Effectiveness of Cognitive Stimulation for Dementia: A Systematic Review and Meta-Analysis

Robin M. T. Cafferata[1], Ben Hicks[2], and Claudia C. von Bastian[1]

[1]Department of Psychology, University of Sheffield

[2]Centre for Dementia Studies, Brighton and Sussex Medical School

© 2021, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/bul0000325

Author Note

We are grateful to all the authors who kindly responded to our queries and who provided additional data for the meta-analysis. We further thank Gabrielle Hardwick for her assistance in the literature search, and Daniel Heck for sharing his scripts for conducting the Bayesian moderator analysis with the metaBMA package. Developing the idea for this research was supported by a financial contribution to BH and CvB from the Ageing and Dementia Research Centre (ADRC), Bournemouth University.

This study has been preregistered on PROSPERO (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018096652). The data and analysis scripts are available on the Open Science Framework (https://osf.io/8mwnt/).

Correspondence concerning this article should be addressed to Claudia C. von Bastian, Department of Psychology, 1 Vicar Lane, Sheffield, S1 2LT, United Kingdom. E-mail: c.c.vonbastian@sheffield.ac.uk

Abstract

Cognitive stimulation (CS) is a non-pharmacological intervention often involving group activities and social interaction used to treat cognitive declines in people with dementia. This pre-registered systematic review and meta-analysis evaluated the effectiveness of CS in producing benefits on cognition (primary outcome) and quality of life, activities of daily living, and psychological symptoms (secondary outcomes) across 44 randomized-controlled trials comprising 45 comparisons including 2,444 participants. A medium-sized effect (g = 0.49) on global cognition was found immediately after the intervention and was supported by decisive Bayesian evidence. Clinical relevance is defined as a reduction of 3 to 4 points on the ADAS-Cog scale; the average attenuation of cognitive decline observed was 2.41 points (after removing one outlier). Therefore, the observed decline was of borderline clinical relevance. CS was also found to significantly improve memory, activities of daily living, depressive symptoms, and dementia ratings; most of these effects were supported by substantial and strong Bayesian evidence. No significant effects were found for global cognition at 1 to 10 months follow-up assessment, for language, quality of life, anxiety, and behavior symptoms. However, evidence for the absence of these effects was ambiguous. A review of study bias highlighted that most studies lacked active, double-blinded controls, potentially leading to an overestimation of the effect, and making it difficult to conclusively attribute the observed improvements to the CS intervention. Hence, although effects are promising, the methodological issues highlight there is still a need for better controlled studies that provide more compelling evidence.

Keywords: cognitive stimulation, dementia, cognitive decline

Public Significance Statement

Cognitive stimulation is a psychosocial, nonpharmacological treatment for people with dementia. This meta-analysis of 44 randomized-controlled trials shows that cognitive stimulation does benefit global cognition, memory, activities of daily living, depressive symptoms, and dementia ratings. However, these benefits are of only borderline clinical relevance. Furthermore, our review identifies several methodological challenges that this research area faces, calling for investment into better controlled studies to provide robust and compelling evidence for the effectiveness of cognitive stimulation as a treatment for people with dementia.

Effectiveness of Cognitive Stimulation for Dementia: A Systematic Review and Meta-Analysis

Dementia is an umbrella term for a range of conditions caused by neurodegeneration, including Alzheimer's disease, vascular dementia, frontotemporal dementia, and Lewy body dementia. It is characterized by a progressive decline in cognitive ability of sufficient severity to interfere with social and/or occupational functioning. It may also include other symptoms such as language problems, deterioration in the ability to perform activities of daily living, and behavior changes (Innes & Manthorpe, 2013; Kitwood, 1997). With 50 million people worldwide currently living with dementia – a number that is bound to triple to 152 million by 2050 – and an estimated associated economic cost of USD 1 trillion thought to double by 2030 (Alzheimer's Disease International [ADI], 2018), it has become a global priority to seek ways to ameliorate the detrimental health and financial impacts of dementia (World Health Organization [WHO], 2018). Yet, drug development efforts have been characterized by a high failure rate (99.6% for clinical trials for drugs targeting Alzheimer's disease, Cummings et al., 2014), and existing pharmacological treatments aimed to alleviate cognitive, behavioral, and psychological symptoms of dementia have raised concerns about adverse side-effects and mortality risks (Ballard et al., 2009; Maust et al., 2015). Consequently, psychosocial, non-pharmacological interventions are increasingly considered a core component of dementia care (Kenigsberg et al., 2016), as they can provide a safe and cost-effective means to support people with dementia (Livingston et al., 2014; McDermott et al., 2019; Nyman & Szymczynska, 2016).

Cognitive stimulation (CS) is a particularly promising psychosocial intervention recommended for people with mild-to-moderate dementia (National Institute for Health and Care Excellence [NICE], 2018). CS aims to improve global cognition and maintain function by stimulating multiple cognitive functions simultaneously, typically with group activities

emphasizing social interaction. This approach is different from cognitive training, which targets isolated cognitive functions (e.g., memory) with individual, repetitive practice of standardized cognitive tasks. CS is also distinguished from cognitive rehabilitation, a person-centered approach that aims at improving everyday life through developing strategies for performing desired functions or tasks (Clare & Woods, 2004). Whereas there is limited evidence for the effectiveness of cognitive training for people with mild to moderate dementia (e.g., Bahar-Fuchs et al., 2013; Bahar-Fuchs et al., 2020; Hill et al., 2017), and only few studies assessing benefits of cognitive rehabilitation (but see Clare et al., 2010), CS has received consistent meta-analytic support for improved cognitive performance (Aguirre et al., 2013; Huntley et al., 2015; Kim et al., 2017; Kurz et al., 2011). Moving beyond previous meta-analyses, in the present work, we compared the effects of different CS approaches considering moderators on a wide range of outcomes.

Cognitive Stimulation Approaches

CS encompasses a variety of approaches including reality orientation, validation, and/or reminiscence. Reality orientation involves consistent repetition of facts of life, for example a person's name, where they live, and the current date. Reality orientation is administered in intensive formal classroom periods of 30-60 minutes. Instead of classes or as a supplement to classes, reality cues are provided throughout everyday life, for example on blackboards that list the name of the hospital, its location, and the current date. Although reality orientation can have measurable benefits on cognitive and independent functioning (Holden & Woods, 1995; Spector et al., 2000), it has also attracted criticism. Powell-Proctor and Miller (1982) pointed out that reality orientation as a technique is defined only vaguely and, thus, its interpretation and translation into practice varies greatly. Moreover, the sole focus of reality orientation on communicating information and instructions can be experienced as distressing. For example, Dietch et al. (1989) reported the case of a woman

who became upset when staff emphasized that her son was not seven years old (as she remembered) but instead much older.

Validation and reminiscence approaches address the potentially distressing aspects of reality orientation. Validation focuses on empathic listening and the person's subjective experience as opposed to objective facts. Reminiscence involves discussing past events and reflecting on a person's life using prompts such as photographs, music and videos.

Incorporating elements of reality orientation, validation, and reminiscence, Spector et al. (2001) developed CS therapy (CST), a structured and manualized psychosocial intervention (Spector et al., 2006) targeting cognitive and social abilities. Typically, CST is delivered over a period of seven weeks, with two sessions per week each lasting around 45 minutes. The 14 sessions are underpinned by 18 guiding principles (Spector et al., 2006), including encouraging mental stimulation; using orientation, both sensitively and implicitly; eliciting opinions rather than facts; using reminiscence as an aid to the here-and-now; providing triggers to aid recall; and building/strengthening relationships. To aid reality orientation, a board is used in every session that shows personal and orientation information. Each session follows a different theme (e.g., physical games, sound, or using money) with a range of activities that can be tailored to the groups' interests.

Prior Reviews

Four previous reviews examined CS benefits for people with dementia. Kurz et al. (2011) reviewed 18 randomized-controlled trials (RCT) of CS interventions that assessed cognitive outcomes in people with a diagnosis of mild cognitive impairment (MCI) or dementia. Compared to control conditions, CS was found to significantly improve indicators of global cognition, with standardized mean differences (SMD) ranging from 0.21, 95% confidence interval (CI) = [0.03, 0.39] for the Mini-Mental State Examination (MMSE; Folstein et al., 1975) to SMD = -0.30, 95% CI [-0.48, -0.13] on the Alzheimer's Disease

Assessment Scale Cognitive subscale (ADAS-Cog; Rosen et al., 1984). No significant improvements of activities of daily living were found across the 9 studies reporting corresponding measures. However, Kurz et al. (2011) did not report the effects of CS specifically for people with dementia (k = 13). Moreover, too few studies were available at the time of their review to test other secondary outcomes such as quality of life, mood (indicators of depression and anxiety), and behavioral and psychological symptoms. Similarly, too little data were available to test for the effects of potential moderators on the benefits of CS.

Aguirre et al. (2013; see also Woods et al., 2012) reviewed 15 RCTs with a duration of at least 4 weeks and where at least one cognitive outcome was assessed. Aguirre et al. focused exclusively on CS interventions and samples of people with a diagnosis of dementia. Notably, their meta-analysis did not include seven of the studies included in Kurz et al. (2011) that, in our reading, would have been eligible. Still, like Kurz et al. (2011), Aguirre et al. (2013) found a significant positive effect of CS on global cognition; SMD = 0.41, 95% CI = [0.25, 0.57]. Moreover, tentative evidence based on three studies suggested that the CS benefits lasted up to 3 months after the intervention, whereas no long-term benefits were found in one study that assessed CS effects after 10 months. The authors also examined a range of secondary outcomes including quality of life, activities of daily living, mood, and behavior. The only significant effect was found for quality of life ratings (SMD = 0.38, 95% CI = [0.11, 0.65]); however, estimates for these secondary outcomes were based on only small sets of studies (ks = 4 to 8).

Huntley et al. (2015) analyzed data from 22 RCTs investigating the benefits of CS interventions for people with dementia and confirmed the findings of previous meta-analyses, with average effects ranging from g = -0.26, 95% CI [-0.44, -0.08] (ADAS-Cog) to g = 0.51, 95% CI [0.35, 0.66] (MMSE) for comparisons to passive control groups. Participants in

passive (or waiting list) control groups typically receive treatment as usual, thereby controlling for effects of repeated testing arising from pre-/post-test designs. In contrast, participants in active control groups undergo an alternative intervention designed to have no effect on the targeted mechanisms while appearing to be a plausible treatment. In addition to retest effects, active control groups control for generic, non-specific intervention effects such as attention from research teams, and expectancy effects (e.g., Oken et al., 2008; see also von Bastian & Oberauer, 2014). At the time of their review, Huntley et al. (2015) identified only three studies that compared effects of CS to active control interventions. Although still significant, the average effect on the MMSE was considerably smaller, g = 0.35, 95% CI [0.06, 0.64]; at the time, none of the studies with active controls evaluated effects on the ADAS-Cog. These findings demonstrate the importance of accounting for risks of biases from suboptimal controls. Huntley et al. further evaluated a number of moderators, including context (inpatient or outpatient), intervention duration, format, and intensity, and baseline MMSE. None of the moderators was significant; importantly though, Huntley et al. tested these moderators only for effect sizes pooled across CS and other cognition-focused interventions (cognitive training and cognitive rehabilitation). Therefore, it remains unclear whether any of the moderators may specifically affect effects of CS.

Most recently, Kim et al.'s (2017) review of 12 RCTs that assessed cognitive outcomes confirmed previous findings with SMD = 0.44, 95% CI [0.27, 0.60]. The authors also examined CS benefits for a range of secondary outcomes. Like Aguirre et al. (2013), they found a significant effect only for quality of life (SMD = 2.05, 95% CI [0.72, 3.38]) but not for activities of daily living, mood, or behavior and psychological symptoms. However, as Kim et al. (2017) included substantially fewer studies than previous reviews (for unclear reasons), estimates for these secondary outcomes were again based on only a few studies (ks between 1 and 6). Furthermore, Kim et al. (2017) noted a lack of double-blinded studies,

which could add to confounds from expectancy effects, and a large heterogeneity in study settings and components of CS interventions.

Present Study

The cumulative evidence from previous reviews suggests positive effects of CS on indicators of global cognition, with tentative evidence supporting benefits also on quality of life (Aguirre et al., 2013; Kim et al., 2017). However, these previous estimates were based on relatively few studies, which is particularly problematic considering the risk of bias identified in previous reports. Moreover, the limited number of studies permitted neither distinguishing between different CS approaches nor examining intervention- and sample-specific factors that could impact the effectiveness specifically of CS and specifically for people with dementia.

The present pre-registered systematic review and meta-analysis updates previous reports by including 44 studies. Furthermore, moving beyond previous reviews, the present work investigates the benefits of CS for people with dementia for a range of outcomes and considering potential moderators. Like Aguirre et al. (2013) and Kim et al. (2017), we considered primary outcomes (global cognition, memory, and language) as well as secondary outcomes (quality of life, activities of daily living, mood, and behavior and psychological symptoms). Critically, the larger number of studies also allowed for examining a range of moderators of CS benefits. Like Huntley et al. (2015), we considered both intervention- and sample-specific moderators including treatment dosage (duration, frequency, and number of sessions of the intervention), intervention context, and baseline MMSE. Different to Huntley et al., though, our analyses focus specifically on CS interventions. In addition, we tested the impact of the type of CS intervention (CST or reality orientation), number of CS components, and demographic factors (age and gender). Given the discussion of the possible impact of the biases identified by Huntley et al. (2015) and Kim et al. (2017), analyses were complemented by a careful risk of bias assessment, and Bayes factors (BF) for each estimate as a measure of

the strength of evidence. Finally, different from the previous meta-analyses – except Huntley et al. (2015) –, we adopted a random-effects modeling approach to account for the large variability in samples and methods of the empirical studies (Field et al., 2010).

Method

The present systematic review and meta-analysis builds directly on the prior review by Aguirre et al. (2013) and was designed to follow PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al., 2009). The study protocol detailing the review question, search strategy, inclusion criteria, and analytical methods was preregistered with PROSPERO on May 23, 2018 (CRD: 42018096652).

Search Method

Figure 1 illustrates the search and selection process. As the article by Aguirre et al. (2013) was the most comprehensive meta-analysis focusing specifically on CS that we were aware of at the time of our search, we updated their set of 15 studies reported by using identical keywords ("cognitive stimulation", "reality orientation", "memory therapy", "memory groups", "memory support", "memory stimulation", "global stimulation", and "cognitive stimulation") to search healthcare databases (MEDLINE, PsycINFO, CINAHL, and LILACS), trial registers (Cochrane Central Register of Controlled Trials [CENTRAL], UK Clinical Trials Gateway, ISRCTN, WHO Portal, and UMIN Japan Trial Register), and gray literature sources (Web of Science, ProQuest, Australian Digital Theses, OpenGrey). For published work (but not for unpublished work and theses), we restricted the search to reports published after 2011 (the date when Aguirre et al., 2013, concluded their search). We carried out the search on June 19, 2018, which resulted in 2,923 references. In addition, we examined the bibliographies and citations of relevant articles, which yielded 8 additional reports. We also considered the overall 21 reports that were not included in Aguirre et al.'s (2013) article but in the previous reviews by Kurz et al. (2011), Huntley et al. (2015), and Kim et al. (2017).

