95% CI)	0.09 [-0.14, 0.32]
1.17.1 Caregiver Burden Inventory	2	151	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.50, 0.75]
1.17.2 Relative's Stress Scale	1	10	Std. Mean Difference (IV, Random, 95% CI)	0.35 [-0.91, 1.60]
1.17.3 Zarit Burden Inventory	2	87	Std. Mean Difference (IV, Random, 95% CI)	0.14 [-0.28, 0.57]
1.17.4 Caregiver Burden Scale	1	40	Std. Mean Difference (IV, Random, 95% CI)	0.23 [-0.40, 0.85]
1.18 Caregiver outcome - caregiving stress/burden (1)	6	286	Std. Mean Difference (IV, Random, 95% CI)	0.09 [-0.15, 0.32]
1.18.1 Caregiver Burden Inventory	2	151	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.50, 0.75]
1.18.2 Relative's Stress Scale	2	50	Std. Mean Difference (IV, Random, 95% CI)	0.19 [-0.37, 0.75]
1.18.3 Zarit Burden Inventory	1	45	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.47, 0.70]
1.18.4 Caregiver Burden Scale	1	40	Std. Mean Difference (IV, Random, 95% CI)	0.23 [-0.40, 0.85]
1.19 Caregiver outcome - health-related quality of life	5	651	Std. Mean Difference (IV, Random, 95% CI)	0.17 [-0.14, 0.49]
1.19.1 EQ-5D	4	514	Std. Mean Difference (IV, Random, 95% CI)	0.24 [-0.18, 0.65]
1.19.2 SF-36	1	137	Std. Mean Difference (IV, Random, 95% CI)	-0.02 [-0.35, 0.32]
1.20 Caregiver outcome - SF-12	3		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
1.20.1 SF-12 PCS	3	461	Std. Mean Difference (IV, Random, 95% CI)	0.07 [-0.11, 0.25]
1.20.2 SF12 - MCS	3	461	Std. Mean Difference (IV, Random, 95% CI)	-0.05 [-0.23, 0.13]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.21 Caregiver outcome - quality of relationship	3	367	Std. Mean Difference (IV, Random, 95% CI)	0.06 [-0.26, 0.37]
1.21.1 QCPR	2	325	Std. Mean Difference (IV, Random, 95% CI)	-0.07 [-0.29, 0.15]
1.21.2 Relationship Satisfaction Scale	1	42	Std. Mean Difference (IV, Random, 95% CI)	0.53 [-0.10, 1.15]
1.22 Caregiver outcome - resilience	2	399	Std. Mean Difference (IV, Random, 95% CI)	0.06 [-0.13, 0.26]
1.22.1 Brief Resilience Scale (Wagnild - RS14)	1	356	Std. Mean Difference (IV, Random, 95% CI)	0.08 [-0.13, 0.29]
1.22.2 Brief Resilience Scale (Smith and colleagues)	1	43	Std. Mean Difference (IV, Random, 95% CI)	-0.07 [-0.68, 0.53]

Analysis 1.1. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 1: Cognition



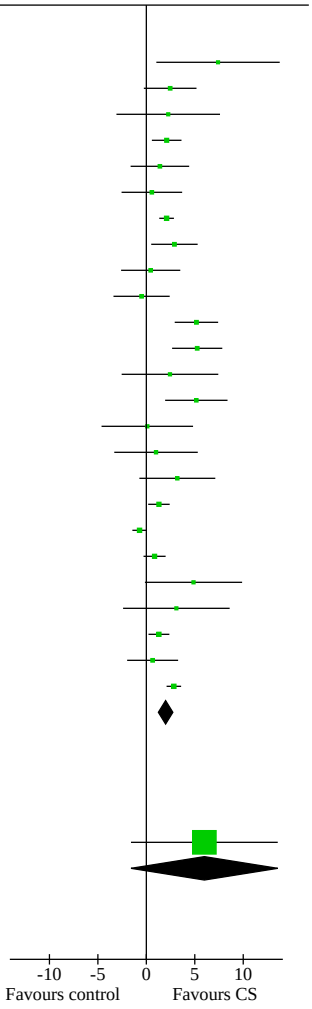
Analysis 1.1. (Continued)



Heterogeneity: Tau² = 0.12; Chi² = 1.98, df = 1 (P = 0.16); I² = 50%								
Test for overall effect: Z = 2.59 (P = 0.009)								
1.1.7 ACE-III								
Leroi 2019	-3.29	8.8518	18	-0.79	8.8518	25	2.9%	-0.28 [-0.89 , 0.33]
Subtotal (95% CI)			**18**			**25**	**2.9%**	**-0.28 [-0.89 , 0.33]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.89 (P = 0.37)								
1.1.8 CAM-COG-DS								
Ali 2021	-2.29	11.0682	20	-0.71	11.0682	20	2.8%	-0.14 [-0.76 , 0.48]
Subtotal (95% CI)			**20**			**20**	**2.8%**	**-0.14 [-0.76 , 0.48]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.44 (P = 0.66)								
Total (95% CI)			**1254**			**1086**	**100.0%**	**0.40 [0.25 , 0.55]**
Heterogeneity: Tau² = 0.10; Chi² = 86.32, df = 33 (P < 0.00001); I² = 62%								
Test for overall effect: Z = 5.28 (P < 0.00001)								
Test for subgroup differences: Chi² = 31.39, df = 7 (P < 0.0001), I² = 77.7%								

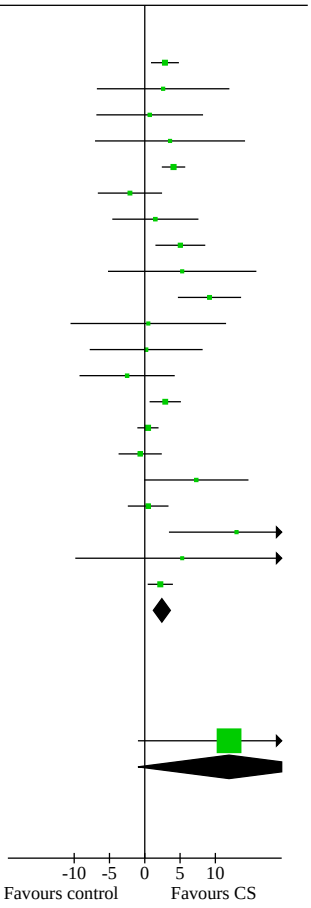
Analysis 1.2. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 2: MMSE

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
1.2.1 One to twelve months of CS								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	1.2%	7.40 [1.03 , 13.77]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	3.9%	2.46 [-0.26 , 5.18]
Bottino 2005	0.83	4.53	6	-1.43	5.3	7	1.6%	2.26 [-3.08 , 7.60]
Breuil 1994	1.4	2.7	29	-0.7	3.1	27	5.9%	2.10 [0.57 , 3.63]
Buschert 2011	0.5	3.14	8	-0.9	2.83	7	3.4%	1.40 [-1.62 , 4.42]
Capotosto 2017	0.18	4.57	20	-0.39	5.34	19	3.3%	0.57 [-2.56 , 3.70]
Carbone 2021	0.8894	2.68673	123	-1.2016	3.01666	101	7.3%	2.09 [1.33 , 2.85]
Coen 2011	0.8	3.6	14	-2.1	2.5	11	4.3%	2.90 [0.50 , 5.30]
Cove 2014	-0.33	6.06	24	-0.78	4.54	23	3.4%	0.45 [-2.60 , 3.50]
Gibbor 2020b	-0.5	3.9626	16	-0.01	3.9626	13	3.6%	-0.49 [-3.39 , 2.41]
Juarez-Cedillo 2020	2.05	4.41	36	-3.12	4.29	24	4.6%	5.17 [2.93 , 7.41]
Justo-Henriques 2022	3.54	3.84	22	-1.71	5.08	24	4.0%	5.25 [2.66 , 7.84]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	1.7%	2.44 [-2.55 , 7.43]
Lok 2020	3	6.39	30	-2.16	6.33	30	3.2%	5.16 [1.94 , 8.38]
Lopez 2020	-1.6	4.82	10	-1.7	5.9	10	1.9%	0.10 [-4.62 , 4.82]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	2.2%	1.00 [-3.31 , 5.31]
Mapelli 2013	2.9	5.03	10	-0.3	3.83	10	2.5%	3.20 [-0.72 , 7.12]
Onder 2005	0.2	3.35	70	-1.1	3.27	67	6.7%	1.30 [0.19 , 2.41]
Orgeta 2015	-2.24	3.5616	180	-1.54	3.5616	176	7.3%	-0.70 [-1.44 , 0.04]
Orrell 2014	-1.46	4.0938	106	-2.31	4.0938	93	6.7%	0.85 [-0.29 , 1.99]
Requena 2006	1.5	7.38	20	-3.37	10.71	30	1.7%	4.87 [-0.14 , 9.88]
Spector 2001	3.1	7.04	17	0	7.04	10	1.5%	3.10 [-2.40 , 8.60]
Spector 2003	0.9	3.5	97	-0.4	3.5	70	6.8%	1.30 [0.22 , 2.38]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	4.0%	0.65 [-1.98 , 3.28]
Young 2019	2.1	2.26	51	-0.74	1.52	50	7.3%	2.84 [2.09 , 3.59]
Subtotal (95% CI)			**1027**			**866**	**100.0%**	**1.99 [1.24 , 2.74]**
Heterogeneity: Tau² = 1.83; Chi² = 85.31, df = 24 (P < 0.00001); I² = 72%								
Test for overall effect: Z = 5.23 (P < 0.00001)								
1.2.2 24 months of CS								
Requena 2006	-1.31	10.3	14	-7.3	10.5	15	100.0%	5.99 [-1.58 , 13.56]
Subtotal (95% CI)			**14**			**15**	**100.0%**	**5.99 [-1.58 , 13.56]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.55 (P = 0.12)								



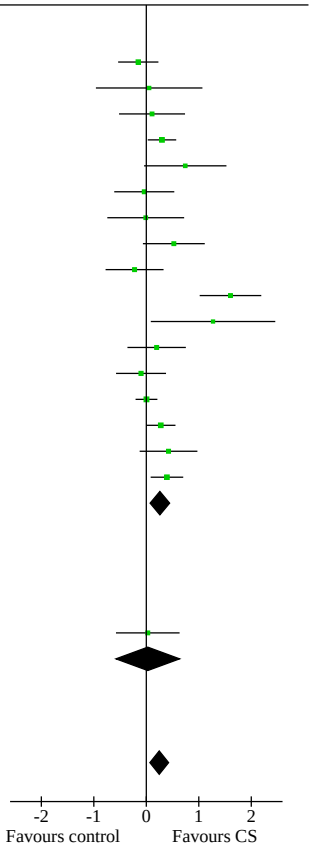
Analysis 1.3. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 3: ADAS-Cog

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
1.3.1 One to 12 months of CS								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	9.8%	2.86 [0.88 , 4.83]
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	1.5%	2.60 [-6.79 , 11.99]
Buschert 2011	0.7	8	8	0	6.93	7	2.1%	0.70 [-6.86 , 8.26]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	1.2%	3.58 [-7.05 , 14.21]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	10.6%	4.08 [2.42 , 5.73]
Coen 2011	0.2	7.2	13	2.3	4.1	12	4.6%	-2.10 [-6.65 , 2.45]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	3.0%	1.50 [-4.60 , 7.60]
Gibbor 2020b	4.8	4.8235	16	-0.24	4.8235	13	6.2%	5.04 [1.51 , 8.57]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	1.2%	5.30 [-5.21 , 15.81]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	4.7%	9.17 [4.69 , 13.65]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	1.1%	0.50 [-10.52 , 11.52]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	1.9%	0.20 [-7.80 , 8.20]
Middelstädt 2016	-0.5	14.42	35	2	13.95	33	2.6%	-2.50 [-9.24 , 4.24]
Onder 2005	0.4	6.69	70	-2.5	6.55	67	9.2%	2.90 [0.68 , 5.12]
Orgeta 2015	-1.52	7.2196	180	-1.97	7.2196	176	11.0%	0.45 [-1.05 , 1.95]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	7.2%	-0.65 [-3.71 , 2.41]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	2.2%	7.30 [-0.09 , 14.69]
Rai 2021	0.51	5.2981	26	0.03	5.2981	26	7.6%	0.48 [-2.40 , 3.36]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	1.4%	13.00 [3.43 , 22.57]
Spector 2001	4.3	17.33	17	-1	20.5	10	0.6%	5.30 [-9.84 , 20.44]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	10.3%	2.20 [0.42 , 3.98]
Subtotal (95% CI)			**918**			**824**	**100.0%**	**2.42 [1.21 , 3.63]**
Heterogeneity: Tau² = 2.85; Chi² = 40.53, df = 20 (P = 0.004); I² = 51%								
Test for overall effect: Z = 3.93 (P < 0.0001)								
1.3.2 24 months of CS								
Requena 2006	3.38	18.26	14	-8.56	17.13	15	100.0%	11.94 [-0.97 , 24.85]
Subtotal (95% CI)			**14**			**15**	**100.0%**	**11.94 [-0.97 , 24.85]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.81 (P = 0.07)								



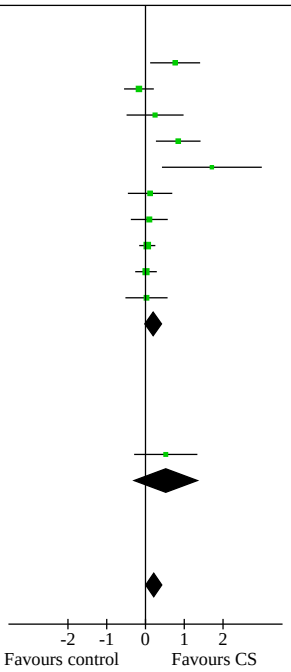
Analysis 1.4. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 4: Quality of Life: self-report

Study or Subgroup	Cognitive Stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.4.1 QoL-AD								
Alvares-Pereira 2021	-0.53	4.63	55	0.27	5.74	50	7.2%	-0.15 [-0.54 , 0.23]
Buschert 2011	-0.4	10.61	8	-0.9	5.52	7	2.4%	0.05 [-0.96 , 1.07]
Capotosto 2017	1.25	11.5	20	0.05	9.91	19	4.6%	0.11 [-0.52 , 0.74]
Carbone 2021	1.6	6.63	118	-0.2	5.12	96	8.7%	0.30 [0.03 , 0.57]
Coen 2011	3.6	3.7	14	0.5	4.4	13	3.5%	0.74 [-0.04 , 1.53]
Cove 2014	0.23	7.97	24	0.54	7.74	23	5.1%	-0.04 [-0.61 , 0.53]
Gibbor 2020b	2.09	5.3427	16	2.15	5.3427	13	3.8%	-0.01 [-0.74 , 0.72]
Justo-Henriques 2022	1.64	7.14	22	-2.5	8.28	24	5.0%	0.52 [-0.06 , 1.11]
Kim 2016	-0.4	0.76	32	-0.23	0.73	21	5.3%	-0.22 [-0.78 , 0.33]
Lok 2020	11.86	9.64	30	-2.16	7.48	30	5.0%	1.60 [1.02 , 2.19]
Maci 2012	12.3	11.78	7	-1.3	7.86	7	1.9%	1.27 [0.09 , 2.46]
Marinho 2021	2	8.77	24	0.4	7.21	26	5.3%	0.20 [-0.36 , 0.75]
Middelstädt 2016	-0.2	5.47	35	0.4	6.36	33	6.1%	-0.10 [-0.58 , 0.38]
Orgeta 2015	-0.25	9.4817	180	-0.28	9.4817	176	9.4%	0.00 [-0.20 , 0.21]
Orrell 2014	-0.67	6.428	106	-2.45	6.428	93	8.5%	0.28 [-0.00 , 0.56]
Rai 2021	0.25	3.9551	26	-1.45	3.9551	26	5.3%	0.42 [-0.13 , 0.97]
Spector 2003	1.3	5.1	97	-0.8	5.6	70	8.1%	0.39 [0.08 , 0.70]
Subtotal (95% CI)			**814**			**727**	**95.2%**	**0.26 [0.07 , 0.44]**
Heterogeneity: Tau² = 0.08; Chi² = 42.68, df = 16 (P = 0.0003); I² = 63%								
Test for overall effect: Z = 2.73 (P = 0.006)								
1.4.2 EQ-5D								
Leroi 2019	0.0075	0.2616	18	0	0.2616	25	4.8%	0.03 [-0.58 , 0.63]
Subtotal (95% CI)			**18**			**25**	**4.8%**	**0.03 [-0.58 , 0.63]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.09 (P = 0.93)								
Total (95% CI)			**832**			**752**	**100.0%**	**0.25 [0.07 , 0.42]**
Heterogeneity: Tau² = 0.08; Chi² = 43.00, df = 17 (P = 0.0005); I² = 60%								
Test for overall effect: Z = 2.71 (P = 0.007)								
Test for subgroup differences: Chi² = 0.51, df = 1 (P = 0.48), I² = 0%								



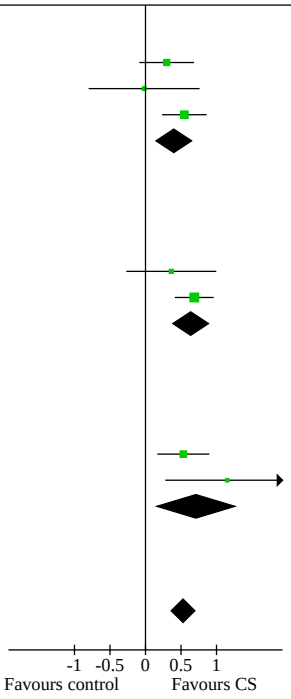
Analysis 1.5. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 5: Quality of Life: proxy-rated

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.5.1 QoL-AD proxy								
Ali 2021	0.26	3.9852	20	-2.85	3.9852	20	7.0%	0.76 [0.12 , 1.41]
Alvares-Pereira 2021	-0.53	5.83	55	0.2	1.096	50	12.3%	-0.17 [-0.55 , 0.21]
Gibbor 2020b	0.68	6.5725	16	-0.99	6.5725	13	5.8%	0.25 [-0.49 , 0.98]
Kim 2016	0.81	2.34	32	-1.09	2.01	21	8.1%	0.84 [0.27 , 1.42]
Maci 2012	10.2	8.08	7	-1.4	3.89	7	2.3%	1.71 [0.43 , 3.00]
Marinho 2021	0.8	7.64	23	-0.2	8.91	24	8.1%	0.12 [-0.45 , 0.69]
Middelstädt 2016	-0.4	6.16	35	-1	5.94	33	10.0%	0.10 [-0.38 , 0.57]
Orgeta 2015	-1.71	5.6313	180	-1.97	5.6313	176	17.5%	0.05 [-0.16 , 0.25]
Orrell 2014	0.62	5.2429	106	0.55	5.2429	93	15.3%	0.01 [-0.27 , 0.29]
Rai 2021	0.04	4.5622	26	-0.08	4.5622	26	8.6%	0.03 [-0.52 , 0.57]
Subtotal (95% CI)			**500**			**463**	**95.0%**	**0.20 [-0.02 , 0.41]**
Heterogeneity: Tau² = 0.06; Chi² = 19.27, df = 9 (P = 0.02); I² = 53%								
Test for overall effect: Z = 1.80 (P = 0.07)								
1.5.2 QoL-D								
Tanaka 2021	1.5	6.9714	15	-1.8	4.4272	10	5.0%	0.52 [-0.29 , 1.34]
Subtotal (95% CI)			**15**			**10**	**5.0%**	**0.52 [-0.29 , 1.34]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.26 (P = 0.21)								
Total (95% CI)			**515**			**473**	**100.0%**	**0.21 [0.00 , 0.42]**
Heterogeneity: Tau² = 0.05; Chi² = 20.22, df = 10 (P = 0.03); I² = 51%								
Test for overall effect: Z = 2.00 (P = 0.05)								
Test for subgroup differences: Chi² = 0.57, df = 1 (P = 0.45), I² = 0%								



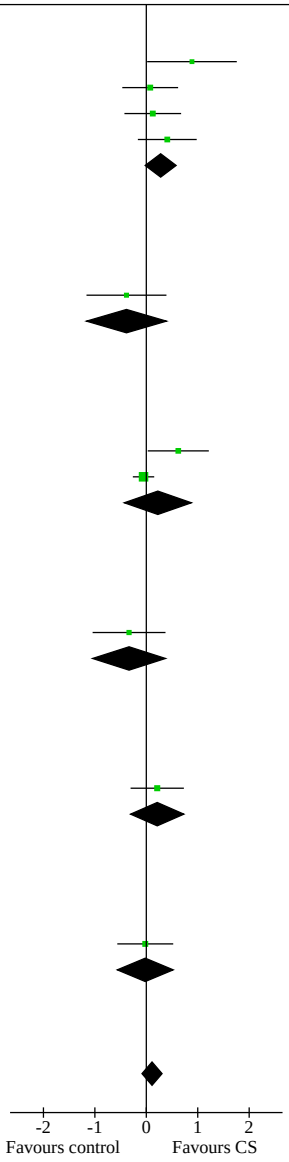
Analysis 1.6. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 6: Communication and social interaction

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.6.1 Holden Communication Scale								
Alvares-Pereira 2021	2.75	6.29	55	0.32	9.67	50	16.3%	0.30 [-0.09 , 0.68]
Spector 2001	-0.7	10.5	17	-0.5	9.4	10	4.4%	-0.02 [-0.80 , 0.76]
Spector 2003	0.2	6.1	97	-3.2	6.3	70	23.0%	0.55 [0.23 , 0.86]
Subtotal (95% CI)			**169**			**130**	**43.7%**	**0.40 [0.15 , 0.65]**
Heterogeneity: Tau² = 0.01; Chi² = 2.22, df = 2 (P = 0.33); I² = 10%								
Test for overall effect: Z = 3.11 (P = 0.002)								
1.6.2 Narrative language - communicative abilities								
Capotosto 2017	2.45	4.73	20	0.53	5.61	19	6.5%	0.36 [-0.27 , 1.00]
Carbone 2021	3.02	4.41	122	0.22	3.58	98	28.5%	0.69 [0.41 , 0.96]
Subtotal (95% CI)			**142**			**117**	**35.1%**	**0.64 [0.38 , 0.89]**
Heterogeneity: Tau² = 0.00; Chi² = 0.84, df = 1 (P = 0.36); I² = 0%								
Test for overall effect: Z = 4.96 (P < 0.00001)								
1.6.3 NOSGER - Social Behaviour								
Graessel 2011	1	3.0993	56	-0.54	2.6603	63	17.7%	0.53 [0.17 , 0.90]
Tanaka 2021	0.8	3.0984	15	-3.4	4.111	10	3.5%	1.15 [0.28 , 2.02]
Subtotal (95% CI)			**71**			**73**	**21.3%**	**0.71 [0.16 , 1.26]**
Heterogeneity: Tau² = 0.07; Chi² = 1.64, df = 1 (P = 0.20); I² = 39%								
Test for overall effect: Z = 2.54 (P = 0.01)								
Total (95% CI)			**382**			**320**	**100.0%**	**0.53 [0.36 , 0.70]**
Heterogeneity: Tau² = 0.01; Chi² = 6.77, df = 6 (P = 0.34); I² = 11%								
Test for overall effect: Z = 6.22 (P < 0.00001)								
Test for subgroup differences: Chi² = 2.15, df = 2 (P = 0.34), I² = 7.0%								



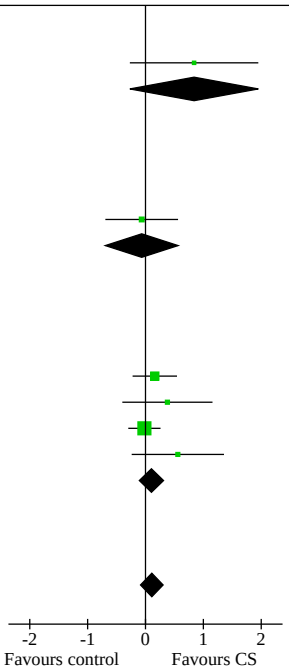
Analysis 1.7. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 7: Mood: Self-reported

Study or Subgroup	Cognitive Stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.7.1 Geriatric Depression Scale (GDS-30) One to twelve months of CS								
Baldelli 1993	2.1	4.61	13	-2.3	4.99	10	4.4%	0.89 [0.02 , 1.76]
Baldelli 2002	3.21	7.98	71	2.57	10	16	9.7%	0.08 [-0.47 , 0.62]
Kim 2016	1.44	10.9	32	0.11	9.62	21	9.5%	0.13 [-0.43 , 0.68]
Requena 2006	5.6	7.87	20	2.03	9.07	30	9.0%	0.41 [-0.16 , 0.98]
Subtotal (95% CI)			**136**			**77**	**32.6%**	**0.28 [-0.02 , 0.58]**
Heterogeneity: Tau² = 0.00; Chi² = 2.91, df = 3 (P = 0.41); I² = 0%								
Test for overall effect: Z = 1.82 (P = 0.07)								
1.7.2 Geriatric Depression Scale (14 item) One to twelve months of CS								
Coen 2011	-0.9	3	13	0.1	1.9	13	5.4%	-0.39 [-1.16 , 0.39]
Subtotal (95% CI)			**13**			**13**	**5.4%**	**-0.39 [-1.16 , 0.39]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.97 (P = 0.33)								
1.7.3 Geriatric Depression Scale (GDS-15) One to twelve months of CS								
Justo-Henriques 2022	1.37	4.62	22	-1.5	4.46	24	8.5%	0.62 [0.03 , 1.22]
Orgeta 2015	-0.3	2.214	180	-0.18	2.214	176	27.1%	-0.05 [-0.26 , 0.15]
Subtotal (95% CI)			**202**			**200**	**35.5%**	**0.22 [-0.43 , 0.88]**
Heterogeneity: Tau² = 0.18; Chi² = 4.44, df = 1 (P = 0.04); I² = 77%								
Test for overall effect: Z = 0.67 (P = 0.50)								
1.7.4 HADS - Depression								
Leroi 2019	0.73	2.7652	12	1.68	2.7652	22	6.4%	-0.34 [-1.04 , 0.37]
Subtotal (95% CI)			**12**			**22**	**6.4%**	**-0.34 [-1.04 , 0.37]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.93 (P = 0.35)								
1.7.5 CESD-R								
Juarez-Cedillo 2020	5.08	16.8	36	1.45	17	24	10.4%	0.21 [-0.31 , 0.73]
Subtotal (95% CI)			**36**			**24**	**10.4%**	**0.21 [-0.31 , 0.73]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.80 (P = 0.42)								
1.7.6 Cornell Scale for Depression in Dementia (self-report)								
Rai 2021	-0.02	3.4033	26	0.05	3.4033	26	9.7%	-0.02 [-0.56 , 0.52]
Subtotal (95% CI)			**26**			**26**	**9.7%**	**-0.02 [-0.56 , 0.52]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.07 (P = 0.94)								
Total (95% CI)			**425**			**362**	**100.0%**	**0.11 [-0.08 , 0.31]**
Heterogeneity: Tau² = 0.03; Chi² = 12.42, df = 9 (P = 0.19); I² = 28%								
Test for overall effect: Z = 1.12 (P = 0.26)								
Test for subgroup differences: Chi² = 4.76, df = 5 (P = 0.45), I² = 0%								



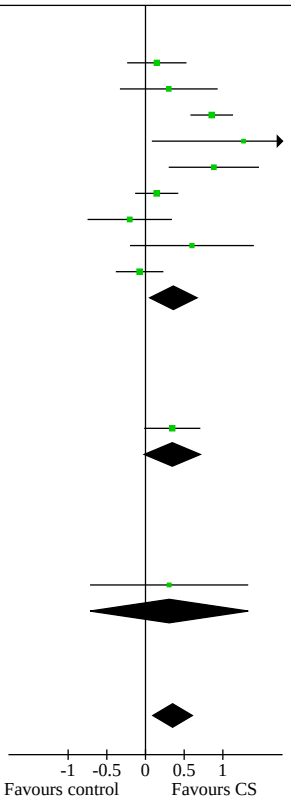
Analysis 1.8. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 8: Anxiety: Interviewer/staff-rated

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.8.1 Hamilton Anxiety Rating Scale								
Maci 2012	5.7	5.61	7	1.3	4.08	7	3.1%	0.84 [-0.27 , 1.95]
Subtotal (95% CI)			**7**			**7**	**3.1%**	**0.84 [-0.27 , 1.95]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.48 (P = 0.14)								
1.8.2 NPI - anxiety subscale								
Capotosto 2017	0.15	1.71	20	0.26	1.58	19	9.7%	-0.07 [-0.69 , 0.56]
Subtotal (95% CI)			**20**			**19**	**9.7%**	**-0.07 [-0.69 , 0.56]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.20 (P = 0.84)								
1.8.3 Rating of Anxiety in Dementia (RAID)								
Alvares-Pereira 2021	3.07	8.26	55	1.9	5.87	50	25.9%	0.16 [-0.22 , 0.54]
Coen 2011	1.1	7.3	14	-1.6	6.4	12	6.3%	0.38 [-0.40 , 1.16]
Orrell 2014	-0.22	4.2208	106	-0.14	4.2208	93	49.1%	-0.02 [-0.30 , 0.26]
Spector 2001	3.1	11.68	17	-3.2	9.45	10	6.0%	0.56 [-0.24 , 1.36]
Subtotal (95% CI)			**192**			**165**	**87.3%**	**0.10 [-0.11 , 0.31]**
Heterogeneity: Tau² = 0.00; Chi² = 2.56, df = 3 (P = 0.46); I² = 0%								
Test for overall effect: Z = 0.96 (P = 0.34)								
Total (95% CI)			**219**			**191**	**100.0%**	**0.11 [-0.09 , 0.30]**
Heterogeneity: Tau² = 0.00; Chi² = 4.52, df = 5 (P = 0.48); I² = 0%								
Test for overall effect: Z = 1.10 (P = 0.27)								
Test for subgroup differences: Chi² = 1.96, df = 2 (P = 0.37), I² = 0%								



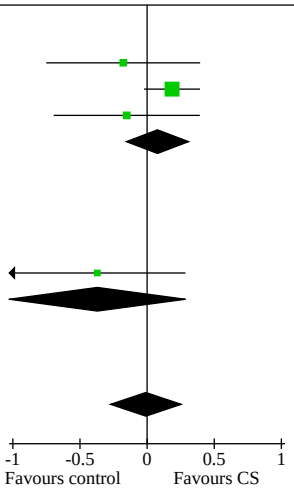
Analysis 1.9. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 9: Mood: Interviewer/staff-rated

Study or Subgroup	Cognitive Stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.9.1 Cornell Scale for Depression in Dementia								
Alvares-Pereira 2021	1.91	4.38	55	1.2	5.25	50	11.2%	0.15 [-0.24 , 0.53]
Capotosto 2017	1.4	2.87	20	0.42	3.49	19	7.8%	0.30 [-0.33 , 0.93]
Carbone 2021	1.9	3.14	123	-0.89	3.37	101	12.8%	0.86 [0.58 , 1.13]
Maci 2012	5.9	6.26	7	-1.3	4.17	7	3.6%	1.27 [0.08 , 2.45]
Marinho 2021	1.3	2.8	24	-1.1	2.55	26	8.4%	0.88 [0.30 , 1.47]
Orrell 2014	-0.38	5.2185	106	-1.14	5.2185	93	12.8%	0.15 [-0.13 , 0.42]
Rai 2021	-0.04	4.0103	26	0.79	4.0103	26	8.9%	-0.20 [-0.75 , 0.34]
Spector 2001	2.6	8.05	17	-2.2	7.19	10	6.1%	0.60 [-0.20 , 1.40]
Spector 2003	0	6.2	97	0.5	7	70	12.4%	-0.08 [-0.38 , 0.23]
Subtotal (95% CI)			**475**			**402**	**84.0%**	**0.36 [0.05 , 0.67]**
Heterogeneity: Tau² = 0.15; Chi² = 33.35, df = 8 (P < 0.0001); I² = 76%								
Test for overall effect: Z = 2.31 (P = 0.02)								
1.9.2 NOSGER - Mood								
Graessel 2011	1	2.5392	56	0.1	2.6206	63	11.6%	0.35 [-0.02 , 0.71]
Subtotal (95% CI)			**56**			**63**	**11.6%**	**0.35 [-0.02 , 0.71]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.87 (P = 0.06)								
1.9.3 Montgomery-Asberg Depression Rating Scale								
Buschert 2011	1.5	5.33	8	-0.4	6.4	7	4.4%	0.31 [-0.72 , 1.33]
Subtotal (95% CI)			**8**			**7**	**4.4%**	**0.31 [-0.72 , 1.33]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.59 (P = 0.56)								
Total (95% CI)			**539**			**472**	**100.0%**	**0.35 [0.09 , 0.61]**
Heterogeneity: Tau² = 0.11; Chi² = 33.36, df = 10 (P = 0.0002); I² = 70%								
Test for overall effect: Z = 2.69 (P = 0.007)								
Test for subgroup differences: Chi² = 0.01, df = 2 (P = 0.99), I² = 0%								