We also contacted the first authors of all included reports to inquire about unpublished work or published work that was not yet included, which resulted in one additional report. Finally, we ran a search update on Google Scholar on June 7, 2020, which resulted in an additional 6 reports. After duplicates removal and initial screening, we assessed 124 full texts for inclusion eligibility.

Inclusion and Exclusion Criteria

We included journal articles, theses, and book chapters written in English and reporting empirical data of at least one measure of cognitive functioning from an RCT. Participants had to be diagnosed with any type of dementia and have received CS for at least four weeks consistent with the definition proposed by Clare and Woods (2004). This included reality orientation, CST, reminiscence therapy, or other CS interventions fulfilling Clare and Wood's criteria, that is, they consisted of a wide range of activities that are completed in a group format with social interaction and aim to improve overall cognition rather than a single cognitive function. Of the 124 full texts assessed, 49 studies met the criteria. However, the data reported in two of these studies (Graessel et al., 2011; Luttenberger, Donath, et al., 2012) overlapped with the data reported in an article already included (Luttenberger, Hofner, & Graessel, 2012) and, thus, were excluded. Three other studies did not report sufficient data to compute effect sizes (D'Amico et al., 2015; Quayhagen et al., 2000; Yamanaka et al., 2013). For two further studies that also did not report sufficient data to compute effect sizes (Baldelli et al., 1993; Chapman et al., 2004) and one study (Ferrario et al., 1991) that could not be accessed, we used the data reported in Aguirre et al. (2013) in the present analysis. The final set of 44 included studies yielded $k = 45$ comparisons.

Data Extraction

Two independent coders (RC and BH) extracted the relevant data from each study based on a coding protocol (see https://osf.io/8mwnt/) that detailed which information to

extract and how (e.g., an explanation of what is regarded an independent comparison; decision-rules which numbers to extract in case the in-text reporting deviated from information listed in tables). The two coders independently entered their data in a mySQL database through a web interface. The web interface – a multipage form – was implemented with the open-source framework Grocery CRUD (version 1.5.0, Skoumbourdis, 2014). Data were stored in a mySQL database to facilitate handling of nested data (e.g., multiple comparisons from the study, multiple outcomes for the same comparisons), and automatized comparison of data entered by different coders. Both coders were involved in piloting and revising the coding protocol based on a set of five studies. Any discrepancies between coders were resolved by discussion and involving a third reviewer (CvB). Interrater reliabilities for categorical moderators were $\kappa = .90$ (type of intervention) and $\kappa = .71$ (sample from urban vs. rural area).

The information extracted from studies included (a) bibliographic information (e.g., publication author, title, year, and publication status), (b) intervention descriptors (e.g., CS content and components, duration and number of sessions, intervention setting)[1], (c) sample descriptors (e.g., age, gender, demographic background, diagnosis, medication, baseline MMSE score/clinical severity), and (d) outcomes (measure, sample size, means and standard deviations, direction of the effect). Where means and standard deviations were not available, effect sizes were estimated based on statistical values reported (e.g., *F*-statistic).

Outcomes

Primary Outcomes

To assess meta-analytic effects of CS on target areas of cognitive function, we extracted effects on global cognition measures (e.g., ADAS-Cog, MMSE). In addition, we

[1] We originally planned to code also the intervention schedule, living situation, and average time since the diagnosis. However, during the literature review we found that too few studies reported this information and, therefore, decided not to code these variables.

coded effects of CS on memory (e.g., word recall) and language (e.g., verbal fluency). Maintenance effects[2] were coded as follow-up data of CS effects on global cognition one month or longer after posttest.

Secondary Outcomes

In addition to cognitive benefits, we preregistered to evaluate effects of CS on quality of life (Logsdon et al., 1999), activities of daily living (Lawton & Brody, 1969), depression (D'Ath et al., 1994) and anxiety (Shankar et al., 1999), and behavior (Pattie & Gilleard, 1975) and psychological (Hughes et al., 1982) symptoms.

Analysis

For all statistical analyses, we used R. Scripts and data are available on the Open Science Framework (https://osf.io/8mwnt/).

Calculation of Effect Sizes

We computed standardized mean between-groups differences (Cohen's *d*) of change scores. We used pooled SDs at pretest where available as they provide least biased effect size estimates (Morris, 2008). In three studies, the pretest *SD*s reported and used for computing the effect size were for a different subsample (Alves et al., 2014; Chapman et al., 2004; Onder et al., 2005). In two studies, pretest *SD*s were not available and posttest *SD*s were used instead (Baglio et al., 2015; Kolanowski et al., 2016). In two studies where neither pretest nor posttest *SD*s were available (Breuil et al., 1994; Ferrario et al., 1991), change score *SD*s were used. As change score *SD*s are typically smaller than pretest and posttest *SD*s, effect sizes from these two studies are likely overestimated. However, because sensitivity analyses excluding these two studies yielded identical conclusions, we decided to report the findings from the analyses including these two studies. Effect size estimates were corrected for small sample bias using Hedges' *g* (Borenstein et al., 2009).

[2] Note that maintenance effects were preregistered as a potential moderator in the preregistration but treated as outcome in the report.

Pooling of Dependent Effect Sizes

Several studies provided multiple measures of an outcome; in particular, effects of CS were often reported for both the ADAS-cog and the MMSE. To avoid biased estimates arising from including multiple measures from the same sample (Borenstein et al., 2009; Cooper et al., 2009), we originally planned to compute composite effect size estimates. However, we later became aware of evidence that the ADAS-Cog measurement is a more precise measure for people with dementia that more accurately discriminates between levels of cognitive dysfunction (Balsis et al., 2015). Moreover, the ADAS-Cog is primarily used in pharmacological interventions (Schneider et al., 2014) and, thus, allows for better comparisons to CS intervention studies. Therefore, we decided to include only the ADAS-Cog scores but discard the MMSE scores where both were reported. For all other outcomes, we classified measures based on the authors' reporting. For the few studies that used multiple memory measures, we selected episodic memory measures over short-term memory measures, and short-term memory measures over working memory measures. For the one study reporting two quality-of-life measures (Gibbor et al., 2020), we selected the self-report measure and discarded the carer-reported measure.

Meta-Analytic Procedures

Mean effect sizes were estimated with random-effects models assuming sampling error within and across studies, thereby allowing for generalization of our findings (Field et al., 2010). Homogeneity of effect sizes was tested conducting Q tests and examining I^2 statistic and the variability parameter τ. Given the low number of studies reporting moderator data, we ran separate meta-regression Q tests for each moderator with at least 10 cases for each category (Deeks et al., 2011). In addition, we estimated the average effect sizes and associated 95% confidence intervals (CIs) for each level of tested moderator. Meta-analysis was conducted with the metafor package (Version 2.0.0., Viechtbauer, 2010), using a

restricted maximum-likelihood estimator with Knapp and Hartung (2003) adjustment to the standard errors of the estimated coefficients (Viechtbauer et al., 2015).

Although the present meta-analysis comprised more comparisons than its predecessors, the number of cases was still relatively low, in particular for conducting meta-regression, which may lead to low statistical power (Schmidt, 2017). However, estimating power for meta-analysis is challenging because it requires a series of assumptions about unknown parameters (Valentine et al., 2010). Similarly, arbitrary rules of thumb (such as the preregistered minimum number of 10 cases used in the present study) are unsatisfying. In addition, the reliabilities of the measures typically used to assess the outcomes evaluated in this meta-analysis can be relatively low. For example, with increasing time interval between tests, reliabilities of change scores derived from the ADAS-Cog tend to be low (e.g., ranging between .53 to .64 after 1 year in Grochowalski et al., 2016). Ideally, the measurement error arising from low reliability coefficients should be corrected (Schmidt & Hunter, 2015). However, reliability estimates are rarely reported in the literature included in this meta-analysis. Moreover, the outcomes assessed in this meta-analysis include a variety of instruments (e.g., ADAS-Cog, MMSE, and the Cognitive Assessment Scale of the Clifton Assessment Procedures for the Elderly [CAPE], Pattie & Gilleard, 1975), each of which is associated with different reliabilities. Therefore, it would have been difficult to justify picking a particular (set of) reliability estimates for correcting the reported effect sizes. Instead, we addressed the uncertainty associated with potentially low statistical power and measurement reliability by computing BFs for average effect-size estimates as well as for the inclusion of moderators. BFs cannot correct for low statistical power or measurement error, but they provide a measure of the strength for the evidence in the given data, thereby allowing for gauging the extent to which effect sizes need to be interpreted more carefully.

Bayes Factors

BFs were estimated for each study, the average effect-size estimates for each outcome and level of each tested moderator, and for the inclusion of moderator variables in the meta-regressions. A BF is the ratio of the probability of the data under one hypothesis relative to the probability of the data under the other hypothesis, with values ranging on a continuous scale from 0 to ∞. As the BF is a ratio, a value of 1 reflects perfect ambiguity. The greater the BF, the stronger the evidence in favor of the hypothesis in the numerator (typically the alternative hypothesis, H_1), with values below 1 reflecting evidence in favor of the hypothesis in the denominator (typically the null hypothesis, H_0). For example, BF = 10 indicates that the data are 10 times more likely under H_1 than under H_0; in contrast, BF = 1/10 reflects that the data are 10 times more likely under H_0 than under H_1. Conventionally, BFs between 1/3 and 3 are considered ambiguous evidence. Table 1 lists verbal labels of BF ranges to facilitate understanding (Wetzels & Wagenmakers, 2012).

Study-level BFs were computed using the BayesFactor package (Version 0.9.12.4.2, Morey & Rouder, 2015), with BFs based on the *t*-statistics (Rouder et al., 2009) derived from the computed Hedges' *g* using the BSDA package (Version 1.2.0, Arnholt & Evans, 2017). Meta-analytic BFs were computed using the metaBMA package (Version 0.3.9, Heck, Gronau, & Wagenmakers, 2017). We pre-registered our priors for both effect-size estimates and the homogeneity parameter τ. For effect-size estimates, we chose a Cauchy distribution with a scaling factor $r = 0.41$ reflecting the average effect size reported by Aguirre et al. (2013). We truncated the distribution at 0 to reflect the one-tailed hypothesis that CS should improve (and not worsen) the outcomes. The Cauchy distribution is a *t* distribution with a single degree of freedom. It is similar to a normal distribution, except that it has fatter tails, thereby allowing more mass on larger effects (Rouder et al., 2009). The Cauchy distribution is widely used to compute BFs since it has been introduced for specifying the prior by Jeffreys (1961), which facilitates the interpretation of the BFs reported in the present article to those

reported in other work. Sensitivity analyses were conducted for $d = 0.20$ and $d = 0.80$ to reflect small and large effect sizes, respectively. For the homogeneity parameter τ, we used an inversed Gamma distribution ($\alpha = 1.23$, $\beta = 0.16$ and boundaries ranging from 0.01 to ∞). This prior distribution is based on the heterogeneity observed in mean-difference effect sizes reported in Psychological Bulletin between 1990 and 2013 (Gronau et al., 2017; van Erp et al., 2017). Using such an informed prior formalizes the existing knowledge about the relative probability of values for parameters, which improves the estimates of the effect sizes and their heterogeneity and, thus, yields also more informative BFs (for more detailed discussions of the benefits of using informed priors, see Lee & Vanpaemel, 2018; Steel et al., 2014). Using the default prior for the homogeneity parameter τ (an inversed Gamma distribution with $\alpha = 1.00$, $\beta = 0.15$) led to same conclusions.

Risk of Bias

For each study, coders independently rated risk of bias from six sources: random sequence generation for random allocation of participants to conditions, allocation concealment, blinding of participants, blinding of assessors, participant attrition, and completeness of reporting using the Cochrane Risk of Bias Tool (Higgins et al., 2011). One point was awarded for low risk of bias, -1 point for high risk of bias, and 0 points for unclear risk of bias for each of the six sources. As for categorical moderators, discrepancies between coders were resolved by discussion and involving a third reviewer (CvB). Interrater reliability was $\kappa = .63$.

As small-sample studies can yield biased results, we visually examined funnel plots depicting effect sizes against their precision and tested for asymmetry using Egger's regression (Egger et al., 1997) and Begg and Mazumdar's (1994) rank correlation. We further pre-registered to conduct a p-curve analysis (Simonsohn et al., 2014); however, 29 out of the 44 studies had to be excluded from *p*-curve analysis for various reasons (e.g., non-significant

results: $k = 4$, not reporting the statistics to compute required *p*-values: $k = 18$). We therefore deemed the analysis not reflective of the full sample of studies and refrained from reporting it here. For the sake of completeness, the *p*-curve disclosure table and results can be found on OSF (https://osf.io/8mwnt/). Finally, we were unable to conduct the pre-registered comparison between published and unpublished studies as a measure for publication bias as we did not identify unpublished data in the search.

Results

Table 2 lists descriptive information of the 44 included studies yielding $k = 45$ comparisons that reported data from 2444 participants (experimental $n = 1{,}300$, age $M = 80.47$ years, $SD = 6.39$; control $n = 1{,}144$, age $M = 80.15$ years, $SD = 6.68$). Most participants were diagnosed with Alzheimer's disease ($k = 22$ comparisons). In two studies, participants were diagnosed with vascular dementia, and in six other studies samples consisted of participants diagnosed with diverse types of dementia. Studies varied in the interventions administered. Twelve administered CST as described by Spector et al. (2001), with one of them also involving exercise. Eleven studies used some other form of CS according to Clare and Woods' (2004) definition, with four of them also involving exercise and one of them also involving reality orientation. Of the 15 studies primarily administering some form of reality orientation, two studies also used exercise, and five studies combined it with reminiscence activities. Six further studies used reminiscence ($k = 7$, as one study included two comparisons), with one of them combining reminiscence with exercise.

Eight studies compared CS to an active control group. Active control activities varied in the degree to which they involved group interactions and/or structured activities. In two studies (Capotosto et al., 2017; Piras et al., 2017), sessions dedicated to reading newspaper articles or stories from books with subsequent group discussions were alternated with sessions focusing on creative activities such as coloring, painting, decorating, or cooking. Woods et al.

(1979) engaged active control participants in unstructured group discussions and occasional board games, and Buschert et al. (2011) administered monthly group meetings in which participants completed paper-pencil exercises for self-study. In the study of Wallis et al. (1983), active control participants could freely choose between a range of not further described group and individual activities. Niu et al. (2010) engaged participants in the active control group in unstructured, individual conversational interactions. Baldelli et al. (2002) administered regular physical rehabilitation, but it is unclear whether it was administered individually or in a group setting. In the study of Quayhagen et al. (1995) the degrees of group interaction and structure are unclear, with the control intervention involving explicitly passive, observational activities (e.g., watching TV).

Most studies were conducted in Europe (k = 29), most of them in Italy (k = 10), followed by the UK (k = 8). Seven studies were conducted in Asia (k = 8, as one study included two comparisons), three in North America, two in South America, and one in Africa; for two studies the information where they conducted the intervention was not available. Table 3 lists the effect sizes and strength of evidence for the primary (global cognition, memory, and language) and secondary outcomes (quality of life, activities of daily living, depression, anxiety, dementia, and behavior). Table 4 lists the results for the moderators tested for each outcome.

Primary Outcomes

Global Cognition

Effects of CS on global cognition (k = 42) were predominantly assessed with the ADAS-Cog (k = 17) or MMSE (k = 19).

Average Effect at Posttest. Figure 2 illustrates the distribution of effect sizes from single studies and the strength of their associated evidence. Like Aguirre et al. (2013), we found a medium-sized effect of CS on global cognition, g = 0. 49, 95% CI [0.35, 0.63], t =

6.95, $p < .001$. Evidence for this effect was decisive, $BF_{H1} = 3{,}345{,}365.05$. Effect size estimates from single studies ranged from $g = -0.19$ (Coen et al., 2011) to $g = 1.71$ (Baglio et al., 2015). The evidence of 69% of studies was ambiguous only (i.e., BFs between 1/3 and 3). Of the studies that did contribute more substantial evidence (i.e., BFs greater than 3 or smaller than 1/3), 92% favored benefits of CS on global cognition. The heterogeneity in effect sizes was significant, $Q(41) = 92.56$, $p < .001$, with an estimated true between-studies variance $\tau^2 = 0.10$ ($SE = 0.05$). Evidence for the heterogeneity in effect sizes was decisive, $\tau^2\ BF_{H1} = 26{,}306.06$. The proportion of variability from true heterogeneity relative to that from sampling error was moderate, $I^2 = 54.68$, 95% CI [28.52, 72.56].

None of the variables significantly moderated the effects of CS on global cognition (ps $\geq .230$), with the residual heterogeneity remaining significant (all ps $< .01$). The evidence consistently supported the absence of moderator effects, with most $BF_{H1} < 1/3$. The only exception was observed for the effect of type of intervention, $BF_{H1} = 1/2.33$. The average effect of interventions involving some form of reality orientation was numerically larger than that for interventions involving CST, with stronger evidence for the effectiveness of reality orientation ($BF_{H1} = 129.07$) than for CST ($BF_{H1} = 13.39$). Because of the practical relevance of the intervention-specific categorical moderators, we illustrate the effect sizes for each level of moderator in Figure 3.

Only eight studies compared the effects of CS to an active control group. Although we preregistered to only test moderators with at least 10 cases per level, we decided to explore the effect of this critical methodological feature on the overall effect size given its clinical importance. For example, a substantially smaller average effect size for actively controlled studies than passively controlled studies would suggest that CS benefits were primarily due to expectancy and non-generic intervention effects (e.g., effects from committing to and participating in a regular activity). However, effect sizes did not differ significantly between

studies with active and passive controls, $F(1, 40) = 0.27$, $p = .606$, with strong evidence supporting the absence of an effect, $BF_{H1} = 1/4.28$ (small-effect prior: $BF_{H1} = 1/2.18$, large-effect prior: $BF_{H1} = 1/6.47$). Numerically, the effect was smaller for actively controlled studies though, $g = 0.41$, 95% CI [0.18, 0.65], $t = 4.14$, $p = .004$, than for passively controlled studies, $g = 0.51$, 95% CI [0.34, 0.68], $t = 6.10$, $p < .001$, and the strength of evidence was weaker in the set of actively controlled studies, $BF_{H1} = 24.40$ (small-effect prior: $BF_{H1} = 22.30$, large-effect prior: $BF_{H1} = 19.04$), than in the set of passively controlled studies, $BF_{H1} = 105{,}213.31$ (small-effect prior: $BF_{H1} = 76{,}025.55$, large-effect prior: $BF_{H1} = 95{,}486.87$). However, the relatively weaker evidence may be due to the smaller number of studies with active controls.

Average Effect at Follow-Up. Aguirre et al. (2013) reported some initial evidence for long-term effects of CS on global cognition based on four studies. Meanwhile, nine of the studies included in the present analysis tested long-term effects at follow-up 1 to 10 months after the intervention. Different from Aguirre et al.'s findings, the average effect was small and non-significant, $g = 0.22$, 95% CI [-0.09, 0.54], $t = 1.64$, $p = .140$. However, the evidence was ambiguous, $BF_{H1} = 1.67$, likely due to the still small number of studies testing for follow-up effects. Heterogeneity in effect sizes was not significant, $Q(8) = 11.18$, $p = 0.192$, $\tau^2 = 0.03$ (0.06), τ^2 $BF_{H1} = 1/1.17$, $I^2 = 19.10$, 95% CI [0.00, 84.37]. No moderators were tested given the small number of follow-up effect sizes and the non-significant heterogeneity in effect sizes.

Memory

The average effect of CS on memory ($k = 15$) was medium-sized, $g = 0.34$, 95% CI [0.06, 0.62], significant, $t = 2.61$, $p = .021$, and supported by substantial evidence $BF_{H1} = 9.74$. The heterogeneity in effect sizes was significant, $Q(14) = 30.16$, $p = .007$, $\tau^2 = 0.12$ ($SE = 0.09$), τ^2 $BF_{H1} = 9.09$, $I^2 = 53.59$, 95% CI [9.27, 83.28]. None of the tested moderators was

significant but, notably, the evidence was mostly inconclusive, except in case of the total number of sessions, BF_{H1} = 1/3.20.

Language

The average effect of CS on language (k = 14) was small and non-significant, g = 0.10, 95% CI [-0.47, 0.67], t = 0.38, p = .710. The strength of evidence for the absence of this effect was ambiguous, BF_{H1} = 1/1.89. The heterogeneity in effect sizes was significant, $Q(13)$ = 86.64, $p < .001$, τ^2 = 0.83 (SE = 0.38), τ^2 BF_{H1} = 2,328,499,419.34, I^2 = 88.87, 95% CI [77.94, 95.94]. None of the tested moderators was significant; however, the evidence was ambiguous in all cases.

Secondary Outcomes

Quality of Life

The average effect of CS on quality of life (k = 11) was small and non-significant, g = 0.16, 95% CI [-0.16, 0.48], t = 1.10, p = .295. Evidence supported the absence of an effect but was ambiguous, BF_{H1} = 1/1.07. Heterogeneity in effect sizes was significant, $Q(10)$ = 18.94, p = .041, τ^2 = 0.02 (SE = 0.05), τ^2 BF_{H1} = 1.31, I^2 = 25.64, 95% CI [0.00, 92.42]. None of the tested moderators was significant, with the evidence largely supporting the absence of effects.

Activities of Daily Living

The average effect of CS on activities of daily living (k = 14) was small yet significant, g = 0.17, 95% CI [0.02, 0.32], t = 2.48, p = .028, but the evidence was ambiguous, BF_{H1} = 2.04. The distribution of effect sizes is illustrated in Figure 4. Effect size estimates from single studies ranged from g = -1.07 (Alves, Alves-Costa, Magalhães, Gonçalves, & Sampaio, 2014) to g = 0.57 (D'Onofrio et al., 2015). These two studies were also the only ones contributing substantial evidence, albeit favoring opposite hypotheses. Evidence from the other studies was ambiguous. Heterogeneity in effect sizes was non-significant, $Q(13)$ =

12.98, p = .456, τ^2 = 0.02 (SE = 0.03), τ^2 BF_{H1} = 1/1.04, I^2 = 20.38, 95% CI [0.00, 55.29]. Thus, no moderators were tested.

Depression

The average effect of CS on depression measures (k = 19) was medium-sized and significant, g = 0.46, 95% CI [0.15, 0.78], t = 3.07, p = .007. The evidence supporting this effect was strong, BF_{H1} = 27.26. The distribution of effect sizes is illustrated in Figure 5. Effect sizes ranged from g = -0.86 (Requena, Maestú, Campo, Fernández, & Ortiz, 2006) to g = 1.97 (Menna, Santaniello, Gerardi, Di Maggio, & Milan, 2016). Evidence from 58% of studies was ambiguous. Of the studies that contributed more substantial evidence, 75% favored benefits of CS on depression scores. The heterogeneity in effect sizes was significant, Q(18) = 68.88, p < .001, τ^2 = 0.28 (SE = 0.14), τ^2 BF_{H1} = 118,219.64, I^2 = 74.73, 95% CI [53.71, 89.65].

Numerically, the average effect size tended to be greater for interventions that were shorter, involved less sessions per week, and entailed a lower number of total sessions, with the latter effect being significant, F(1, 17) = 4.90, p = .041. However, evidence for the effects of dosage was ambiguous, with BF_{H1} ranging between 1/1.14 and 2.35. Figure 6 visualizes the counterintuitive relationship between dosage and average effect sizes. Visual inspection suggested that the negative relationship might be driven by a single study that administered five sessions per week over the course of 104 weeks (Requena et al., 2006). Indeed, after excluding the study from analysis[3], the association between duration and the average effect size became positive and significant, F(1, 16) = 5.90, p = .027, BF_{H1} = 3.41. The associations between frequency and total number of sessions and average effect size were also positive, albeit non-significant (ps ≥ .563), with the evidence supporting the absence of moderator effects (BF_{H1} = 1/2.69 and 1/3.41, respectively). Evidence was ambiguous regarding the

[3]We reran the moderator analyses for testing the effects of intervention duration, intensity, and dosage for all outcomes when excluding the study by Requena et al. (2006). Results yielded identical conclusions.

remaining tested moderators age, gender, and baseline MMSE, with BF_{H1} ranging between 1/2.43 and 1.05. The evidence may be ambiguous due to the relatively low number of cases for a meta-regression.

Anxiety

The average effect of CS on anxiety (k = 5) was small and non-significant, g = 0.25, 95% CI [-0.28, 0.77], t = 1.31, p = .259. Evidence favoring an effect of CS on anxiety scores was ambiguous, BF_{H1} = 1.23. Heterogeneity in effect sizes was non-significant, $Q(4)$ = 6.89, p = .142, τ^2 = 0.07 (SE = 0.12), τ^2 BF_{H1} = 1/1.30, I^2 = 43.10, 95% CI [0.00, 94.16]. No moderators were tested given the small number of studies available and the non-significant heterogeneity in effect sizes.

Dementia Ratings

The average effect of CS on dementia ratings (k = 7) was medium-sized and significant, g = 0.66, 95% CI [0.02, 1.29], t = 2.54, p = .044. The effect was supported by strong evidence, BF_{H1} = 10.17. The distribution of effect sizes is illustrated in Figure 7. Effect size estimates from single studies ranged from g = -0.17 (Coen et al., 2011) to g = 1.75 (Spector et al., 2001). Three studies reported substantial evidence favoring an effect of CS on dementia ratings, and four studies contributed ambiguous evidence. Data were missing from one study that did not report group size data (Kolanowski et al., 2016). Heterogeneity in effect sizes was significant, $Q(6)$ = 26.13, p < .001, τ^2 = 0.34 (SE = 0.25), τ^2 BF_{H1} = 36.81, I^2 = 80.63, 95% CI [50.28, 96.71], but the small number of studies prevented us from testing moderator effects.

Behavior

The average effect of CS on behavior (k = 11) was small and non-significant, g = 0.28, 95% CI [-0.60, 1.17], t = 0.71, p = .492. Evidence favoring the absence of an effect of CS on behavior measures was ambiguous, BF_{H1} = 1/1.19. Heterogeneity in effect sizes was

significant, $Q(10) = 95.09$, $p < .001$, $\tau^2 = 1.53$ ($SE = 0.75$), τ^2 $BF_{H1} = 54{,}654{,}671{,}997.77$, $I^2 = 95.10$, 95% CI [89.54, 98.55], but, again, the small number of studies prevented us from testing moderator effects.

Assessment of Bias

Risk of Bias

Figure 8 displays the results of the risk of bias assessment using the Cochrane Risk of Bias Tool (Higgins et al., 2011).

Selection Bias. Participants were allocated to conditions based on a random, concealed allocation method in 20 of the 45 included comparisons (44%). In two studies (Camargo, Justus, & Retzlaff, 2015; Requena et al., 2006), participants were allocated to conditions using the day of arrival, which may result in selection bias if a researcher can anticipate when potential participants will be assigned (Higgins et al., 2011).

Performance Bias. In 37 studies (82%), the effectiveness of CS was compared to some form of passive control group. In addition to not controlling for non-specific intervention and expectancy effects, participants in a passive control group can unambiguously determine that they are not receiving treatment. Thus, the use of a passive control also prevents true participant blinding. While Baldelli et al. (2002) did include a control activity, we deemed it as not comparable enough to effectively blind participants. Baldelli et al.'s (2002) control group engaged only in physical therapy, while the active group engaged in physical therapy as well as reality orientation sessions[4]. On a positive note, most studies ensured and reported to have blinded assessors to conditions.

Attrition and Reporting Bias. We identified overall little bias from attrition and reporting, with a low risk of bias in 40 comparisons (89%). The highest attrition rate during treatment was 39% observed in Wallis et al. (1983), where participants were excluded if

[4] Reclassifying Baldelli et al. (2002) as "unclear risk of bias" for participant blinding did not affect the conclusions.

completing less than 20% of the study. Importantly though, attrition rates were similar across groups. Most studies (82%) reported data on all outcomes mentioned in the respective method sections.

Impact of Risk of Bias

For each outcome with at least 10 cases, we assessed the impact of the risk of bias on the average effect sizes using meta-regression (not preregistered). As for the above analyses, we computed BFs to gauge the strength of evidence for each analysis. The sum score of the risk of bias ratings was not significantly associated with the average effect size estimates for any of those outcomes (global cognition: $F(1, 40) = 0.76$, $p = .388$; memory: $F(1, 13) = 0.52$, $p = .483$; language: $F(1, 12) = 0.05$, $p = .832$; quality of life: $F(1, 9) = 0.22$, $p = .649$; activities of daily living: $F(1, 12) = 0.59$, $p = .457$; depression: $F(1, 17) = 0.32$, $p = .578$, behavior, $F(1, 9) = 2.48$, $p = .150$. The evidence substantially supported the absence of an association with risk of bias ratings for global cognition ($BF_{H1} = 1/3.96$) and activities of daily living ($BF_{H1} = 1/3.18$), but it was less clear cut for memory ($BF_{H1} = 1/2.84$), language ($BF_{H1} = 1/2.07$), quality of life ($BF_{H1} = 1/2.88$), and depression symptoms ($BF_{H1} = 1/2.66$). The evidence was particularly ambiguous for behavior, $BF_{H1} = 1.20$. Evidence was likely ambiguous for the effects of bias on the average effect size estimates for these outcomes because of the relatively low number of cases for meta-regression analyses.