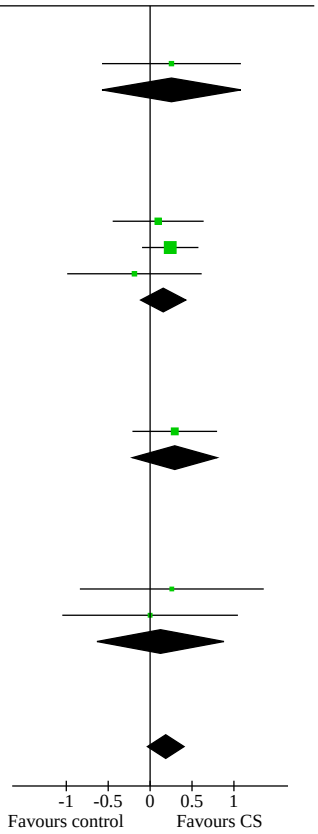
Analysis 1.10. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 10: Quality of Relationship: self-report

Study or Subgroup	CS Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.10.1 QCPR								
Cove 2014	-1.44	9.72	24	0.28	9.23	23	16.6%	-0.18 [-0.75 , 0.39]
Orgeta 2015	0.65	7.4121	180	-0.73	7.4121	176	52.2%	0.19 [-0.02 , 0.39]
Rai 2021	0.71	5.39	26	1.54	5.39	26	18.0%	-0.15 [-0.70 , 0.39]
Subtotal (95% CI)			**230**			**225**	**86.7%**	**0.08 [-0.15 , 0.31]**
Heterogeneity: Tau² = 0.01; Chi² = 2.37, df = 2 (P = 0.31); I² = 16%								
Test for overall effect: Z = 0.65 (P = 0.51)								
1.10.2 Relationship Satisfaction Scale								
Leroi 2019	-0.42	8.5747	16	2.83	8.5747	21	13.3%	-0.37 [-1.03 , 0.29]
Subtotal (95% CI)			**16**			**21**	**13.3%**	**-0.37 [-1.03 , 0.29]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.11 (P = 0.27)								
Total (95% CI)			**246**			**246**	**100.0%**	**-0.01 [-0.27 , 0.25]**
Heterogeneity: Tau² = 0.02; Chi² = 4.28, df = 3 (P = 0.23); I² = 30%								
Test for overall effect: Z = 0.07 (P = 0.95)								
Test for subgroup differences: Chi² = 1.59, df = 1 (P = 0.21), I² = 37.2%								



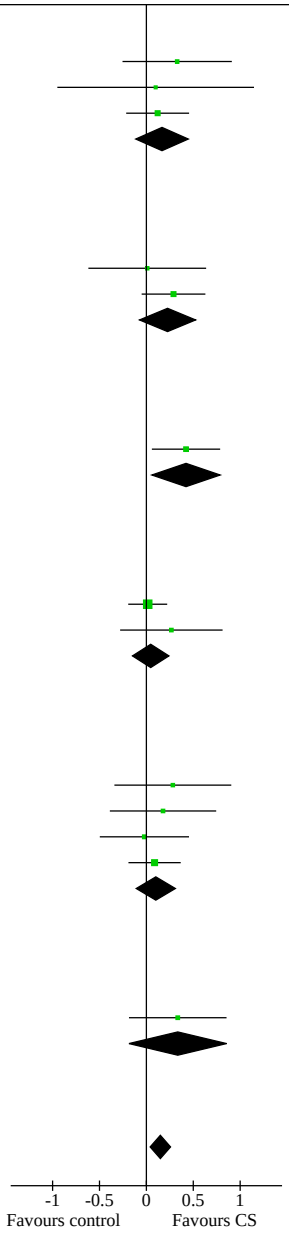
Analysis 1.11. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 11: ADL scales

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.11.1 Stewart ADL scale								
Baldelli 1993	1.5	39.47	13	-8.9	39.2	10	7.0%	0.25 [-0.57 , 1.08]
Subtotal (95% CI)			**13**			**10**	**7.0%**	**0.25 [-0.57 , 1.08]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.60 (P = 0.55)								
1.11.2 Barthel ADL scale								
Baldelli 2002	15.37	34.94	71	11.88	40.48	16	16.2%	0.10 [-0.45 , 0.64]
Onder 2005	-0.9	8.37	70	-2.9	8.19	67	42.3%	0.24 [-0.10 , 0.58]
Tanaka 2021	0.7	5.0349	15	1.7	5.3759	10	7.4%	-0.19 [-0.99 , 0.62]
Subtotal (95% CI)			**156**			**93**	**66.0%**	**0.16 [-0.11 , 0.43]**
Heterogeneity: Tau² = 0.00; Chi² = 0.99, df = 2 (P = 0.61); I² = 0%								
Test for overall effect: Z = 1.14 (P = 0.25)								
1.11.3 Erlangen Test of ADL								
Graessel 2011	-0.3	7.43	31	-2.8	9.28	30	18.8%	0.29 [-0.21 , 0.80]
Subtotal (95% CI)			**31**			**30**	**18.8%**	**0.29 [-0.21 , 0.80]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.14 (P = 0.25)								
1.11.4 Katz ADL scale								
Bottino 2005	1	3.27	6	0.15	2.86	7	4.0%	0.26 [-0.84 , 1.36]
Maci 2012	0	1.27	7	0	1.84	7	4.4%	0.00 [-1.05 , 1.05]
Subtotal (95% CI)			**13**			**14**	**8.3%**	**0.12 [-0.63 , 0.88]**
Heterogeneity: Tau² = 0.00; Chi² = 0.11, df = 1 (P = 0.74); I² = 0%								
Test for overall effect: Z = 0.32 (P = 0.75)								
Total (95% CI)			**213**			**147**	**100.0%**	**0.19 [-0.03 , 0.41]**
Heterogeneity: Tau² = 0.00; Chi² = 1.38, df = 6 (P = 0.97); I² = 0%								
Test for overall effect: Z = 1.67 (P = 0.09)								
Test for subgroup differences: Chi² = 0.27, df = 3 (P = 0.96), I² = 0%								

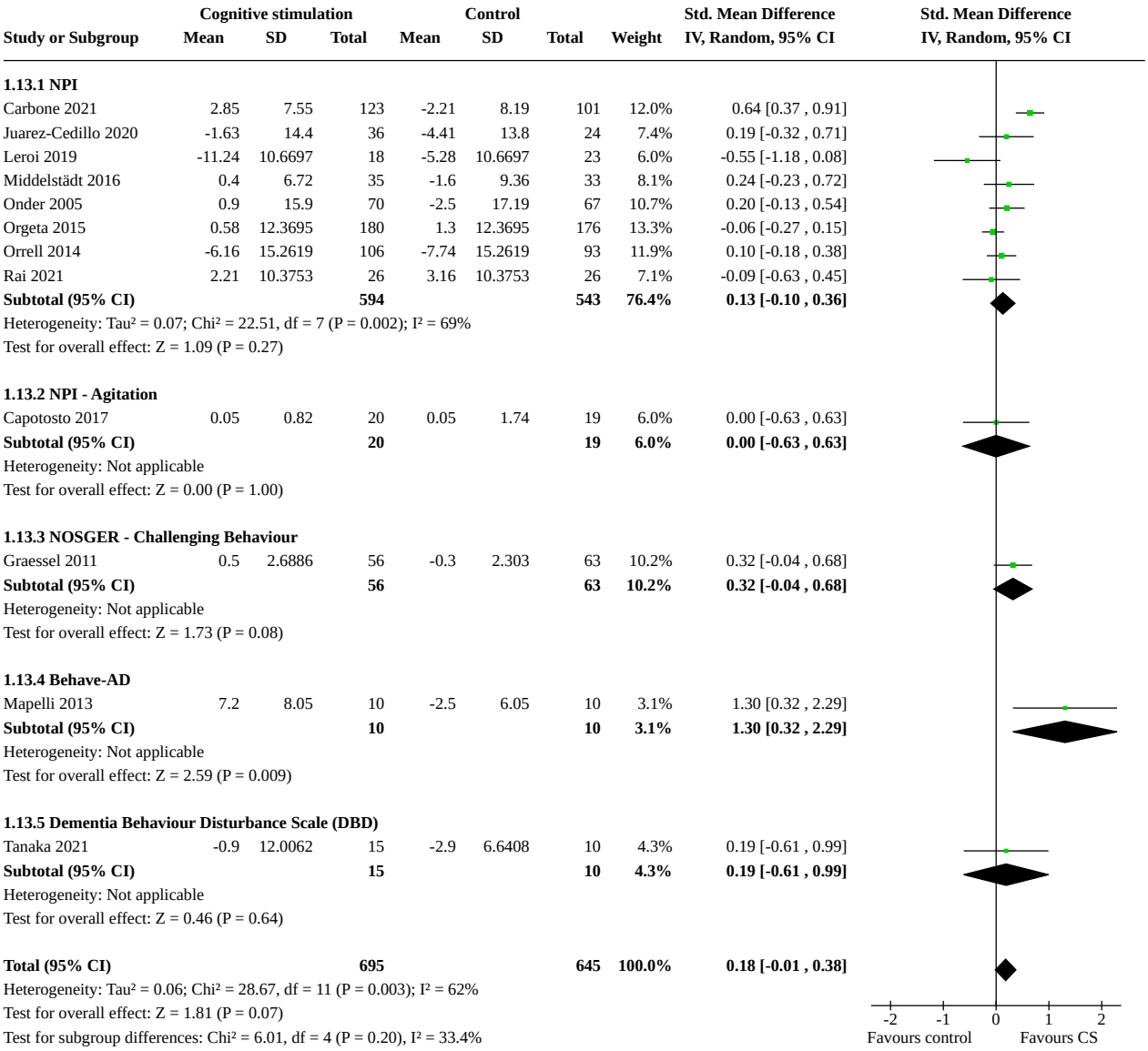


Analysis 1.12. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 12: Instrumental ADL

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.12.1 Lawton Brody Instrumental ADL								
Justo-Henriques 2022	1.09	8.32	22	-1.66	8.15	24	3.5%	0.33 [-0.25 , 0.91]
Maci 2012	0.1	1.03	7	0	0.85	7	1.1%	0.10 [-0.95 , 1.15]
Onder 2005	0	1.67	70	-0.2	1.64	67	10.5%	0.12 [-0.22 , 0.46]
Subtotal (95% CI)			**99**			**98**	**15.1%**	**0.17 [-0.11 , 0.45]**
Heterogeneity: Tau² = 0.00; Chi² = 0.39, df = 2 (P = 0.82); I² = 0%								
Test for overall effect: Z = 1.17 (P = 0.24)								
1.12.2 Disability Assessment for Dementia								
Capotosto 2017	-0.35	27.6	20	-0.63	31	19	3.0%	0.01 [-0.62 , 0.64]
Carbone 2021	-0.77	9.98	84	-4.41	15.57	56	10.2%	0.29 [-0.05 , 0.63]
Subtotal (95% CI)			**104**			**75**	**13.2%**	**0.23 [-0.07 , 0.52]**
Heterogeneity: Tau² = 0.00; Chi² = 0.59, df = 1 (P = 0.44); I² = 0%								
Test for overall effect: Z = 1.48 (P = 0.14)								
1.12.3 NOSGER IADL subscale								
Graessel 2011	0.7	3.174	56	-0.5	2.465	63	8.9%	0.42 [0.06 , 0.79]
Subtotal (95% CI)			**56**			**63**	**8.9%**	**0.42 [0.06 , 0.79]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.28 (P = 0.02)								
1.12.4 Bristol Activities of Daily Living Scale (BADLS)								
Orgeta 2015	-8.55	6.3051	180	-8.64	6.3051	176	27.4%	0.01 [-0.19 , 0.22]
Rai 2021	-0.08	3.7344	26	-1.09	3.7344	26	4.0%	0.27 [-0.28 , 0.81]
Subtotal (95% CI)			**206**			**202**	**31.3%**	**0.05 [-0.15 , 0.24]**
Heterogeneity: Tau² = 0.00; Chi² = 0.72, df = 1 (P = 0.40); I² = 0%								
Test for overall effect: Z = 0.47 (P = 0.64)								
1.12.5 Alzheimer's Disease Cooperative Study - ADL Scale								
Ali 2021	-3.72	7.7445	20	-5.95	7.7445	20	3.0%	0.28 [-0.34 , 0.91]
Marinho 2021	2	21.02	24	-1.8	21.1	24	3.7%	0.18 [-0.39 , 0.74]
Middelstädt 2016	-0.2	21.78	35	0.3	25.04	33	5.2%	-0.02 [-0.50 , 0.45]
Orrell 2014	1.17	10.7013	106	0.23	10.7013	93	15.2%	0.09 [-0.19 , 0.37]
Subtotal (95% CI)			**185**			**170**	**27.1%**	**0.10 [-0.11 , 0.31]**
Heterogeneity: Tau² = 0.00; Chi² = 0.66, df = 3 (P = 0.88); I² = 0%								
Test for overall effect: Z = 0.95 (P = 0.34)								
1.12.6 Rapid Disability Rating Scale								
Juarez-Cedillo 2020	-0.5	5.31	36	-2.66	7.71	24	4.4%	0.33 [-0.19 , 0.85]
Subtotal (95% CI)			**36**			**24**	**4.4%**	**0.33 [-0.19 , 0.85]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.26 (P = 0.21)								
Total (95% CI)			**686**			**632**	**100.0%**	**0.15 [0.04 , 0.26]**
Heterogeneity: Tau² = 0.00; Chi² = 6.56, df = 12 (P = 0.89); I² = 0%								
Test for overall effect: Z = 2.69 (P = 0.007)								
Test for subgroup differences: Chi² = 4.22, df = 5 (P = 0.52), I² = 0%								



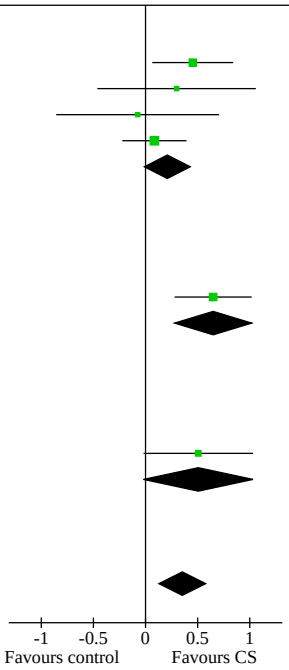
Analysis 1.13. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 13: Behaviour that challenges



Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.13.1 NPI								
Carbone 2021	2.85	7.55	123	-2.21	8.19	101	12.0%	0.64 [0.37 , 0.91]
Juarez-Cedillo 2020	-1.63	14.4	36	-4.41	13.8	24	7.4%	0.19 [-0.32 , 0.71]
Leroi 2019	-11.24	10.6697	18	-5.28	10.6697	23	6.0%	-0.55 [-1.18 , 0.08]
Middelstädt 2016	0.4	6.72	35	-1.6	9.36	33	8.1%	0.24 [-0.23 , 0.72]
Onder 2005	0.9	15.9	70	-2.5	17.19	67	10.7%	0.20 [-0.13 , 0.54]
Orgeta 2015	0.58	12.3695	180	1.3	12.3695	176	13.3%	-0.06 [-0.27 , 0.15]
Orrell 2014	-6.16	15.2619	106	-7.74	15.2619	93	11.9%	0.10 [-0.18 , 0.38]
Rai 2021	2.21	10.3753	26	3.16	10.3753	26	7.1%	-0.09 [-0.63 , 0.45]
Subtotal (95% CI)			**594**			**543**	**76.4%**	**0.13 [-0.10 , 0.36]**
Heterogeneity: Tau² = 0.07; Chi² = 22.51, df = 7 (P = 0.002); I² = 69%								
Test for overall effect: Z = 1.09 (P = 0.27)								
1.13.2 NPI - Agitation								
Capotosto 2017	0.05	0.82	20	0.05	1.74	19	6.0%	0.00 [-0.63 , 0.63]
Subtotal (95% CI)			**20**			**19**	**6.0%**	**0.00 [-0.63 , 0.63]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.00 (P = 1.00)								
1.13.3 NOSGER - Challenging Behaviour								
Graessel 2011	0.5	2.6886	56	-0.3	2.303	63	10.2%	0.32 [-0.04 , 0.68]
Subtotal (95% CI)			**56**			**63**	**10.2%**	**0.32 [-0.04 , 0.68]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.73 (P = 0.08)								
1.13.4 Behave-AD								
Mapelli 2013	7.2	8.05	10	-2.5	6.05	10	3.1%	1.30 [0.32 , 2.29]
Subtotal (95% CI)			**10**			**10**	**3.1%**	**1.30 [0.32 , 2.29]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.59 (P = 0.009)								
1.13.5 Dementia Behaviour Disturbance Scale (DBD)								
Tanaka 2021	-0.9	12.0062	15	-2.9	6.6408	10	4.3%	0.19 [-0.61 , 0.99]
Subtotal (95% CI)			**15**			**10**	**4.3%**	**0.19 [-0.61 , 0.99]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.46 (P = 0.64)								
Total (95% CI)			**695**			**645**	**100.0%**	**0.18 [-0.01 , 0.38]**
Heterogeneity: Tau² = 0.06; Chi² = 28.67, df = 11 (P = 0.003); I² = 62%								
Test for overall effect: Z = 1.81 (P = 0.07)								
Test for subgroup differences: Chi² = 6.01, df = 4 (P = 0.20), I² = 33.4%								

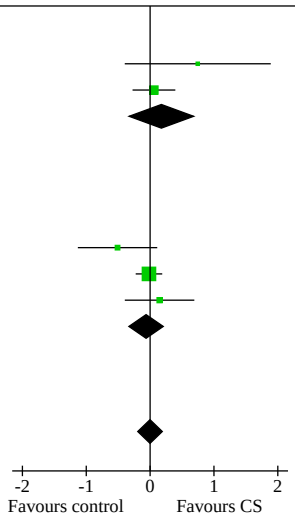
Analysis 1.14. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 14: Behaviour, General Rating Scales

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.14.1 CAPE Behaviour Rating Scale								
Alvares-Pereira 2021	1.473	3.43	55	-0.36	4.58	50	21.2%	0.45 [0.06 , 0.84]
Coen 2011	0	3.6	14	-1.4	5.4	13	7.6%	0.30 [-0.46 , 1.06]
Spector 2001	-1.1	6.08	17	-0.6	7.07	10	7.2%	-0.08 [-0.86 , 0.71]
Spector 2003	-0.2	6.1	97	-0.7	5.5	70	27.7%	0.08 [-0.22 , 0.39]
Subtotal (95% CI)			**183**			**143**	**63.6%**	**0.21 [-0.01 , 0.43]**
Heterogeneity: Tau² = 0.00; Chi² = 2.70, df = 3 (P = 0.44); I² = 0%								
Test for overall effect: Z = 1.86 (P = 0.06)								
1.14.2 NOSGER								
Graessel 2011	5.3	11.5757	56	-1	7.5443	63	22.5%	0.65 [0.28 , 1.02]
Subtotal (95% CI)			**56**			**63**	**22.5%**	**0.65 [0.28 , 1.02]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 3.44 (P = 0.0006)								
1.14.3 Blessed Dementia Rating Scale								
Juarez-Cedillo 2020	-0.52	5.12	36	-3.77	7.84	24	13.9%	0.51 [-0.02 , 1.03]
Subtotal (95% CI)			**36**			**24**	**13.9%**	**0.51 [-0.02 , 1.03]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.89 (P = 0.06)								
Total (95% CI)			**275**			**230**	**100.0%**	**0.35 [0.13 , 0.58]**
Heterogeneity: Tau² = 0.02; Chi² = 7.12, df = 5 (P = 0.21); I² = 30%								
Test for overall effect: Z = 3.08 (P = 0.002)								
Test for subgroup differences: Chi² = 4.42, df = 2 (P = 0.11), I² = 54.7%								



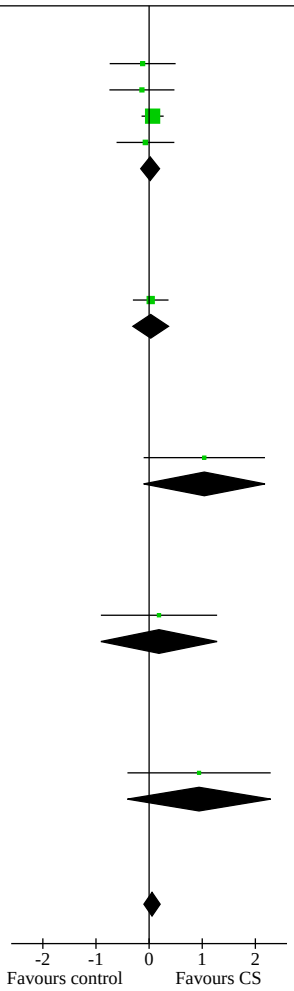
Analysis 1.15. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 15: Caregiver outcome - anxiety

Study or Subgroup	CS Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.15.1 Hamilton Anxiety Scale								
Bottino 2005	4.83	6.33	6	0.14	5.41	7	2.8%	0.75 [-0.40 , 1.89]
Onder 2005	-0.3	3.35	70	-0.5	3.27	67	26.1%	0.06 [-0.27 , 0.40]
Subtotal (95% CI)			**76**			**74**	**28.8%**	**0.18 [-0.33 , 0.68]**
Heterogeneity: Tau² = 0.05; Chi² = 1.27, df = 1 (P = 0.26); I² = 22%								
Test for overall effect: Z = 0.69 (P = 0.49)								
1.15.2 Hospital Anxiety & Depression Scale - Anxiety								
Leroi 2019	-1.57	3.1352	18	0.06	3.1352	24	8.9%	-0.51 [-1.13 , 0.11]
Orgeta 2015	-0.11	3.8986	180	-0.04	3.8986	176	50.9%	-0.02 [-0.23 , 0.19]
Rai 2021	-0.08	2.925	26	-0.52	2.925	26	11.3%	0.15 [-0.40 , 0.69]
Subtotal (95% CI)			**224**			**226**	**71.2%**	**-0.06 [-0.33 , 0.21]**
Heterogeneity: Tau² = 0.02; Chi² = 2.70, df = 2 (P = 0.26); I² = 26%								
Test for overall effect: Z = 0.46 (P = 0.65)								
Total (95% CI)			**300**			**300**	**100.0%**	**-0.00 [-0.19 , 0.19]**
Heterogeneity: Tau² = 0.01; Chi² = 4.66, df = 4 (P = 0.32); I² = 14%								
Test for overall effect: Z = 0.01 (P = 0.99)								
Test for subgroup differences: Chi² = 0.67, df = 1 (P = 0.41), I² = 0%								



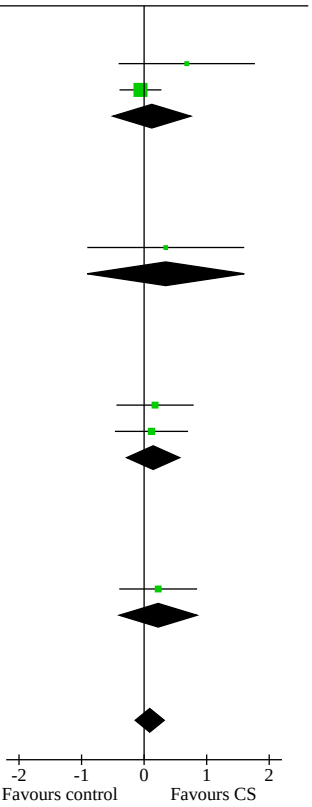
Analysis 1.16. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 16: Caregiver outcome - depressed mood

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.16.1 Hospital Anxiety & Depression Scales - depression								
Ali 2021	0.76	4.7596	20	1.35	4.7596	20	6.1%	-0.12 [-0.74 , 0.50]
Leroi 2019	-0.53	2.8292	18	-0.14	2.8292	24	6.2%	-0.14 [-0.75 , 0.48]
Orgeta 2015	-0.74	3.0322	180	-0.94	3.0322	176	54.0%	0.07 [-0.14 , 0.27]
Rai 2021	0	2.0052	26	0.14	2.0052	26	7.9%	-0.07 [-0.61 , 0.48]
Subtotal (95% CI)			**244**			**246**	**74.2%**	**0.02 [-0.16 , 0.20]**
Heterogeneity: Tau² = 0.00; Chi² = 0.74, df = 3 (P = 0.86); I² = 0%								
Test for overall effect: Z = 0.21 (P = 0.83)								
1.16.2 Hamilton Depression Scale								
Onder 2005	-0.9	3.35	70	-1	3.27	67	20.8%	0.03 [-0.30 , 0.37]
Subtotal (95% CI)			**70**			**67**	**20.8%**	**0.03 [-0.30 , 0.37]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.18 (P = 0.86)								
1.16.3 Beck Depression Inventory								
Maci 2012	6.9	9.26	7	-1.4	5.11	7	1.8%	1.04 [-0.10 , 2.18]
Subtotal (95% CI)			**7**			**7**	**1.8%**	**1.04 [-0.10 , 2.18]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.78 (P = 0.07)								
1.16.4 Montgomery-Asberg Depression Rating Scale								
Bottino 2005	3.83	7.71	6	2.29	7.69	7	1.9%	0.19 [-0.91 , 1.28]
Subtotal (95% CI)			**6**			**7**	**1.9%**	**0.19 [-0.91 , 1.28]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.33 (P = 0.74)								
1.16.5 General Health Questionnaire (GHQ-12)								
Spector 2001	3.8	2.91	5	-0.3	4.75	5	1.3%	0.94 [-0.41 , 2.29]
Subtotal (95% CI)			**5**			**5**	**1.3%**	**0.94 [-0.41 , 2.29]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.37 (P = 0.17)								
Total (95% CI)			**332**			**332**	**100.0%**	**0.05 [-0.10 , 0.21]**
Heterogeneity: Tau² = 0.00; Chi² = 5.48, df = 7 (P = 0.60); I² = 0%								
Test for overall effect: Z = 0.70 (P = 0.48)								
Test for subgroup differences: Chi² = 4.74, df = 4 (P = 0.31), I² = 15.6%								



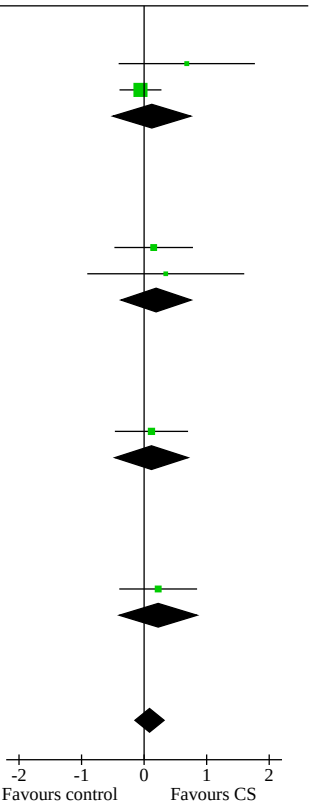
Analysis 1.17. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 17: Caregiver outcome - caregiving stress/burden (2)

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.17.1 Caregiver Burden Inventory								
Maci 2012	8	18.54	7	-2	5.75	7	4.6%	0.68 [-0.41 , 1.77]
Onder 2005	-2	11.7	70	-1.3	12.3	67	48.1%	-0.06 [-0.39 , 0.28]
Subtotal (95% CI)			**77**			**74**	**52.7%**	**0.12 [-0.50 , 0.75]**
Heterogeneity: Tau² = 0.10; Chi² = 1.62, df = 1 (P = 0.20); I² = 38%								
Test for overall effect: Z = 0.39 (P = 0.70)								
1.17.2 Relative's Stress Scale								
Spector 2001	-1	12.8	5	-9	26.6	5	3.4%	0.35 [-0.91 , 1.60]
Subtotal (95% CI)			**5**			**5**	**3.4%**	**0.35 [-0.91 , 1.60]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.54 (P = 0.59)								
1.17.3 Zarit Burden Inventory								
Leroi 2019	2.38	8.748	17	0.82	8.748	25	14.2%	0.17 [-0.44 , 0.79]
Marinho 2021	2	20.4	22	-0.4	19.67	23	15.8%	0.12 [-0.47 , 0.70]
Subtotal (95% CI)			**39**			**48**	**29.9%**	**0.14 [-0.28 , 0.57]**
Heterogeneity: Tau² = 0.00; Chi² = 0.02, df = 1 (P = 0.90); I² = 0%								
Test for overall effect: Z = 0.67 (P = 0.50)								
1.17.4 Caregiver Burden Scale								
Ali 2021	-0.1	1.8716	20	-0.53	1.8716	20	14.0%	0.23 [-0.40 , 0.85]
Subtotal (95% CI)			**20**			**20**	**14.0%**	**0.23 [-0.40 , 0.85]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.71 (P = 0.48)								
Total (95% CI)			**141**			**147**	**100.0%**	**0.09 [-0.14 , 0.32]**
Heterogeneity: Tau² = 0.00; Chi² = 2.31, df = 5 (P = 0.81); I² = 0%								
Test for overall effect: Z = 0.76 (P = 0.45)								
Test for subgroup differences: Chi² = 0.14, df = 3 (P = 0.99), I² = 0%								



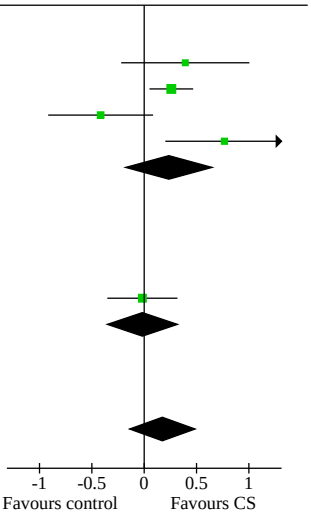
Analysis 1.18. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 18: Caregiver outcome - caregiving stress/burden (1)

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.18.1 Caregiver Burden Inventory								
Maci 2012	8	18.54	7	-2	5.75	7	4.6%	0.68 [-0.41 , 1.77]
Onder 2005	-2	11.7	70	-1.3	12.3	67	48.3%	-0.06 [-0.39 , 0.28]
Subtotal (95% CI)			**77**			**74**	**52.9%**	**0.12 [-0.50 , 0.75]**
Heterogeneity: Tau² = 0.10; Chi² = 1.62, df = 1 (P = 0.20); I² = 38%								
Test for overall effect: Z = 0.39 (P = 0.70)								
1.18.2 Relative's Stress Scale								
Leroi 2019	0.78	6.5259	17	-0.24	6.5259	23	13.8%	0.15 [-0.47 , 0.78]
Spector 2001	-1	12.8	5	-9	26.6	5	3.4%	0.35 [-0.91 , 1.60]
Subtotal (95% CI)			**22**			**28**	**17.2%**	**0.19 [-0.37 , 0.75]**
Heterogeneity: Tau² = 0.00; Chi² = 0.07, df = 1 (P = 0.79); I² = 0%								
Test for overall effect: Z = 0.67 (P = 0.50)								
1.18.3 Zarit Burden Inventory								
Marinho 2021	2	20.4	22	-0.4	19.67	23	15.9%	0.12 [-0.47 , 0.70]
Subtotal (95% CI)			**22**			**23**	**15.9%**	**0.12 [-0.47 , 0.70]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.39 (P = 0.69)								
1.18.4 Caregiver Burden Scale								
Ali 2021	-0.1	1.8716	20	-0.53	1.8716	20	14.0%	0.23 [-0.40 , 0.85]
Subtotal (95% CI)			**20**			**20**	**14.0%**	**0.23 [-0.40 , 0.85]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.71 (P = 0.48)								
Total (95% CI)			**141**			**145**	**100.0%**	**0.09 [-0.15 , 0.32]**
Heterogeneity: Tau² = 0.00; Chi² = 2.27, df = 5 (P = 0.81); I² = 0%								
Test for overall effect: Z = 0.73 (P = 0.47)								
Test for subgroup differences: Chi² = 0.09, df = 3 (P = 0.99), I² = 0%								



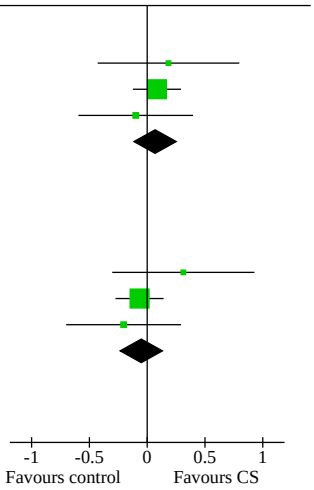
Analysis 1.19. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 19: Caregiver outcome - health-related quality of life

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.19.1 EQ-5D								
Leroi 2019	0.05	0.175	18	-0.02	0.175	25	14.5%	0.39 [-0.22 , 1.00]
Orgeta 2015	-0.03	0.1925	180	-0.08	0.1925	176	28.4%	0.26 [0.05 , 0.47]
Orrell 2014	-0.05	0.1187	29	0	0.1187	34	17.7%	-0.42 [-0.92 , 0.09]
Rai 2021	3.01	9.8787	26	-4.68	9.8787	26	15.8%	0.77 [0.20 , 1.33]
Subtotal (95% CI)			**253**			**261**	**76.3%**	**0.24 [-0.18 , 0.65]**
Heterogeneity: Tau² = 0.12; Chi² = 10.20, df = 3 (P = 0.02); I² = 71%								
Test for overall effect: Z = 1.12 (P = 0.26)								
1.19.2 SF-36								
Onder 2005	-1.3	11.7132	70	-1.1	11.4595	67	23.7%	-0.02 [-0.35 , 0.32]
Subtotal (95% CI)			**70**			**67**	**23.7%**	**-0.02 [-0.35 , 0.32]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.10 (P = 0.92)								
Total (95% CI)			**323**			**328**	**100.0%**	**0.17 [-0.14 , 0.49]**
Heterogeneity: Tau² = 0.08; Chi² = 11.93, df = 4 (P = 0.02); I² = 66%								
Test for overall effect: Z = 1.09 (P = 0.28)								
Test for subgroup differences: Chi² = 0.87, df = 1 (P = 0.35), I² = 0%								