Publication Bias

To assess small-study bias, we first visually inspected contour-enhanced funnel plots for outcomes with at least 10 cases (see Figure 9). For global cognition, few effect sizes fell into the areas of negative effect sizes; however, because many effect sizes also fell into the area of non-significance for positive effect sizes, publication bias seems an unlikely cause for the pattern. Overall, the funnel plots did not suggest censoring of non-significant effect sizes.

Both Egger's and Begg's and Mazumdar's regression were non-significant for these outcomes ($ps \geq .116$ and $ps \geq .108$, respectively).

Discussion

This pre-registered systematic review and meta-analysis examined the effectiveness of CS for people with dementia on cognitive outcomes, quality of life, activities of daily living, mood, and behavior and psychological symptoms. Pooling data from 44 published RCTs and considering the Bayesian strength of evidence supporting the presence or absence of average effects, the present meta-analysis is the most comprehensive to date.

CS Improves Performance on Indicators of Global Cognition

We found a medium effect of CS on global cognition ($g = 0.49$, 95% CI [0.35, 0.63]) that was slightly larger than that reported in previous meta-analyses (e.g., SMD = 0.41 in Aguirre et al., 2013) and was supported by decisive evidence (BF > 3,000,000). The effect was no longer significant at follow-up assessments 1 to 10 months after the intervention; however, estimates at follow-up were based on only nine studies, and the evidence was ambiguous. Thus, additional data are needed to gauge long-term benefits of CS on global cognition. Whereas we also found substantial evidence for CS benefits on memory, the effect on language measures was small, non-significant, and based on ambiguous evidence.

None of the moderators reached statistical significance; however, we found a numerically larger pooled effect size for interventions involving reality orientation than for CST. Moreover, evidence decisively supported the benefits of reality orientation, and only strongly the effectiveness of CST. We also observed an, albeit non-significant, tendency that interventions with multiple components (e.g., additional physical exercise) were more effective than those with single components. Only one CST intervention (Buschert et al., 2011), but four reality orientation interventions (Baglio et al., 2015; Baines et al., 1987; Baldelli et al., 2002; Bottino et al., 2005), incorporated additional components, which to some

degree confounds effects of multiple components and the difference between CST and reality orientation. There was also no significant effect of whether interventions were conducted in care or in another context, although effects were numerically larger for the latter. Age, gender, and baseline cognitive status did not moderate the effect. However, although the absence of moderator effects was supported by largely substantial evidence, we cannot rule out the possibility of false-negative findings where the evidence was more ambiguous, especially given the overall still relatively low number of studies.

Evidence for Benefits of CS on Some Secondary Outcomes

Previous meta-analyses reported positive effects of CS on quality of life (Aguirre et al., 2013; Kim et al., 2017) but not for activities of daily living (Aguirre et al., 2013; Kim et al., 2017; Kurz et al., 2011). We observed the opposite pattern: there were no CS benefits on quality of life but a small yet significant effect on activities of daily living. However, the evidence was ambiguous for both outcomes, which may explain the discrepancy. Multiple factors can contribute to evidence being ambiguous, for example low statistical power or low reliability of the outcome measures. In this case, we could only speculate why the evidence is ambiguous. Moreover, we found a positive effect of CS on depressive symptoms and on dementia ratings. The evidence was strong in both cases, and relatively consistently positive effects were observed in those single studies that contributed at least substantial evidence. There was also some tentative evidence that interventions with a longer duration (i.e., higher number of weeks) attenuated depressive symptoms more than shorter interventions. Therefore, intervention dosage may be an important design factor to consider when targeting depression-related outcomes in this population. No significant benefits were found for anxiety and behavior symptoms, but the evidence was ambiguous for these outcomes. Hence, a larger and stronger evidence base is needed to allow for drawing firm conclusions on the benefits of CS on most secondary outcomes and their potential moderators.

Lack of Double-Blinded, Actively Controlled Studies

We detected no publication bias and, overall, risk of biases in the included studies was relatively moderate and not systematically associated with study outcomes. However, in most studies, participant blinding was inadequate. Of the 44 published RCTs, only 8 compared effects of CS to active control groups. As participants in a passive control group can never be truly blinded, effects observed could be confounded with effects of participant- or experimenter-driven expectancy. Moreover, of those, only five studies designed alternative interventions that allowed for distinguishing between CS-specific and generic intervention effects such as from engaging in social activities. We identified two double-blinded (but strongly underpowered: $n = 9$ and $n = 38$) studies where both participants and assessors were blind to group allocation (Wallis et al., 1983; Woods, 1979). This is highly problematic as it implies that the sizable effect of CS benefits on global cognition could be mainly driven by expectancy and non-generic intervention effects. Notably, our additional analyses did not suggest a difference in average effect sizes between actively and passively controlled studies; however, the absence of a difference on the meta-analytic level cannot rule out such undesirable effects in any single study. For example, awareness of being or not being selected for a new treatment that includes regular, social activities could affect motivation and mood when completing the outcome measures. We do not suggest disregarding the overall positive effects of CS, but it warrants caution as to what the mechanisms of these benefits are. For example, if regular group activities regardless of the specific contents were sufficient to improve cognition and mental health in people with dementia, then there would be no need to run full-blown CS programs. Therefore, it is important to evaluate the effectiveness of CS over and above generic group activities.

To address this, future RCTs should include an active control group that completes activities comparable to those in the CS group but are assumed not to yield the cognitive

benefits of CS. Specifically, (1) the active control group should complete activities with the same frequency and duration as the treatment group; (2) the activities should be distinct from those included in CS but plausible enough to control for expectancy effects; and (3), as social interactions and networks have been found to protect against dementia (Fratiglioni et al., 2000), the treatment and the control group should have similar levels of staff and peer interaction, though the form of staff interaction may differ. For example, in Wallis et al.'s (1983) study, the control group was offered activities each day and general conversation was encouraged, while orientation topics were only mentioned if it occurred in ordinary conversation.

Arguably, randomly allocating participants to complete activities believed to be ineffective is ethically questionable. To avoid this problem, benefits of CS should, ideally, be compared to an alternative intervention that is assumed to be potentially effective through other mechanisms than CS (see also Bahar-Fuchs et al., 2020). However, there is currently no consensus of what aspects of CS are likely to benefit people with dementia. For example, the benefits of reality orientation may come from its structured implementation (Bowlby, 1991; Voelkel, 1978), but other forms of CS focus on reminiscence or social interaction in a specified way. Critically, our moderator analysis did not unambiguously show that either approach is superior to the other.

Are the Benefits Clinically Relevant and Generalizable?

Cost-effectiveness analysis for CST found that, in the best-case scenario, it is a cost-effective therapy (Knapp et al., 2006). However, as any intervention, CS inevitably generates not only financial but also opportunity costs. Thus, it is important to consider whether any effects observed are clinically relevant and generalizable to contexts other than the diagnostic settings. The U.S. Food and Drug Administration (FDA, 1989, as cited in Hansen et al., 2012; see also Qaseem et al., 2008) define a reduction of ≥ 4 points on the ADAS-Cog as clinically

relevant, with distribution-based analyses suggesting ≥ 3 points (Schrag, Schott, & Alzheimer's Disease Neuroimaging Initiative, 2012). Of the 17 studies that used the ADAS-Cog, scores on average decreased by 3.50 points more than in the control group after the intervention. However, this decrease was primarily due to one outlier (Requena et al., 2006): after removing this study, the average difference in the decrease in ADAS-Cog scores was 2.41. Thus, despite the overall medium effect size, the cognitive changes after CS are borderline clinically relevant. Nonetheless, the positive effects of CS on activities of daily living, depressive symptoms, and dementia ratings are promising. Except for the effect on activities of daily living, these benefits of CS were supported by substantial and strong evidence. However, none of the studies assessed whether the cognitive benefits of CS were in any way linked to reduced dementia progression, which would be a critical indicator of its effectiveness (Ito et al., 2011). Finally, due to the variation in CS interventions, it can be difficult to compare interventions even of the same type. Whereas CST is manualized (Spector et al., 2006) and has been adapted to multiple cultures (Capotosto et al., 2017; Yamanaka et al., 2013), other CS interventions such as reality orientation lack operational guides (Bowlby, 1991). To enable the identification of the most effective components of successful CS interventions, it is critical for future research to describe the interventions administered in greater detail.

Limitations

The findings of this systematic review and meta-analysis should be interpreted in the context of its limitations. First, our approach of selecting one effect size where multiple measures for the same type of outcome were reported (e.g., selecting the ADAS-Cog over the MMSE) arguably yielded some loss of information. We made the decision to select one of the measures and discarded the other measure(s) reported based on theoretical or empirical considerations (e.g., evidence that the ADAS-Cog is a more precise measure of cognition in

dementia; Balsis et al., 2015). Alternative approaches would have been to create composite scores or to use a multi-level modeling approach that accounts for dependencies in effect sizes. However, we opted for our approach as we believe that the benefits of greater construct validity outweigh the loss of information from discarding the (overall relatively few) less adequate measures.

Second, although the present study included a considerably larger number of studies than previous meta-analyses examining the effectiveness of CS, the overall sample size was still relatively small, especially for the analysis of secondary outcomes and moderator variables. Thus, statistical power was potentially low for some analyses. Moreover, the outcome measures administered in the included studies may vary in their reliability and, thus, measurement error may have attenuated the effect sizes. Only few studies reported reliabilities of the measures used and, therefore, we were unable to correct for measurement error. To address potentially low statistical power and varying reliabilities of outcome measures, we reported BFs for all analyses. BFs cannot correct for small sample sizes or measurement error, but they provide an estimate of the sensitivity of the given data distinguished between the presence or absence of effects by quantifying the evidence available. BFs indicated that the evidence was sufficiently strong to support either the presence of average effects (global cognition, memory, depression, and dementia ratings; see Table 3) or the absence of effects of most potential moderators on global cognition and quality of life (see Table 4). However, where the evidence was ambiguous only (i.e., close to 1), the available data were not sufficiently sensitive to draw firm conclusions, possibly due to noise from the relatively small number of studies, measurement error, or a combination of both. Therefore, findings with low BFs should be interpreted cautiously until more data are available.

Third, studies varied widely in the instruments administered to assess the primary and secondary outcomes of interest for the present meta-analysis. For example, to assess activities

of daily living, five studies administered the Instrumental Activities of Daily Living scale (Lawton & Brody, 1969), three studies used the Barthel Index (Mahoney & Barthel, 1965), and the remaining studies used either the MOSES self-care scale (Helmes et al., 1987; two studies), the Disability Assessment for Dementia (Gélinas et al., 1999; De Vreese et al., 2008; two studies), Stewart's Activities of Daily Living (Stewart, 1980; one study), or a performance-based test (Graessel et al., 2009; one study). Although these instruments all have been designed to assess broadly the same construct – activities of daily living – they may differ in what specific activities they place emphasis on. Thus, between-study heterogeneity may partly be driven by the variability in outcome measures. Different instruments may also vary in terms of internal and test-retest reliability, adding further noise to the meta-analytic estimates. Therefore, it is possible that the heterogeneity in outcome measures contributed to the ambiguity in evidence we observed for some of the meta-analytic average effects.

Finally, the reporting in the original studies was often insufficient to extract the relevant data. For example, incomplete statistical reporting prevented us from conducting a *p*-curve analysis we pre-registered. Moreover, because of the lack of detail in descriptions of samples and interventions, data were missing for several potentially critical moderators. For example, descriptive data on ethnicity was missing in 31 of 44 studies; thus, it is impossible to gauge whether the data are representative for the population of people with dementia. Moreover, 16 of the included studies did not report clinical severity of their sample (i.e., baseline MMSE scores). The lack of data limits the representativeness of the absence of a moderating effect of clinical severity on global cognition and depression, and it prevented us from examining effect of this moderator on memory, language, and quality of life. Furthermore, the MMSE provides only information of clinical severity of cognitive impairments; ideally, and especially given that the present meta-analysis found evidence for CS benefits on non-cognitive outcomes, clinical severity should be assessed more holistically

through clinical dementia ratings. However, only seven of the included studies provided such data. Knowing who benefits most from CS interventions is critical both for practitioners treating people with dementia as well as for further developing maximally effective interventions.

Conclusions

In this systematic review and meta-analysis, we examined the evidence for the effectiveness of CS, a group of interventions recommended by NICE (2018) for treating dementia. We found a medium average effect of CS on indicators of global cognition (e.g., ADAS-Cog, MMSE) supported by decisive Bayesian evidence. Additional positive effects on memory and secondary outcomes – activities of daily living, depression, and dementia rating – were mostly supported by substantial, and sometimes strong, evidence. However, only few studies tested the longer-term benefits of CS, yielding inconclusive evidence at this point. Furthermore, the review highlighted a lack of double-blinded, actively controlled studies, and relatively low clinical relevance of the effects. To optimize treatment of dementia, future RCTs should evaluate longer-term benefits of CS, will need better controls for non-specific intervention effects, and more directly test generalization of effects beyond diagnostic measures of global cognition. This study concludes that the benefits of CS for people with dementia are promising. Practitioners may find that the medium average effect size, relatively low cost and absence of negative side-effects of CS outweigh that the effects are borderline clinically relevant. Importantly, however, a stronger evidence basis for the long-term effectiveness of CS beyond generic group activities is urgently needed before definitive practice recommendations can be made.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

Aguirre, E., Woods, R. T., Spector, A., & Orrell, M. (2013). Cognitive stimulation for dementia: A systematic review of the evidence of effectiveness from randomised controlled trials. *Ageing Research Reviews*, *12*(1), 253–262. https://doi.org/10.1016/j.arr.2012.07.001

*Alves, J., Alves-Costa, F., Magalhães, R., Gonçalves, Ó. F., & Sampaio, A. (2014). Cognitive stimulation for Portuguese older adults with cognitive impairment. *American Journal of Alzheimer's Disease & Other Dementias*, *29*(6), 503–512. https://doi.org/10.1177/1533317514522541

Alzheimer's Disease International [ADI]. (2018). *World Alzheimer Report 2018*. London, UK. https://www.alz.co.uk/research/WorldAlzheimerReport2018.pdf

Arnholt, A. T., & Evans, B. (2017). BSDA: Basic Statistics and Data Analysis. https://cran.r-project.org/package=BSDA

*Baglio, F., Griffanti, L., Saibene, F. L., Ricci, C., Alberoni, M., Critelli, R., Villanelli, F., Fioravanti, R., Mantovani, F., D'amico, A., Cabinio, M., Preti, M. G., Nemmi, R., & Farina, E. (2015). Multistimulation group therapy in Alzheimer's disease promotes changes in brain functioning. *Neurorehabilitation and Neural Repair*, *29*(1), 13–24. https://doi.org/10.1177/1545968314532833

Bahar-Fuchs, A., Clare, L., & Woods, B. (2013). Cognitive training and cognitive rehabilitation for mild to moderate Alzheimer's disease and vascular dementia. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD003260.pub2

Bahar-Fuchs, A., Martyr, A., Goh, A. M. Y., Sabates, J., & Clare, L. (2020). Cognitive training for people with mild to moderate dementia. *Cochrane Database of Systematic*

Reviews. https://doi.org/10.1002/14651858.CD013069.pub2

*Baines, S., Saxby, P., & Ehlert, K. (1987). Reality orientation and reminiscence therapy. A controlled cross-over study of elderly confused people. *British Journal of Psychiatry*, *151*(2), 222–231. https://doi.org/10.1192/bjp.151.2.222

*Baldelli, M. V., Boiardi, R., Fabbo, A., Pradelli, J. M., & Neri, M. (2002). The role of reality orientaion therapy in restorative care of elderly patients with dementia plus stroke in the subacute nursing home setting. *Archives of Gerontology and Geriatrics*, *35*, 15–22. https://doi.org/10.1016/S0167-4943(02)00098-5

*Baldelli, M. V., Pirani, A., Motta, M., Abati, E., Mariani, E., & Manzi, V. (1993). Effects of reality orientation therapy on elderly patients in the community. *Archives of Gerontology and Geriatrics*, *17*(3), 211–218. https://doi.org/10.1016/0167-4943(93)90052-J

Ballard, C. G., Gauthier, S., Cummings, J. L., Brodaty, H., Grossberg, G. T., Robert, P., & Lyketsos, C. G. (2009). Management of agitation and aggression associated with Alzheimer disease. *Nature Reviews Neurology*, *5*(5), 245–255. https://doi.org/10.1038/nrneurol.2009.39

Balsis, S., Benge, J. F., Lowe, D. A., Geraci, L., & Doody, R. S. (2015). How do scores on the ADAS-Cog, MMSE, and CDR-SOB correspond? *The Clinical Neuropsychologist*, *29*(7), 1002–1009. https://doi.org/10.1080/13854046.2015.1119312

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446

Borenstein, M., Hedges, L. V, Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Wiley. https://doi.org/10.1002/9780470743386

*Bottino, C. M. C., Carvalho, I. A. M., Alvarez, A. M. M. A., Avila, R., Zukauskas, P. R., Bustamante, S. E. Z., Andrade, F. C., Hototian, S. R., Saffi, F., & Camargo, C. H. P. (2005). Cognitive rehabilitation combined with drug treatment in Alzheimer's disease

patients: A pilot study. *Clinical Rehabilitation, 19*(8), 861–869. https://doi.org/10.1191/0269215505cr911oa

Bowlby, M. C. (1991). Reality orientation thirty years later: Are we still confused? *Canadian Journal of Occupational Therapy, 58*(3), 114–122. https://doi.org/10.1177/000841749105800303

*Breuil, V., De Rotrou, J., Forette, F., Tortrat, D., Ganansia-Ganem, A., Frambourt, A., Moulin, F., & Boller, F. (1994). Cognitive stimulation of patients with dementia: Preliminary results. *International Journal of Geriatric Psychiatry*, *9*(3), 211–217. https://doi.org/10.1002/gps.930090306

*Buschert, V. C., Friese, U., Teipel, S. J., Schneider, P., Merensky, W., Rujescu, D., Möller, H.-J., Hampel, H., & Buerger, K. (2011). 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*, *25*(4), 679–694. https://doi.org/10.3233/JAD-2011-100999

*Camargo, C. H. F., Justus, F. F., & Retzlaff, G. (2015). The effectiveness of reality orientation in the treatment of Alzheimer's Disease. *American Journal of Alzheimer's Disease and Other Dementias*, *30*(5), 527–532. https://doi.org/10.1177/1533317514568004

*Capotosto, E., Belacchi, C., Gardini, S., Faggian, S., Piras, F., Mantoan, V., Salvalaio, E., Pradelli, S., & Borella, E. (2017). 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*, *32*(3), 331–340. https://doi.org/10.1002/gps.4521

*Chapman, S. B., Weiner, M. F., Rackley, A., Hynan, L. S., & Zientz, J. (2004). Effects of cognitive-communication stimulation for Alzheimer's disease patients treated with

donepezil. *Journal of Speech, Language, and Hearing Research : JSLHR*, *47*(5), 1149–1163. https://doi.org/10.1044/1092-4388(2004/085)

Clare, L., Linden, D. E. J., Woods, R. T., Whitaker, R., Evans, S. J., Parkinson, C. H., van Paaschen, J., Nelis, S. M., Hoare, Z., Yuen, K. S. L., & Rugg, M. D. (2010). Goal-oriented cognitive rehabilitation for people with early-stage Alzheimer disease: A single-blind randomized controlled trial of clinical efficacy. *The American Journal of Geriatric Psychiatry*, *18*(10), 928–939. https://doi.org/10.1097/JGP.0b013e3181d5792a

Clare, L., & Woods, R. T. (2004). Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: A review. *Neuropsychological Rehabilitation*, *14*(4), 385–401. https://doi.org/10.1080/09602010443000074

*Coen, R. F., Flynn, B., Rigney, E., O'Connor, E., Fitzgerald, L., Murray, C., Dunleavy, C., McDonald, M., Delaney, D., Merriman, N., & Edgeworth, J. (2011). Efficacy of a cognitive stimulation therapy programme for people with dementia. *Irish Journal of Psychological Medicine*, *28*(3), 145–147. https://doi.org/10.1017/S0790966700012131

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed.). Russell Sage.

*Cove, J., Jacobi, N., Donovan, H., Orrell, M., Stott, J., & Spector, A. (2014). 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*, 2143–2150. https://doi.org/10.2147/CIA.S66232

Cummings, J. L., Morstorf, T., & Zhong, K. (2014). Alzheimer's disease drug-development pipeline: Few candidates, frequent failures. *Alzheimer's Research & Therapy*, *6*(4), Article 37. https://doi.org/10.1186/alzrt269

D'Amico, F., Rehill, A., Knapp, M., Aguirre, E., Donovan, H., Hoare, Z., Hoe, J., Russell, I., Spector, A., Streater, A., Whitaker, C., Woods, R. T., & Orrell, M. (2015). Maintenance

cognitive stimulation therapy: An economic evaluation within a randomized controlled trial. *Journal of the American Medical Directors Association, 16*(1), 63–70. https://doi.org/10.1016/j.jamda.2014.10.020

D'Ath, P., Katona, P., Mullan, E., Evans, S., & Katona, C. (1994). Screening, detection and management of depression in elderly primary care attenders. I: The acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. *Family Practice*, *11*(3), 260–266. https://doi.org/10.1093/fampra/11.3.260

*D'Onofrio, G., Sancarlo, D., Addante, F., Ciccone, F., Cascavilla, L., Paris, F., Elia, A. C., Nuzzaci, C., Picoco, M., Greco, A., Panza, F., & Pilotto, A. (2015). A pilot randomized controlled trial evaluating an integrated treatment of rivastigmine transdermal patch and cognitive stimulation in patients with Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *30*(9), 965–975. https://doi.org/10.1002/gps.4247

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2011). Chapter 9: Analysing data and undertaking meta-analyses. In J. P. T. Higgins & S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions: online version (Version 5.1.0)*. The Cochrane Collaboration. http://www.cochrane-handbook.org/

De Vreese, L. P., Caffarra, P., Savarè, R., Cerutti, R., Franceschi M., & Grossi, E. (2008). Functional disability in early Alzheimer's disease – a validation study of the Italian version of the Disability Assessment for Dementia scale. *Dementia and Geriatric Disorders, 25*, 186–194. https://doi.org/10.1159/000113415

Dietch, J. T., Hewett, L. J., & Jones, S. (1989). Adverse effects of reality orientation. *Journal of the American Geriatrics Society*, *37*(10), 974–976. https://doi.org/10.1111/j.1532-5415.1989.tb07284.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634.

https://doi.org/10.1136/bmj.315.7109.629

*Ferrario, E., Cappa, G., Molaschi, M., Rocco, M., & Fabris, F. (1991). Reality orientation therapy in institutionalized elderly patients: Preliminary results. *Archives of Gerontology and Geriatrics, Suppl 2*, 139–142.

Field, A. P., Gillet, R., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*, 665–694. https://doi.org/10.1348/000711010x502733

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fratiglioni, L., Wang, H.-X., Ericsson, K., Maytan, M., & Winblad, B. (2000). Influence of social network on occurrence of dementia: a community-based longitudinal study. *The Lancet*, *355*(9212), 1315–1319. https://doi.org/10.1016/S0140-6736(00)02113-9

Gélinas, I., Gauthier, I., McIntyre, M. C., & Gauthier, S. (1999). Development of a functional measure for persons with Alzheimer disease: The disability assessment for dementia. *The American Journal of Occupational Therapy*, *53*, 471–481. https://doi.org/10.5014/ajot.53.5.471

*Gibbor, L., Forde, L., Yates, L., Orfanos, S., Komodromos, C., Page, H., Harvey, K., & Spector, A. (in press). A feasibility randomised control trial of individual cognitive stimulation therapy for dementia: impact on cognition, quality of life and positive psychology. *Aging & Mental Health.* https://doi.org/10.1080/13607863.2020.1747048

Graessel, E., Stemmer, R., Eichenseer, B., Pickel, S., Donath, C., Kornhuber, J., & Luttenberger, K. (2011). *BMC Medicine, 9,* Article 129. https://doi.org/10.1186/1741-7015-9-129

Graessel, E., Viegas, R., Stemmer, R., Küchly, B., Kornhuber, J., & Donath, C. (2008). The

Erlangen Test of Activities of Daily Living.: First results on reliability and validity of a short performance test to measure fundamental activities of daily living in dementia patients. *International Psychogeriatrics, 21*(1), 103–112. https://doi.org/10.1017/S1041610208007710

Grochowalski, J. H., Liu, Y., & Siedlecki, K. L. (2016) Examining the reliability of ADAS-Cog change scores. *Aging, Neuropsychology, and Cognition, 23*(5), 513–529. https://doi.org/10.1080/13825585.2015.1127320

Gronau, Q. F., Van Erp, S., Heck, D. W., Cesario, J., Jonas, K. J., & Wagenmakers, E. (2017). A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: the case of felt power. *Comprehensive Results in Social Psychology*, *2*(1), 123–138. https://doi.org/10.1080/23743603.2017.1326760

*Haight, B. K., Gibson, F., & Michel, Y. (2006). The Northern Ireland life review/life storybook project for people with dementia. *Alzheimer's & Dementia, 2*(1), 56–58. https://doi.org/10.1016/j.jalz.2005.12.003

Hansen, R. A., Gartlehner, G., Lohr, K. N., & Kaufer, D. I. (2007). Functional outcomes of drug treatment in Alzheimer's Disease. *Drugs & Aging, 24*, 155–167. https://doi.org/10.2165/00002512-200724020-00007

Heck, D. W., Gronau, Q. F., & Wagenmakers, E.-J. (2017). metaBMA: Bayesian model averaging for random and fixed effects meta-analysis. https://cran.r-project.org/package=metaBMA

Helmes, E., Csapo, K. G., & Short, J.-A. (1987). Standardization and validation of the multidimensional observation scale for elderly subjects (MOSES). *Journal of Gerontology, 42*(4), 394–405. https://doi.org/10.1093/geronj/42.4.395

Higgins, J. P. T., Altman, D. G., & Sterne, J. A. C. (2011). Chapter 8: Assessing risk of bias in included studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane Handbook for*

Systematic Reviews of Interventions Version 5.1. The Cochrane Collaboration. http://handbook-5-1.cochrane.org/

Hill, N. T. M., Mowszowski, L., Naismith, S. L., Chadwick, V. L., Valenzuela, M., & Lampit, A. (2017). Computerized cognitive training in older adults with mild cognitive impairment or dementia: A systematic review and meta-analysis. *American Journal of Psychiatry*, *174*(4), 329–340. https://doi.org/10.1176/appi.ajp.2016.16030360

Holden, U. P., & Woods, R. T. (1995). *Positive Approaches to Dementia Care* (3rd ed.). Churchill Livingstone.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, *140*(6), 566–572. https://doi.org/10.1192/bjp.140.6.566

Huntley, J. D., Gould, R. L., Liu, K., Smith, M., & Howard, R. J. (2015). Do cognitive interventions improve general cognition in dementia? A meta-analysis and meta-regression. *BMJ Open*, *5*, e005247. https://doi.org/10.1136/bmjopen-2014-005247

Innes, A., & Manthorpe, J. (2013). Developing theoretical understandings of dementia and their application to dementia care policy in the UK. *Dementia*, *12*(6), 682–696. https://doi.org/10.1177/1471301212442583

*Ishizaki, J., Meguro, K., Ohe, K., Kimura, E., Tsuchiya, E., Ishii, H., Sekita, Y., & Yamadori, A. (2002). Therapeutic psychosocial intervention for elderly subjects with very mild Alzheimer disease in a community: the tajiri project. *Alzheimer Disease and Associated Disorders*, *16*(4), 261–269. https://doi.org/10.1097/00002093-200210000-00008

Ito, K., Corrigan, B., Zhao, Q., French, J., Miller, R., Soares, H., Nicholas, T., Billing, B., Anziano, R., Fullerton, T, & Alzheimer's Disease Neuroimaging Initiative. (2011). Disease progression model for cognitive deterioration from Alzheimer's Disease

Neuroimaging Initiative database. *Alzheimer's and Dementia*, *7*(2), 151–160. https://doi.org/10.1016/j.jalz.2010.03.018

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Kenigsberg, P.-A., Aquino, J.-P., Bérard, A., Gzil, F., Andrieu, S., Banerjee, S., Brémond, F., Buée, L., Cohens-Mansfield, J., Mangialasche, F., Platel, H., Salmon, E., & Robert, P. (2016). Dementia beyond 2025: Knowledge and uncertainties. *Dementia*, *15*(1), 6–21. https://doi.org/10.1177/1471301215574785

*Kim, H.-J., Yang, Y., Oh, J.-G., Oh, S., Choi, H., Kim, K. H., & Kim, S. H. (2016). Effectiveness of a community-based multidomain cognitive intervention program in patients with Alzheimer's disease. *Geriatrics & Gerontology International*, *16*(2), 191–199. https://doi.org/10.1111/ggi.12453

Kim, K., Han, J. W., So, Y., Seo, J., Kim, Y. J., Park, J. H., Lee, S. B., Lee, J. J., Jeong, H.-G., Kim, T. H., & Kim, K. W. (2017). Cognitive stimulation as a therapeutic modality for dementia: A meta-analysis. *Psychiatry Investigation*, *14*(5), 626–639. https://doi.org/10.4306/pi.2017.14.5.626

Kitwood, T. (1997). *Dementia reconsidered: The person comes first*. Buckingham, UK: Open University Press.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Knapp, M., Thorgrimsen, L., Patel, A., Spector, A., Hallam, A., Woods, B., & Orrell, M. (2006). Cognitive stimulation therapy for people with dementia: Cost-effectiveness analysis. *British Journal of Psychiatry*, *188*(6), 574–580. https://doi.org/10.1192/bjp.bp.105.010561

*Kolanowski, A., Fick, D., Litaker, M., Mulhall, P., Clare, L., Hill, N., Mogle, J., Boustani,

M., Gill, D., & Yevchak-Sillner, A. (2016). Effect of cognitively stimulating activities on symptom management of delirium superimposed on dementia: A randomized controlled trial. *Journal of the American Geriatrics Society*, *64*(12), 2424–2432. https://doi.org/10.1111/jgs.14511

Kurz, A. F., Leucht, S., & Lautenschlager, N. T. (2011). The clinical significance of cognition-focused interventions for cognitively impaired older adults: A systematic review of randomized controlled trials. *International Psychogeriatrics*, *23*(9), 1364–1375. https://doi.org/10.1017/S1041610211001001

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *The Gerontologist*, *9*, 179–186. https://doi.org/10.1093/geront/9.3_Part_1.179

Lee, M. D., & Vanpaemel, W. (2018). Determining informative priors for cognitive models. *Psychonomic Bulletin & Review*, *25*(1), 114–127. https://doi.org/10.3758/s13423-017-1238-3

Livingston, G., Kelly, L., Lewis-Holmes, E., Baio, G., Morris, S., Patel, N., Omar, R. Z., Katona, C., & Cooper, C. (2014). A systematic review of the clinical effectiveness and cost-effectiveness of sensory, psychological and behavioural interventions for managing agitation in older adults with dementia. *Health Technology Assessment*, *18*(39), 1–226. https://doi.org/10.3310/hta18390

Logsdon, R. G., Gibbons, L. E., McCurry, S. M., & Teri, L. (1999). Quality of Life in Alzheimer's disease: Patient and caregiver reports. *Journal of Mental Health and Aging*, *5*(1), 21–32.