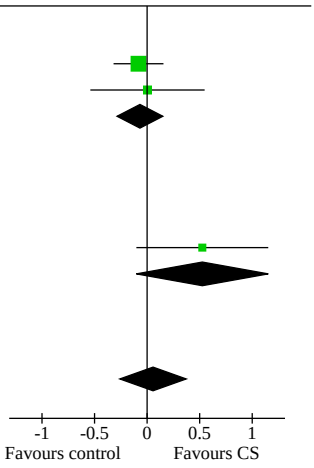
Analysis 1.20. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 20: Caregiver outcome - SF-12

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.20.1 SF-12 PCS								
Leroi 2019	1.79	8.0393	24	0.28	8.0393	18	8.9%	0.18 [-0.43 , 0.80]
Orgeta 2015	-3.03	9.578	180	-3.84	9.578	176	77.5%	0.08 [-0.12 , 0.29]
Orrell 2014	-0.95	5.499	29	-0.4	5.499	34	13.6%	-0.10 [-0.59 , 0.40]
Subtotal (95% CI)			**233**			**228**	**100.0%**	**0.07 [-0.11 , 0.25]**
Heterogeneity: Tau² = 0.00; Chi² = 0.60, df = 2 (P = 0.74); I² = 0%								
Test for overall effect: Z = 0.73 (P = 0.46)								
1.20.2 SF12 - MCS								
Leroi 2019	2.11	10.2597	24	-1.16	10.2597	18	8.9%	0.31 [-0.30 , 0.93]
Orgeta 2015	-1.94	10.2999	180	-1.26	10.2999	176	77.6%	-0.07 [-0.27 , 0.14]
Orrell 2014	-0.45	4.9847	29	0.58	4.9847	34	13.6%	-0.20 [-0.70 , 0.29]
Subtotal (95% CI)			**233**			**228**	**100.0%**	**-0.05 [-0.23 , 0.13]**
Heterogeneity: Tau² = 0.00; Chi² = 1.73, df = 2 (P = 0.42); I² = 0%								
Test for overall effect: Z = 0.55 (P = 0.58)								

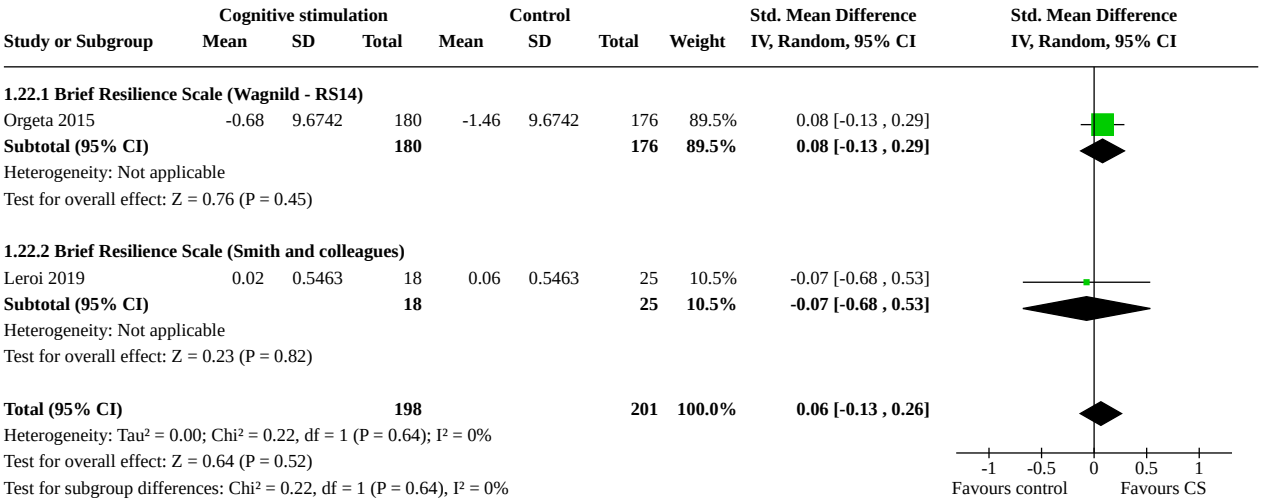


Analysis 1.21. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 21: Caregiver outcome - quality of relationship

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.21.1 QCPR								
Orgeta 2015	-0.51	5.6051	134	-0.05	5.6051	139	56.6%	-0.08 [-0.32 , 0.16]
Rai 2021	0.82	5.4268	26	0.8	5.4268	26	24.0%	0.00 [-0.54 , 0.55]
Subtotal (95% CI)			**160**			**165**	**80.6%**	**-0.07 [-0.29 , 0.15]**
Heterogeneity: Tau² = 0.00; Chi² = 0.08, df = 1 (P = 0.78); I² = 0%								
Test for overall effect: Z = 0.61 (P = 0.54)								
1.21.2 Relationship Satisfaction Scale								
Leroi 2019	2.46	6.4628	17	-1	6.4628	25	19.4%	0.53 [-0.10 , 1.15]
Subtotal (95% CI)			**17**			**25**	**19.4%**	**0.53 [-0.10 , 1.15]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.64 (P = 0.10)								
Total (95% CI)			**177**			**190**	**100.0%**	**0.06 [-0.26 , 0.37]**
Heterogeneity: Tau² = 0.03; Chi² = 3.15, df = 2 (P = 0.21); I² = 36%								
Test for overall effect: Z = 0.35 (P = 0.73)								
Test for subgroup differences: Chi² = 3.07, df = 1 (P = 0.08), I² = 67.4%								



Analysis 1.22. Comparison 1: Cognitive stimulation versus no cognitive stimulation: post-treatment, Outcome 22: Caregiver outcome - resilience



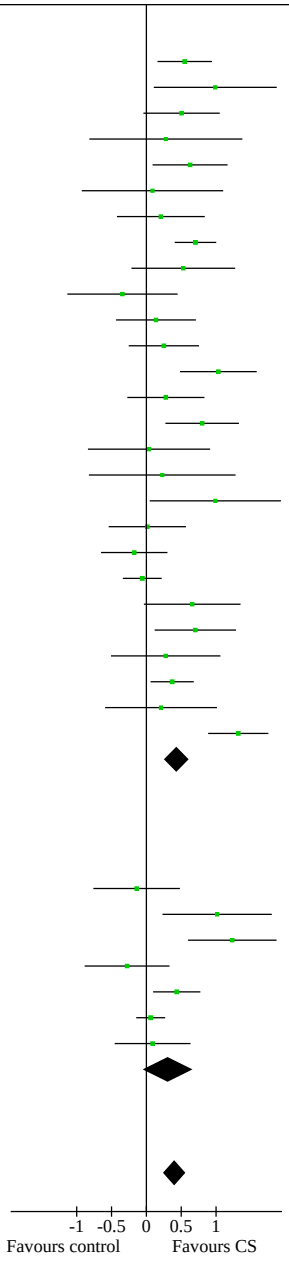
Comparison 2. Cognitive stimulation versus no cognitive stimulation (modality): post-treatment

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 Cognition - modality	34	2340	Std. Mean Difference (IV, Random, 95% CI)	0.40 [0.25, 0.55]
2.1.1 One to twelve months of group CS	27	1637	Std. Mean Difference (IV, Random, 95% CI)	0.43 [0.26, 0.59]
2.1.2 One to twelve months of individual CS	7	703	Std. Mean Difference (IV, Random, 95% CI)	0.30 [-0.03, 0.64]
2.2 MMSE - modality	25	1893	Mean Difference (IV, Random, 95% CI)	1.99 [1.24, 2.74]
2.2.1 One to twelve months of group CS	21	1325	Mean Difference (IV, Random, 95% CI)	2.16 [1.58, 2.74]
2.2.2 One to twelve months of individual CS	4	568	Mean Difference (IV, Random, 95% CI)	1.18 [-0.91, 3.28]
2.3 ADAS-Cog - modality	21	1742	Mean Difference (IV, Random, 95% CI)	2.42 [1.21, 3.63]
2.3.1 One to twelve months of group CS	17	1168	Mean Difference (IV, Random, 95% CI)	2.66 [1.12, 4.20]
2.3.2 One to twelve months of individual CS	4	574	Mean Difference (IV, Random, 95% CI)	1.92 [-0.00, 3.85]
2.4 Quality of Life: self-report	18	1584	Std. Mean Difference (IV, Random, 95% CI)	0.25 [0.07, 0.42]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.4.1 One to twelve months of group CS	13	1058	Std. Mean Difference (IV, Random, 95% CI)	0.28 [0.05, 0.52]
2.4.2 One to twelve months of individual CS	5	526	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.09, 0.30]
2.5 Quality of Life: proxy-rated	11	988	Std. Mean Difference (IV, Random, 95% CI)	0.21 [0.00, 0.42]
2.5.1 One to twelve months of group CS	7	511	Std. Mean Difference (IV, Random, 95% CI)	0.26 [-0.06, 0.58]
2.5.2 One to twelve months of individual CS	4	477	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.11, 0.47]
2.6 Mood: self-reported	10	787	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.08, 0.31]
2.6.1 One to twelve months of group cognitive stimulation	6	299	Std. Mean Difference (IV, Random, 95% CI)	0.20 [-0.06, 0.45]
2.6.2 One to twelve months of individual cognitive stimulation	4	488	Std. Mean Difference (IV, Random, 95% CI)	0.04 [-0.28, 0.35]
2.7 Instrumental ADL	13	1318	Std. Mean Difference (IV, Random, 95% CI)	0.15 [0.04, 0.26]
2.7.1 One to twelve months of group cognitive stimulation	8	687	Std. Mean Difference (IV, Random, 95% CI)	0.20 [0.05, 0.35]
2.7.2 One to twelve months of individual cognitive stimulation	5	631	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.06, 0.25]
2.8 Behaviour that challenges	12	1340	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.01, 0.38]
2.8.1 One to twelve months of group cognitive stimulation	8	754	Std. Mean Difference (IV, Random, 95% CI)	0.33 [0.11, 0.54]
2.8.2 One to twelve months of individual cognitive stimulation	4	586	Std. Mean Difference (IV, Random, 95% CI)	-0.04 [-0.28, 0.19]
2.9 Caregiver outcome - depressed mood	8	664	Std. Mean Difference (IV, Random, 95% CI)	0.05 [-0.10, 0.21]
2.9.1 One to twelve months of group cognitive stimulation	3	37	Std. Mean Difference (IV, Random, 95% CI)	0.68 [0.00, 1.36]
2.9.2 One to twelve months of individual cognitive stimulation	5	627	Std. Mean Difference (IV, Random, 95% CI)	0.02 [-0.14, 0.18]

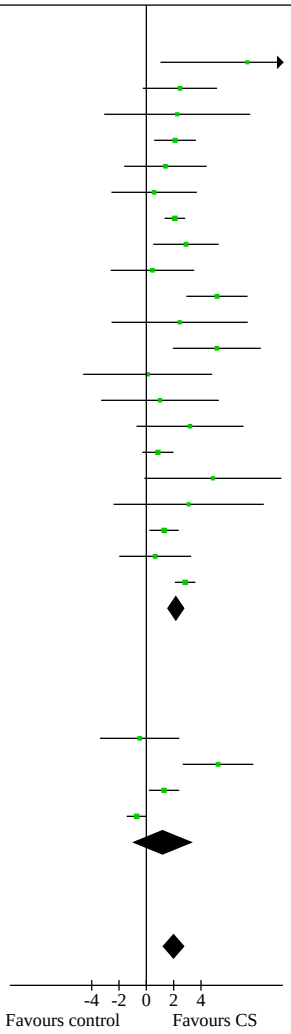
Analysis 2.1. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 1: Cognition - modality

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 One to twelve months of group CS								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	4.0%	0.55 [0.16 , 0.94]
Baldelli 1993	3	5.32	13	-4.4	9.15	10	1.9%	0.99 [0.11 , 1.87]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	3.1%	0.50 [-0.04 , 1.05]
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	1.4%	0.28 [-0.82 , 1.38]
Breuil 1994	5.8	7.3	29	1	7.8	27	3.2%	0.63 [0.09 , 1.17]
Buschert 2011	0.7	8	8	0	6.93	7	1.5%	0.09 [-0.93 , 1.10]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	2.8%	0.21 [-0.42 , 0.84]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	4.5%	0.70 [0.41 , 1.00]
Cheung 2019	0.94	2.6745	18	-0.5	2.5969	12	2.3%	0.53 [-0.21 , 1.27]
Coen 2011	0.2	7.2	13	2.3	4.1	12	2.1%	-0.34 [-1.13 , 0.45]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	3.0%	0.14 [-0.43 , 0.71]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	3.4%	0.25 [-0.25 , 0.75]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	3.1%	1.03 [0.48 , 1.58]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	3.1%	0.28 [-0.27 , 0.83]
Lok 2020	3	6.39	30	-2.16	6.33	30	3.2%	0.80 [0.27 , 1.33]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	1.9%	0.04 [-0.84 , 0.91]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.5%	0.23 [-0.82 , 1.28]
Mapelli 2013	8.7	10.88	10	-2.2	10.21	10	1.7%	0.99 [0.05 , 1.93]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	3.1%	0.01 [-0.54 , 0.57]
Middelstädt 2016	-0.5	14.42	35	2	13.94	33	3.5%	-0.17 [-0.65 , 0.30]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	4.6%	-0.06 [-0.34 , 0.22]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	2.5%	0.66 [-0.04 , 1.35]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	3.0%	0.70 [0.12 , 1.29]
Spector 2001	4.3	17.33	17	-1	20.5	10	2.2%	0.28 [-0.51 , 1.06]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	4.4%	0.37 [0.06 , 0.68]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	2.1%	0.21 [-0.59 , 1.01]
Young 2019	10.65	11.07	51	-1.4	6.41	50	3.8%	1.32 [0.89 , 1.75]
Subtotal (95% CI)			**902**			**735**	**77.0%**	**0.43 [0.26 , 0.59]**
Heterogeneity: Tau² = 0.09; Chi² = 59.26, df = 26 (P = 0.0002); I² = 56%								
Test for overall effect: Z = 5.12 (P < 0.00001)								
2.1.2 One to twelve months of individual CS								
Ali 2021	-2.29	11.0682	20	-0.71	11.0682	20	2.8%	-0.14 [-0.76 , 0.48]
Gibbor 2020b	4.8	4.8235	16	-0.24	4.8235	13	2.2%	1.02 [0.23 , 1.80]
Justo-Henriques 2022	5.18	6.57	22	-1.83	4.51	24	2.7%	1.23 [0.60 , 1.87]
Leroi 2019	-3.29	8.8518	18	-0.79	8.8518	25	2.9%	-0.28 [-0.89 , 0.33]
Onder 2005	0.4	6.69	70	-2.5	6.55	67	4.3%	0.44 [0.10 , 0.77]
Orgeta 2015	-1.52	7.2196	180	-1.97	7.2196	176	5.0%	0.06 [-0.15 , 0.27]
Rai 2021	0.51	5.2981	26	0.03	5.2981	26	3.2%	0.09 [-0.45 , 0.63]
Subtotal (95% CI)			**352**			**351**	**23.0%**	**0.30 [-0.03 , 0.64]**
Heterogeneity: Tau² = 0.13; Chi² = 21.51, df = 6 (P = 0.001); I² = 72%								
Test for overall effect: Z = 1.77 (P = 0.08)								
Total (95% CI)			**1254**			**1086**	**100.0%**	**0.40 [0.25 , 0.55]**
Heterogeneity: Tau² = 0.10; Chi² = 86.32, df = 33 (P < 0.00001); I² = 62%								
Test for overall effect: Z = 5.28 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.43, df = 1 (P = 0.51), I² = 0%								



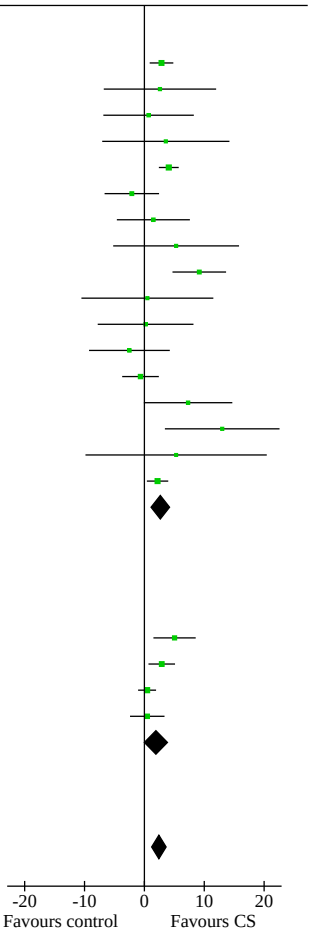
Analysis 2.2. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 2: MMSE - modality

Study or Subgroup	Control Mean	Control SD	Control Total	CS Mean	CS SD	CS Total	Weight	Mean Difference IV, Random, 95% CI
2.2.1 One to twelve months of group CS								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	1.2%	7.40 [1.03 , 13.77]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	3.9%	2.46 [-0.26 , 5.18]
Bottino 2005	0.83	4.53	6	-1.43	5.3	7	1.6%	2.26 [-3.08 , 7.60]
Breuil 1994	1.4	2.7	29	-0.7	3.1	27	5.9%	2.10 [0.57 , 3.63]
Buschert 2011	0.5	3.14	8	-0.9	2.83	7	3.4%	1.40 [-1.62 , 4.42]
Capotosto 2017	0.18	4.57	20	-0.39	5.34	19	3.3%	0.57 [-2.56 , 3.70]
Carbone 2021	0.8894	2.68673	123	-1.2016	3.01666	101	7.3%	2.09 [1.33 , 2.85]
Coen 2011	0.8	3.6	14	-2.1	2.5	11	4.3%	2.90 [0.50 , 5.30]
Cove 2014	-0.33	6.06	24	-0.78	4.54	23	3.4%	0.45 [-2.60 , 3.50]
Juarez-Cedillo 2020	2.05	4.41	36	-3.12	4.29	24	4.6%	5.17 [2.93 , 7.41]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	1.7%	2.44 [-2.55 , 7.43]
Lok 2020	3	6.39	30	-2.16	6.33	30	3.2%	5.16 [1.94 , 8.38]
Lopez 2020	-1.6	4.82	10	-1.7	5.9	10	1.9%	0.10 [-4.62 , 4.82]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	2.2%	1.00 [-3.31 , 5.31]
Mapelli 2013	2.9	5.03	10	-0.3	3.83	10	2.5%	3.20 [-0.72 , 7.12]
Orrell 2014	-1.46	4.0938	106	-2.31	4.0938	93	6.7%	0.85 [-0.29 , 1.99]
Requena 2006	1.5	7.38	20	-3.37	10.71	30	1.7%	4.87 [-0.14 , 9.88]
Spector 2001	3.1	7.04	17	0	7.04	10	1.5%	3.10 [-2.40 , 8.60]
Spector 2003	0.9	3.5	97	-0.4	3.5	70	6.8%	1.30 [0.22 , 2.38]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	4.0%	0.65 [-1.98 , 3.28]
Young 2019	2.1	2.26	51	-0.74	1.52	50	7.3%	2.84 [2.09 , 3.59]
Subtotal (95% CI)			**739**			**586**	**78.3%**	**2.16 [1.58 , 2.74]**
Heterogeneity: Tau² = 0.45; Chi² = 30.24, df = 20 (P = 0.07); I² = 34%								
Test for overall effect: Z = 7.27 (P < 0.00001)								
2.2.2 One to twelve months of individual CS								
Gibbor 2020b	-0.5	3.9626	16	-0.01	3.9626	13	3.6%	-0.49 [-3.39 , 2.41]
Justo-Henriques 2022	3.54	3.84	22	-1.71	5.08	24	4.0%	5.25 [2.66 , 7.84]
Onder 2005	0.2	3.35	70	-1.1	3.27	67	6.7%	1.30 [0.19 , 2.41]
Orgeta 2015	-2.24	3.5616	180	-1.54	3.5616	176	7.3%	-0.70 [-1.44 , 0.04]
Subtotal (95% CI)			**288**			**280**	**21.7%**	**1.18 [-0.91 , 3.28]**
Heterogeneity: Tau² = 3.64; Chi² = 24.28, df = 3 (P < 0.0001); I² = 88%								
Test for overall effect: Z = 1.11 (P = 0.27)								
Total (95% CI)			**1027**			**866**	**100.0%**	**1.99 [1.24 , 2.74]**
Heterogeneity: Tau² = 1.83; Chi² = 85.31, df = 24 (P < 0.00001); I² = 72%								
Test for overall effect: Z = 5.23 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.77, df = 1 (P = 0.38), I² = 0%								



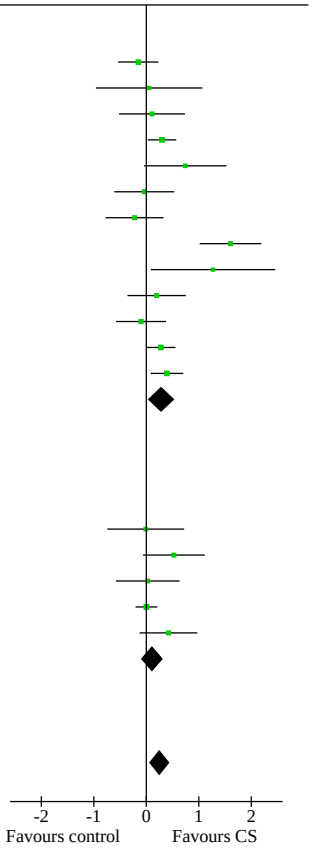
Analysis 2.3. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 3: ADAS-Cog - modality

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
2.3.1 One to twelve months of group CS								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	9.8%	2.86 [0.88 , 4.83]
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	1.5%	2.60 [-6.79 , 11.99]
Buschert 2011	0.7	8	8	0	6.93	7	2.1%	0.70 [-6.86 , 8.26]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	1.2%	3.58 [-7.05 , 14.21]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	10.6%	4.08 [2.42 , 5.73]
Coen 2011	0.2	7.2	13	2.3	4.1	12	4.6%	-2.10 [-6.65 , 2.45]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	3.0%	1.50 [-4.60 , 7.60]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	1.2%	5.30 [-5.21 , 15.81]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	4.7%	9.17 [4.69 , 13.65]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	1.1%	0.50 [-10.52 , 11.52]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	1.9%	0.20 [-7.80 , 8.20]
Middelstädt 2016	-0.5	14.42	35	2	13.95	33	2.6%	-2.50 [-9.24 , 4.24]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	7.2%	-0.65 [-3.71 , 2.41]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	2.2%	7.30 [-0.09 , 14.69]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	1.4%	13.00 [3.43 , 22.57]
Spector 2001	4.3	17.33	17	-1	20.5	10	0.6%	5.30 [-9.84 , 20.44]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	10.3%	2.20 [0.42 , 3.98]
Subtotal (95% CI)			**626**			**542**	**66.0%**	**2.66 [1.12 , 4.20]**
Heterogeneity: Tau² = 3.32; Chi² = 29.38, df = 16 (P = 0.02); I² = 46%								
Test for overall effect: Z = 3.38 (P = 0.0007)								
2.3.2 One to twelve months of individual CS								
Gibbor 2020b	4.8	4.8235	16	-0.24	4.8235	13	6.2%	5.04 [1.51 , 8.57]
Onder 2005	0.4	6.69	70	-2.5	6.55	67	9.2%	2.90 [0.68 , 5.12]
Orgeta 2015	-1.52	7.2196	180	-1.97	7.2196	176	11.0%	0.45 [-1.05 , 1.95]
Rai 2021	0	5.2981	26	-0.48	5.2981	26	7.6%	0.48 [-2.40 , 3.36]
Subtotal (95% CI)			**292**			**282**	**34.0%**	**1.92 [-0.00 , 3.85]**
Heterogeneity: Tau² = 2.28; Chi² = 7.76, df = 3 (P = 0.05); I² = 61%								
Test for overall effect: Z = 1.96 (P = 0.05)								
Total (95% CI)			**918**			**824**	**100.0%**	**2.42 [1.21 , 3.63]**
Heterogeneity: Tau² = 2.85; Chi² = 40.53, df = 20 (P = 0.004); I² = 51%								
Test for overall effect: Z = 3.93 (P < 0.0001)								
Test for subgroup differences: Chi² = 0.34, df = 1 (P = 0.56), I² = 0%								



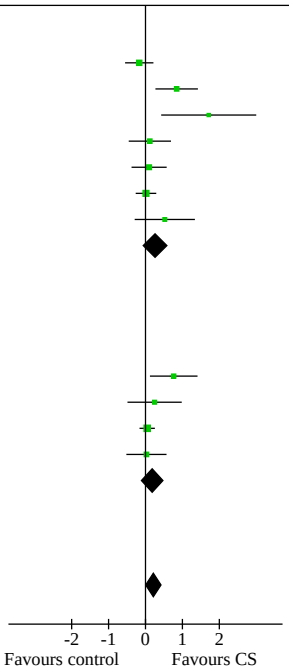
Analysis 2.4. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 4: Quality of Life: self-report

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.4.1 One to twelve months of group CS								
Alvares-Pereira 2021	-0.53	4.63	55	0.27	5.74	50	7.2%	-0.15 [-0.54 , 0.23]
Buschert 2011	-0.4	10.61	8	-0.9	5.52	7	2.4%	0.05 [-0.96 , 1.07]
Capotosto 2017	1.25	11.5	20	0.05	9.91	19	4.6%	0.11 [-0.52 , 0.74]
Carbone 2021	1.6	6.63	118	-0.2	5.12	96	8.7%	0.30 [0.03 , 0.57]
Coen 2011	3.6	3.7	14	0.5	4.4	13	3.5%	0.74 [-0.04 , 1.53]
Cove 2014	0.23	7.97	24	0.54	7.74	23	5.1%	-0.04 [-0.61 , 0.53]
Kim 2016	-0.4	0.76	32	-0.23	0.73	21	5.3%	-0.22 [-0.78 , 0.33]
Lok 2020	11.86	9.64	30	-2.16	7.48	30	5.0%	1.60 [1.02 , 2.19]
Maci 2012	12.3	11.78	7	-1.3	7.86	7	1.9%	1.27 [0.09 , 2.46]
Marinho 2021	2	8.77	24	0.4	7.21	26	5.3%	0.20 [-0.36 , 0.75]
Middelstädt 2016	-0.2	5.47	35	0.4	6.36	33	6.1%	-0.10 [-0.58 , 0.38]
Orrell 2014	-0.67	6.428	106	-2.45	6.428	93	8.5%	0.28 [-0.00 , 0.56]
Spector 2003	1.3	5.1	97	-0.8	5.6	70	8.1%	0.39 [0.08 , 0.70]
Subtotal (95% CI)			**570**			**488**	**71.7%**	**0.28 [0.05 , 0.52]**
Heterogeneity: Tau² = 0.11; Chi² = 36.28, df = 12 (P = 0.0003); I² = 67%								
Test for overall effect: Z = 2.34 (P = 0.02)								
2.4.2 One to twelve months of individual CS								
Gibbor 2020b	2.09	5.3427	16	2.15	5.3427	13	3.8%	-0.01 [-0.74 , 0.72]
Justo-Henriques 2022	1.64	7.14	22	-2.5	8.28	24	5.0%	0.52 [-0.06 , 1.11]
Leroi 2019	0.0075	0.2616	18	0	0.2616	25	4.8%	0.03 [-0.58 , 0.63]
Orgeta 2015	-0.25	9.4817	180	-0.28	9.4817	176	9.4%	0.00 [-0.20 , 0.21]
Rai 2021	0.25	3.9551	26	-1.45	3.9551	26	5.3%	0.42 [-0.13 , 0.97]
Subtotal (95% CI)			**262**			**264**	**28.3%**	**0.11 [-0.09 , 0.30]**
Heterogeneity: Tau² = 0.00; Chi² = 4.28, df = 4 (P = 0.37); I² = 7%								
Test for overall effect: Z = 1.08 (P = 0.28)								
Total (95% CI)			**832**			**752**	**100.0%**	**0.25 [0.07 , 0.42]**
Heterogeneity: Tau² = 0.08; Chi² = 43.00, df = 17 (P = 0.0005); I² = 60%								
Test for overall effect: Z = 2.71 (P = 0.007)								
Test for subgroup differences: Chi² = 1.27, df = 1 (P = 0.26), I² = 21.5%								



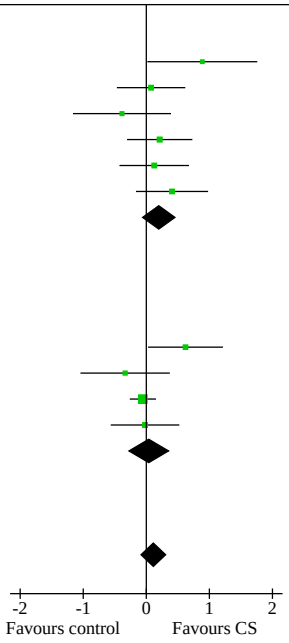
Analysis 2.5. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 5: Quality of Life: proxy-rated

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.5.1 One to twelve months of group CS								
Alvares-Pereira 2021	-0.53	5.83	55	0.2	1.096	50	12.3%	-0.17 [-0.55 , 0.21]
Kim 2016	0.81	2.34	32	-1.09	2.01	21	8.1%	0.84 [0.27 , 1.42]
Maci 2012	10.2	8.08	7	-1.4	3.89	7	2.3%	1.71 [0.43 , 3.00]
Marinho 2021	0.8	7.64	23	-0.2	8.91	24	8.1%	0.12 [-0.45 , 0.69]
Middelstädt 2016	-0.4	6.16	35	-1	5.94	33	10.0%	0.10 [-0.38 , 0.57]
Orrell 2014	0.62	5.2429	106	0.55	5.2429	93	15.3%	0.01 [-0.27 , 0.29]
Tanaka 2021	1.5	6.9714	15	-1.8	4.4272	10	5.0%	0.52 [-0.29 , 1.34]
Subtotal (95% CI)			**273**			**238**	**61.1%**	**0.26 [-0.06 , 0.58]**
Heterogeneity: Tau² = 0.10; Chi² = 15.65, df = 6 (P = 0.02); I² = 62%								
Test for overall effect: Z = 1.59 (P = 0.11)								
2.5.2 One to twelve months of individual CS								
Ali 2021	0.26	3.9852	20	-2.85	3.9852	20	7.0%	0.76 [0.12 , 1.41]
Gibbor 2020b	0.68	6.5725	16	-0.99	6.5725	13	5.8%	0.25 [-0.49 , 0.98]
Orgeta 2015	-1.71	5.6313	180	-1.97	5.6313	176	17.5%	0.05 [-0.16 , 0.25]
Rai 2021	0.04	4.5622	26	-0.08	4.5622	26	8.6%	0.03 [-0.52 , 0.57]
Subtotal (95% CI)			**242**			**235**	**38.9%**	**0.18 [-0.11 , 0.47]**
Heterogeneity: Tau² = 0.03; Chi² = 4.56, df = 3 (P = 0.21); I² = 34%								
Test for overall effect: Z = 1.22 (P = 0.22)								
Total (95% CI)			**515**			**473**	**100.0%**	**0.21 [0.00 , 0.42]**
Heterogeneity: Tau² = 0.05; Chi² = 20.22, df = 10 (P = 0.03); I² = 51%								
Test for overall effect: Z = 2.00 (P = 0.05)								
Test for subgroup differences: Chi² = 0.13, df = 1 (P = 0.72), I² = 0%								



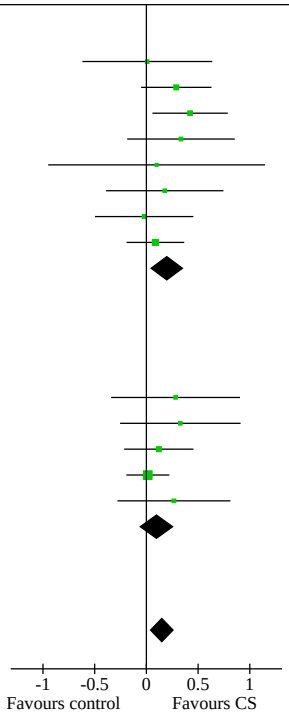
Analysis 2.6. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 6: Mood: self-reported