*López, C. Sánchez, J. L., & Martín, J. (in press) The effect of cognitive stimulation on the progression of cognitive impairment in subjects with Alzheimer's disease. *Applied Neuropsychology: Adult*. https://doi.org/10.1080/23279095.2019.1710510

Luttenberger, K., Donath, C., Uter, W., & Graessel, E. (2012). 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*, *60*(5), 830–840. https://doi.org/10.1111/j.1532-5415.2012.03938.x

*Luttenberger, K., Hofner, B., & Graessel, E. (2012). 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*, *12*, Article 151. https://doi.org/10.1186/1471-2377-12-151

*Maci, T., Pira, F. Le, Quattrocchi, G., Nuovo, S. Di, Perciavalle, V., & Zappia, M. (2012). Physical and cognitive stimulation in Alzheimer disease. The GAIA project: A pilot study. *American Journal of Alzheimer's Disease and Other Dementias*, *27*(2), 107–113. https://doi.org/10.1177/1533317512440493

Mahoney, F. I., & Barthel, D. W. (1965). Functional evaluation: The Barthel Index. *Maryland State Medical Journal, 14*, 61–65.

Maust, D. T., Kim, H. M., Seyfried, L. S., Chiang, C., Kavanagh, J., Schneider, L. S., & Kales, H. C. (2015). Antipsychotics, other psychotropics, and the risk of death in patients with dementia. *JAMA Psychiatry, 72*(5), 438–445. https://doi.org/10.1001/jamapsychiatry.2014.3018

McDermott, O., Charlesworth, G., Hogervorst, E., Stoner, C., Moniz-Cook, E., Spector, A., Csipke, E., & Orrell, M. (2019). Psychosocial interventions for people with dementia: A synthesis of systematic reviews. *Aging & Mental Health, 23*(4), 393–403. https://doi.org/10.1080/13607863.2017.1423031

*Menna, L. F., Santaniello, A., Gerardi, F., Di Maggio, A., & Milan, G. (2016). Evaluation of the efficacy of animal-assisted therapy based on the reality orientation therapy protocol

in Alzheimer's disease patients: a pilot study. *Psychogeriatrics*, *16*(4), 240–246. https://doi.org/10.1111/psyg.12145

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/ journal.pmed.1000097

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. https://cran.r-project.org/package=BayesFactor

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control hroup designs. *Organization Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

National Institute for Health and Care Excellence [NICE]. (2018). *Dementia: Assessment, management and support for people living with dementia and their carers*. https://www.nice.org.uk/guidance/ng97

*Niu, Y.-X., Tan, J.-P., Guan, J.-Q., Zhang, Z.-Q., & Wang, L.-N. (2010). Cognitive stimulation therapy in the treatment of neuropsychiatric symptoms in Alzheimer's disease: A randomized controlled trial. *Clinical Rehabilitation*, *24*(12), 1102–1111. https://doi.org/10.1177/0269215510376004

Nyman, S. R., & Szymczynska, P. (2016). Meaningful activities for improving the wellbeing of people with dementia: Beyond mere pleasure to meeting fundamental psychological needs. *Perspectives in Public Health*, *136*(2), 99–107. https://doi.org/10.1177/1757913915626193

Oken, B. S., Flegal, K., Zajdel, D., Kishiyama, S., Haas, M., & Peters, D. (2008). Expectancy effect: Impact of pill administration on cognitive performance in healthy seniors. *Journal of Clinical and Experimental Neuropsychology*, *30*(1), 7–17. https://doi.org/10.1080/13803390701775428

*Onder, G., Zanetti, O., Giacobini, E., Frisoni, G. B., Bartorelli, L., Carbone, G., Lambertucci, P., Silveri, M. C., & Bernabei, R. (2005). Reality orientation therapy combined with cholinesterase inhibitors in Alzheimer's disease: Randomised controlled trial. *British Journal of Psychiatry, 187*(5), 450–455. https://doi.org/10.1192/bjp.187.5.450

*Onor, M. L., Trevisiol, M., Negro, C., Alessandra, S., Saina, M., & Aguglia, E. (2007). Impact of multimodal rehabilitative intervention on demented patients and their caregivers. *American Journal of Alzheimer's Disease & Other Dementias, 22*(4), 261–272. https://doi.org/10.1177/1533317507302071

*Orrell, M., Yates, L., Leung, P., Kang, S., Hoare, Z., Whitaker, C., Burns, A., Knapp, M., Leroi, I., Moniz-Cook, E., Pearson, S., Simpson, S., Spector, A., Roberts, S., Russell, I., de Waal, H., Woods, R. T., & Orgeta, V. (2017). 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, 14*(3), Article e1002269. https://doi.org/10.1371/journal.pmed.1002269

*Paddick, S.-M., Mkenda, S., Mbowe, G., Kisoli, A., Gray, W. K., Dotchin, C. L., Ternent, L., Ogunniyi, A., Kissima, J., Olakehinde, O., Mushi, D., & Walker, R. W. (2017). Cognitive stimulation therapy as a sustainable intervention for dementia in sub-Saharan Africa: Feasibility and clinical efficacy using a stepped-wedge design. *International Psychogeriatrics, 29*(6), 979–989. https://doi.org/10.1017/S1041610217000163

Pattie, A. H., & Gilleard, C. J. (1975). *Clifton Assessment Procedures for the Elderly (CAPE)*. Hodder & Stoughton. https://doi.org/10.1037/t28520-000

*Piras, F., Carbone, E., Faggian, S., Salvalaio, E., Gardini, S., & Borella, E. (2017). Efficacy of cognitive stimulation therapy for older adults with vascular dementia. *Dementia & Neuropsychologia, 11*(4), 434–441. https://doi.org/10.1590/1980-57642016dn11-040014

Powell-Proctor, L., & Miller, E. (1982). Reality orientation: A critical appraisal. *British Journal of Psychiatry*, *140*(5), 457–463. https://doi.org/10.1192/bjp.140.5.457

Qaseem, A., Snow, V., Cross, T., Jr., Forcia M. A., Hopkins, R., Jr., Shekelle, P., Adelman, A., Mehr, D., Schellhase, K., Campos-Outcalt, D., Santaguida, P., Owens, D. K., Joint American College of Physicians/American Academy of Family Physicians Panel on Dementia (2008). Current pharmacological treatment of dementia: A clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. *Annals of Internal Medicine, 148*(5), 370–378. https://doi.org/10.7326/0003-4819-148-5-200803040-00008

Quayhagen, M. P., Quayhagen, M., Corbeil, R. R., Hendrix, R. C., Jackson, J. E., Snyder, L., & Bower, D. (2000). Coping with dementia: Evaluation of four nonpharmacologic interventions. *International Psychogeriatrics, 12*(2), 249–265. https://doi.org/10.1017/S1041610200006360

*Quayhagen, M. P., Quayhagen, M., Corbeil, R. R., Roth, P. A., & Rodgers, J. A. (1995). A dyadic remediation program for care recipients with dementia. *Nursing Research, 44*(3), 153–159.

*Quintana-Hernández, D. J., Miró-Barrachina, M. T., Ibáñez-Fernández, I. J., Santana-Del Pino, A., Quintana-Montesdeoca, M. P., Rodríguez-De Vera, B., Morales-Casanova, D., Pérez-Vieitez, Mdel. C., Rodríguez-de Vera, J., & Bravo-Caraduje, N. (2016). Mindfulness in the maintenance of cognitive capacities in Alzheimer's disease: A randomized clinical trial. *Journal of Alzheimer's Disease, 50*(1), 217–232. https://doi.org/10.3233/JAD-143009

*Requena, C., Maestú, F., Campo, P., Fernández, A., & Ortiz, T. (2006). Effects of cholinergic drugs and cognitive training on dementia: 2-Year follow-up. *Dementia and Geriatric Cognitive Disorders*, *22*(4), 339–345. https://doi.org/10.1159/000095600

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, *141*(11), 1356–1364. https://doi.org/10.1176/ajp.141.11.1356

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

*Schecker, M., Pirnay-Dummer, P., Schmidtke, K., Hentrich-Hesse, T., & Borchardt, D. (2013). 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*, *3*(1), 301–311. https://doi.org/10.1159/000354190

Schmidt, F. L. (2017). Statistical and measurement pitfalls in the use of meta-regression in meta-analysis. *Career Development International*, *22*(5), 469–476. https://doi.org/10.1108/CDI-08-2017-0136

Schmidt, F. L., & Hunger, J. E. (2015). *Methods of meta-analysis: Correcting for error and bias in research findings* (3rd ed.). Sage.

Schneider, L. S., Mangialasche, F., Andreasen, N., Feldman, H., Giacobini, E., Jones, R., Mantua, V., Mecocci, P., Pani, L., Winblad, B., & Kivipelto, M. (2014). Clinical trials and late-stage drug development for Alzheimer's disease: An appraisal from 1984 to 2014. *Journal of Internal Medicine*, *275*(3), 251–283. https://doi.org/10.1111/joim.12191

Schrag, A., Schott, J. M., & Alzheimer's Disease Neuroimaging Initiative. (2012). What is the clinically relevant change on the ADAS-Cog? *Journal of Neurology, Neurosurgery & Psychiatry*, *83*(2), 171–173. https://doi.org/10.1136/jnnp-2011-300881

Shankar, K. K., Walker, M., Frost, D., & Orrell, M. W. (1999). The development of a valid and reliable scale for rating anxiety in dementia (RAID). *Aging and Mental Health*, *3*(1), 39–49. https://doi.org/10.1080/13607869956424

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-Curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143*(2), 534-547. http://doi.org/10.1037/a0033242

Skoumbourdis, J. (2014). *Grocery CRUD*. http://www.grocerycrud.com

Spector, A., Davies, S., Woods, B., & Orrell, M. (2000). Reality orientation fordementia: A systematic review of the evidence of effectiveness from randomized controlled trials. *The Gerontologist*, *40*(2), 206–212. https://doi.org/10.1093/geront/40.2.206

*Spector, A., Orrell, M., Davies, S., & Woods, B. (2001). Can reality orientation be rehabilitated? Development and piloting of an evidence-based programme of cognition-based therapies for people with dementia. *Neuropsychological Rehabilitation*, *11*(3–4), 377–397. https://doi.org/10.1080/09602010143000068

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

*Spector, A., Thorgrimsen, L., Woods, B., Royan, L., Davies, S., Butterworth, M., & Orrell, M. (2003). Efficacy of an evidence-based cognitive stimulation therapy programme for people with dementia. *British Journal of Psychiatry*, *183*, 248–254. https://doi.org/10.1192/bjp.183.3.248

Steel, P., Kammeyer-Mueller, J., & Paterson, T. A. (2015). Improving the meta-analytic assessment of effect size variance with an informed Bayesian prior. *Journal of Management*, *41*(2), 718–743. https://doi.org/10.1177%2F0149206314551964

Stewart, C. P. U. (1980). A prediction score for geriatric rehabilitation prospects. *Rheumatology, 19*(4), 239–245. https://doi.org/10.1093/rheumatology/19.4.239

*Tadaka, E., Kanagawa, K. (2007). Effects of reminisence group in elderly people with Alzheimer disease and vascular dementia in a community setting. *Geriatrics &*

Gerontology International, *7*(2), 167–173. https://doi.org/10.1111/j.1447-0594.2007.00381.x

*Tárraga, L., Boada, M., Modinos, G., Espinosa, A., Diego, S., Morera, A., Guitart, M., Balcells, J., López, O. L., Becker, J. T. (2006). 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*, *77*(10), 1116–1121. https://doi.org/10.1136/jnnp.2005.086074

*Tsantali, E., Economidis, D., & Rigopoulou, S. (2017). Testing the benefits of cognitive training vs. cognitive stimulation in mild Alzheimer's Disease: A randomised controlled trial. *Brain Impairment*, *18*(2), 188–196. https://doi.org/10.1017/BrImp.2017.6

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247. https://doi.org/10.3102/1076998609346961

van Erp, S., Verhagen, J., Grasman, R., & Wagenmakers, E. J. E.-J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in Psychological Bulletin from 1990 to 2013. *Journal of Open Psychology Data*, *5*(1), Article 4. https://doi.org/10.5334/jopd.33

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48. http://dx.doi.org/10.18637/jss.v036.i03

Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., & Marín-Martínez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods*, *20*(3), 360–374. https://doi.org/10.1037/met0000023

Voelkel, D. (1978). A study of reality orientation and resocialization groups with confused elderly. *Journal of Gerontological Nursing*, *4*(3), 13–18. https://doi.org/10.3928/0098-9134-19780501-04

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: A review. *Psychological Research*, *78*(6), 803–820. https://doi.org/10.1007/s00426-013-0524-6

*Wallis, G. G., Baldwin, M., & Higginbotham, P. (1983). Reality orientation therapy — a controlled trial. *British Journal of Medical Psychology*, *56*(3), 271–277. https://doi.org/10.1111/j.2044-8341.1983.tb01556.x

*Wang, J.-J. (2007). Group reminiscence therapy for cognitive and affective function of demented elderly in Tawain. *International Journal of Geriatric Psychiatry, 22*(12), 1235–1240. https://doi.org/10.1002/gps.1821

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 1057–1064. https://doi.org/10.3758/s13423-012-0295-x

Woods, B., Aguirre, E., Spector, A. E., & Orrell, M. (2012). Cognitive stimulation to improve cognitive functioning in people with dementia. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD005562.pub2

*Woods, R. T. (1979). Reality orientation and staff attention: A controlled study. *British Journal of Psychiatry*, *134*(5), 502–507. https://doi.org/10.1192/bjp.134.5.502

World Health Organization [WHO]. (2018). *Towards a dementia plan: A WHO guide*. https://www.who.int/publications-detail/towards-a-dementia-plan-a-who-guide

*Yamagami, T., Takayama, Y., Maki, Y., & Yamaguchi, H. (2012). A randomized controlled trial of brain-activating rehabilitation for elderly participants with dementia in residential care homes. *Dementia and Geriatric Cognitive Disorders Extra*, *2*(1), 372–380. https://doi.org/10.1159/000342614

Yamanaka, K., Kawano, Y., Noguchi, D., Nakaaki, S., Watanabe, N., Amano, T., & Spector, A. (2013). Effects of cognitive stimulation therapy Japanese version (CST-J) for people

with dementia: A single-blind, controlled clinical trial. *Aging and Mental Health*, *17*(5), 579–586. https://doi.org/10.1080/13607863.2013.777395

*Young, D. K., Ng, P. Y., Kwok, T., Ho, F., Cheng, D., Mak, V., & Lau, A. (2019). 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*, *23*(7), 855–862. https://doi.org/10.1080/13607863.2018.1471586

Table 1

Verbal Labels for Interpreting Bayes Factors

Bayes factor		
H1	H0	Interpretation of evidence
> 100	< 1/100	Decisive
30 to 100	1/100 to 1/30	Very strong
10 to 30	1/30 to 1/10	Strong
3 to 10	1/10 to 1/3	Substantial
1 to 3	1/3 to 1	Ambiguous
1	1	No evidence

Note. Adapted from Wetzels and Wagenmakers (2012).

Table 2

Sample and Intervention Characteristics, and Effect Sizes and Strength of Evidence of Cognitive Change for Studies Included in the Review

Study	*N*	Age	Gender	MMSE	Type	Components	Duration	Frequency	Context	*g* [95% CI]	BF [sensitivity]
Alves et al. (2014)	17	79.60	70.00	-	other	multiple	6	3	in care	0.20 [-0.72, 1.12]	1/1.38 [1/1.14, 1/1.92]
Baglio et al. (2015)	50	75.61	53.57	19.50	RO+	multiple	10	3	other	1.71 [1.07, 2.36]	**>100** [>100, >100]
Baines et al. (1987)	10	81.40	-	-	RO+	multiple	4	5	in care	-0.02 [-1.14, 1.10]	1/1.59 [1/1.28, 1/2.23]
Baldelli et al. (1993)	23	-	100	18.50	RO+	single	12	3	in care	1.46 [0.56, 2.36]	**24.07** [14.44, 33.21]
Baldelli et al. (2002)	87	79.60	70.42	-	RO+	multiple	4	5	in care	0.83 [0.28, 1.38]	**19.52** [13.69, 21.10]
Bottino et al. (2005)	13	74.67	83.30	-	RO+	multiple	20	1	other	0.38 [-0.64, 1.41]	1/1.11 [1.02, 1/1.42]
Breuil et al. (1994)	56	76.10	68.75	19.50	other	single	5	2	other	0.71 [0.18, 1.25]	**9.37** [7.28, 9.21]
Buschert et al. (2011)	15	77.30	50.00	24.00	CST+	multiple	20	1	other	0.13 [-0.83, 1.09]	1/1.49 [1/1.20, 1/2.08]
Camargo et al. (2015)	14	80.86	42.86	20.50	RO+	single	26	1	other	0.82 [-0.20, 1.85]	1.71 [1.60, 1.60]
Capotosto et al. (2017)	39	88.25	75.00	22.00	CST+	single	7	2	in care	0.29 [-0.33, 0.90]	1/1.14 [1.06, 1/1.62]
Chapman et al. (2004)[a]	54	-	-	21	other	single	8	1	other	0.12 [-0.48, 0.72]	1/1.72 [1/1.28, 1/2.65]
Coen et al. (2011)	25	78.40	35.71	16.50	CST+	single	7	2	in care	-0.19 [-0.95, 0.57]	1/2.47 [1/1.72, 1/3.99]
Cove et al. (2014)	59	76.15	42.21	24.00	CST+	single	14	1	other	0.15 [-0.37, 0.68]	1/1.63 [1/1.21, 1/2.54]
D'Onofrio et al. (2015)	90	76.67	51.10	20.00	RO+	single	16	1	other	0.24 [-0.17, 0.66]	1/1.05 [1.22, 1/1.60]
Ferrario et al. (1991)	19	82.5	42.00	-	RO+	single	21	5	in care	0.34 [-0.59, 1.27]	1/1.15 [1.00, 1/1.51]
Gibbor et al. (in press)	29	86.24	42.90	20.94	CST+	single	7	2	in care	0.40 [-0.32, 1.12]	1.09 [1.22, 1/1.20]
Haight et al. (2006)	30	-	-	16.93	other	single	6	1	in care	1.25 [0.49, 2.02]	**28.48** [17.45, 37.37]
Ishizaki et al. (2002)	25	75.9	64.29	26.4	RO+	multiple	24	1	other	0.80 [0.01, 1.60]	2.86 [2.47, 2.76]
Kim et al. (2016)	53	78.44	75.00	-	other	single	24	5	other	0.42 [-0.13, 0.97]	1.58 [1.70, 1.20]
Kolanowski et al. (2016)[b]	283	85.28	62.40	-	other	single	4	5	other	-0.03 [-0.26, 0.20]	**1/5.48** [1/3.04, 1/10.25]
Lopez et al. (2020)	20	83.3	80.00	18.1	other	multiple	24	3	other	0.06 [-0.78, 0.90]	1/1.70 [1/1.31, 1/2.50]
Luttenberger et al. (2012)	61	84.10	90.00	12.00	other	multiple	52	6	in care	0.40 [-0.10, 0.90]	1.62 [1.77, 1.21]
Maci et al. (2012)	14	75.00	57.14	20.00	other	multiple	12	5	other	0.33 [-0.65, 1.32]	1/1.17 [1/1.02, 1/1.52]
Menna et al. (2016)	30	74.95	70.00	-	RO+	single	24	1	other	1.36 [0.54, 2.17]	**32.17** [19.24, 43.76]
Niu et al. (2010)	32	80.56	12.50	16.94	CST+	single	10	2	in care	0.31 [-0.37, 0.99]	1/1.10 [1.07, 1/1.53]

Onder et al. (2005)	137	75.70	73.00	20.50	RO+	single	25	3	other	0.21 [-0.13, 0.55]	1/1.09 [1.23, 1/1.74]
Onor et al. (2007)	16	68	37.50	23.12	RO+	multiple	24	3	other	0.00 [-0.93, 0.93]	1/1.73 [1/1.34, 1/2.54]
Orrell et al. (2017)	273	-	46.00	20.00	CST+	single	25	3	other	-0.06 [-0.30, 0.18]	**1/6.39** [1/3.49, 1/12.01]
Paddick et al. (2017)	34	-	87.50	-	CST+	single	7	2	other	0.95 [0.26, 1.65]	**9.44** [6.76, 10.49]
Piras et al. (2017)	35	83.81	71.43	22.00	CST+	single	7	2	other	0.35 [-0.32, 1.01]	1.00 [1.16, 1/1.36]
Quayhagen et al. (1995)	53	-	-	-	other	single	12	6	other	0.41 [-0.13, 0.94]	1.53 [1.66, 1.15]
Quintana-Hernandez et al. (2016)	62	-	-	24.00	other	single	96	3	other	0.74 [0.23, 1.25]	**15.40** [11.39, 15.66]
Requena et al. (2006)	36	77.00	72.22	-	CST+	single	104	5	other	1.43 [0.71, 2.15]	**>100** [>100, >100]
Schecker et al. (2013)	27	76.50	75.00	18.50	other	single	24	0	other	0.51 [-0.24, 1.26]	1.38 [1.44, 1.12]
Spector et al. (2001)	27	-	-	-	CST+	single	7	2	other	0.40 [-0.37, 1.16]	1.04 [1.17, 1/1.26]
Spector et al. (2003)	201	85.70	83.48	17.00	CST+	single	7	2	in care	0.29 [0.01, 0.57]	2.90 [3.38, 1.94]
Tadaka et al. (2007), comp. 1	21	82.5	66.67	14.6	other	single	8	1	other	0.02 [-0.80, 0.84]	1/1.80 [1/1.37, 1/2.70]
Tadaka et al. (2007), comp. 2	36	85.3	72.22	19.8	other	single	8	1	other	0.68 [0.02, 1.33]	**3.11** [2.76, 2.84]
Tarraga et al. (2015)	28	77.4	87.50	22.5	other	multiple	24	5	other	0.13 [-0.59, 0.86]	1/1.60 [1/1.24, 1/2.37]
Tsantali et al. (2017)[a]	38	73.3	-	-	other	single	16	3	other	-0.43 [-1.06, 0.21]	1/4.05 [1/2.49, 1/7.13]
Wallis et al. (1983)	38	71.80	27.78	-	RO+	single	13	5	in care	0.12 [-0.50, 0.74]	1/1.71 [1/1.28, 1/2.62]
Wang et al. (2007)	102	79.76	52.90	14.33	other	single	8	1	in care	0.46 [0.07, 0.85]	**5.38** [5.07, 4.31]
Woods et al. (1979)	9	-	-	-	RO+	single	20	5	in care	1.12 [-0.15, 2.40]	1.83 [1.62, 1.85]
Yamagami et al. (2012)	53	85.50	96.40	-	RO+	multiple	12	2	in care	-0.02 [-0.55, 0.51]	1/2.53 [1/1.70, 1/4.21]
Young et al. (2019)	101	80.53	40.60	19	other	multiple	7	2	other	1.22 [0.80, 1.64]	**>100** [>100, >100]

Note. Age (years), gender (percentage of females) and Mini-Mental State Examination score are given in means at baseline. Duration is given in weeks; frequency is given in sessions per week. Bayes factors unambiguously supporting the presence of a cognitive stimulation effect (i.e., Bayes factors greater than 3) or unambiguously supporting the absence of a cognitive stimulation effect (i.e., Bayes factors smaller than 1/3) are printed in bold. Bayes factors from sensitivity analyses for small-effect priors (d = 0.2) and large-effect priors (d = 0.8) are provided in angular brackets. MMSE = Mini-Mental State Examination; CST+ = cognitive stimulation therapy with or without additional cognitive stimulation components (e.g., physical exercise); RO+ = reality orientation with or without additional cognitive stimulation components (e.g., physical exercise); BF = Bayes factor.

[a] Effect size from follow-up. [b] Effect size from memory measure.

Table 3

Cognitive Stimulation Effects on Primary and Secondary Outcomes

Outcome	*k*	*g*	*SE*	CI	BF [sensitivity]	I^2 [95% CI]	τ^2	BF τ2
			Primary Outcomes					
Global cognition								
Posttest	42	0.49	0.07	[0.35, 0.63]	**>100** [>100, >100]	54.68 [28.52, 72.56]	0.1 (0.05)	< 100
Follow-up	9	0.22	0.14	[-0.09, 0.54]	1.67 [2.19, 1.05]	19.10 [0.00, 84.37]	0.03 (0.07)	1/1.17
Memory	15	0.34	0.13	[0.06, 0.62]	**9.74** [10.44, 6.83**]**	53.59 [9.27, 83.28]	0.12 (0.09)	9.09
Language	14	0.10	0.27	[-0.47, 0.67]	1/1.89 [1/1.35, 1/3.03]	88.87 [77.94, 95.94]	0.83 (0.38)	< 100
			Secondary Outcomes					
Quality of life	11	0.16	0.15	[-0.16, 0.48]	1.07 [1.54, 1/1.59]	25.64 [0.00, 92.42]	0.04 (0.06)	1.31
Activities of daily living	14	0.17	0.07	[0.02, 0.32]	2.04 [2.99, 1.19]	20.38 [0.00, 55.29]	0.02 (0.03)	1/1.04
Depression	19	0.46	0.15	[0.15, 0.78]	**27.26** [23.57, 22.38]	74.73 [53.71, 89.65]	0.28 (0.14)	< 100
Anxiety	5	0.25	0.19	[-0.28, 0.77]	1/1.23 [1.11, 1/2.00]	43.10 [0.00, 94.16]	0.07 (0.12)	1/1.30
Dementia	7	0.66	0.26	[0.02, 1.29]	**10.17** [7.69, 9.53]	80.63 [50.28, 96.71]	0.34 (0.25)	36.81
Behavior	11	0.28	0.40	[-0.60, 1.17]	1/1.19 [1.01, 1/1.67]	95.10 [89.54, 98.55]	1.53 (0.75)	< 100

Note. Bayes factors unambiguously supporting the presence of a cognitive stimulation effect (i.e., Bayes factors greater than 3) or unambiguously supporting the absence of a cognitive stimulation effect (i.e., Bayes factors smaller than 1/3) are printed in bold. Bayes factors from sensitivity analyses for small-effect priors (d = 0.2) and large-effect priors (d = 0.8) are provided in angular brackets. CI = confidence interval; BF = Bayes factor.