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.6.1 One to twelve months of group cognitive stimulation								
Baldelli 1993	2.1	4.61	13	-2.3	4.99	10	4.4%	0.89 [0.02 , 1.76]
Baldelli 2002	3.21	7.98	71	2.57	10	16	9.7%	0.08 [-0.47 , 0.62]
Coen 2011	-0.9	3	13	0.1	1.9	13	5.4%	-0.39 [-1.16 , 0.39]
Juarez-Cedillo 2020	5.08	16.8	36	1.45	17	24	10.4%	0.21 [-0.31 , 0.73]
Kim 2016	1.44	10.9	32	0.11	9.62	21	9.5%	0.13 [-0.43 , 0.68]
Requena 2006	5.6	7.87	20	2.03	9.07	30	9.0%	0.41 [-0.16 , 0.98]
Subtotal (95% CI)			**185**			**114**	**48.4%**	**0.20 [-0.06 , 0.45]**
Heterogeneity: Tau² = 0.01; Chi² = 5.36, df = 5 (P = 0.37); I² = 7%								
Test for overall effect: Z = 1.51 (P = 0.13)								
2.6.2 One to twelve months of individual cognitive stimulation								
Justo-Henriques 2022	1.37	4.62	22	-1.5	4.46	24	8.5%	0.62 [0.03 , 1.22]
Leroi 2019	0.73	2.7652	12	1.68	2.7652	22	6.4%	-0.34 [-1.04 , 0.37]
Orgeta 2015	-0.3	2.214	180	-0.18	2.214	176	27.1%	-0.05 [-0.26 , 0.15]
Rai 2021	-0.02	3.4033	26	0.05	3.4033	26	9.7%	-0.02 [-0.56 , 0.52]
Subtotal (95% CI)			**240**			**248**	**51.6%**	**0.04 [-0.28 , 0.35]**
Heterogeneity: Tau² = 0.04; Chi² = 5.34, df = 3 (P = 0.15); I² = 44%								
Test for overall effect: Z = 0.24 (P = 0.81)								
Total (95% CI)			**425**			**362**	**100.0%**	**0.11 [-0.08 , 0.31]**
Heterogeneity: Tau² = 0.03; Chi² = 12.42, df = 9 (P = 0.19); I² = 28%								
Test for overall effect: Z = 1.12 (P = 0.26)								
Test for subgroup differences: Chi² = 0.59, df = 1 (P = 0.44), I² = 0%								



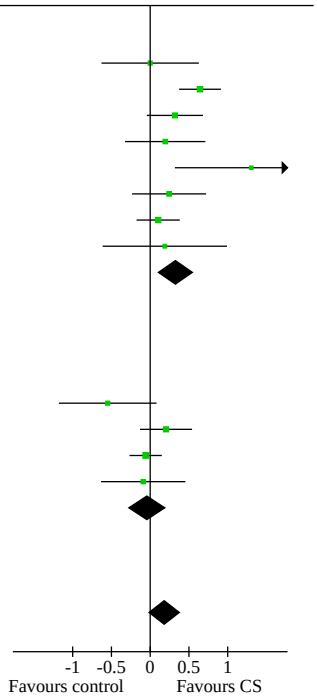
Analysis 2.7. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 7: Instrumental ADL

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.7.1 One to twelve months of group cognitive stimulation								
Capotosto 2017	-0.35	27.6	20	-0.63	31	19	3.0%	0.01 [-0.62 , 0.64]
Carbone 2021	-0.77	9.98	84	-4.41	15.57	56	10.2%	0.29 [-0.05 , 0.63]
Graessel 2011	0.7	3.174	56	-0.5	2.465	63	8.9%	0.42 [0.06 , 0.79]
Juarez-Cedillo 2020	-0.5	5.31	36	-2.66	7.71	24	4.4%	0.33 [-0.19 , 0.85]
Maci 2012	0.1	1.03	7	0	0.85	7	1.1%	0.10 [-0.95 , 1.15]
Marinho 2021	2	21.02	24	-1.8	21.1	24	3.7%	0.18 [-0.39 , 0.74]
Middelstädt 2016	-0.2	21.78	35	0.3	25.04	33	5.2%	-0.02 [-0.50 , 0.45]
Orrell 2014	1.17	10.7013	106	0.23	10.7013	93	15.2%	0.09 [-0.19 , 0.37]
Subtotal (95% CI)			**368**			**319**	**51.7%**	**0.20 [0.05 , 0.35]**
Heterogeneity: Tau² = 0.00; Chi² = 3.81, df = 7 (P = 0.80); I² = 0%								
Test for overall effect: Z = 2.56 (P = 0.01)								
2.7.2 One to twelve months of individual cognitive stimulation								
Ali 2021	-3.72	7.7445	20	-5.95	7.7445	20	3.0%	0.28 [-0.34 , 0.91]
Justo-Henriques 2022	1.09	8.32	22	-1.66	8.15	24	3.5%	0.33 [-0.25 , 0.91]
Onder 2005	0	1.67	70	-0.2	1.64	67	10.5%	0.12 [-0.22 , 0.46]
Orgeta 2015	-8.55	6.3051	180	-8.64	6.3051	176	27.4%	0.01 [-0.19 , 0.22]
Rai 2021	-0.08	3.7344	26	-1.09	3.7344	26	4.0%	0.27 [-0.28 , 0.81]
Subtotal (95% CI)			**318**			**313**	**48.3%**	**0.10 [-0.06 , 0.25]**
Heterogeneity: Tau² = 0.00; Chi² = 1.94, df = 4 (P = 0.75); I² = 0%								
Test for overall effect: Z = 1.22 (P = 0.22)								
Total (95% CI)			**686**			**632**	**100.0%**	**0.15 [0.04 , 0.26]**
Heterogeneity: Tau² = 0.00; Chi² = 6.56, df = 12 (P = 0.89); I² = 0%								
Test for overall effect: Z = 2.69 (P = 0.007)								
Test for subgroup differences: Chi² = 0.81, df = 1 (P = 0.37), I² = 0%								



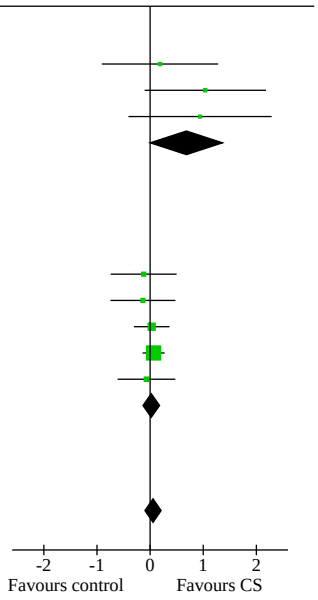
Analysis 2.8. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 8: Behaviour that challenges

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.8.1 One to twelve months of group cognitive stimulation								
Capotosto 2017	0.05	0.82	20	0.05	1.74	19	6.0%	0.00 [-0.63 , 0.63]
Carbone 2021	2.85	7.55	123	-2.21	8.19	101	12.0%	0.64 [0.37 , 0.91]
Graessel 2011	0.5	2.6886	56	-0.3	2.303	63	10.2%	0.32 [-0.04 , 0.68]
Juarez-Cedillo 2020	-1.63	14.4	36	-4.41	13.8	24	7.4%	0.19 [-0.32 , 0.71]
Mapelli 2013	7.2	8.05	10	-2.5	6.05	10	3.1%	1.30 [0.32 , 2.29]
Middelstädt 2016	0.4	6.72	35	-1.6	9.36	33	8.1%	0.24 [-0.23 , 0.72]
Orrell 2014	-6.16	15.2619	106	-7.74	15.2619	93	11.9%	0.10 [-0.18 , 0.38]
Tanaka 2021	-0.9	12.0062	15	-2.9	6.6408	10	4.3%	0.19 [-0.61 , 0.99]
Subtotal (95% CI)			**401**			**353**	**63.0%**	**0.33 [0.11 , 0.54]**
Heterogeneity: Tau² = 0.04; Chi² = 13.03, df = 7 (P = 0.07); I² = 46%								
Test for overall effect: Z = 2.91 (P = 0.004)								
2.8.2 One to twelve months of individual cognitive stimulation								
Leroi 2019	-11.24	10.6697	18	-5.28	10.6697	23	6.0%	-0.55 [-1.18 , 0.08]
Onder 2005	0.9	15.9	70	-2.5	17.19	67	10.7%	0.20 [-0.13 , 0.54]
Orgeta 2015	0.58	12.3695	180	1.3	12.3695	176	13.3%	-0.06 [-0.27 , 0.15]
Rai 2021	0	10.3753	26	0.95	10.3753	26	7.1%	-0.09 [-0.63 , 0.45]
Subtotal (95% CI)			**294**			**292**	**37.0%**	**-0.04 [-0.28 , 0.19]**
Heterogeneity: Tau² = 0.02; Chi² = 4.59, df = 3 (P = 0.20); I² = 35%								
Test for overall effect: Z = 0.37 (P = 0.71)								
Total (95% CI)			**695**			**645**	**100.0%**	**0.18 [-0.01 , 0.38]**
Heterogeneity: Tau² = 0.06; Chi² = 28.67, df = 11 (P = 0.003); I² = 62%								
Test for overall effect: Z = 1.81 (P = 0.07)								
Test for subgroup differences: Chi² = 5.14, df = 1 (P = 0.02), I² = 80.5%								



Analysis 2.9. Comparison 2: Cognitive stimulation versus no cognitive stimulation (modality): post-treatment, Outcome 9: Caregiver outcome - depressed mood

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.9.1 One to twelve months of group cognitive stimulation								
Bottino 2005	3.83	7.71	6	2.29	7.69	7	1.9%	0.19 [-0.91 , 1.28]
Maci 2012	6.9	9.26	7	-1.4	5.11	7	1.8%	1.04 [-0.10 , 2.18]
Spector 2001	3.8	2.91	5	-0.3	4.75	5	1.3%	0.94 [-0.41 , 2.29]
Subtotal (95% CI)			**18**			**19**	**5.0%**	**0.68 [0.00 , 1.36]**
Heterogeneity: Tau² = 0.00; Chi² = 1.31, df = 2 (P = 0.52); I² = 0%								
Test for overall effect: Z = 1.96 (P = 0.05)								
2.9.2 One to twelve months of individual cognitive stimulation								
Ali 2021	0.76	4.7596	20	1.35	4.7596	20	6.1%	-0.12 [-0.74 , 0.50]
Leroi 2019	-0.53	2.8292	18	-0.14	2.8292	24	6.2%	-0.14 [-0.75 , 0.48]
Onder 2005	-0.9	3.35	70	-1	3.27	67	20.8%	0.03 [-0.30 , 0.37]
Orgeta 2015	-0.74	3.0322	180	-0.94	3.0322	176	54.0%	0.07 [-0.14 , 0.27]
Rai 2021	0	2.0052	26	0.14	2.0052	26	7.9%	-0.07 [-0.61 , 0.48]
Subtotal (95% CI)			**314**			**313**	**95.0%**	**0.02 [-0.14 , 0.18]**
Heterogeneity: Tau² = 0.00; Chi² = 0.74, df = 4 (P = 0.95); I² = 0%								
Test for overall effect: Z = 0.27 (P = 0.79)								
Total (95% CI)			**332**			**332**	**100.0%**	**0.05 [-0.10 , 0.21]**
Heterogeneity: Tau² = 0.00; Chi² = 5.48, df = 7 (P = 0.60); I² = 0%								
Test for overall effect: Z = 0.70 (P = 0.48)								
Test for subgroup differences: Chi² = 3.43, df = 1 (P = 0.06), I² = 70.9%								

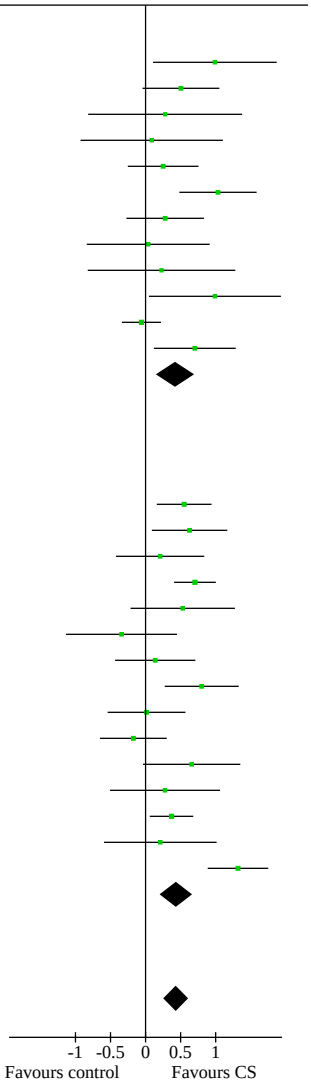


Comparison 3. Group cognitive stimulation versus no cognitive stimulation (number of sessions): post-treatment

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 Cognition	27	1637	Std. Mean Difference (IV, Random, 95% CI)	0.43 [0.26, 0.59]
3.1.1 20 or more group sessions of CS during one to twelve months	12	615	Std. Mean Difference (IV, Random, 95% CI)	0.42 [0.16, 0.67]
3.1.2 Less than 20 group sessions of CS during one to twelve months	15	1022	Std. Mean Difference (IV, Random, 95% CI)	0.43 [0.22, 0.65]
3.2 ADAS-Cog	17	1168	Mean Difference (IV, Random, 95% CI)	2.66 [1.12, 4.20]
3.2.1 20 or more group sessions of CS during one to twelve months	7	418	Mean Difference (IV, Random, 95% CI)	4.18 [-0.28, 8.64]
3.2.2 Less than 20 group sessions of CS during one to twelve months	10	750	Mean Difference (IV, Random, 95% CI)	2.52 [1.22, 3.83]
3.3 MMSE	21	1325	Mean Difference (IV, Random, 95% CI)	2.16 [1.58, 2.74]
3.3.1 20 or more group sessions of CS during one to twelve months	11	554	Mean Difference (IV, Random, 95% CI)	2.51 [1.20, 3.83]
3.3.2 Less than 20 group sessions of CS during one to twelve months	10	771	Mean Difference (IV, Random, 95% CI)	2.11 [1.51, 2.72]
3.4 Quality of Life: self-report	13	1058	Mean Difference (IV, Random, 95% CI)	1.97 [0.47, 3.47]
3.4.1 20 or more group sessions of CS during one to twelve months	4	281	Mean Difference (IV, Random, 95% CI)	1.33 [-1.14, 3.79]
3.4.2 Less than 20 group sessions of CS during one to twelve months	9	777	Mean Difference (IV, Random, 95% CI)	2.24 [0.13, 4.36]

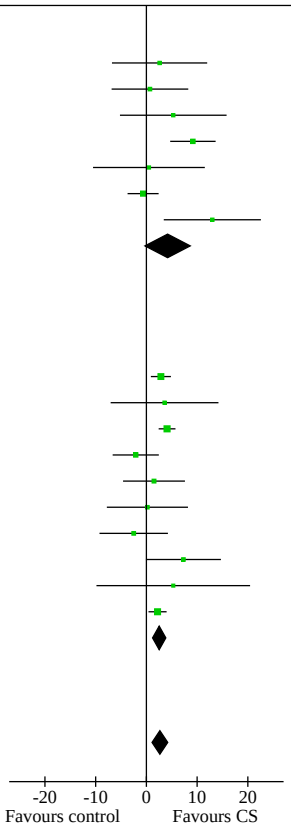
Analysis 3.1. Comparison 3: Group cognitive stimulation versus no cognitive stimulation (number of sessions): post-treatment, Outcome 1: Cognition

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.1.1 20 or more group sessions of CS during one to twelve months								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	2.4%	0.99 [0.11 , 1.87]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	4.1%	0.50 [-0.04 , 1.05]
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	1.7%	0.28 [-0.82 , 1.38]
Buschert 2011	0.7	8	8	0	6.93	7	1.9%	0.09 [-0.93 , 1.10]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	4.4%	0.25 [-0.25 , 0.75]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	4.1%	1.03 [0.48 , 1.58]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	4.0%	0.28 [-0.27 , 0.83]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	2.4%	0.04 [-0.84 , 0.91]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.8%	0.23 [-0.82 , 1.28]
Mapelli 2013	8.7	10.88	10	-2.2	10.21	10	2.2%	0.99 [0.05 , 1.93]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	6.2%	-0.06 [-0.34 , 0.22]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	3.8%	0.70 [0.12 , 1.29]
Subtotal (95% CI)			**350**			**265**	**39.0%**	**0.42 [0.16 , 0.67]**
Heterogeneity: Tau² = 0.08; Chi² = 20.51, df = 11 (P = 0.04); I² = 46%								
Test for overall effect: Z = 3.22 (P = 0.001)								
3.1.2 Less than 20 group sessions of CS during one to twelve months								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	5.3%	0.55 [0.16 , 0.94]
Breuil 1994	5.8	7.3	29	1	7.8	27	4.2%	0.63 [0.09 , 1.17]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	3.6%	0.21 [-0.42 , 0.84]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	6.0%	0.70 [0.41 , 1.00]
Cheung 2019	0.94	2.6745	18	-0.5	2.5969	12	3.0%	0.53 [-0.21 , 1.27]
Coen 2011	0.2	7.2	13	2.3	4.1	12	2.7%	-0.34 [-1.13 , 0.45]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	3.9%	0.14 [-0.43 , 0.71]
Lok 2020	3	6.39	30	-2.16	6.33	30	4.2%	0.80 [0.27 , 1.33]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	4.0%	0.01 [-0.54 , 0.57]
Middelstädt 2016	-0.5	14.42	35	2	13.94	33	4.6%	-0.17 [-0.65 , 0.30]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	3.2%	0.66 [-0.04 , 1.35]
Spector 2001	4.3	17.33	17	-1	20.5	10	2.8%	0.28 [-0.51 , 1.06]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	5.9%	0.37 [0.06 , 0.68]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	2.7%	0.21 [-0.59 , 1.01]
Young 2019	10.65	11.07	51	-1.4	6.41	50	4.9%	1.32 [0.89 , 1.75]
Subtotal (95% CI)			**552**			**470**	**61.0%**	**0.43 [0.22 , 0.65]**
Heterogeneity: Tau² = 0.10; Chi² = 36.13, df = 14 (P = 0.0010); I² = 61%								
Test for overall effect: Z = 3.92 (P < 0.0001)								
Total (95% CI)			**902**			**735**	**100.0%**	**0.43 [0.26 , 0.59]**
Heterogeneity: Tau² = 0.09; Chi² = 59.26, df = 26 (P = 0.0002); I² = 56%								
Test for overall effect: Z = 5.12 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.01, df = 1 (P = 0.94), I² = 0%								



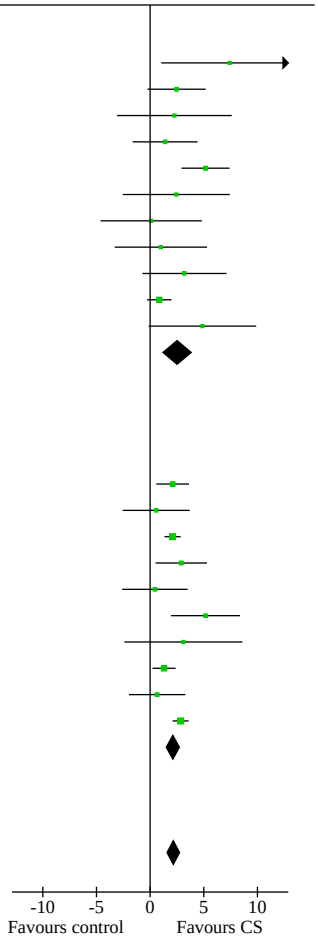
Analysis 3.2. Comparison 3: Group cognitive stimulation versus no cognitive stimulation (number of sessions): post-treatment, Outcome 2: ADAS-Cog

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
3.2.1 20 or more group sessions of CS during one to twelve months								
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	2.4%	2.60 [-6.79 , 11.99]
Buschert 2011	0.7	8	8	0	6.93	7	3.4%	0.70 [-6.86 , 8.26]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	1.9%	5.30 [-5.21 , 15.81]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	7.3%	9.17 [4.69 , 13.65]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	1.8%	0.50 [-10.52 , 11.52]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	10.8%	-0.65 [-3.71 , 2.41]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	2.3%	13.00 [3.43 , 22.57]
Subtotal (95% CI)			**217**			**201**	**29.8%**	**4.18 [-0.28 , 8.64]**
Heterogeneity: Tau² = 20.92; Chi² = 17.69, df = 6 (P = 0.007); I² = 66%								
Test for overall effect: Z = 1.84 (P = 0.07)								
3.2.2 Less than 20 group sessions of CS during one to twelve months								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	14.3%	2.86 [0.88 , 4.83]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	1.9%	3.58 [-7.05 , 14.21]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	15.4%	4.08 [2.42 , 5.73]
Coen 2011	0.2	7.2	13	2.3	4.1	12	7.1%	-2.10 [-6.65 , 2.45]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	4.8%	1.50 [-4.60 , 7.60]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	3.1%	0.20 [-7.80 , 8.20]
Middelstädt 2016	-0.5	14.42	35	2	13.95	33	4.1%	-2.50 [-9.24 , 4.24]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	3.5%	7.30 [-0.09 , 14.69]
Spector 2001	4.3	17.33	17	-1	20.5	10	1.0%	5.30 [-9.84 , 20.44]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	15.0%	2.20 [0.42 , 3.98]
Subtotal (95% CI)			**409**			**341**	**70.2%**	**2.52 [1.22 , 3.83]**
Heterogeneity: Tau² = 0.87; Chi² = 11.67, df = 9 (P = 0.23); I² = 23%								
Test for overall effect: Z = 3.79 (P = 0.0001)								
Total (95% CI)			**626**			**542**	**100.0%**	**2.66 [1.12 , 4.20]**
Heterogeneity: Tau² = 3.32; Chi² = 29.38, df = 16 (P = 0.02); I² = 46%								
Test for overall effect: Z = 3.38 (P = 0.0007)								
Test for subgroup differences: Chi² = 0.49, df = 1 (P = 0.48), I² = 0%								

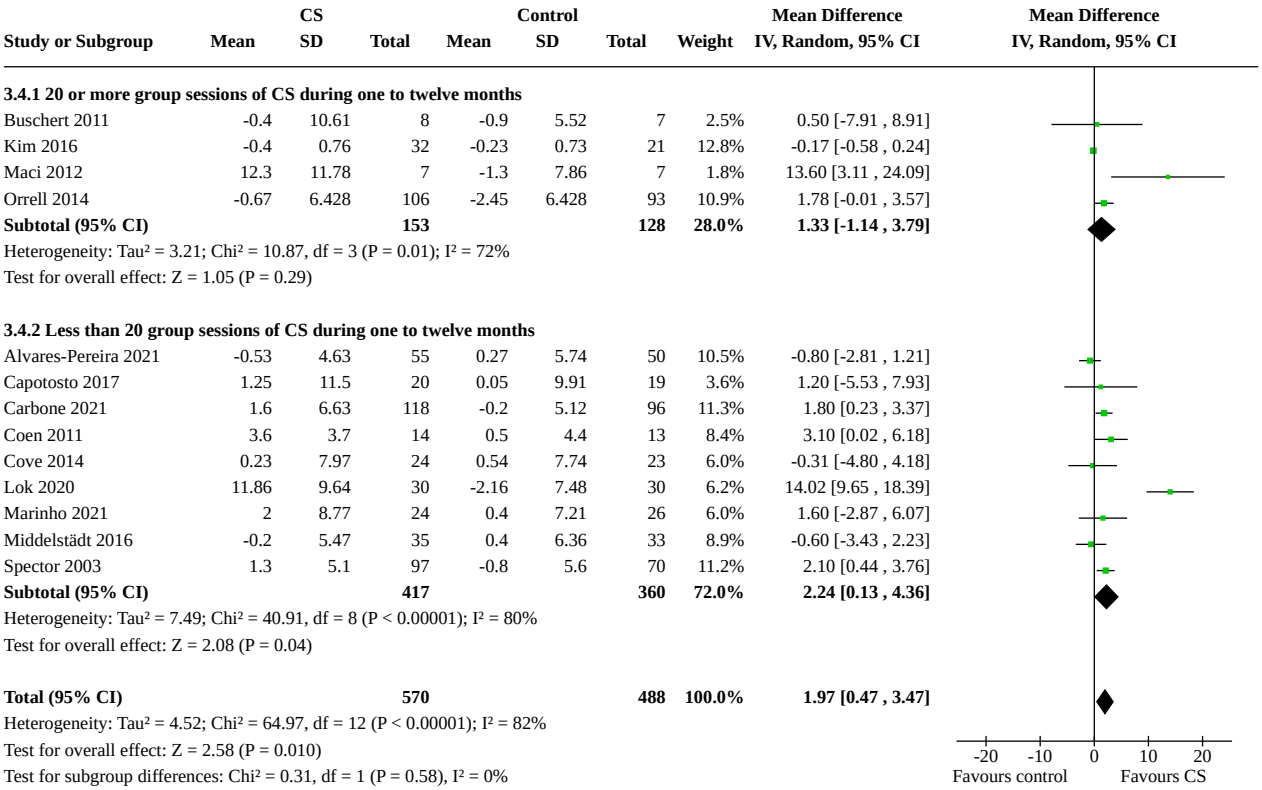


Analysis 3.3. Comparison 3: Group cognitive stimulation versus no cognitive stimulation (number of sessions): post-treatment, Outcome 3: MMSE

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
3.3.1 20 or more group sessions of CS during one to twelve months								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	0.8%	7.40 [1.03 , 13.77]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	3.7%	2.46 [-0.26 , 5.18]
Bottino 2005	0.83	4.53	6	-1.43	5.3	7	1.1%	2.26 [-3.08 , 7.60]
Buschert 2011	0.5	3.14	8	-0.9	2.83	7	3.1%	1.40 [-1.62 , 4.42]
Juarez-Cedillo 2020	2.05	4.41	36	-3.12	4.29	24	5.0%	5.17 [2.93 , 7.41]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	1.3%	2.44 [-2.55 , 7.43]
Lopez 2020	-1.6	4.82	10	-1.7	5.9	10	1.4%	0.10 [-4.62 , 4.82]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.7%	1.00 [-3.31 , 5.31]
Mapelli 2013	2.9	5.03	10	-0.3	3.83	10	2.0%	3.20 [-0.72 , 7.12]
Orrell 2014	-1.46	4.0938	106	-2.31	4.0938	93	11.1%	0.85 [-0.29 , 1.99]
Requena 2006	1.5	7.38	20	-3.37	10.71	30	1.3%	4.87 [-0.14 , 9.88]
Subtotal (95% CI)			**319**			**235**	**32.5%**	**2.51 [1.20 , 3.83]**
Heterogeneity: Tau² = 1.73; Chi² = 17.10, df = 10 (P = 0.07); I² = 42%								
Test for overall effect: Z = 3.75 (P = 0.0002)								
3.3.2 Less than 20 group sessions of CS during one to twelve months								
Breuil 1994	1.4	2.7	29	-0.7	3.1	27	8.3%	2.10 [0.57 , 3.63]
Capotosto 2017	0.18	4.57	20	-0.39	5.34	19	2.9%	0.57 [-2.56 , 3.70]
Carbone 2021	0.8894	2.68673	123	-1.2016	3.01666	101	14.6%	2.09 [1.33 , 2.85]
Coen 2011	0.8	3.6	14	-2.1	2.5	11	4.5%	2.90 [0.50 , 5.30]
Cove 2014	-0.33	6.06	24	-0.78	4.54	23	3.1%	0.45 [-2.60 , 3.50]
Lok 2020	3	6.39	30	-2.16	6.33	30	2.8%	5.16 [1.94 , 8.38]
Spector 2001	3.1	7.04	17	0	7.04	10	1.1%	3.10 [-2.40 , 8.60]
Spector 2003	0.9	3.5	97	-0.4	3.5	70	11.7%	1.30 [0.22 , 2.38]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	3.9%	0.65 [-1.98 , 3.28]
Young 2019	2.1	2.26	51	-0.74	1.52	50	14.7%	2.84 [2.09 , 3.59]
Subtotal (95% CI)			**420**			**351**	**67.5%**	**2.11 [1.51 , 2.72]**
Heterogeneity: Tau² = 0.24; Chi² = 12.90, df = 9 (P = 0.17); I² = 30%								
Test for overall effect: Z = 6.81 (P < 0.00001)								
Total (95% CI)			**739**			**586**	**100.0%**	**2.16 [1.58 , 2.74]**
Heterogeneity: Tau² = 0.45; Chi² = 30.24, df = 20 (P = 0.07); I² = 34%								
Test for overall effect: Z = 7.27 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.29, df = 1 (P = 0.59), I² = 0%								



Analysis 3.4. Comparison 3: Group cognitive stimulation versus no cognitive stimulation (number of sessions): post-treatment, Outcome 4: Quality of Life: self-report



Comparison 4. Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment

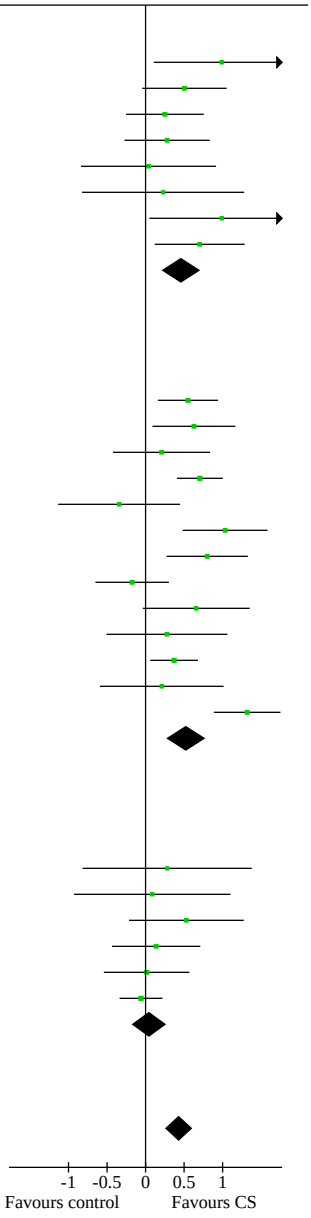
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.1 Cognition (3 levels)	27	1637	Std. Mean Difference (IV, Random, 95% CI)	0.43 [0.26, 0.59]
4.1.1 Three or more group sessions of CS per week during one to twelve months	8	328	Std. Mean Difference (IV, Random, 95% CI)	0.46 [0.22, 0.69]
4.1.2 Two group sessions of CS per week during one to twelve months	13	955	Std. Mean Difference (IV, Random, 95% CI)	0.52 [0.28, 0.76]
4.1.3 One group session of CS per week during one to twelve months	6	354	Std. Mean Difference (IV, Random, 95% CI)	0.04 [-0.17, 0.25]
4.2 Cognition (2 levels)	27	1637	Std. Mean Difference (IV, Random, 95% CI)	0.43 [0.26, 0.59]
4.2.1 Two or more group sessions of CS per week during one to twelve months	21	1283	Std. Mean Difference (IV, Random, 95% CI)	0.51 [0.34, 0.69]
4.2.2 One group session of CS per week during one to twelve months	6	354	Std. Mean Difference (IV, Random, 95% CI)	0.04 [-0.17, 0.25]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.3 ADAS-Cog (3 levels)	17	1168	Mean Difference (IV, Random, 95% CI)	2.66 [1.12, 4.20]
4.3.1 Three or more group sessions of CS per week during one to twelve months	3	131	Mean Difference (IV, Random, 95% CI)	6.68 [-0.57, 13.92]
4.3.2 Two group sessions of CS per week during one to twelve months	9	713	Mean Difference (IV, Random, 95% CI)	3.10 [1.24, 4.96]
4.3.3 One group session of CS per week during one to twelve months	5	324	Mean Difference (IV, Random, 95% CI)	0.09 [-2.28, 2.46]
4.4 ADAS-Cog (2 levels)	17	1168	Mean Difference (IV, Random, 95% CI)	2.66 [1.12, 4.20]
4.4.1 Two or more group sessions of CS per week during one to twelve months	12	844	Mean Difference (IV, Random, 95% CI)	3.41 [1.58, 5.24]
4.4.2 One group session of CS per week during one to twelve months	5	324	Mean Difference (IV, Random, 95% CI)	0.09 [-2.28, 2.46]
4.5 MMSE (3 levels)	21	1325	Mean Difference (IV, Random, 95% CI)	2.16 [1.58, 2.74]
4.5.1 Three or more group sessions of CS per week during one to twelve months	7	267	Mean Difference (IV, Random, 95% CI)	2.66 [1.09, 4.23]
4.5.2 Two group sessions of CS per week during one to twelve months	10	784	Mean Difference (IV, Random, 95% CI)	2.40 [1.66, 3.13]
4.5.3 One group session of CS per week during one to twelve months	4	274	Mean Difference (IV, Random, 95% CI)	0.92 [-0.07, 1.90]
4.6 MMSE (2 levels)	21	1325	Mean Difference (IV, Random, 95% CI)	2.16 [1.58, 2.74]
4.6.1 Two or more group sessions of CS per week during one to twelve months	17	1051	Mean Difference (IV, Random, 95% CI)	2.42 [1.80, 3.03]
4.6.2 One group session of CS per week during one to twelve months	4	274	Mean Difference (IV, Random, 95% CI)	0.92 [-0.07, 1.90]
4.7 Quality of Life: self-report	13	1058	Mean Difference (IV, Random, 95% CI)	1.97 [0.47, 3.47]
4.7.1 Two or more group sessions of CS per week during one to twelve months	9	747	Mean Difference (IV, Random, 95% CI)	2.40 [0.48, 4.33]
4.7.2 One group session per week of CS during one to twelve months	4	311	Mean Difference (IV, Random, 95% CI)	1.47 [-0.06, 3.01]
4.8 Mood: interviewer/staff-rated	10	959	Std. Mean Difference (IV, Random, 95% CI)	0.40 [0.14, 0.67]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.8.1 Two or more group sessions of CS per week during one to twelve months	7	695	Std. Mean Difference (IV, Random, 95% CI)	0.41 [0.07, 0.75]
4.8.2 One group session per week of CS during one to twelve months	3	264	Std. Mean Difference (IV, Random, 95% CI)	0.42 [-0.10, 0.94]
4.9 Anxiety - interviewer/staff-rated	6	410	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.09, 0.30]
4.9.1 Two or more group sessions of CS per week during one to twelve months	5	211	Std. Mean Difference (IV, Random, 95% CI)	0.23 [-0.04, 0.51]
4.9.2 One group session per week of CS during one to twelve months	1	199	Std. Mean Difference (IV, Random, 95% CI)	-0.02 [-0.30, 0.26]

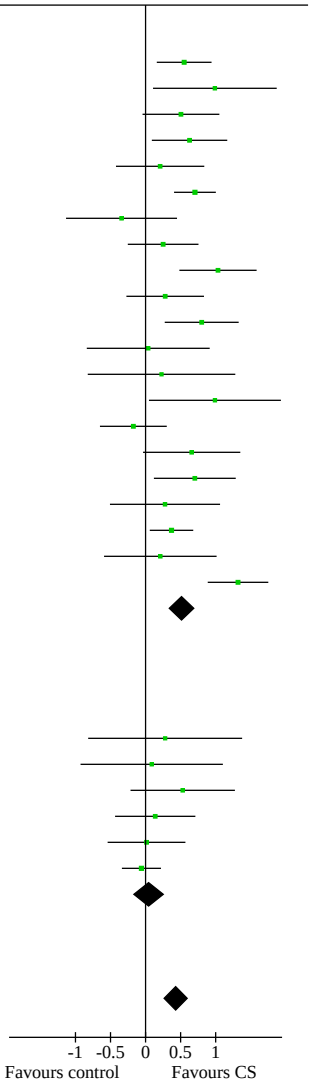
Analysis 4.1. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 1: Cognition (3 levels)

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
4.1.1 Three or more group sessions of CS per week during one to twelve months								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	2.4%	0.99 [0.11 , 1.87]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	4.1%	0.50 [-0.04 , 1.05]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	4.4%	0.25 [-0.25 , 0.75]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	4.0%	0.28 [-0.27 , 0.83]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	2.4%	0.04 [-0.84 , 0.91]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.8%	0.23 [-0.82 , 1.28]
Mapelli 2013	8.7	10.88	10	-2.2	10.21	10	2.2%	0.99 [0.05 , 1.93]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	3.8%	0.70 [0.12 , 1.29]
Subtotal (95% CI)			**194**			**134**	**25.1%**	**0.46 [0.22 , 0.69]**
Heterogeneity: Tau² = 0.00; Chi² = 5.43, df = 7 (P = 0.61); I² = 0%								
Test for overall effect: Z = 3.82 (P = 0.0001)								
4.1.2 Two group sessions of CS per week during one to twelve months								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	5.3%	0.55 [0.16 , 0.94]
Breuil 1994	5.8	7.3	29	1	7.8	27	4.2%	0.63 [0.09 , 1.17]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	3.6%	0.21 [-0.42 , 0.84]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	6.0%	0.70 [0.41 , 1.00]
Coen 2011	0.2	7.2	13	2.3	4.1	12	2.7%	-0.34 [-1.13 , 0.45]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	4.1%	1.03 [0.48 , 1.58]
Lok 2020	3	6.39	30	-2.16	6.33	30	4.2%	0.80 [0.27 , 1.33]
Middelstädt 2016	-0.5	14.42	35	2	13.94	33	4.6%	-0.17 [-0.65 , 0.30]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	3.2%	0.66 [-0.04 , 1.35]
Spector 2001	4.3	17.33	17	-1	20.5	10	2.8%	0.28 [-0.51 , 1.06]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	5.9%	0.37 [0.06 , 0.68]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	2.7%	0.21 [-0.59 , 1.01]
Young 2019	10.65	11.07	51	-1.4	6.41	50	4.9%	1.32 [0.89 , 1.75]
Subtotal (95% CI)			**522**			**433**	**54.1%**	**0.52 [0.28 , 0.76]**
Heterogeneity: Tau² = 0.11; Chi² = 34.52, df = 12 (P = 0.0006); I² = 65%								
Test for overall effect: Z = 4.31 (P < 0.0001)								
4.1.3 One group session of CS per week during one to twelve months								
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	1.7%	0.28 [-0.82 , 1.38]
Buschert 2011	0.7	8	8	0	6.93	7	1.9%	0.09 [-0.93 , 1.10]
Cheung 2019	0.94	2.6745	18	-0.5	2.5969	12	3.0%	0.53 [-0.21 , 1.27]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	3.9%	0.14 [-0.43 , 0.71]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	4.0%	0.01 [-0.54 , 0.57]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	6.2%	-0.06 [-0.34 , 0.22]
Subtotal (95% CI)			**186**			**168**	**20.7%**	**0.04 [-0.17 , 0.25]**
Heterogeneity: Tau² = 0.00; Chi² = 2.46, df = 5 (P = 0.78); I² = 0%								
Test for overall effect: Z = 0.40 (P = 0.69)								
Total (95% CI)			**902**			**735**	**100.0%**	**0.43 [0.26 , 0.59]**
Heterogeneity: Tau² = 0.09; Chi² = 59.26, df = 26 (P = 0.0002); I² = 56%								
Test for overall effect: Z = 5.12 (P < 0.00001)								
Test for subgroup differences: Chi² = 10.82, df = 2 (P = 0.004), I² = 81.5%								

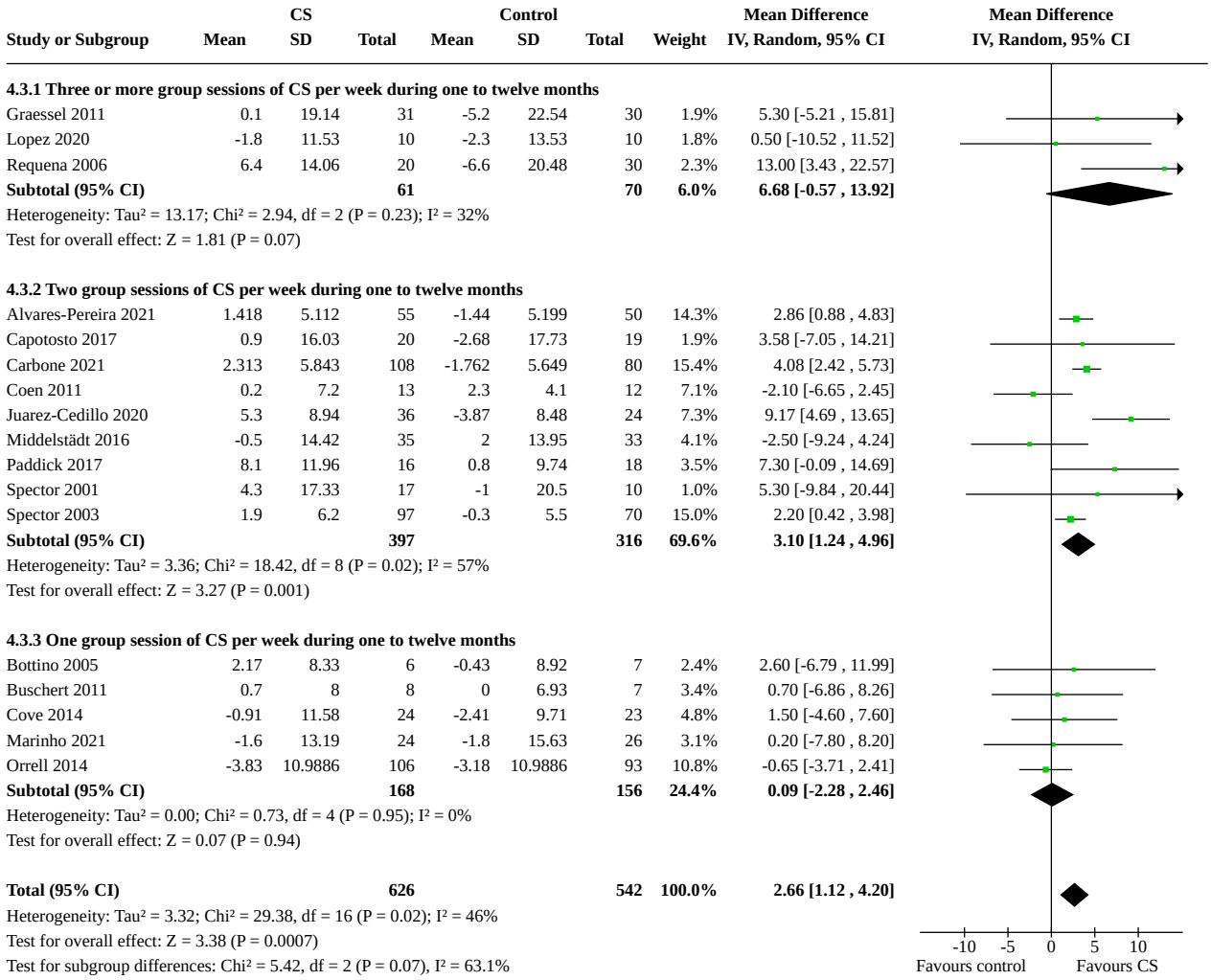


Analysis 4.2. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 2: Cognition (2 levels)

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
4.2.1 Two or more group sessions of CS per week during one to twelve months								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	5.3%	0.55 [0.16 , 0.94]
Baldelli 1993	3	5.32	13	-4.4	9.15	10	2.4%	0.99 [0.11 , 1.87]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	4.1%	0.50 [-0.04 , 1.05]
Breuil 1994	5.8	7.3	29	1	7.8	27	4.2%	0.63 [0.09 , 1.17]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	3.6%	0.21 [-0.42 , 0.84]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	6.0%	0.70 [0.41 , 1.00]
Coen 2011	0.2	7.2	13	2.3	4.1	12	2.7%	-0.34 [-1.13 , 0.45]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	4.4%	0.25 [-0.25 , 0.75]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	4.1%	1.03 [0.48 , 1.58]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	4.0%	0.28 [-0.27 , 0.83]
Lok 2020	3	6.39	30	-2.16	6.33	30	4.2%	0.80 [0.27 , 1.33]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	2.4%	0.04 [-0.84 , 0.91]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.8%	0.23 [-0.82 , 1.28]
Mapelli 2013	8.7	10.88	10	-2.2	10.21	10	2.2%	0.99 [0.05 , 1.93]
Middelstädt 2016	-0.5	14.42	35	2	13.94	33	4.6%	-0.17 [-0.65 , 0.30]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	3.2%	0.66 [-0.04 , 1.35]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	3.8%	0.70 [0.12 , 1.29]
Spector 2001	4.3	17.33	17	-1	20.5	10	2.8%	0.28 [-0.51 , 1.06]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	5.9%	0.37 [0.06 , 0.68]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	2.7%	0.21 [-0.59 , 1.01]
Young 2019	10.65	11.07	51	-1.4	6.41	50	4.9%	1.32 [0.89 , 1.75]
Subtotal (95% CI)			**716**			**567**	**79.3%**	**0.51 [0.34 , 0.69]**
Heterogeneity: Tau² = 0.08; Chi² = 40.49, df = 20 (P = 0.004); I² = 51%								
Test for overall effect: Z = 5.76 (P < 0.00001)								
4.2.2 One group session of CS per week during one to twelve months								
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	1.7%	0.28 [-0.82 , 1.38]
Buschert 2011	0.7	8	8	0	6.93	7	1.9%	0.09 [-0.93 , 1.10]
Cheung 2019	0.94	2.6745	18	-0.5	2.5969	12	3.0%	0.53 [-0.21 , 1.27]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	3.9%	0.14 [-0.43 , 0.71]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	4.0%	0.01 [-0.54 , 0.57]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	6.2%	-0.06 [-0.34 , 0.22]
Subtotal (95% CI)			**186**			**168**	**20.7%**	**0.04 [-0.17 , 0.25]**
Heterogeneity: Tau² = 0.00; Chi² = 2.46, df = 5 (P = 0.78); I² = 0%								
Test for overall effect: Z = 0.40 (P = 0.69)								
Total (95% CI)			**902**			**735**	**100.0%**	**0.43 [0.26 , 0.59]**
Heterogeneity: Tau² = 0.09; Chi² = 59.26, df = 26 (P = 0.0002); I² = 56%								
Test for overall effect: Z = 5.12 (P < 0.00001)								
Test for subgroup differences: Chi² = 11.41, df = 1 (P = 0.0007), I² = 91.2%								

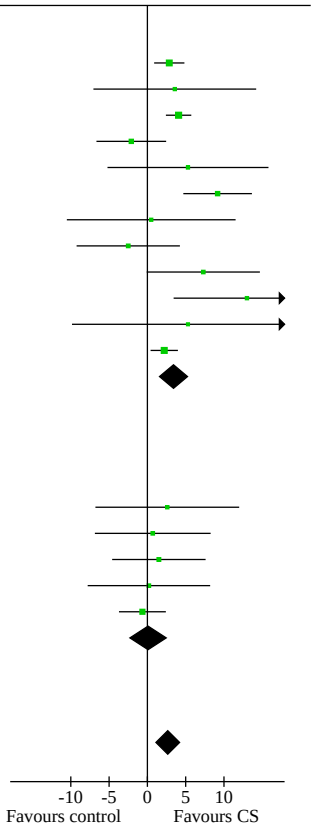


Analysis 4.3. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 3: ADAS-Cog (3 levels)



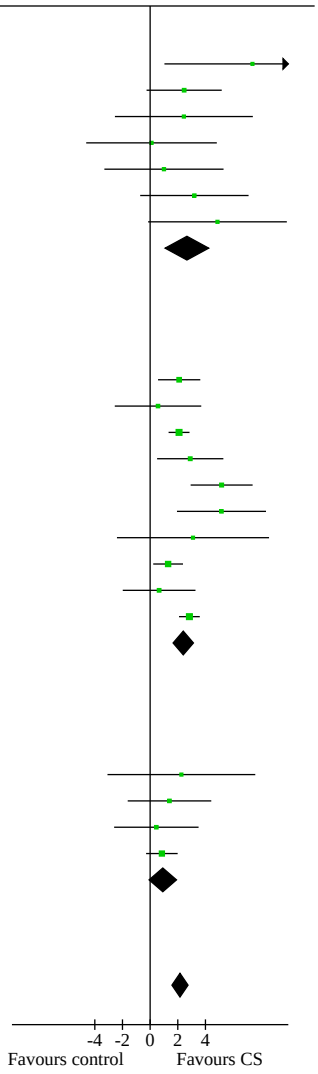
Analysis 4.4. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 4: ADAS-Cog (2 levels)

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
4.4.1 Two or more group sessions of CS per week during one to twelve months								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	14.3%	2.86 [0.88 , 4.83]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	1.9%	3.58 [-7.05 , 14.21]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	15.4%	4.08 [2.42 , 5.73]
Coen 2011	0.2	7.2	13	2.3	4.1	12	7.1%	-2.10 [-6.65 , 2.45]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	1.9%	5.30 [-5.21 , 15.81]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	7.3%	9.17 [4.69 , 13.65]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	1.8%	0.50 [-10.52 , 11.52]
Middelstädt 2016	-0.5	14.42	35	2	13.95	33	4.1%	-2.50 [-9.24 , 4.24]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	3.5%	7.30 [-0.09 , 14.69]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	2.3%	13.00 [3.43 , 22.57]
Spector 2001	4.3	17.33	17	-1	20.5	10	1.0%	5.30 [-9.84 , 20.44]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	15.0%	2.20 [0.42 , 3.98]
Subtotal (95% CI)			**458**			**386**	**75.6%**	**3.41 [1.58 , 5.24]**
Heterogeneity: Tau² = 3.66; Chi² = 22.85, df = 11 (P = 0.02); I² = 52%								
Test for overall effect: Z = 3.66 (P = 0.0003)								
4.4.2 One group session of CS per week during one to twelve months								
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	2.4%	2.60 [-6.79 , 11.99]
Buschert 2011	0.7	8	8	0	6.93	7	3.4%	0.70 [-6.86 , 8.26]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	4.8%	1.50 [-4.60 , 7.60]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	3.1%	0.20 [-7.80 , 8.20]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	10.8%	-0.65 [-3.71 , 2.41]
Subtotal (95% CI)			**168**			**156**	**24.4%**	**0.09 [-2.28 , 2.46]**
Heterogeneity: Tau² = 0.00; Chi² = 0.73, df = 4 (P = 0.95); I² = 0%								
Test for overall effect: Z = 0.07 (P = 0.94)								
Total (95% CI)			**626**			**542**	**100.0%**	**2.66 [1.12 , 4.20]**
Heterogeneity: Tau² = 3.32; Chi² = 29.38, df = 16 (P = 0.02); I² = 46%								
Test for overall effect: Z = 3.38 (P = 0.0007)								
Test for subgroup differences: Chi² = 4.73, df = 1 (P = 0.03), I² = 78.9%								



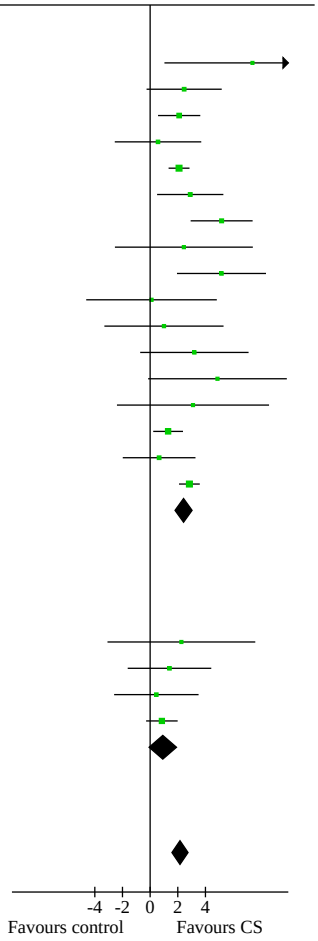
Analysis 4.5. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 5: MMSE (3 levels)

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
4.5.1 Three or more group sessions of CS per week during one to twelve months								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	0.8%	7.40 [1.03 , 13.77]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	3.7%	2.46 [-0.26 , 5.18]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	1.3%	2.44 [-2.55 , 7.43]
Lopez 2020	-1.6	4.82	10	-1.7	5.9	10	1.4%	0.10 [-4.62 , 4.82]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.7%	1.00 [-3.31 , 5.31]
Mapelli 2013	2.9	5.03	10	-0.3	3.83	10	2.0%	3.20 [-0.72 , 7.12]
Requena 2006	1.5	7.38	20	-3.37	10.71	30	1.3%	4.87 [-0.14 , 9.88]
Subtotal (95% CI)			**163**			**104**	**12.1%**	**2.66 [1.09 , 4.23]**
Heterogeneity: Tau² = 0.00; Chi² = 4.68, df = 6 (P = 0.59); I² = 0%								
Test for overall effect: Z = 3.31 (P = 0.0009)								
4.5.2 Two group sessions of CS per week during one to twelve months								
Breuil 1994	1.4	2.7	29	-0.7	3.1	27	8.3%	2.10 [0.57 , 3.63]
Capotosto 2017	0.18	4.57	20	-0.39	5.34	19	2.9%	0.57 [-2.56 , 3.70]
Carbone 2021	0.8894	2.68673	123	-1.2016	3.01666	101	14.6%	2.09 [1.33 , 2.85]
Coen 2011	0.8	3.6	14	-2.1	2.5	11	4.5%	2.90 [0.50 , 5.30]
Juarez-Cedillo 2020	2.05	4.41	36	-3.12	4.29	24	5.0%	5.17 [2.93 , 7.41]
Lok 2020	3	6.39	30	-2.16	6.33	30	2.8%	5.16 [1.94 , 8.38]
Spector 2001	3.1	7.04	17	0	7.04	10	1.1%	3.10 [-2.40 , 8.60]
Spector 2003	0.9	3.5	97	-0.4	3.5	70	11.7%	1.30 [0.22 , 2.38]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	3.9%	0.65 [-1.98 , 3.28]
Young 2019	2.1	2.26	51	-0.74	1.52	50	14.7%	2.84 [2.09 , 3.59]
Subtotal (95% CI)			**432**			**352**	**69.5%**	**2.40 [1.66 , 3.13]**
Heterogeneity: Tau² = 0.55; Chi² = 17.99, df = 9 (P = 0.04); I² = 50%								
Test for overall effect: Z = 6.39 (P < 0.00001)								
4.5.3 One group session of CS per week during one to twelve months								
Bottino 2005	0.83	4.53	6	-1.43	5.3	7	1.1%	2.26 [-3.08 , 7.60]
Buschert 2011	0.5	3.14	8	-0.9	2.83	7	3.1%	1.40 [-1.62 , 4.42]
Cove 2014	-0.33	6.06	24	-0.78	4.54	23	3.1%	0.45 [-2.60 , 3.50]
Orrell 2014	-1.46	4.0938	106	-2.31	4.0938	93	11.1%	0.85 [-0.29 , 1.99]
Subtotal (95% CI)			**144**			**130**	**18.4%**	**0.92 [-0.07 , 1.90]**
Heterogeneity: Tau² = 0.00; Chi² = 0.44, df = 3 (P = 0.93); I² = 0%								
Test for overall effect: Z = 1.81 (P = 0.07)								
Total (95% CI)			**739**			**586**	**100.0%**	**2.16 [1.58 , 2.74]**
Heterogeneity: Tau² = 0.45; Chi² = 30.24, df = 20 (P = 0.07); I² = 34%								
Test for overall effect: Z = 7.27 (P < 0.00001)								
Test for subgroup differences: Chi² = 6.39, df = 2 (P = 0.04), I² = 68.7%								



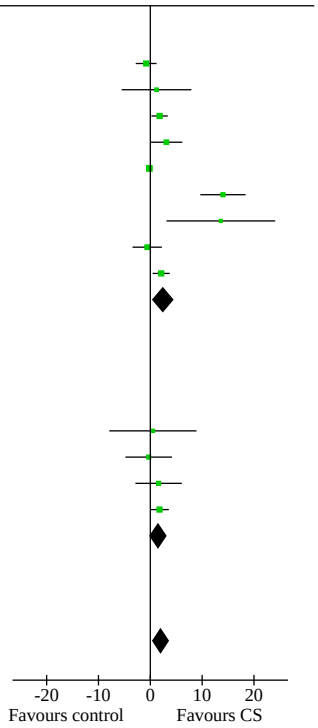
Analysis 4.6. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 6: MMSE (2 levels)

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
4.6.1 Two or more group sessions of CS per week during one to twelve months								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	0.8%	7.40 [1.03 , 13.77]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	3.7%	2.46 [-0.26 , 5.18]
Breuil 1994	1.4	2.7	29	-0.7	3.1	27	8.3%	2.10 [0.57 , 3.63]
Capotosto 2017	0.18	4.57	20	-0.39	5.34	19	2.9%	0.57 [-2.56 , 3.70]
Carbone 2021	0.8894	2.68673	123	-1.2016	3.01666	101	14.6%	2.09 [1.33 , 2.85]
Coen 2011	0.8	3.6	14	-2.1	2.5	11	4.5%	2.90 [0.50 , 5.30]
Juarez-Cedillo 2020	2.05	4.41	36	-3.12	4.29	24	5.0%	5.17 [2.93 , 7.41]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	1.3%	2.44 [-2.55 , 7.43]
Lok 2020	3	6.39	30	-2.16	6.33	30	2.8%	5.16 [1.94 , 8.38]
Lopez 2020	-1.6	4.82	10	-1.7	5.9	10	1.4%	0.10 [-4.62 , 4.82]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.7%	1.00 [-3.31 , 5.31]
Mapelli 2013	2.9	5.03	10	-0.3	3.83	10	2.0%	3.20 [-0.72 , 7.12]
Requena 2006	1.5	7.38	20	-3.37	10.71	30	1.3%	4.87 [-0.14 , 9.88]
Spector 2001	3.1	7.04	17	0	7.04	10	1.1%	3.10 [-2.40 , 8.60]
Spector 2003	0.9	3.5	97	-0.4	3.5	70	11.7%	1.30 [0.22 , 2.38]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	3.9%	0.65 [-1.98 , 3.28]
Young 2019	2.1	2.26	51	-0.74	1.52	50	14.7%	2.84 [2.09 , 3.59]
Subtotal (95% CI)			**595**			**456**	**81.6%**	**2.42 [1.80 , 3.03]**
Heterogeneity: Tau² = 0.37; Chi² = 22.81, df = 16 (P = 0.12); I² = 30%								
Test for overall effect: Z = 7.69 (P < 0.00001)								
4.6.2 One group session of CS per week during one to twelve months								
Bottino 2005	0.83	4.53	6	-1.43	5.3	7	1.1%	2.26 [-3.08 , 7.60]
Buschert 2011	0.5	3.14	8	-0.9	2.83	7	3.1%	1.40 [-1.62 , 4.42]
Cove 2014	-0.33	6.06	24	-0.78	4.54	23	3.1%	0.45 [-2.60 , 3.50]
Orrell 2014	-1.46	4.0938	106	-2.31	4.0938	93	11.1%	0.85 [-0.29 , 1.99]
Subtotal (95% CI)			**144**			**130**	**18.4%**	**0.92 [-0.07 , 1.90]**
Heterogeneity: Tau² = 0.00; Chi² = 0.44, df = 3 (P = 0.93); I² = 0%								
Test for overall effect: Z = 1.81 (P = 0.07)								
Total (95% CI)			**739**			**586**	**100.0%**	**2.16 [1.58 , 2.74]**
Heterogeneity: Tau² = 0.45; Chi² = 30.24, df = 20 (P = 0.07); I² = 34%								
Test for overall effect: Z = 7.27 (P < 0.00001)								
Test for subgroup differences: Chi² = 6.38, df = 1 (P = 0.01), I² = 84.3%								



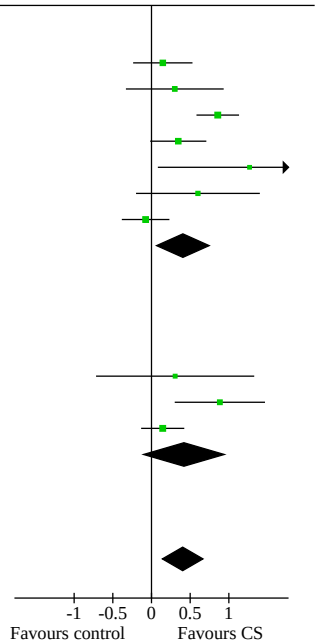
Analysis 4.7. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 7: Quality of Life: self-report

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
4.7.1 Two or more group sessions of CS per week during one to twelve months								
Alvares-Pereira 2021	-0.53	4.63	55	0.27	5.74	50	10.5%	-0.80 [-2.81 , 1.21]
Capotosto 2017	1.25	11.5	20	0.05	9.91	19	3.6%	1.20 [-5.53 , 7.93]
Carbone 2021	1.6	6.63	118	-0.2	5.12	96	11.3%	1.80 [0.23 , 3.37]
Coen 2011	3.6	3.7	14	0.5	4.4	13	8.4%	3.10 [0.02 , 6.18]
Kim 2016	-0.4	0.76	32	-0.23	0.73	21	12.8%	-0.17 [-0.58 , 0.24]
Lok 2020	11.86	9.64	30	-2.16	7.48	30	6.2%	14.02 [9.65 , 18.39]
Maci 2012	12.3	11.78	7	-1.3	7.86	7	1.8%	13.60 [3.11 , 24.09]
Middelstädt 2016	-0.2	5.47	35	0.4	6.36	33	8.9%	-0.60 [-3.43 , 2.23]
Spector 2003	1.3	5.1	97	-0.8	5.6	70	11.2%	2.10 [0.44 , 3.76]
Subtotal (95% CI)			**408**			**339**	**74.5%**	**2.40 [0.48 , 4.33]**
Heterogeneity: Tau² = 5.91; Chi² = 61.67, df = 8 (P < 0.00001); I² = 87%								
Test for overall effect: Z = 2.45 (P = 0.01)								
4.7.2 One group session per week of CS during one to twelve months								
Buschert 2011	-0.4	10.61	8	-0.9	5.52	7	2.5%	0.50 [-7.91 , 8.91]
Cove 2014	0.23	7.97	24	0.54	7.74	23	6.0%	-0.31 [-4.80 , 4.18]
Marinho 2021	2	8.77	24	0.4	7.21	26	6.0%	1.60 [-2.87 , 6.07]
Orrell 2014	-0.67	6.428	106	-2.45	6.428	93	10.9%	1.78 [-0.01 , 3.57]
Subtotal (95% CI)			**162**			**149**	**25.5%**	**1.47 [-0.06 , 3.01]**
Heterogeneity: Tau² = 0.00; Chi² = 0.77, df = 3 (P = 0.86); I² = 0%								
Test for overall effect: Z = 1.88 (P = 0.06)								
Total (95% CI)			**570**			**488**	**100.0%**	**1.97 [0.47 , 3.47]**
Heterogeneity: Tau² = 4.52; Chi² = 64.97, df = 12 (P < 0.00001); I² = 82%								
Test for overall effect: Z = 2.58 (P = 0.010)								
Test for subgroup differences: Chi² = 0.55, df = 1 (P = 0.46), I² = 0%								

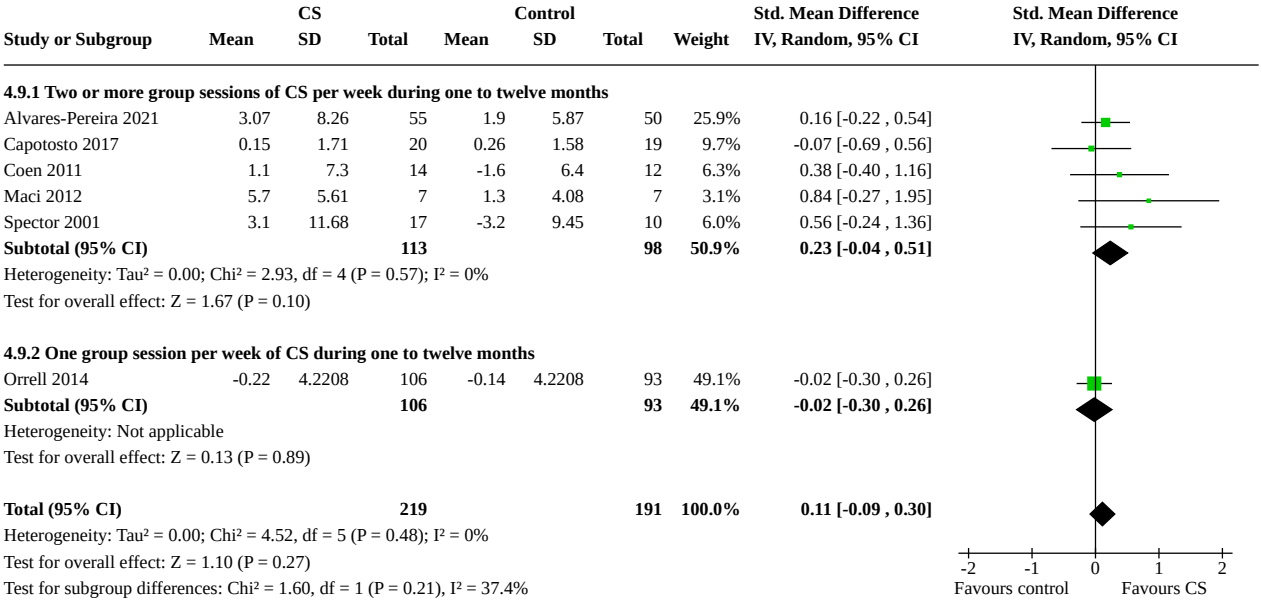


Analysis 4.8. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 8: Mood: interviewer/staff-rated

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
4.8.1 Two or more group sessions of CS per week during one to twelve months								
Alvares-Pereira 2021	1.91	4.38	55	1.2	5.25	50	12.4%	0.15 [-0.24 , 0.53]
Capotosto 2017	1.4	2.87	20	0.42	3.49	19	8.5%	0.30 [-0.33 , 0.93]
Carbone 2021	1.9	3.14	123	-0.89	3.37	101	14.2%	0.86 [0.58 , 1.13]
Graessel 2011	1	2.5392	56	0.1	2.6206	63	12.7%	0.35 [-0.02 , 0.71]
Maci 2012	5.9	6.26	7	-1.3	4.17	7	3.8%	1.27 [0.08 , 2.45]
Spector 2001	2.6	8.05	17	-2.2	7.19	10	6.6%	0.60 [-0.20 , 1.40]
Spector 2003	0	6.2	97	0.5	7	70	13.7%	-0.08 [-0.38 , 0.23]
Subtotal (95% CI)			**375**			**320**	**71.9%**	**0.41 [0.07 , 0.75]**
Heterogeneity: Tau² = 0.14; Chi² = 23.93, df = 6 (P = 0.0005); I² = 75%								
Test for overall effect: Z = 2.34 (P = 0.02)								
4.8.2 One group session per week of CS during one to twelve months								
Buschert 2011	1.5	5.33	8	-0.4	6.4	7	4.7%	0.31 [-0.72 , 1.33]
Marinho 2021	1.3	2.8	24	-1.1	2.55	26	9.2%	0.88 [0.30 , 1.47]
Orrell 2014	-0.38	5.2185	106	-1.14	5.2185	93	14.2%	0.15 [-0.13 , 0.42]
Subtotal (95% CI)			**138**			**126**	**28.1%**	**0.42 [-0.10 , 0.94]**
Heterogeneity: Tau² = 0.12; Chi² = 5.02, df = 2 (P = 0.08); I² = 60%								
Test for overall effect: Z = 1.58 (P = 0.12)								
Total (95% CI)			**513**			**446**	**100.0%**	**0.40 [0.14 , 0.67]**
Heterogeneity: Tau² = 0.11; Chi² = 29.47, df = 9 (P = 0.0005); I² = 69%								
Test for overall effect: Z = 3.01 (P = 0.003)								
Test for subgroup differences: Chi² = 0.00, df = 1 (P = 0.97), I² = 0%								



Analysis 4.9. Comparison 4: Group cognitive stimulation versus no cognitive stimulation (frequency): post-treatment, Outcome 9: Anxiety - interviewer/staff-rated



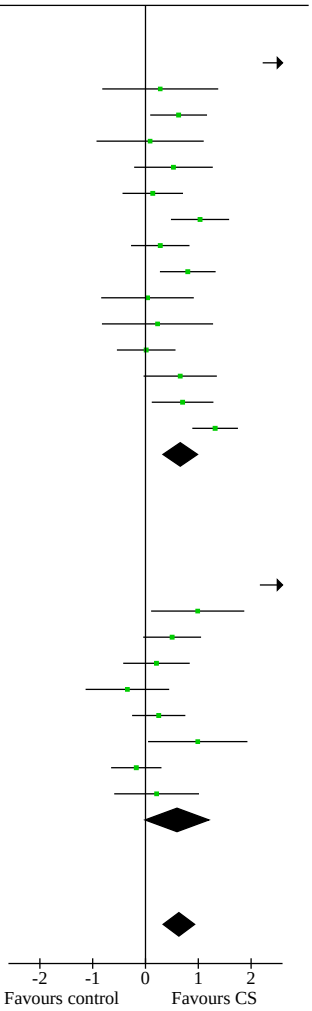
Comparison 5. Cognitive stimulation versus no cognitive stimulation (setting): post-treatment

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5.1 Cognition	23	1056	Std. Mean Difference (IV, Random, 95% CI)	0.63 [0.33, 0.93]
5.1.1 Community	15	642	Std. Mean Difference (IV, Random, 95% CI)	0.66 [0.33, 0.99]
5.1.2 Care home	9	414	Std. Mean Difference (IV, Random, 95% CI)	0.60 [-0.01, 1.20]
5.2 ADAS-Cog	13	587	Mean Difference (IV, Random, 95% CI)	3.16 [1.85, 4.47]
5.2.1 Community	9	328	Mean Difference (IV, Random, 95% CI)	4.49 [2.18, 6.80]
5.2.2 Care home	5	259	Mean Difference (IV, Random, 95% CI)	1.15 [-1.71, 4.01]
5.3 MMSE	17	708	Mean Difference (IV, Random, 95% CI)	2.57 [1.87, 3.27]
5.3.1 Community	11	489	Mean Difference (IV, Random, 95% CI)	2.72 [1.85, 3.59]
5.3.2 Care home	6	219	Mean Difference (IV, Random, 95% CI)	2.16 [0.83, 3.48]
5.4 Quality of Life: self-report	9	373	Mean Difference (IV, Random, 95% CI)	2.98 [0.06, 5.90]
5.4.1 Community	6	239	Mean Difference (IV, Random, 95% CI)	4.33 [-0.75, 9.42]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5.4.2 Care home	3	134	Mean Difference (IV, Random, 95% CI)	1.17 [-1.44, 3.78]
5.5 Quality of Life: proxy-rated	5	207	Std. Mean Difference (IV, Random, 95% CI)	0.50 [0.05, 0.94]
5.5.1 Community	3	114	Std. Mean Difference (IV, Random, 95% CI)	0.74 [-0.02, 1.49]
5.5.2 Care home	2	93	Std. Mean Difference (IV, Random, 95% CI)	0.21 [-0.21, 0.62]
5.6 Mood: self-reported	6	299	Std. Mean Difference (IV, Random, 95% CI)	0.20 [-0.06, 0.45]
5.6.1 One to twelve months of group cognitive stimulation: community	3	163	Std. Mean Difference (IV, Random, 95% CI)	0.24 [-0.07, 0.56]
5.6.2 One to twelve months of group cognitive stimulation: care home	3	136	Std. Mean Difference (IV, Random, 95% CI)	0.16 [-0.48, 0.79]
5.7 Mood: Interviewer/staff-rated	5	237	Std. Mean Difference (IV, Random, 95% CI)	0.50 [0.21, 0.79]
5.7.1 One to twelve months of group cognitive stimulation: community	3	79	Std. Mean Difference (IV, Random, 95% CI)	0.82 [0.36, 1.29]
5.7.2 One to twelve months of group cognitive stimulation: care home	2	158	Std. Mean Difference (IV, Random, 95% CI)	0.34 [0.02, 0.65]

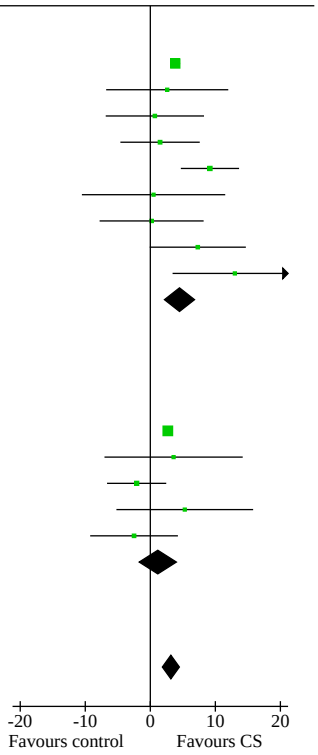
Analysis 5.1. Comparison 5: Cognitive stimulation versus no cognitive stimulation (setting): post-treatment, Outcome 1: Cognition

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
5.1.1 Community								
Alvares-Pereira 2021	2.789	1.192	19	-1.05	1.162	20	3.4%	3.20 [2.22 , 4.17]
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	3.1%	0.28 [-0.82 , 1.38]
Breuil 1994	5.8	7.3	29	1	7.8	27	4.6%	0.63 [0.09 , 1.17]
Buschert 2011	0.7	8	8	0	6.93	7	3.3%	0.09 [-0.93 , 1.10]
Cheung 2019	0.94	2.6745	18	-0.5	2.5969	12	4.1%	0.53 [-0.21 , 1.27]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	4.5%	0.14 [-0.43 , 0.71]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	4.6%	1.03 [0.48 , 1.58]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	4.6%	0.28 [-0.27 , 0.83]
Lok 2020	3	6.39	30	-2.16	6.33	30	4.7%	0.80 [0.27 , 1.33]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	3.7%	0.04 [-0.84 , 0.91]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	3.2%	0.23 [-0.82 , 1.28]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	4.6%	0.01 [-0.54 , 0.57]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	4.2%	0.66 [-0.04 , 1.35]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	4.5%	0.70 [0.12 , 1.29]
Young 2019	10.65	11.07	51	-1.4	6.41	50	4.9%	1.32 [0.89 , 1.75]
Subtotal (95% CI)			**330**			**312**	**62.2%**	**0.66 [0.33 , 0.99]**
Heterogeneity: Tau² = 0.29; Chi² = 51.46, df = 14 (P < 0.00001); I² = 73%								
Test for overall effect: Z = 3.95 (P < 0.0001)								
5.1.2 Care home								
Alvares-Pereira 2021	0.694	0.866	36	-2	1	30	4.2%	2.87 [2.17 , 3.56]
Baldelli 1993	3	5.32	13	-4.4	9.15	10	3.7%	0.99 [0.11 , 1.87]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	4.6%	0.50 [-0.04 , 1.05]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	4.4%	0.21 [-0.42 , 0.84]
Coen 2011	0.2	7.2	13	2.3	4.1	12	3.9%	-0.34 [-1.13 , 0.45]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	4.7%	0.25 [-0.25 , 0.75]
Mapelli 2013	8.7	10.88	10	-2.2	10.21	10	3.5%	0.99 [0.05 , 1.93]
Middelstädt 2016	-0.5	14.42	35	2	13.94	33	4.8%	-0.17 [-0.65 , 0.30]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	3.9%	0.21 [-0.59 , 1.01]
Subtotal (95% CI)			**244**			**170**	**37.8%**	**0.60 [-0.01 , 1.20]**
Heterogeneity: Tau² = 0.72; Chi² = 60.46, df = 8 (P < 0.00001); I² = 87%								
Test for overall effect: Z = 1.94 (P = 0.05)								
Total (95% CI)			**574**			**482**	**100.0%**	**0.63 [0.33 , 0.93]**
Heterogeneity: Tau² = 0.42; Chi² = 114.02, df = 23 (P < 0.00001); I² = 80%								
Test for overall effect: Z = 4.16 (P < 0.0001)								
Test for subgroup differences: Chi² = 0.03, df = 1 (P = 0.86), I² = 0%								



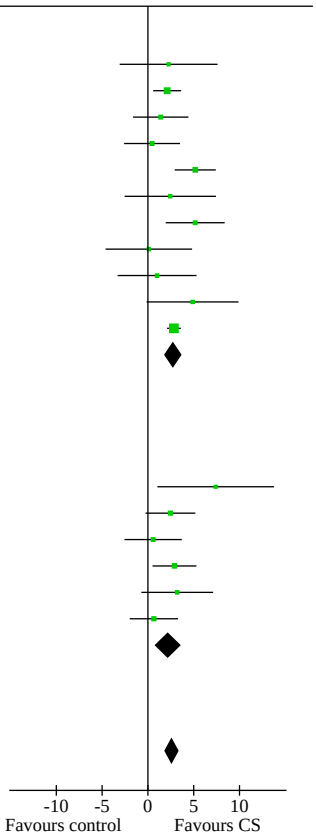
Analysis 5.2. Comparison 5: Cognitive stimulation versus no cognitive stimulation (setting): post-treatment, Outcome 2: ADAS-Cog

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
5.2.1 Community								
Alvares-Pereira 2021	2.789	1.192	19	-1.05	1.162	20	30.6%	3.84 [3.10 , 4.58]
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	1.8%	2.60 [-6.79 , 11.99]
Buschert 2011	0.7	8	8	0	6.93	7	2.7%	0.70 [-6.86 , 8.26]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	4.0%	1.50 [-4.60 , 7.60]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	6.8%	9.17 [4.69 , 13.65]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	1.4%	0.50 [-10.52 , 11.52]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	2.5%	0.20 [-7.80 , 8.20]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	2.9%	7.30 [-0.09 , 14.69]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	1.8%	13.00 [3.43 , 22.57]
Subtotal (95% CI)			**163**			**165**	**54.5%**	**4.49 [2.18 , 6.80]**
Heterogeneity: Tau² = 3.58; Chi² = 12.16, df = 8 (P = 0.14); I² = 34%								
Test for overall effect: Z = 3.81 (P = 0.0001)								
5.2.2 Care home								
Alvares-Pereira 2021	0.694	0.866	36	-2	1	30	32.6%	2.69 [2.24 , 3.15]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	1.4%	3.58 [-7.05 , 14.21]
Coen 2011	0.2	7.2	13	2.3	4.1	12	6.6%	-2.10 [-6.65 , 2.45]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	1.5%	5.30 [-5.21 , 15.81]
Middelstädt 2016	-0.5	14.42	35	2	13.95	33	3.4%	-2.50 [-9.24 , 4.24]
Subtotal (95% CI)			**135**			**124**	**45.5%**	**1.15 [-1.71 , 4.01]**
Heterogeneity: Tau² = 4.08; Chi² = 6.73, df = 4 (P = 0.15); I² = 41%								
Test for overall effect: Z = 0.79 (P = 0.43)								
Total (95% CI)			**298**			**289**	**100.0%**	**3.16 [1.85 , 4.47]**
Heterogeneity: Tau² = 1.31; Chi² = 28.30, df = 13 (P = 0.008); I² = 54%								
Test for overall effect: Z = 4.74 (P < 0.00001)								
Test for subgroup differences: Chi² = 3.17, df = 1 (P = 0.08), I² = 68.4%								



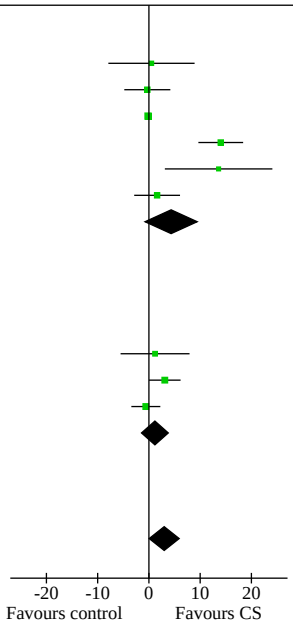
Analysis 5.3. Comparison 5: Cognitive stimulation versus no cognitive stimulation (setting): post-treatment, Outcome 3: MMSE

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
5.3.1 Community								
Bottino 2005	0.83	4.53	6	-1.43	5.3	7	1.7%	2.26 [-3.08 , 7.60]
Breuil 1994	1.4	2.7	29	-0.7	3.1	27	13.8%	2.10 [0.57 , 3.63]
Buschert 2011	0.5	3.14	8	-0.9	2.83	7	4.8%	1.40 [-1.62 , 4.42]
Cove 2014	-0.33	6.06	24	-0.78	4.54	23	4.7%	0.45 [-2.60 , 3.50]
Juarez-Cedillo 2020	2.05	4.41	36	-3.12	4.29	24	7.9%	5.17 [2.93 , 7.41]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	1.9%	2.44 [-2.55 , 7.43]
Lok 2020	3	6.39	30	-2.16	6.33	30	4.3%	5.16 [1.94 , 8.38]
Lopez 2020	-1.6	4.82	10	-1.7	5.9	10	2.1%	0.10 [-4.62 , 4.82]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	2.5%	1.00 [-3.31 , 5.31]
Requena 2006	1.5	7.38	20	-3.37	10.71	30	1.9%	4.87 [-0.14 , 9.88]
Young 2019	2.1	2.26	51	-0.74	1.52	50	27.2%	2.84 [2.09 , 3.59]
Subtotal (95% CI)			**253**			**236**	**72.5%**	**2.72 [1.85 , 3.59]**
Heterogeneity: Tau² = 0.43; Chi² = 12.91, df = 10 (P = 0.23); I² = 23%								
Test for overall effect: Z = 6.13 (P < 0.00001)								
5.3.2 Care home								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	1.2%	7.40 [1.03 , 13.77]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	5.7%	2.46 [-0.26 , 5.18]
Capotosto 2017	0.18	4.57	20	-0.39	5.34	19	4.5%	0.57 [-2.56 , 3.70]
Coen 2011	0.8	3.6	14	-2.1	2.5	11	7.1%	2.90 [0.50 , 5.30]
Mapelli 2013	2.9	5.03	10	-0.3	3.83	10	3.0%	3.20 [-0.72 , 7.12]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	6.0%	0.65 [-1.98 , 3.28]
Subtotal (95% CI)			**143**			**76**	**27.5%**	**2.16 [0.83 , 3.48]**
Heterogeneity: Tau² = 0.28; Chi² = 5.55, df = 5 (P = 0.35); I² = 10%								
Test for overall effect: Z = 3.19 (P = 0.001)								
Total (95% CI)			**396**			**312**	**100.0%**	**2.57 [1.87 , 3.27]**
Heterogeneity: Tau² = 0.33; Chi² = 19.25, df = 16 (P = 0.26); I² = 17%								
Test for overall effect: Z = 7.17 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.49, df = 1 (P = 0.49), I² = 0%								



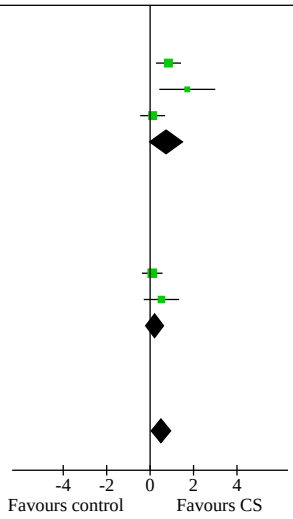
Analysis 5.4. Comparison 5: Cognitive stimulation versus no cognitive stimulation (setting): post-treatment, Outcome 4: Quality of Life: self-report

Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
5.4.1 Community								
Buschert 2011	-0.4	10.61	8	-0.9	5.52	7	6.9%	0.50 [-7.91 , 8.91]
Cove 2014	0.23	7.97	24	0.54	7.74	23	11.7%	-0.31 [-4.80 , 4.18]
Kim 2016	-0.4	0.76	32	-0.23	0.73	21	16.1%	-0.17 [-0.58 , 0.24]
Lok 2020	11.86	9.64	30	-2.16	7.48	30	11.9%	14.02 [9.65 , 18.39]
Maci 2012	12.3	11.78	7	-1.3	7.86	7	5.2%	13.60 [3.11 , 24.09]
Marinho 2021	2	8.77	24	0.4	7.21	26	11.7%	1.60 [-2.87 , 6.07]
Subtotal (95% CI)			**125**			**114**	**63.6%**	**4.33 [-0.75 , 9.42]**
Heterogeneity: Tau² = 31.95; Chi² = 47.25, df = 5 (P < 0.00001); I² = 89%								
Test for overall effect: Z = 1.67 (P = 0.09)								
5.4.2 Care home								
Capotosto 2017	1.25	11.5	20	0.05	9.91	19	8.7%	1.20 [-5.53 , 7.93]
Coen 2011	3.6	3.7	14	0.5	4.4	13	13.7%	3.10 [0.02 , 6.18]
Middelstädt 2016	-0.2	5.47	35	0.4	6.36	33	14.0%	-0.60 [-3.43 , 2.23]
Subtotal (95% CI)			**69**			**65**	**36.4%**	**1.17 [-1.44 , 3.78]**
Heterogeneity: Tau² = 1.82; Chi² = 3.01, df = 2 (P = 0.22); I² = 34%								
Test for overall effect: Z = 0.88 (P = 0.38)								
Total (95% CI)			**194**			**179**	**100.0%**	**2.98 [0.06 , 5.90]**
Heterogeneity: Tau² = 13.75; Chi² = 51.42, df = 8 (P < 0.00001); I² = 84%								
Test for overall effect: Z = 2.00 (P = 0.05)								
Test for subgroup differences: Chi² = 1.18, df = 1 (P = 0.28), I² = 15.1%								

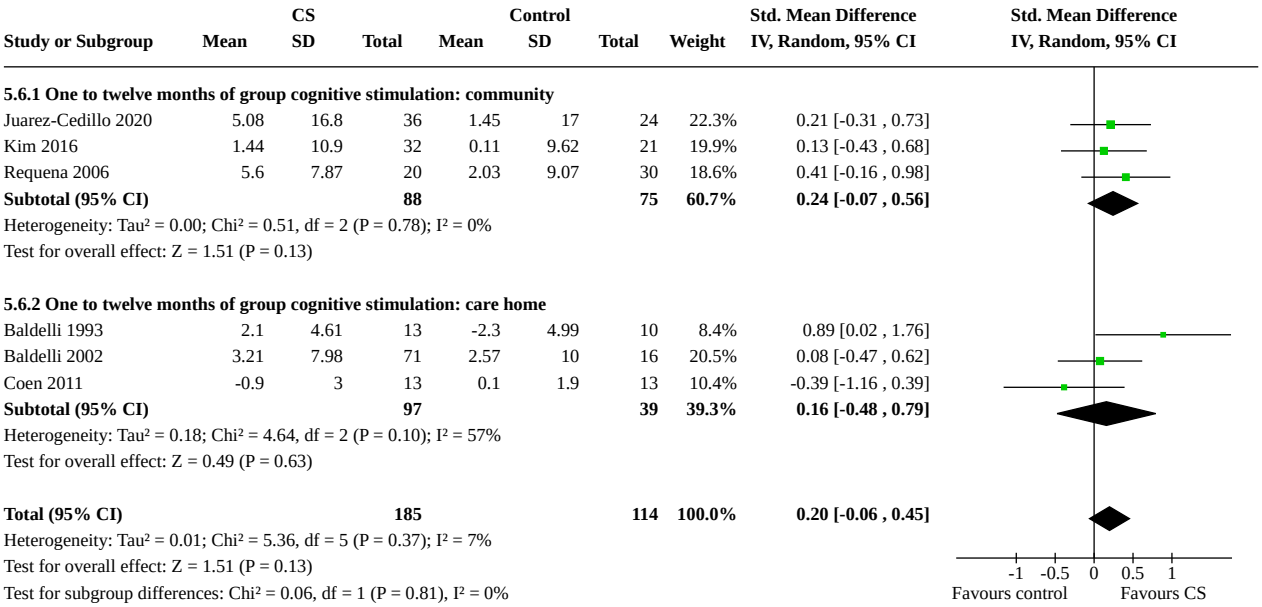


Analysis 5.5. Comparison 5: Cognitive stimulation versus no cognitive stimulation (setting): post-treatment, Outcome 5: Quality of Life: proxy-rated

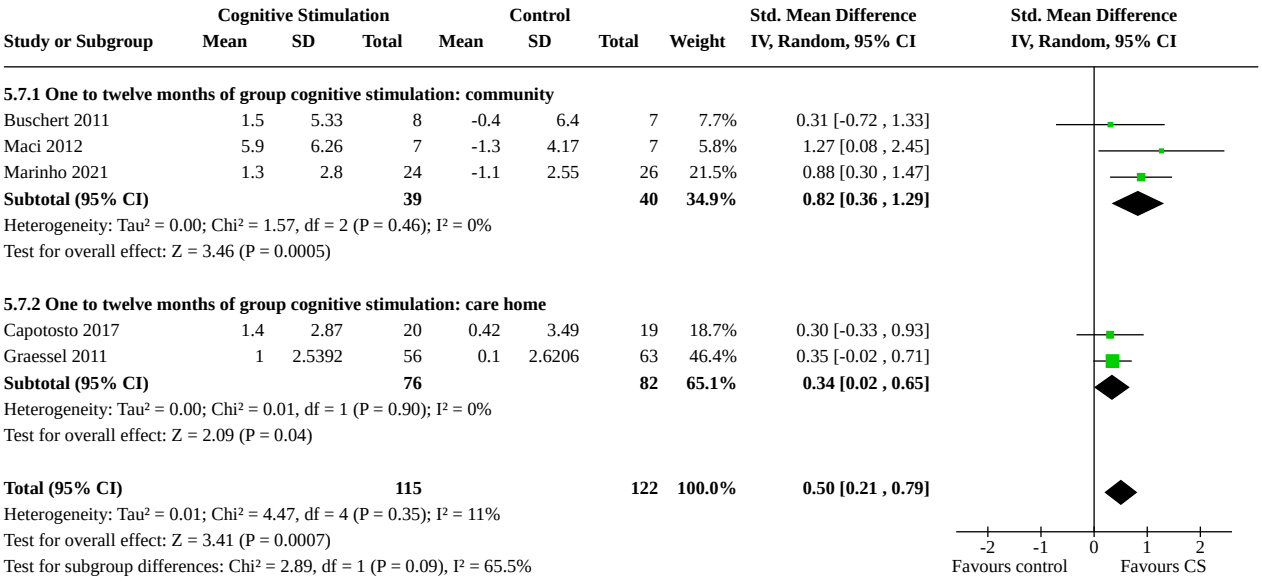
Study or Subgroup	CS Mean	CS SD	CS Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
5.5.1 Community								
Kim 2016	0.81	2.34	32	-1.09	2.01	21	23.5%	0.84 [0.27 , 1.42]
Maci 2012	10.2	8.08	7	-1.4	3.89	7	9.2%	1.71 [0.43 , 3.00]
Marinho 2021	0.8	7.64	23	-0.2	8.91	24	23.6%	0.12 [-0.45 , 0.69]
Subtotal (95% CI)			**62**			**52**	**56.3%**	**0.74 [-0.02 , 1.49]**
Heterogeneity: Tau² = 0.29; Chi² = 6.28, df = 2 (P = 0.04); I² = 68%								
Test for overall effect: Z = 1.91 (P = 0.06)								
5.5.2 Care home								
Middelstädt 2016	-0.4	6.16	35	-1	5.94	33	26.8%	0.10 [-0.38 , 0.57]
Tanaka 2021	1.5	6.9714	15	-1.8	4.4272	10	16.9%	0.52 [-0.29 , 1.34]
Subtotal (95% CI)			**50**			**43**	**43.7%**	**0.21 [-0.21 , 0.62]**
Heterogeneity: Tau² = 0.00; Chi² = 0.78, df = 1 (P = 0.38); I² = 0%								
Test for overall effect: Z = 0.98 (P = 0.33)								
Total (95% CI)			**112**			**95**	**100.0%**	**0.50 [0.05 , 0.94]**
Heterogeneity: Tau² = 0.13; Chi² = 8.85, df = 4 (P = 0.07); I² = 55%								
Test for overall effect: Z = 2.19 (P = 0.03)								
Test for subgroup differences: Chi² = 1.45, df = 1 (P = 0.23), I² = 31.2%								



Analysis 5.6. Comparison 5: Cognitive stimulation versus no cognitive stimulation (setting): post-treatment, Outcome 6: Mood: self-reported



Analysis 5.7. Comparison 5: Cognitive stimulation versus no cognitive stimulation (setting): post-treatment, Outcome 7: Mood: Interviewer/staff-rated



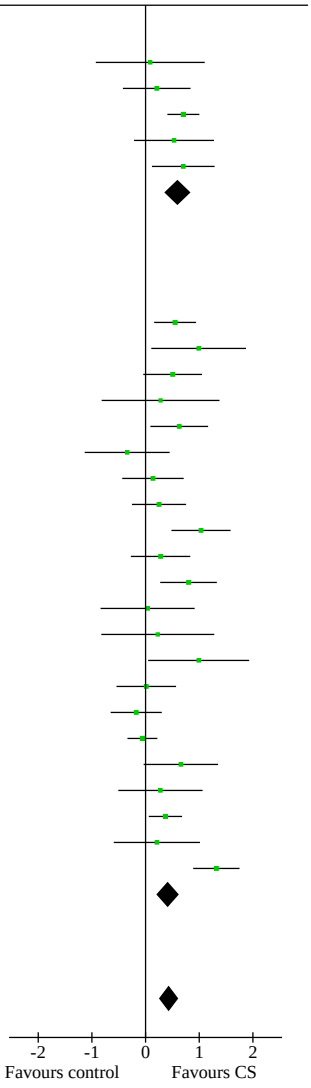
Comparison 6. Group cognitive stimulation versus no treatment post-treatment (Active control group)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6.1 Cognition	27	1637	Std. Mean Difference (IV, Random, 95% CI)	0.43 [0.26, 0.59]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6.1.1 Alternate activity control group	5	322	Std. Mean Difference (IV, Random, 95% CI)	0.59 [0.37, 0.82]
6.1.2 Treatment-as-usual control group	22	1315	Std. Mean Difference (IV, Random, 95% CI)	0.41 [0.22, 0.60]

Analysis 6.1. Comparison 6: Group cognitive stimulation versus no treatment post-treatment (Active control group), Outcome 1: Cognition

Study or Subgroup	Control Mean	Control SD	Control Total	Cognitive Stimulation Mean	Cognitive Stimulation SD	Cognitive Stimulation Total	Weight	Std. Mean Difference IV, Random, 95% CI
6.1.1 Alternate activity control group								
Buschert 2011	0.7	8	8	0	6.93	7	1.9%	0.09 [-0.93 , 1.10]
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	3.6%	0.21 [-0.42 , 0.84]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	6.0%	0.70 [0.41 , 1.00]
Cheung 2019	0.94	2.6745	18	-0.5	2.5969	12	3.0%	0.53 [-0.21 , 1.27]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	3.8%	0.70 [0.12 , 1.29]
Subtotal (95% CI)			**174**			**148**	**18.3%**	**0.59 [0.37 , 0.82]**
Heterogeneity: Tau² = 0.00; Chi² = 3.09, df = 4 (P = 0.54); I² = 0%								
Test for overall effect: Z = 5.13 (P < 0.00001)								
6.1.2 Treatment-as-usual control group								
Alvares-Pereira 2021	1.418	5.112	55	-1.44	5.199	50	5.3%	0.55 [0.16 , 0.94]
Baldelli 1993	3	5.32	13	-4.4	9.15	10	2.4%	0.99 [0.11 , 1.87]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	4.1%	0.50 [-0.04 , 1.05]
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	1.7%	0.28 [-0.82 , 1.38]
Breuil 1994	5.8	7.3	29	1	7.8	27	4.2%	0.63 [0.09 , 1.17]
Coen 2011	0.2	7.2	13	2.3	4.1	12	2.7%	-0.34 [-1.13 , 0.45]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	3.9%	0.14 [-0.43 , 0.71]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	4.4%	0.25 [-0.25 , 0.75]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	4.1%	1.03 [0.48 , 1.58]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	4.0%	0.28 [-0.27 , 0.83]
Lok 2020	3	6.39	30	-2.16	6.33	30	4.2%	0.80 [0.27 , 1.33]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	2.4%	0.04 [-0.84 , 0.91]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.8%	0.23 [-0.82 , 1.28]
Mapelli 2013	8.7	10.88	10	-2.2	10.21	10	2.2%	0.99 [0.05 , 1.93]
Marinho 2021	-1.6	13.19	24	-1.8	15.63	26	4.0%	0.01 [-0.54 , 0.57]
Middelstädt 2016	-0.5	14.42	35	2	13.94	33	4.6%	-0.17 [-0.65 , 0.30]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	6.2%	-0.06 [-0.34 , 0.22]
Paddick 2017	8.1	11.96	16	0.8	9.74	18	3.2%	0.66 [-0.04 , 1.35]
Spector 2001	4.3	17.33	17	-1	20.5	10	2.8%	0.28 [-0.51 , 1.06]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	5.9%	0.37 [0.06 , 0.68]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	2.7%	0.21 [-0.59 , 1.01]
Young 2019	10.65	11.07	51	-1.4	6.41	50	4.9%	1.32 [0.89 , 1.75]
Subtotal (95% CI)			**728**			**587**	**81.7%**	**0.41 [0.22 , 0.60]**
Heterogeneity: Tau² = 0.12; Chi² = 53.41, df = 21 (P = 0.0001); I² = 61%								
Test for overall effect: Z = 4.17 (P < 0.0001)								
Total (95% CI)			**902**			**735**	**100.0%**	**0.43 [0.26 , 0.59]**
Heterogeneity: Tau² = 0.09; Chi² = 59.26, df = 26 (P = 0.0002); I² = 56%								
Test for overall effect: Z = 5.12 (P < 0.00001)								
Test for subgroup differences: Chi² = 1.45, df = 1 (P = 0.23), I² = 30.9%								

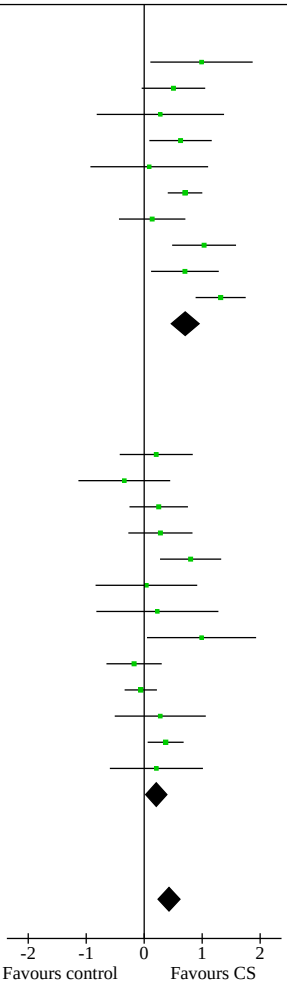


Comparison 7. Group cognitive stimulation versus no cognitive stimulation post-treatment (dementia severity)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
7.1 Cognition	23	1418	Std. Mean Difference (IV, Random, 95% CI)	0.43 [0.24, 0.62]
7.1.1 Mild impairment	10	640	Std. Mean Difference (IV, Random, 95% CI)	0.71 [0.47, 0.95]
7.1.2 Moderate impairment	13	778	Std. Mean Difference (IV, Random, 95% CI)	0.21 [0.03, 0.39]
7.2 MMSE	21	1325	Mean Difference (IV, Random, 95% CI)	2.16 [1.58, 2.74]
7.2.1 Mild impairment	10	676	Mean Difference (IV, Random, 95% CI)	2.59 [1.87, 3.31]
7.2.2 Moderate impairment	11	649	Mean Difference (IV, Random, 95% CI)	1.42 [0.78, 2.07]
7.3 ADAS-Cog	14	979	Mean Difference (IV, Random, 95% CI)	2.52 [0.59, 4.46]
7.3.1 Mild impairment	6	373	Mean Difference (IV, Random, 95% CI)	4.98 [1.95, 8.00]
7.3.2 Moderate impairment	8	606	Mean Difference (IV, Random, 95% CI)	1.10 [-0.29, 2.48]
7.4 Quality of life (self-report)	11	903	Mean Difference (IV, Random, 95% CI)	2.41 [0.66, 4.17]
7.4.1 Mild impairment	3	276	Mean Difference (IV, Random, 95% CI)	1.54 [0.07, 3.00]
7.4.2 Moderate impairment	8	627	Mean Difference (IV, Random, 95% CI)	3.07 [0.77, 5.36]
7.5 Quality of life (proxy rated)	5	359	Std. Mean Difference (IV, Random, 95% CI)	0.45 [-0.01, 0.90]
7.5.1 Moderate impairment	5	359	Std. Mean Difference (IV, Random, 95% CI)	0.45 [-0.01, 0.90]
7.6 Mood: self-reported	6	299	Std. Mean Difference (IV, Random, 95% CI)	0.20 [-0.06, 0.45]
7.6.1 Mild impairment	4	220	Std. Mean Difference (IV, Random, 95% CI)	0.30 [0.01, 0.60]
7.6.2 Moderate impairment	2	79	Std. Mean Difference (IV, Random, 95% CI)	-0.05 [-0.53, 0.42]

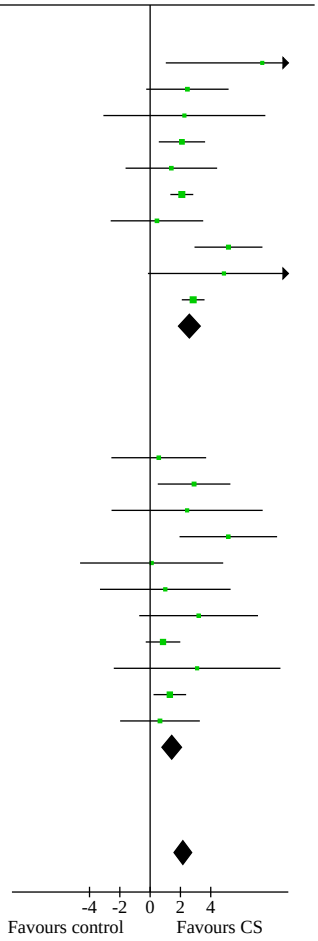
Analysis 7.1. Comparison 7: Group cognitive stimulation versus no cognitive stimulation post-treatment (dementia severity), Outcome 1: Cognition

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
7.1.1 Mild impairment								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	2.9%	0.99 [0.11 , 1.87]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	4.8%	0.50 [-0.04 , 1.05]
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	2.2%	0.28 [-0.82 , 1.38]
Breuil 1994	5.8	7.3	29	1	7.8	27	4.9%	0.63 [0.09 , 1.17]
Buschert 2011	0.7	8	8	0	6.93	7	2.4%	0.09 [-0.93 , 1.10]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	6.8%	0.70 [0.41 , 1.00]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	4.7%	0.14 [-0.43 , 0.71]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	4.8%	1.03 [0.48 , 1.58]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	4.6%	0.70 [0.12 , 1.29]
Young 2019	10.65	11.07	51	-1.4	6.41	50	5.7%	1.32 [0.89 , 1.75]
Subtotal (95% CI)			**366**			**274**	**43.8%**	**0.71 [0.47 , 0.95]**
Heterogeneity: Tau² = 0.06; Chi² = 15.75, df = 9 (P = 0.07); I² = 43%								
Test for overall effect: Z = 5.78 (P < 0.00001)								
7.1.2 Moderate impairment								
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	4.3%	0.21 [-0.42 , 0.84]
Coen 2011	0.2	7.2	13	2.3	4.1	12	3.4%	-0.34 [-1.13 , 0.45]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	5.2%	0.25 [-0.25 , 0.75]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	4.8%	0.28 [-0.27 , 0.83]
Lok 2020	3	6.39	30	-2.16	6.33	30	5.0%	0.80 [0.27 , 1.33]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	3.0%	0.04 [-0.84 , 0.91]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	2.3%	0.23 [-0.82 , 1.28]
Mapelli 2013	8.7	10.88	10	-2.2	10.21	10	2.7%	0.99 [0.05 , 1.93]
Middelstädt 2016	-0.5	14.42	35	2	13.94	33	5.4%	-0.17 [-0.65 , 0.30]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	6.9%	-0.06 [-0.34 , 0.22]
Spector 2001	4.3	17.33	17	-1	20.5	10	3.4%	0.28 [-0.51 , 1.06]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	6.7%	0.37 [0.06 , 0.68]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	3.3%	0.21 [-0.59 , 1.01]
Subtotal (95% CI)			**423**			**355**	**56.2%**	**0.21 [0.03 , 0.39]**
Heterogeneity: Tau² = 0.03; Chi² = 16.61, df = 12 (P = 0.17); I² = 28%								
Test for overall effect: Z = 2.23 (P = 0.03)								
Total (95% CI)			**789**			**629**	**100.0%**	**0.43 [0.24 , 0.62]**
Heterogeneity: Tau² = 0.11; Chi² = 56.23, df = 22 (P < 0.0001); I² = 61%								
Test for overall effect: Z = 4.44 (P < 0.00001)								
Test for subgroup differences: Chi² = 10.53, df = 1 (P = 0.001), I² = 90.5%								



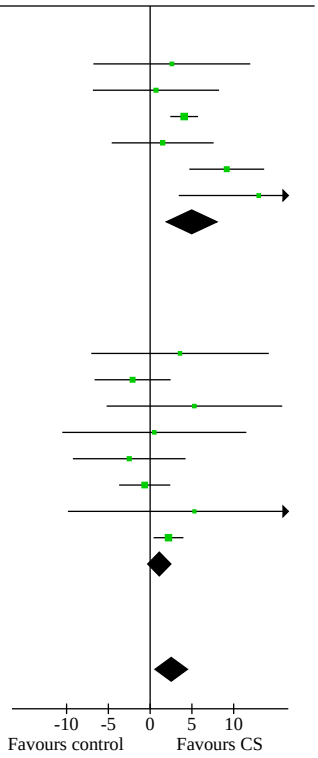
Analysis 7.2. Comparison 7: Group cognitive stimulation versus no cognitive stimulation post-treatment (dementia severity), Outcome 2: MMSE

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
7.2.1 Mild impairment								
Baldelli 1993	3	5.32	13	-4.4	9.15	10	0.8%	7.40 [1.03 , 13.77]
Baldelli 2002	2.34	4.78	71	-0.12	5.06	16	3.7%	2.46 [-0.26 , 5.18]
Bottino 2005	0.83	4.53	6	-1.43	5.3	7	1.1%	2.26 [-3.08 , 7.60]
Breuil 1994	1.4	2.7	29	-0.7	3.1	27	8.3%	2.10 [0.57 , 3.63]
Buschert 2011	0.5	3.14	8	-0.9	2.83	7	3.1%	1.40 [-1.62 , 4.42]
Carbone 2021	0.8894	2.68673	123	-1.2016	3.01666	101	14.6%	2.09 [1.33 , 2.85]
Cove 2014	-0.33	6.06	24	-0.78	4.54	23	3.1%	0.45 [-2.60 , 3.50]
Juarez-Cedillo 2020	2.05	4.41	36	-3.12	4.29	24	5.0%	5.17 [2.93 , 7.41]
Requena 2006	1.5	7.38	20	-3.37	10.71	30	1.3%	4.87 [-0.14 , 9.88]
Young 2019	2.1	2.26	51	-0.74	1.52	50	14.7%	2.84 [2.09 , 3.59]
Subtotal (95% CI)			**381**			**295**	**55.7%**	**2.59 [1.87 , 3.31]**
Heterogeneity: Tau² = 0.33; Chi² = 13.01, df = 9 (P = 0.16); I² = 31%								
Test for overall effect: Z = 7.07 (P < 0.00001)								
7.2.2 Moderate impairment								
Capotosto 2017	0.18	4.57	20	-0.39	5.34	19	2.9%	0.57 [-2.56 , 3.70]
Coen 2011	0.8	3.6	14	-2.1	2.5	11	4.5%	2.90 [0.50 , 5.30]
Kim 2016	0.53	7.64	32	-1.91	9.88	21	1.3%	2.44 [-2.55 , 7.43]
Lok 2020	3	6.39	30	-2.16	6.33	30	2.8%	5.16 [1.94 , 8.38]
Lopez 2020	-1.6	4.82	10	-1.7	5.9	10	1.4%	0.10 [-4.62 , 4.82]
Maci 2012	-0.2	4.26	7	-1.2	3.96	7	1.7%	1.00 [-3.31 , 5.31]
Mapelli 2013	2.9	5.03	10	-0.3	3.83	10	2.0%	3.20 [-0.72 , 7.12]
Orrell 2014	-1.46	4.0938	106	-2.31	4.0938	93	11.1%	0.85 [-0.29 , 1.99]
Spector 2001	3.1	7.04	17	0	7.04	10	1.1%	3.10 [-2.40 , 8.60]
Spector 2003	0.9	3.5	97	-0.4	3.5	70	11.7%	1.30 [0.22 , 2.38]
Tanaka 2021	1.9	2.3238	15	1.25	3.7947	10	3.9%	0.65 [-1.98 , 3.28]
Subtotal (95% CI)			**358**			**291**	**44.3%**	**1.42 [0.78 , 2.07]**
Heterogeneity: Tau² = 0.00; Chi² = 9.92, df = 10 (P = 0.45); I² = 0%								
Test for overall effect: Z = 4.31 (P < 0.0001)								
Total (95% CI)			**739**			**586**	**100.0%**	**2.16 [1.58 , 2.74]**
Heterogeneity: Tau² = 0.45; Chi² = 30.24, df = 20 (P = 0.07); I² = 34%								
Test for overall effect: Z = 7.27 (P < 0.00001)								
Test for subgroup differences: Chi² = 5.58, df = 1 (P = 0.02), I² = 82.1%								



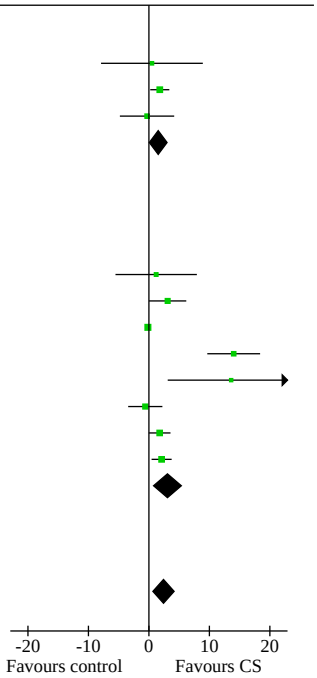
Analysis 7.3. Comparison 7: Group cognitive stimulation versus no cognitive stimulation post-treatment (dementia severity), Outcome 3: ADAS-Cog

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
7.3.1 Mild impairment								
Bottino 2005	2.17	8.33	6	-0.43	8.92	7	3.5%	2.60 [-6.79 , 11.99]
Buschert 2011	0.7	8	8	0	6.93	7	4.9%	0.70 [-6.86 , 8.26]
Carbone 2021	2.313	5.843	108	-1.762	5.649	80	17.1%	4.08 [2.42 , 5.73]
Cove 2014	-0.91	11.58	24	-2.41	9.71	23	6.6%	1.50 [-4.60 , 7.60]
Juarez-Cedillo 2020	5.3	8.94	36	-3.87	8.48	24	9.6%	9.17 [4.69 , 13.65]
Requena 2006	6.4	14.06	20	-6.6	20.48	30	3.4%	13.00 [3.43 , 22.57]
Subtotal (95% CI)			**202**			**171**	**45.1%**	**4.98 [1.95 , 8.00]**
Heterogeneity: Tau² = 5.97; Chi² = 9.52, df = 5 (P = 0.09); I² = 47%								
Test for overall effect: Z = 3.22 (P = 0.001)								
7.3.2 Moderate impairment								
Capotosto 2017	0.9	16.03	20	-2.68	17.73	19	2.8%	3.58 [-7.05 , 14.21]
Coen 2011	0.2	7.2	13	2.3	4.1	12	9.4%	-2.10 [-6.65 , 2.45]
Graessel 2011	0.1	19.14	31	-5.2	22.54	30	2.9%	5.30 [-5.21 , 15.81]
Lopez 2020	-1.8	11.53	10	-2.3	13.53	10	2.7%	0.50 [-10.52 , 11.52]
Middelstädt 2016	-0.5	14.42	35	2	13.95	33	5.8%	-2.50 [-9.24 , 4.24]
Orrell 2014	-3.83	10.9886	106	-3.18	10.9886	93	13.1%	-0.65 [-3.71 , 2.41]
Spector 2001	4.3	17.33	17	-1	20.5	10	1.5%	5.30 [-9.84 , 20.44]
Spector 2003	1.9	6.2	97	-0.3	5.5	70	16.7%	2.20 [0.42 , 3.98]
Subtotal (95% CI)			**329**			**277**	**54.9%**	**1.10 [-0.29 , 2.48]**
Heterogeneity: Tau² = 0.00; Chi² = 6.84, df = 7 (P = 0.45); I² = 0%								
Test for overall effect: Z = 1.55 (P = 0.12)								
Total (95% CI)			**531**			**448**	**100.0%**	**2.52 [0.59 , 4.46]**
Heterogeneity: Tau² = 5.02; Chi² = 27.52, df = 13 (P = 0.01); I² = 53%								
Test for overall effect: Z = 2.55 (P = 0.01)								
Test for subgroup differences: Chi² = 5.21, df = 1 (P = 0.02), I² = 80.8%								



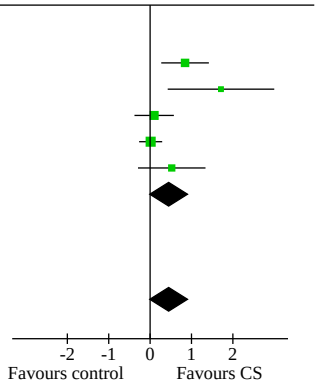
Analysis 7.4. Comparison 7: Group cognitive stimulation versus no cognitive stimulation post-treatment (dementia severity), Outcome 4: Quality of life (self-report)

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
7.4.1 Mild impairment								
Buschert 2011	-0.4	10.61	8	-0.9	5.52	7	3.4%	0.50 [-7.91 , 8.91]
Carbone 2021	1.6	6.63	118	-0.2	5.12	96	13.2%	1.80 [0.23 , 3.37]
Cove 2014	0.23	7.97	24	0.54	7.74	23	7.5%	-0.31 [-4.80 , 4.18]
Subtotal (95% CI)			**150**			**126**	**24.0%**	**1.54 [0.07 , 3.00]**
Heterogeneity: Tau² = 0.00; Chi² = 0.82, df = 2 (P = 0.67); I² = 0%								
Test for overall effect: Z = 2.06 (P = 0.04)								
7.4.2 Moderate impairment								
Capotosto 2017	1.25	11.5	20	0.05	9.91	19	4.6%	1.20 [-5.53 , 7.93]
Coen 2011	3.6	3.7	14	0.5	4.4	13	10.1%	3.10 [0.02 , 6.18]
Kim 2016	-0.4	0.76	32	-0.23	0.73	21	14.6%	-0.17 [-0.58 , 0.24]
Lok 2020	11.86	9.64	30	-2.16	7.48	30	7.7%	14.02 [9.65 , 18.39]
Maci 2012	12.3	11.78	7	-1.3	7.86	7	2.3%	13.60 [3.11 , 24.09]
Middelstädt 2016	-0.2	5.47	35	0.4	6.36	33	10.7%	-0.60 [-3.43 , 2.23]
Orrell 2014	-0.67	6.428	106	-2.45	6.428	93	12.8%	1.78 [-0.01 , 3.57]
Spector 2003	1.3	5.1	97	-0.8	5.6	70	13.0%	2.10 [0.44 , 3.76]
Subtotal (95% CI)			**341**			**286**	**76.0%**	**3.07 [0.77 , 5.36]**
Heterogeneity: Tau² = 7.62; Chi² = 59.74, df = 7 (P < 0.00001); I² = 88%								
Test for overall effect: Z = 2.62 (P = 0.009)								
Total (95% CI)			**491**			**412**	**100.0%**	**2.41 [0.66 , 4.17]**
Heterogeneity: Tau² = 5.42; Chi² = 63.53, df = 10 (P < 0.00001); I² = 84%								
Test for overall effect: Z = 2.70 (P = 0.007)								
Test for subgroup differences: Chi² = 1.21, df = 1 (P = 0.27), I² = 17.6%								

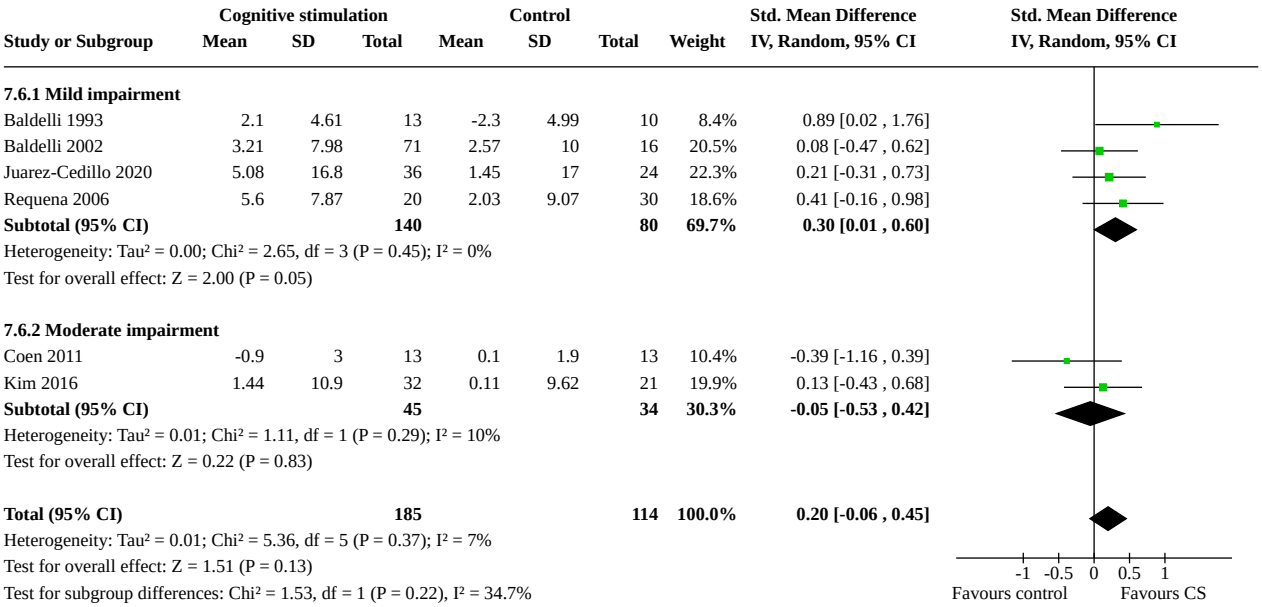


Analysis 7.5. Comparison 7: Group cognitive stimulation versus no cognitive stimulation post-treatment (dementia severity), Outcome 5: Quality of life (proxy rated)

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
7.5.1 Moderate impairment								
Kim 2016	0.81	2.34	32	-1.09	2.01	21	21.5%	0.84 [0.27 , 1.42]
Maci 2012	10.2	8.08	7	-1.4	3.89	7	9.0%	1.71 [0.43 , 3.00]
Middelstädt 2016	-0.4	6.16	35	-1	5.94	33	24.2%	0.10 [-0.38 , 0.57]
Orrell 2014	0.62	5.2429	106	0.55	5.2429	93	29.3%	0.01 [-0.27 , 0.29]
Tanaka 2021	1.5	6.9714	15	-1.8	4.4272	10	16.0%	0.52 [-0.29 , 1.34]
Subtotal (95% CI)			**195**			**164**	**100.0%**	**0.45 [-0.01 , 0.90]**
Heterogeneity: Tau² = 0.16; Chi² = 12.59, df = 4 (P = 0.01); I² = 68%								
Test for overall effect: Z = 1.93 (P = 0.05)								
Total (95% CI)			**195**			**164**	**100.0%**	**0.45 [-0.01 , 0.90]**
Heterogeneity: Tau² = 0.16; Chi² = 12.59, df = 4 (P = 0.01); I² = 68%								
Test for overall effect: Z = 1.93 (P = 0.05)								
Test for subgroup differences: Not applicable								



Analysis 7.6. Comparison 7: Group cognitive stimulation versus no cognitive stimulation post-treatment (dementia severity), Outcome 6: Mood: self-reported



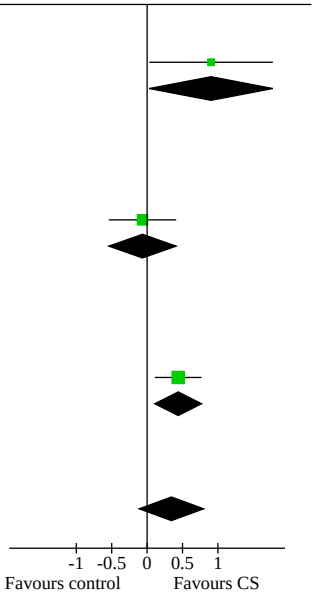
Comparison 8. Cognitive stimulation versus no cognitive stimulation: follow-up

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
8.1 Cognition - 6 to 12-week follow-up	3	242	Std. Mean Difference (IV, Random, 95% CI)	0.34 [-0.11, 0.80]
8.1.1 Three months follow-up MMSE	1	23	Std. Mean Difference (IV, Random, 95% CI)	0.90 [0.03, 1.78]
8.1.2 Six week follow-up ADAS-Cog	1	68	Std. Mean Difference (IV, Random, 95% CI)	-0.07 [-0.54, 0.41]
8.1.3 Three-months follow-up ADAS-Cog	1	151	Std. Mean Difference (IV, Random, 95% CI)	0.44 [0.11, 0.77]
8.2 Cognition - 8 to 12-month follow-up	4	194	Std. Mean Difference (IV, Random, 95% CI)	0.13 [-0.16, 0.42]
8.2.1 Ten - twelve months follow-up ADAS-Cog	3	156	Std. Mean Difference (IV, Random, 95% CI)	0.23 [-0.09, 0.55]
8.2.2 Eight months follow-up CAM-COG	1	38	Std. Mean Difference (IV, Random, 95% CI)	-0.25 [-0.90, 0.39]
8.3 Cognition: MMSE - Three-month follow-up	2	210	Mean Difference (IV, Random, 95% CI)	3.16 [-0.99, 7.31]
8.4 Cognition: MMSE - 8 to 12-month follow-up	3	142	Mean Difference (IV, Random, 95% CI)	0.53 [-0.63, 1.70]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
8.5 Quality of Life: self-report & proxy measures	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
8.5.1 Six to twelve week follow-up QoL-AD	2	254	Mean Difference (IV, Random, 95% CI)	0.36 [-2.47, 3.20]
8.5.2 Six week follow-up QoL-AD proxy	1	68	Mean Difference (IV, Random, 95% CI)	-0.30 [-3.15, 2.55]
8.5.3 Ten months follow-up QoL-AD	1	54	Mean Difference (IV, Random, 95% CI)	2.15 [-1.12, 5.42]
8.5.4 Ten months follow-up QoL-AD proxy	1	54	Mean Difference (IV, Random, 95% CI)	-0.28 [-3.14, 2.58]
8.6 Communication and social interaction	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.6.1 Twelve week follow-up Narrative Language - communicative abilities	1	182	Std. Mean Difference (IV, Random, 95% CI)	0.33 [0.04, 0.63]
8.6.2 Ten month follow-up 'Relevance of discourse'	1	54	Std. Mean Difference (IV, Random, 95% CI)	0.15 [-0.38, 0.69]
8.7 Mood	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.7.1 Twelve week follow-up: interviewer / staff-rated	1	187	Std. Mean Difference (IV, Random, 95% CI)	0.54 [0.24, 0.83]
8.7.2 Twelve month follow-up: self-report	1	50	Std. Mean Difference (IV, Random, 95% CI)	0.36 [-0.23, 0.94]
8.8 ADL/IADL scales	5		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
8.8.1 Six to twelve weeks follow-up IADL scales	2	176	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.19, 0.42]
8.8.2 Ten to twelve months follow-up ADL / IADL scales	3	156	Std. Mean Difference (IV, Random, 95% CI)	0.40 [0.07, 0.72]
8.9 Behaviour that challenges	4		Std. Mean Difference (IV, Fixed, 95% CI)	Subtotals only
8.9.1 Six to twelve weeks follow-up NPI	2	255	Std. Mean Difference (IV, Fixed, 95% CI)	0.19 [-0.06, 0.44]
8.9.2 Ten to twelve months follow-up NPI severity	2	104	Std. Mean Difference (IV, Fixed, 95% CI)	0.43 [0.03, 0.83]
8.9.3 Ten-month follow up NPI (Caregiver Distress)	1	54	Std. Mean Difference (IV, Fixed, 95% CI)	0.41 [-0.13, 0.95]

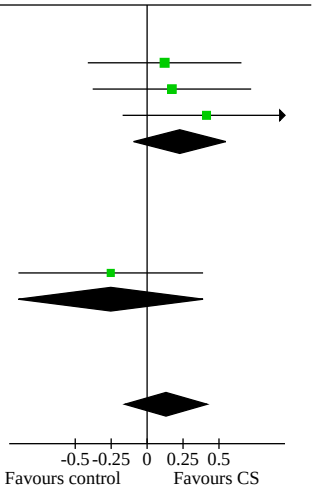
Analysis 8.1. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 1: Cognition - 6 to 12-week follow-up

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.1.1 Three months follow-up MMSE								
Baldelli 1993	0.9	5.37	13	-5.6	8.6	10	18.6%	0.90 [0.03 , 1.78]
Subtotal (95% CI)			**13**			**10**	**18.6%**	**0.90 [0.03 , 1.78]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.03 (P = 0.04)								
8.1.2 Six week follow-up ADAS-Cog								
Middelstädt 2016	-0.2	13.87	35	0.7	13.41	33	36.0%	-0.07 [-0.54 , 0.41]
Subtotal (95% CI)			**35**			**33**	**36.0%**	**-0.07 [-0.54 , 0.41]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.27 (P = 0.79)								
8.1.3 Three-months follow-up ADAS-Cog								
Carbone 2021	0.939	6.106	91	-2.0389	7.631	60	45.3%	0.44 [0.11 , 0.77]
Subtotal (95% CI)			**91**			**60**	**45.3%**	**0.44 [0.11 , 0.77]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.61 (P = 0.009)								
Total (95% CI)			**139**			**103**	**100.0%**	**0.34 [-0.11 , 0.80]**
Heterogeneity: Tau² = 0.09; Chi² = 4.72, df = 2 (P = 0.09); I² = 58%								
Test for overall effect: Z = 1.48 (P = 0.14)								
Test for subgroup differences: Chi² = 4.72, df = 2 (P = 0.09), I² = 57.7%								



Analysis 8.2. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 2: Cognition - 8 to 12-month follow-up

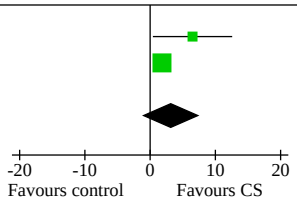
Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.2.1 Ten - twelve months follow-up ADAS-Cog								
Chapman 2004	-4.89	5.78	26	-5.62	6.02	28	28.8%	0.12 [-0.41 , 0.66]
Graessel 2011	-7.46	20.56	30	-11.32	23.87	22	27.1%	0.17 [-0.38 , 0.72]
Juarez-Cedillo 2020	-2.47	9.45	32	-6.68	11	18	24.2%	0.41 [-0.17 , 1.00]
Subtotal (95% CI)			**88**			**68**	**80.1%**	**0.23 [-0.09 , 0.55]**
Heterogeneity: Tau² = 0.00; Chi² = 0.58, df = 2 (P = 0.75); I² = 0%								
Test for overall effect: Z = 1.39 (P = 0.17)								
8.2.2 Eight months follow-up CAM-COG								
Tsantali 2017	-6.1	12.48	17	-3.2	10.11	21	19.9%	-0.25 [-0.90 , 0.39]
Subtotal (95% CI)			**17**			**21**	**19.9%**	**-0.25 [-0.90 , 0.39]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.77 (P = 0.44)								
Total (95% CI)			**105**			**89**	**100.0%**	**0.13 [-0.16 , 0.42]**
Heterogeneity: Tau² = 0.00; Chi² = 2.29, df = 3 (P = 0.51); I² = 0%								
Test for overall effect: Z = 0.90 (P = 0.37)								
Test for subgroup differences: Chi² = 1.72, df = 1 (P = 0.19), I² = 41.8%								



Analysis 8.3. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 3: Cognition: MMSE - Three-month follow-up

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Baldelli 1993	0.9	5.37	13	-5.6	8.6	10	28.6%	6.50 [0.42 , 12.58]
Carbone 2021	0.3464	2.82566	106	-1.4691	3.52167	81	71.4%	1.82 [0.88 , 2.75]
Total (95% CI)			**119**			**91**	**100.0%**	**3.16 [-0.99 , 7.31]**

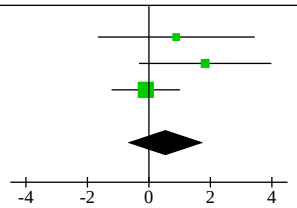
Heterogeneity: Tau² = 6.05; Chi² = 2.23, df = 1 (P = 0.14); I² = 55%
Test for overall effect: Z = 1.49 (P = 0.14)
Test for subgroup differences: Not applicable



Analysis 8.4. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 4: Cognition: MMSE - 8 to 12-month follow-up

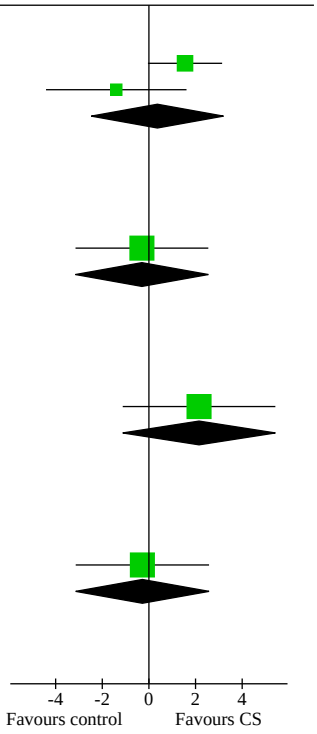
Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Chapman 2004	-1.25	3.98	26	-2.14	5.51	28	17.9%	0.89 [-1.66 , 3.44]
Juarez-Cedillo 2020	-1.17	3.93	32	-3	3.61	18	23.7%	1.83 [-0.32 , 3.98]
Tsantali 2017	-1.6	1.35	17	-1.5	2.13	21	58.4%	-0.10 [-1.21 , 1.01]
Total (95% CI)			**75**			**67**	**100.0%**	**0.53 [-0.63 , 1.70]**

Heterogeneity: Tau² = 0.28; Chi² = 2.61, df = 2 (P = 0.27); I² = 23%
Test for overall effect: Z = 0.90 (P = 0.37)
Test for subgroup differences: Not applicable



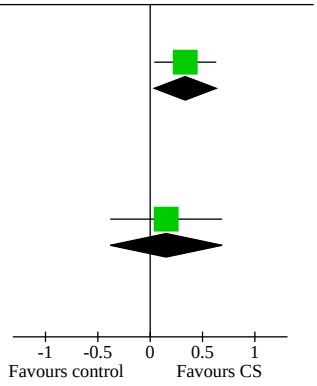
Analysis 8.5. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 5: Quality of Life: self-report & proxy measures

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
8.5.1 Six to twelve week follow-up QoL-AD								
Carbone 2021	1.009	6.8	106	-0.54	4.22	80	59.8%	1.55 [-0.04 , 3.14]
Middelstädt 2016	-1.1	5.8	35	0.3	6.8	33	40.2%	-1.40 [-4.41 , 1.61]
Subtotal (95% CI)			**141**			**113**	**100.0%**	**0.36 [-2.47 , 3.20]**
Heterogeneity: Tau² = 2.84; Chi² = 2.88, df = 1 (P = 0.09); I² = 65%								
Test for overall effect: Z = 0.25 (P = 0.80)								
8.5.2 Six week follow-up QoL-AD proxy								
Middelstädt 2016	-1.7	5.8	35	-1.4	6.16	33	100.0%	-0.30 [-3.15 , 2.55]
Subtotal (95% CI)			**35**			**33**	**100.0%**	**-0.30 [-3.15 , 2.55]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.21 (P = 0.84)								
8.5.3 Ten months follow-up QoL-AD								
Chapman 2004	2.05	5.98	26	-0.1	6.29	28	100.0%	2.15 [-1.12 , 5.42]
Subtotal (95% CI)			**26**			**28**	**100.0%**	**2.15 [-1.12 , 5.42]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.29 (P = 0.20)								
8.5.4 Ten months follow-up QoL-AD proxy								
Chapman 2004	1.05	5.8677	26	1.33	4.7452	28	100.0%	-0.28 [-3.14 , 2.58]
Subtotal (95% CI)			**26**			**28**	**100.0%**	**-0.28 [-3.14 , 2.58]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.19 (P = 0.85)								



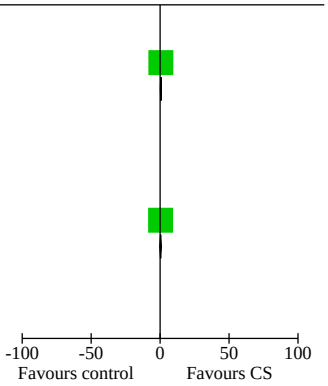
Analysis 8.6. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 6: Communication and social interaction

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.6.1 Twelve week follow-up Narrative Language - communicative abilities								
Carbone 2021	1.27	4.68	105	-0.27	4.44	77	100.0%	0.33 [0.04 , 0.63]
Subtotal (95% CI)			**105**			**77**	**100.0%**	**0.33 [0.04 , 0.63]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.22 (P = 0.03)								
8.6.2 Ten month follow-up 'Relevance of discourse'								
Chapman 2004	-3.35	10.5717	26	-5.05	11.3214	28	100.0%	0.15 [-0.38 , 0.69]
Subtotal (95% CI)			**26**			**28**	**100.0%**	**0.15 [-0.38 , 0.69]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.56 (P = 0.58)								



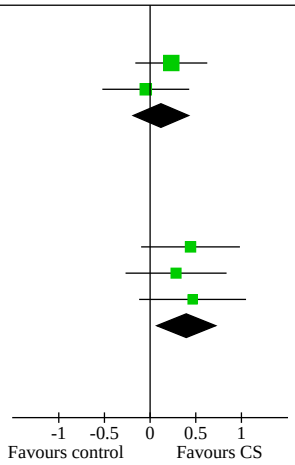
Analysis 8.7. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 7: Mood

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.7.1 Twelve week follow-up: interviewer / staff-rated								
Carbone 2021	0.887	3.36	106	-0.901	3.25	81	100.0%	0.54 [0.24 , 0.83]
Subtotal (95% CI)			**106**			**81**	**100.0%**	**0.54 [0.24 , 0.83]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 3.58 (P = 0.0003)								
8.7.2 Twelve month follow-up: self-report								
Juarez-Cedillo 2020	12.9	20.1	32	6.31	14.1	18	100.0%	0.36 [-0.23 , 0.94]
Subtotal (95% CI)			**32**			**18**	**100.0%**	**0.36 [-0.23 , 0.94]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.20 (P = 0.23)								

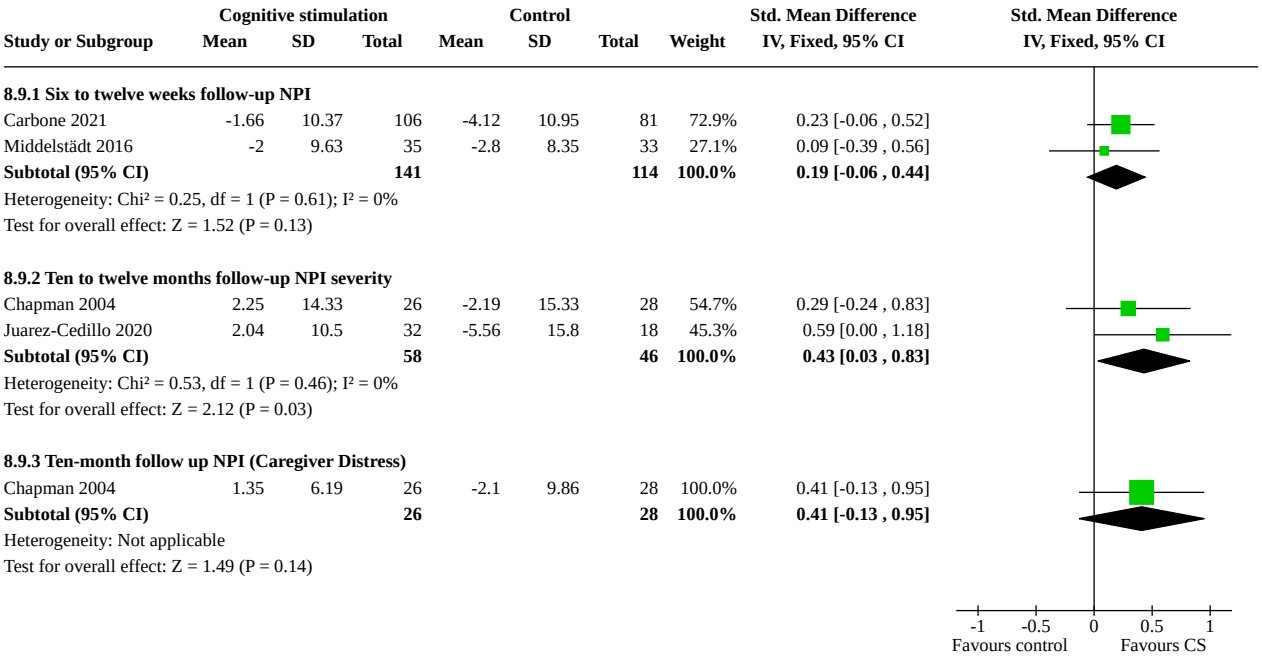


Analysis 8.8. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 8: ADL/IADL scales

Study or Subgroup	Cognitive stimulation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
8.8.1 Six to twelve weeks follow-up IADL scales								
Carbone 2021	-2.91	12.89	69	-6.55	19.48	39	59.3%	0.23 [-0.16 , 0.63]
Middelstädt 2016	-1.8	21.35	35	-0.7	24.55	33	40.7%	-0.05 [-0.52 , 0.43]
Subtotal (95% CI)			**104**			**72**	**100.0%**	**0.12 [-0.19 , 0.42]**
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.79$, df = 1 (P = 0.38); $I^2 = 0\%$								
Test for overall effect: Z = 0.76 (P = 0.44)								
8.8.2 Ten to twelve months follow-up ADL / IADL scales								
Chapman 2004	-2.89	7.4522	26	-6.86	9.9546	28	35.6%	0.44 [-0.10 , 0.98]
Graessel 2011	-5.7	11.3	30	-8.92	10.97	22	34.0%	0.28 [-0.27 , 0.84]
Juarez-Cedillo 2020	-2.21	7.3	32	-5.81	8.16	18	30.4%	0.47 [-0.12 , 1.05]
Subtotal (95% CI)			**88**			**68**	**100.0%**	**0.40 [0.07 , 0.72]**
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.24$, df = 2 (P = 0.89); $I^2 = 0\%$								
Test for overall effect: Z = 2.40 (P = 0.02)								



Analysis 8.9. Comparison 8: Cognitive stimulation versus no cognitive stimulation: follow-up, Outcome 9: Behaviour that challenges



ADDITIONAL TABLES

Table 1. Risk of bias assessment table

Domain	Risk of bias judgement		
Selection bias	**Low**	**High**	**Unclear**
Random sequence generation	Assigned if simple randomisation was used (e.g. computer-generated random sequence, coin tossing). Assigned if restricted randomisation was used (e.g. block randomisation, provided that within groups randomisation was not affected).	Assigned if study reported an inadequate randomisation method (e.g. using date of birth or odd/even numbers).	Assigned if there was insufficient detail to judge the risk of bias as low or high.
Allocation Concealment	Assigned if there was evidence of concealed allocation sequence in which allocations could not have been foreseen in advance of, or during, enrolment.	Assigned if those enrolling participants were aware of the group (or period in a cross-over trial) to which the next enrolled participant would be allocated.	Assigned if there was insufficient detail to judge the risk of bias as low or high.
Detection bias	**Low**	**High**	**Unclear**
Blinding of outcome assessors (blinding of participants and facilitators is not possible in psychosocial interventions).	Assigned if outcome assessors were blind to treatment allocation.	Assigned if the outcome assessors were aware of treatment allocation (e.g. if the cognitive stimulation group leader was also an outcome assessor).	Assigned if there was insufficient detail to judge the risk of bias as low or high.

Table 1. Risk of bias assessment table *(Continued)*

Attrition bias	**Low**	**High**	**Unclear**
Incomplete outcome data	Assigned if the study reported levels of attrition, reasons for attrition and how missing data were dealt with. Assigned if the impact of missing data was not believed to alter the conclusions and there were acceptable reasons for the missing data.	Assigned if there was inadequate information regarding the level of attrition in each group, reasons for attrition and if missing data were not handled correctly.	Assigned if there was insufficient detail to judge the risk of bias as low or high.
Reporting bias	**Low**	**High**	**Unclear**
Selective reporting	Assigned if study reported results of all outcome measures that were detailed in the methods section. If a study protocol was available, low risk of bias was assigned if the outcome assessments reported in the trial paper matched those detailed in the protocol.	Assigned if study did not report results of all outcome measures that were detailed in the methods section. Assigned if all outcome measures detailed in the protocol (if available) were not reported in the study.	Assigned if there was insufficient detail to judge the risk of bias as low or high.
Other bias	**Low**	**High**	**Unclear**
Availability of training and supervision	Assigned if cognitive stimulation sessions were facilitated by people who had received some form of training to ensure the necessary principles of cognitive stimulation were adhered to. The definition of training was inclusive and could range from a brief session to a longer, more intensive course. This also applied to interventions delivered by trained family carers. The opportunity for facilitators to access appropriate supervision was also desirable.	Assigned if there was no evidence of facilitator training or supervision.	Assigned if there was insufficient detail to judge the risk of bias as low or high.
Availability of manual, structure or protocol	Assigned if there was evidence of a documented intervention protocol, structure or manual outlining the content of each session to ensure the principles of cognitive stimulation were adhered to.	Assigned if there was no evidence of a treatment protocol, structure or manual for facilitators to follow.	Assigned if there was insufficient detail to judge the risk of bias as low or high.

Table 2. Summary of key characteristics of included studies

Study ID	Intervention/modality	Setting	Frequency (per week)	Duration (weeks)	Total number of sessions	Session length (minutes)	MMSE Mean (SD)	Follow-up?	Age Mean (SD)	Relevant sample size intervention/control
Ali 2021	Individual iCST adapted from 'Making a Difference' manuals delivered by paid staff or family/friends	Mixed community/supported housing/residential care, UK	2	20	40	30	N/A	None	60.4 (8.2)	20/20
Alvares-Pereira 2021	CST groups using Portugese version of 'Making a Difference' manual	Mixed day-centre/residential settings, Portugal	2	7	14	45	N/A	None	83.6 (7.6)	55/50
Baldelli 1993	RO group sessions	Institution, Italy	3	12	36	60	20.6 (4.9)	3-month	84.5 (6.4)	13/10
Baldelli 2002	RO group sessions	Nursing home, Italy	5	4	20	60	20.7 (3.0)	None	80.0 (7.4)	71 /16
Bottino 2005	'Cognitive rehabilitation' group sessions/carer support group	Outpatients, Brazil	1	20	20	90	22.3 (3.6)	None	73.7 (6.6)	6/7
Breuil 1994	Cognitive stimulation groups	Outpatients, France	2	5	10	60	21.5	None	77.1 (7.1)	29/27
Buschert 2011	Cognitive stimulation groups	Outpatients, all on AChEIs/memantine, Germany	1	26	20	120	24.9 (1.6)	None	75.9 (8.1)	8/7
Capotosto 2017	CST groups using 'Making a Difference' manual	Residential homes, Italy	2	7	14	45	18.2 (3.4)	None	87.4 (5.4)	20/19
Carbone 2021	CST groups using Italian version of 'Making a Difference' manual	Residential homes and day-centres, Italy	2	7	14	45	20.1 (4.0)	3-month follow-up	83.6 (8.1)	123/102

Table 2. Summary of key characteristics of included studies *(Continued)*

Chapman 2004	Cognitive-communication stimulation groups	Outpatients – all on donepezil, USA	1	8	8	90	20.9 (3.6)	6-month and 10-month follow-up	76.4 (7.9)	26/28
Cheung 2019	Cognitive stimulating play intervention groups	2 daycare centres, Hong Kong	1	8	8	45-60	(MoCA 7.9 (4.4))	None	83.2 (7.2)	18/12
Coen 2011	CST Groups using 'Making a Difference' manual	Long-term care, nursing home, Ireland	2	7	14	45	16.9 (5.0)	None	79.8 (5.6)	14/13
Cove 2014	CST Groups using 'Making a Difference' manual	Community, UK	1	14	14	45	22.8 (3.4)	None	77.3 (7.0)	24/23
Gibbor 2020b	Individual iCST adapted from 'Making a Difference' manuals delivered by researchers	Care homes, UK	2	7	14	45	21.7 (3.5)	None	81.9 (10.3)	17/16
Graessel 2011	'MAKS' groups	Nursing homes, Germany	6	52	300	120	14.6 (5.4)	10-month follow-up	85.1 (5.1)	71/68 (6 months) 50/46 (12 months)
Juarez-Cedillo 2020	'SADEM' cognitive stimulation groups	Outpatients, Mexico	2	48	96	90	22.6 (0.9)	12-month follow-up	77.7 (8.2)	39/28
Justo-Henriques 2022	Home-based individual cognitive stimulation delivered by clinical psychologist	Community, Portugal	1	47	47	45	23.2 (3.2)	None	78.9 (7.5)	30/29
Kim 2016	Multi-domain cognitive stimulation groups	Community – all receiving pharmacotherapy, South Korea	5	26	130	60	18.0 (5.8)	None	78.5 (1.5)	32/21
Leroi 2019	Individual cognitive stimulation delivered by informal carers (CST-PD)	Community, UK	2-3	12	24-36	30	N/A	None	Median 75 (range 55-90)	31 /30

Table 2. Summary of key characteristics of included studies *(Continued)*

Lin 2018	Cognitive stimulation groups	Long-term care institutions, Taiwan	1	10	10	50	14.9 (3.7)	3-month follow-up	79.5 (7.7)	30/32
Lok 2020	RAM-based CST groups (using 'Making a Difference' themes and structure)	Community - all receiving AChEIs, Turkey	2	7	14	45	16.9 (4.3)	None	Not stated	30/30
Lopez 2020	Cognitive stimulation groups	Community (daycare centre) - all receiving AChEIs, Spain	3	26	78	60	17.9 (3.9)	None	81.9 (5.5)	10/10
Maci 2012	Cognitive stimulation and physical activity groups	Community (gymnasium) - all receiving AChEIs/memantine/anti-depressants, Italy	5	12	60	120	17.8 (2.8)	None	72.6 (9.5)	7/7
Mapelli 2013	Cognitive stimulation groups	Nursing home, Italy	5	8	40	60	19.5 (3.4)	None	83.7 (4.6)	10/10
Marinho 2021	CST groups using Brazilian version of 'Making a Difference' manual	Outpatients - all receiving AChEIs, Brazil	2 (but both sessions on same day)	7	14	45	N/A	None	77.8 (8.4)	23/24
Middelstädt 2016	NEUROvitalis senseful cognitive stimulation groups	Nursing homes, Germany	2	8	16	60	16.9 (4.5)	6-week follow-up	86.4 (4.5)	36/35
Onder 2005	Individual reality orientation delivered by family carers	Community – all on donepezil, Italy	3	25	75	30	20.1 (3.1)	None	75.8 (7.1)	79/77
Orgeta 2015	Individual cognitive stimulation delivered by informal carers; 'Making a Difference' manual	Community, UK	3	25	75	30	21.2 (4.3)	None	78.2 (7.5)	180/176

Table 2. Summary of key characteristics of included studies *(Continued)*

Orrell 2014	Maintenance cognitive stimulation groups; 'Making a Difference' manual	Care homes and community, UK	1	24	24	45	17.8 (5.5)	None	83.1 (7.6)	123/113
Paddick 2017	Cognitive stimulation groups using adapted 'Making a Difference' manual	Community, Tanzania	2	7	14	45	Mean Clinical Dementia Rating 1.65	8-week follow-up (uncontrolled)	Median 80 (IQR 76.5,85.3)	16/18
Rai 2021	Individual cognitive stimulation app delivered by informal carers based on 'Making a Difference' manual	Community, UK	2-3	11	22-33	30	N/A	None	73.0 (7.7)	31/30
Requena 2006	Cognitive stimulation groups using computer-controlled visual stimuli on TV screen	Community – all on donepezil, Spain	5	52 and 104	250 and 500	45	21.9 (6.3)	None	77.0 (7.5)	20/30
Spector 2001	Cognitive stimulation groups using 'Making a Difference' manual	Mixed community & care home, UK	2	7	14	45	13.1 (4.4)	None	85.7 (6.7)	21 /14
Spector 2003	Cognitive stimulation groups using 'Making a Difference' manual	Mixed community & care home, UK	2	7	14	45	14.4 (3.8)	None	85.3 (7.0)	115/86
Tanaka 2021	Group exercise and cognitive stimulation	Residential geriatric rehabilitation facility	2	8	16	45	15.5 (5.8)	None	86.2 (7.8)	16/15
Tsantali 2017	Individual cognitive stimulation delivered by psychologists	Community – all receiving AChEIs, Greece	3	16	48	90	23.0 (1.3)	8-month follow-up	73.7 (5.3)	17/21
Young 2019	Cognitive stimulation groups plus Tai Chi (using adapted 'Making a Difference' manual)	Community, Hong Kong	2	7	14	60	20.7 (2.3)	None	80.2 (6.4)	51/50

AChEI: acetylcholinesterase Inhibitor
CST: cognitive stimulation therapy
CST-PD: cognitive stimulation therapy – Parkinson's Disease
iCST: individual cognitive stimulation therapy
IQR: interquartile range
MAKS: motor stimulation; activities of daily living; cognitive stimulation; spiritual element
MMSE: Mini Mental State Examination
MoCA: Montreal Cognitive Assessment
N/A: not applicable
RAM: Roy's adaptation model
RO: reality orientation
SADEM: study on ageing and dementia in Mexico

Table 3. Summary of exploratory subgroup analyses: cognition

Cognition	Effect size (SMD)	95% CI	I^2	Number of studies	Number of participants	Quality of the evidence
Cognitive stimulation (CS)	0.40	0.25, 0.55	62%	34	2340	Moderate
Group CS	0.43	0.26, 0.59	56%	27	1637	Moderate
Individual CS	0.30	-0.03, 0.64	72%	7	703	Low
20 or more group sessions	0.42	0.16, 0.67	46%	12	615	Moderate
Fewer than 20 group sessions	0.43	0.22, 0.65	61%	19	1022	Moderate
3 or more group sessions per week	0.46[a]	0.22, 0.69	0%	8	328	Moderate
2 or more group sessions per week	0.51[b]	0.34, 0.69	51%	21	1283	Moderate
1 group session per week	0.04[c]	-0.17, 0.25	0%	6	354	Low
Community setting (group CS)	0.66	0.33, 0.99	77%	15	642	Moderate
Care home setting (group CS)	0.60	-0.01, 1.20	87%	9	323	Very low
Mild dementia severity	0.71[d]	0.47, 0.95	43%	10	640	Moderate
Moderate dementia severity	0.21	0.03, 0.39	28%	13	778	High
Active control	0.59	0.37, 0.82	0%	5	322	Moderate
Treatment-as-usual	0.41	0.22, 0.60	61%	22	1315	Moderate

[a]3 group sessions > 1 group session (P = 0.01).
[b]2 or more > 1 group session (P = 0.0007).
[c]2 group sessions > 1 group session (P = 0.003).
[d]Mild dementia severity > moderate dementia severity (P = 0.001).
All other subgroup comparisons shown were not statistically significant.

APPENDICES

Appendix 1. Sources searched and search strategies

Source	Search strategy	Hits retrieved
1. CDCIG Register (CRSWEB) [Date of most recent search: 3 March 2022]	cognitive stimulation OR reality orientation OR memory therapy OR memory groups OR memory support OR memory stimulation OR global stimulation OR cognitive psychostimulation	Jul 2018: 22 Jul 2019: 18 May 2020: 27 March 2022: 81

(Continued)

2. MEDLINE In-process and other non-indexed citations and MEDLINE 1950-present (Ovid SP) [Date of most recent search: 3 March 2022]	1. exp Dementia/	Jul 2018: 209
	2. Delirium/	Jul 2019: 50
	3. Wernicke Encephalopathy/	May 2020: 45
	4. Delirium, Dementia, Amnestic, Cognitive Disorders/	March 2022: 133
	5. dement*.mp.	
	6. alzheimer*.mp.	
	7. (lewy* adj2 bod*).mp.	
	8. deliri*.mp.	
	9. (chronic adj2 cerebrovascular).mp.	
	10. ("organic brain disease" or "organic brain syndrome").mp.	
	11. ("normal pressure hydrocephalus" and "shunt*").mp.	
	12. "benign senescent forgetfulness".mp.	
	13. (cerebr* adj2 deteriorat*).mp.	
	14. (cerebral* adj2 insufficient*).mp.	
	15. (pick* adj2 disease).mp.	
	16. (creutzfeldt or jcd or cjd).mp.	
	17. huntington*.mp.	
	18. binswanger*.mp.	
	19. korsako*.mp.	
	20. or/1-19	
	21. "cognitiv* stimul*".mp.	
	22. "reality orientation".mp.	
	23. (memory adj2 therapy).mp.	
	24. "memory group*".mp.	
	25. "memory support".mp.	
	26. (memory adj2 stimulat*).mp.	
	27. "global stimulation".mp.	
	28. ("cognitive psycho-stimulation" or "cognitive psychostimulation").mp.	
	29. *Psychomotor Performance/	
	30. or/21-29	
	31. 20 and 30	
	32. (2010* OR 2011*).ed.	
	33. 31 and 32	
	34. randomized controlled trial.pt.	

(Continued)

	35. controlled clinical trial.pt.	
	36. randomized.ab.	
	37. placebo.ab.	
	38. drug therapy.fs.	
	39. randomly.ab.	
	40. trial.ab.	
	41. groups.ab.	
	42. or/34-41	
	43. (animals not (humans and animals)).sh.	
	44. 42 not 43	
	45. 33 and 44	
3. Embase 1980-present (Ovid SP) [Date of most recent search: 3 March 2022]	1. exp dementia/	Jul 2018: 1277 Jul 2019: 50 May 2020: 179 March 2022: 305
	2. Lewy body/	
	3. delirium/	
	4. Wernicke encephalopathy/	
	5. cognitive defect/	
	6. dement*.mp.	
	7. alzheimer*.mp.	
	8. (lewy* adj2 bod*).mp.	
	9. deliri*.mp.	
	10. (chronic adj2 cerebrovascular).mp.	
	11. ("organic brain disease" or "organic brain syndrome").mp.	
	12. "supranuclear palsy".mp.	
	13. ("normal pressure hydrocephalus" and "shunt*").mp.	
	14. "benign senescent forgetfulness".mp.	
	15. (cerebr* adj2 deteriorat*).mp.	
	16. (cerebral* adj2 insufficient*).mp.	
	17. (pick* adj2 disease).mp.	
	18. (creutzfeldt or jcd or cjd).mp.	
	19. huntington*.mp.	
	20. binswanger*.mp.	
	21. korsako*.mp.	
	22. CADASIL.mp.	
	23. or/1-22	

(Continued)

	24. "cognitiv* stimul*".mp. 25. "reality orientation".mp. 26. (memory adj2 therapy).mp. 27. "memory group*".mp. 28. "memory support".mp. 29. (memory adj2 stimulat*).mp. 30. "global stimulation".mp. 31. ("cognitive psycho-stimulation" or "cognitive psychostimulation").mp. 32. *psychomotor performance/ 33. or/24-32 34. 23 and 33 35. (2010* OR 2011*).em. 36. 34 and 35	
4. PsycINFO 1806-July week 1 2019 (Ovid SP) [Date of most recent search: 3 March 2022]	1. exp Dementia/ 2. exp Delirium/ 3. exp Huntingtons Disease/ 4. exp Kluver Bucy Syndrome/ 5. exp Wernickes Syndrome/ 6. exp Cognitive Impairment/ 7. dement*.mp. 8. alzheimer*.mp. 9. (lewy* adj2 bod*).mp. 10. deliri*.mp. 11. (chronic adj2 cerebrovascular).mp. 12. ("organic brain disease" or "organic brain syndrome").mp. 13. "supranuclear palsy".mp. 14. ("normal pressure hydrocephalus" and "shunt*").mp. 15. "benign senescent forgetfulness".mp. 16. (cerebr* adj2 deteriorat*).mp. 17. (cerebral* adj2 insufficient*).mp. 18. (pick* adj2 disease).mp. 19. (creutzfeldt or jcd or cjd).mp. 20. huntington*.mp. 21. binswanger*.mp.	Jul 2018: 116 Jul 2019: 40 May 2020: 31 March 2022: 84

(Continued)

	22. korsako*.mp. 23. ("parkinson* disease dementia" or PDD or "parkinson* dementia").mp. 24. or/1-23 25. "cognitiv* stimul*".mp. 26. "reality orientation".mp. 27. (memory adj2 therapy).mp. 28. "memory group*".mp. 29. "memory support".mp. 30. (memory adj2 stimulat*).mp. 31. "global stimulation".mp. 32. ("cognitive psycho-stimulation" or "cognitive psychostimulation").mp. 33. "psychomotor performance".mp. 34. or/25-33 35. 24 and 34 36. random*.mp. 37. trial.mp. 38. placebo.mp. 39. group*.mp. 40. exp Clinical Trials/ 41. or/36-40 42. 35 and 41 43. (2010* OR 2011*).up. 44. 42 and 43	
5. CINAHL (EBSCOhost) [Date of most recent search: 3 March 2022]	S1 (MH "Dementia+") S2 (MH "Delirium") or (MH "Delirium, Dementia, Amnestic, Cognitive Disorders") S3 (MH "Wernicke's Encephalopathy") S4 TX dement* S5 TX alzheimer* S6 TX lewy* N2 bod* S7 TX deliri* S8 TX chronic N2 cerebrovascular S9 TX "organic brain disease" or "organic brain syndrome" S10 TX "normal pressure hydrocephalus" and "shunt*"	Jul 2018: 70 Jul 2019: 45 May 2020: 16 March 2022: 45

(Continued)

	S11 TX "benign senescent forgetfulness"	
	S12 TX cerebr* N2 deteriorat*	
	S13 TX cerebral* N2 insufficient*	
	S14 TX pick* N2 disease	
	S15 TX creutzfeldt or jcd or cjd	
	S16 TX huntington*	
	S17 TX binswanger*	
	S18 TX korsako*	
	S19 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or S12 or S13 or S14 or S15 or S16 or S17 or S18	
	S20 TX "cognitiv* stimul*"	
	S21 TX "reality orientation"	
	S22 TX memory N2 therapy	
	S23 TX "memory group*"	
	S24 TX "memory support"	
	S25 TX memory N2 stimulat*	
	S26 TX "global stimulation"	
	S27 TX "cognitive psycho-stimulation" OR "cognitive psychostimulation"	
	S28 (MM "Psychomotor Performance")	
	S29 S20 or S21 or S22 or S23 or S24 or S25 or S26 or S27 or S28	
	S30 S19 and S29	
	S31 EM 2010	
	S32 EM 2011	
	S33 S31 or S32	
	S34 S30 and S33	
	S35 TX random*	
	S36 (MH "Clinical Trials+")	
	S37 AB group	
	S38 TI study	
	S39 S35 or S36 or S37 or S38	
	S40 S34 and S39	
6. Web of Science Core Collection (1945-present) (Clarivate) [Date of most recent search: 3 March 2022]	Topic=(dement* OR alzheimer* OR "lew* bod*" OR deliri* OR creutzfeldt OR cjd OR jcd OR huntington* OR binswanger* OR korsako*) AND Topic=("cogniti* stimul*" OR CST OR "reality orienation" OR "memory therapy" OR "memory group*" OR "memory support" OR "psychomotor performance" OR "global stimulation" OR "cognitive performance") AND Topic=(random* OR trial* OR RCT OR "cross-over" OR cross-over) AND Year Published=(2010-2011)	Jul 2018: 1349 Jul 2019: 193 May 2020: 165

(Continued)

	Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-SSH. Lemmatization=On	March 2022: 357
7. LILACS (BIREME) [Date of most recent search: 3 March 2022]	"cognitiv$ stimul$" OR "reality orienation" OR "memory therapy" OR "memory group$" OR "memory support" OR "psychomotor performance" OR "global stimulation" OR "cognitive performance" [Words] and dementia OR alzheimer $ OR demenc$ OR AD OR demência [Words]	Jul 2018: 9 Jul 2019: 0 May 2020: 1 March 2022: 0
8. CENTRAL (*The Cochrane Library*) [Date of most recent search: 3 March 2022]	#1 dement* #2 alzheimer* #3 deliri* #4 chronic adj2 cerebrovascular #5 (lewy* bod*) #6 "organic brain disease" or "organic brain syndrome" #7 (pick* disease) #8 creutzfeldt or jcd or cjd #9 huntington* #10 binswanger* #11 korsako* #12 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11) #13 "cognitiv* stimul*" #14 "reality orientation" #15 "memory therapy" #16 "memory group*" #17 "memory support" #18 "memory stimulat*" #19 "global stimulation" #20 "cognitive psycho-stimulation" #21 "cognitive psychostimulation" #22 MeSH descriptor Psychomotor Performance explode all trees #23 (#13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22) #24 (#12 AND #23) #25 (#24), from 2010 to 2011	Jul 2018: 370 Jul 2019: 30 May 2020: 80 March 2022: 75
9. Clinicaltrials.gov (www.clinicaltrials.gov)	#1 Intervention: Cognitive stimulation AND Interventional studies AND First rec: 01/01/2010-12/05/2011 = 30	Jul 2018: 86 Jul 2019: 0

(Continued)

[Date of most recent search: 3 March 2022]	#2 Intervention: reality orientation AND Interventional studies AND First rec: 01/01/2010-12/05/2011 = 1 #3 Interventional Studies | dementia OR alzheimers OR AD OR alzheimer's OR alzheimer OR lewy OR FTLD OR FLD | memory therapy OR memory training | received from 01/01/2010 to 12/05/2011=20	May 2020: 8 March 2022: 19
10. ICTRP Search Portal (http://apps.who.int/trialsearch) [includes: Australian New Zealand Clinical Trials Registry; ClinicalTrilas.gov; ISRCTN; Chinese Clinical Trial Registry; Clinical Trials Registry – India; Clinical Research Information Service – Republic of Korea; German Clinical Trials Register; Iranian Registry of Clinical Trials; Japan Primary Registries Network; Pan African Clinical Trial Registry; Sri Lanka Clinical Trials Registry; The Netherlands National Trial Register] [Date of most recent search: 3 March 2022]	Advanced search: (dementia OR alzheimer OR alzheimers OR alzheimers) AND (cognitive stimulation OR reality orientation OR memory therapy OR memory training OR cognitive training) AND (2010-2011)	Jul 2018: 15 Jul 2019: 1 March 2022: 1
TOTAL before de-duplication		Jul 2018:3523 Jul 2019: 427 May 2020: 552 March 2022: 1100 TOTAL: 5602
TOTAL after de-duplication and first-assessment by CDCIG information specialist		Jul 2018: 528 Jul 2019: 155 May 2020: 406 March 2022: 775 TOTAL: 1864

WHAT'S NEW

Date	Event	Description
30 January 2023	New search has been performed	New search performed, new studies for inclusion

Date	Event	Description
30 January 2023	New citation required and conclusions have changed	New search performed. New studies for inclusion; content revised. New authors added.

HISTORY

Protocol first published: Issue 4, 2005
Review first published: Issue 2, 2012

CONTRIBUTIONS OF AUTHORS

BW: lead author; drafting updated review; selection of trials; extraction of data; risk of bias assessments; GRADE assessments; entry of data; data analysis; interpretation of data analyses

HKR: selection of trials; extraction of data; risk of bias assessments; GRADE assessments; commenting on drafts of the review

EE: selection of trials; extraction of data; risk of bias assessments; commenting on drafts of the review

EA: For 2012 update: search for trials; obtaining copies of trial reports; entry of data; data analysis; interpretation of data analyses. For 2022 updated review: commenting on all sections

AS: For previous versions of the review: selection of trials; extraction of data; interpretation of data analyses. For updated 2022 review: commenting on all sections

MO: For previous versions of the review: selection of trials; extraction of data; interpretation of data analyses. For updated 2022 review: commenting on all sections

DECLARATIONS OF INTEREST

The authors have produced various training materials in dementia care, including cognitive stimulation therapy manuals, in order to disseminate research findings to care workers, family caregivers and others. Royalties for the manuals are received by the International Cognitive Stimulation Therapy Centre, based at University College London. AS receives fees for providing training in cognitive stimulation approaches.

SOURCES OF SUPPORT

Internal sources

- Bangor University, UK

  For the 2012 review, BW was employed by Bangor University; for the 2022 review BW was affiliated with Bangor University. No additional funding was provided.
- University College London, UK

  For the 2012 review, MO and AS were employed by University College London. For the 2022 update, AS was employed by University College London. No additional funding was provided for the 2022 update.
- University of Nottingham, UK

  For the 2022 update, HR was a PhD student at and MO was employed by the University of Nottingham.

External sources

- NIHR, UK

  For the 2012 review, EA was supported by the Support at Home - Interventions to Enhance Life in Dementia (SHIELD) project (Application No RP-PG-0606-1083) which is funded by the NIHR Programme Grants for Applied Research funding scheme.
- NIHR, UK

  This update was supported by the National Institute for Health and Care Research (NIHR), via Cochrane Infrastructure funding to the Cochrane Dementia and Cognitive Improvement group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, National Health Service or the Department of Health

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

In previous versions of this review, studies were included from the original 2000 review of Reality Orientation (Spector 2000a) which used terms other than 'dementia' for their participants. These were studies where the review authors were satisfied that the included population would now be described as having a dementia. However, the development of the field means that, for this update, it is possible to exclude those early studies with unclear diagnostic categorisation.

Previously, fixed-effect models were used in meta-analyses unless heterogeneity was high, where random-effects models were utilised. In this review, with greater diversity of studies, random-effects models have been used throughout, in order to provide a consistent approach.

The current review includes an updating and extension of the use of the risk of bias evaluation tool, adopting a similar framework to our Cochrane Review of reminiscence therapy for people with dementia (Woods 2018a).

The current review includes the use of the GRADE approach throughout and the inclusion of summary of findings tables.

The current review includes the use of subgroup analyses to explore potential factors leading to heterogeneity. To compare results from studies with different overall levels of dementia severity, we followed NICE-SCIE 2006 and considered MMSE scores of 20 and below as indicating 'moderate' and above 20 as 'mild' dementia.

NOTES

This review replaces the review of Reality Orientation for dementia (Spector A, Orrell M, Davies S, Woods B. Reality orientation for dementia. The Cochrane Database of Systematic Reviews 2000, Issue 3. Art. No.: CD001119. DOI: 10.1002/14651858.CD001119).

INDEX TERMS

Medical Subject Headings (MeSH)

Cognition [*physiology]; Dementia [*therapy]; Memory [physiology]; Orientation [*physiology]; Psychotherapy [*methods]; Randomized Controlled Trials as Topic

MeSH check words

Aged; Humans