Table 4

Effects of Moderators on the Benefits of Cognitive Stimulation

Moderator	*k*	*F*	*dfs*	*p*	BF [sensitivity]
		Cognition			
Type of intervention	27	1.51	1, 25	.230	1/2.33 [1/1.35, 1/3.45]
Number of CS components	42	0.09	1, 40	.766	**1/5.56** [1/2.56, 1/8.33]
Intervention context	34	0.64	1, 32	.429	**1/4.00** [1/2.08, 1/6.25]
Intervention duration	42	1.32	1, 40	.258	**1/3.23** [1/1.59, 1/5.00]
Intervention frequency	42	0.02	1, 40	.889	**1/5.56** [1/2.33, 1/9.09]
Intervention sessions	42	1.41	1, 40	.243	**1/3.03** [1/1.49, 1/4.76]
Age	34	0.61	1, 32	.439	**1/4.00** [1/1.85, 1/6.25]
Gender distribution	36	0.57	1, 34	.454	**1/4.00** [1/1.89, 1/6.25]
Clinical severity (baeline MMSE)	28	< 0.01	1, 26	.971	**1/4.76** [1/2.08, 1/7.69]
		Memory			
Intervention duration	15	0.44	1, 13	.520	**1/3.03** [1/1. 61, 1/4.55]
Intervention frequency	15	0.69	1, 13	.422	1/2.00 [1/1.20, ½.94]
Intervention sessions	15	0.16	1, 13	.692	**1/3.23** [1/1.75, 1/5.26]
Age	12	1.69	1, 10	.223	1/1.69 [1/1.25, 1/2.38]
Gender distribution	12	0.19	1, 10	.676	1/2.37 [1/1.42, 1/3.35]
		Language			
Intervention duration	14	0.06	1, 12	.813	1/1.96 [1/1.30, 1/2.63]
Intervention frequency	14	0.49	1, 12	.497	1/1.75 [1/1.16, 1/2.38]
Intervention sessions	14	0.08	1, 12	.788	1/1.96 [1/1.30, 1/2.78]
Age	11	3.75	1, 9	.085	1.99 [1.59, 1.84]
Gender distribution	10	2.25	1, 8	.172	1.28 [1.27, 1.14]
		Quality of Life			
Intervention duration	11	0.09	1, 9	.774	**1/3.18** [1/1.59, 1/5.00]
Intervention frequency	11	0.07	1, 9	.804	**1/3.18** [1/1.60, 1/4.93]
Intervention sessions	11	0.02	1, 9	.904	**1/3.77** [1/1.83, 1/5.74]
Age	11	0.05	1, 9	.822	1/2.80 [1/1.51, 1/4.19]
Gender distribution	10	> 0.01	1, 8	.958	**1/3.27** [1/1.71, 1/4.81]
		Depression			
Intervention duration	19	2.29	1, 17	.149	1/1.14 [1.13, 1/1. 54]
Intervention frequency	19	1.89	1, 17	.187	1/1.22 [1.09, 1/1.64]
Intervention sessions	19	4.90	1, 17	.041[a]	2.53 [2.35, 2.11]
Age	16	2.80	1, 14	.116	1.05 [1.23, 1/1.15]
Gender distribution	17	0.27	1, 15	.614	1/2.44 [1/1.39, 1/3.57]
Clinical severity (baeline MMSE)	13	2.06	1, 11	.179	1/1.30 [1.03, 1/1.75]

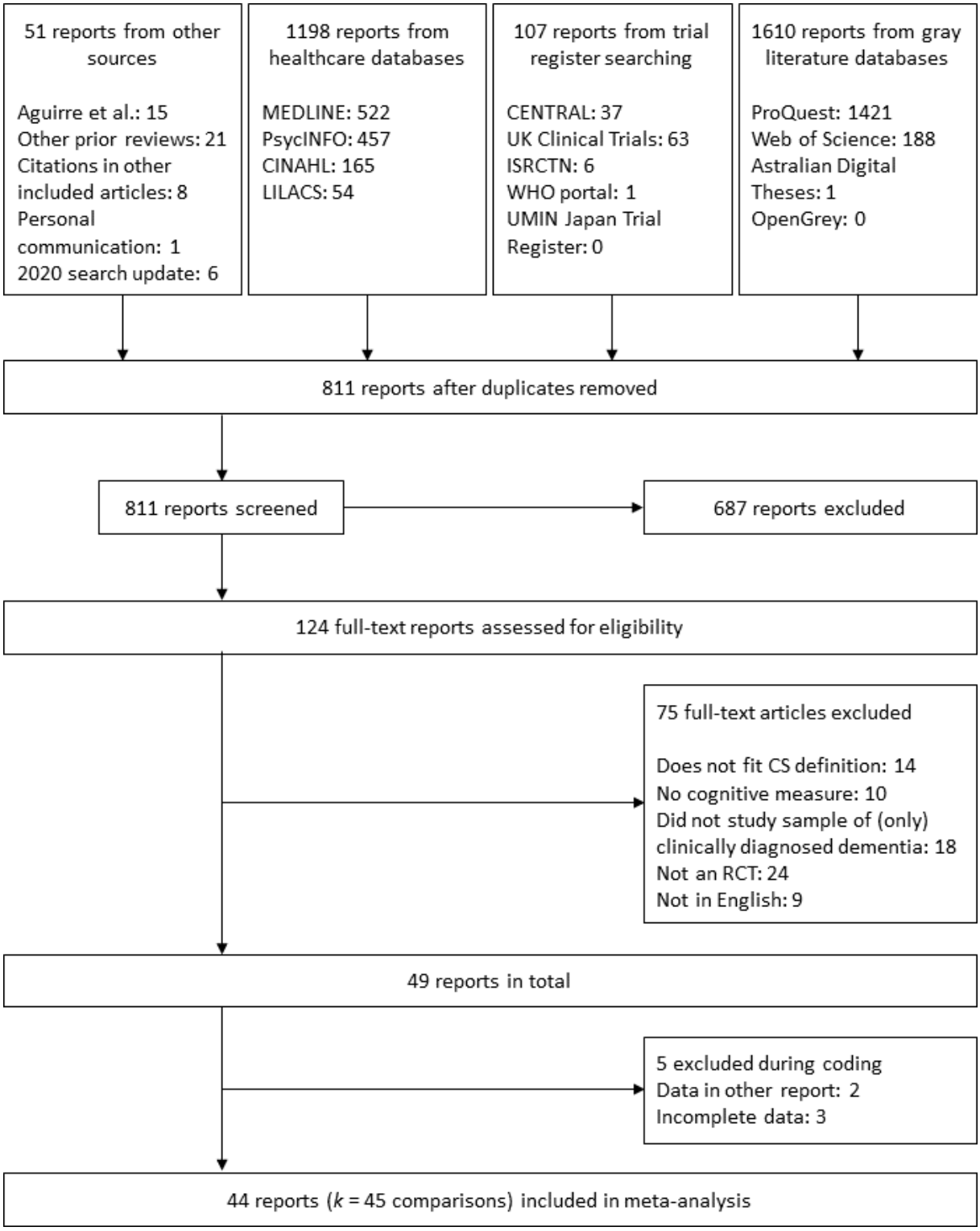
Note. Bayes factors unambiguously supporting the presence of a cognitive stimulation effect (i.e., Bayes factors greater than 3) or unambiguously supporting the absence of a cognitive stimulation effect (i.e., Bayes factors smaller than 1/3) are printed in bold. Bayes factors from sensitivity analyses for small-effect priors (d = 0.2) and large-effect priors (d = 0.8) are

provided in angular brackets. MMSE = Mini-Mental State Examination (Folstein et al., 1975) BF = Bayes factor.

[a] After removal of one outlier study (Requena et al., 2006; see text for details), this association reversed from being significantly negative to being significantly positive ($p = .027$).

Figure 1

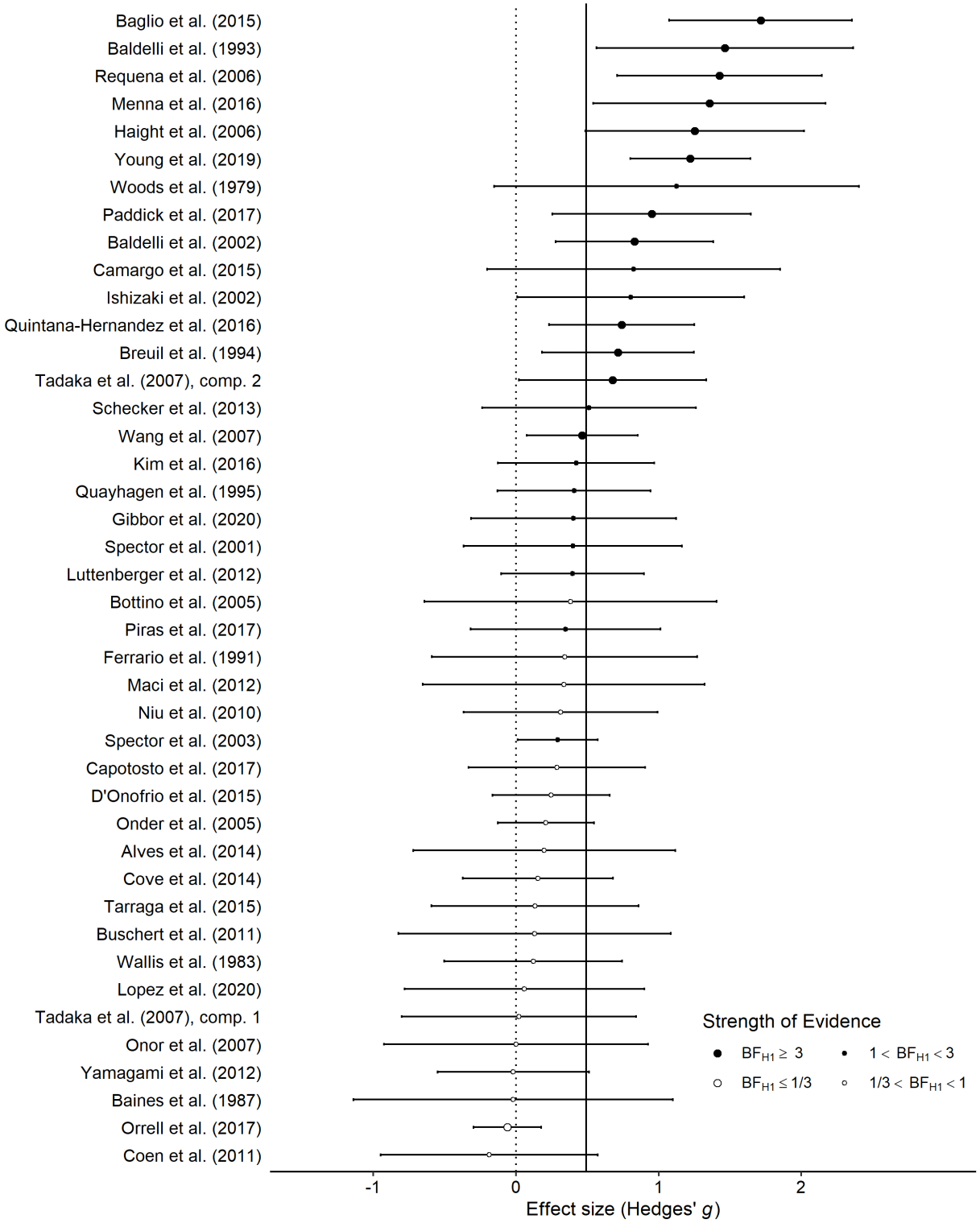
Flowchart of Systematic Search and Study Selection



Note. CS = cognitive stimulation; RCT = randomized-controlled trial.

Figure 2

Average Effects of Cognitive Stimulation on Global Cognition

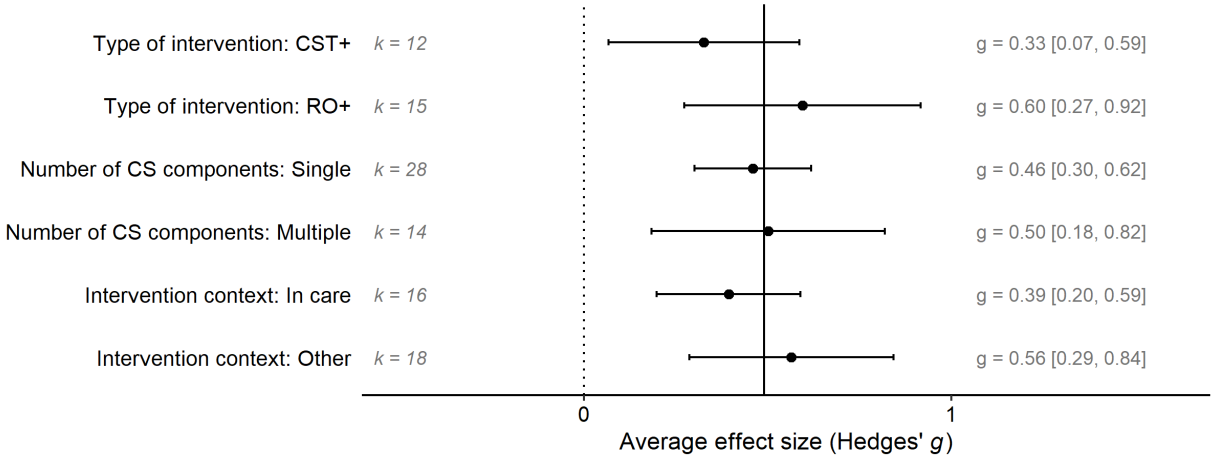


Note. Effect sizes are weighted averages; error bars represent 95% confidence intervals. The filling of the circles reflects the direction of evidence (black: alternative hypothesis; white: null hypothesis) and their size the strength of evidence in terms of Bayes factors (BFs; small: ambiguous evidence with BFs between 1/3 and 3; large: at least substantial evidence with BFs

≥ 3 or ≤ 1/3). The solid vertical line represents the weighted average effect size (g = 0.49), the dotted line represents the reference line at zero.

Figure 3

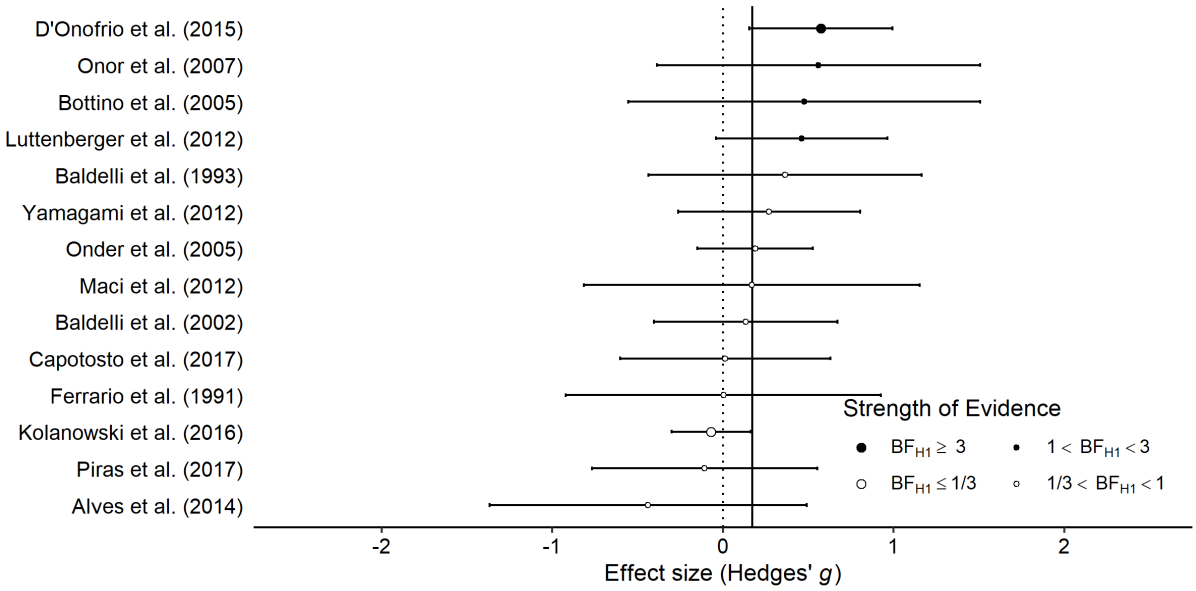
Moderator Effects on the Benefits of Cognitive Stimulation for Global Cognition



Note. Effect size estimates are weighted averages for the subgroups of each level of tested categorical moderator. Error bars represent 95% confidence intervals. Bayes factors were ≥ 3 for all subgroup effects. The solid vertical line represents the weighted average effect size (g = 0.49) across all comparisons, the dotted line represents the reference line at zero. CST+ = cognitive stimulation therapy with or without additional cognitive stimulation components (e.g., physical exercise); RO+ = reality orientation with or without additional cognitive stimulation components (e.g., physical exercise); CS = cognitive stimulation.

Figure 4

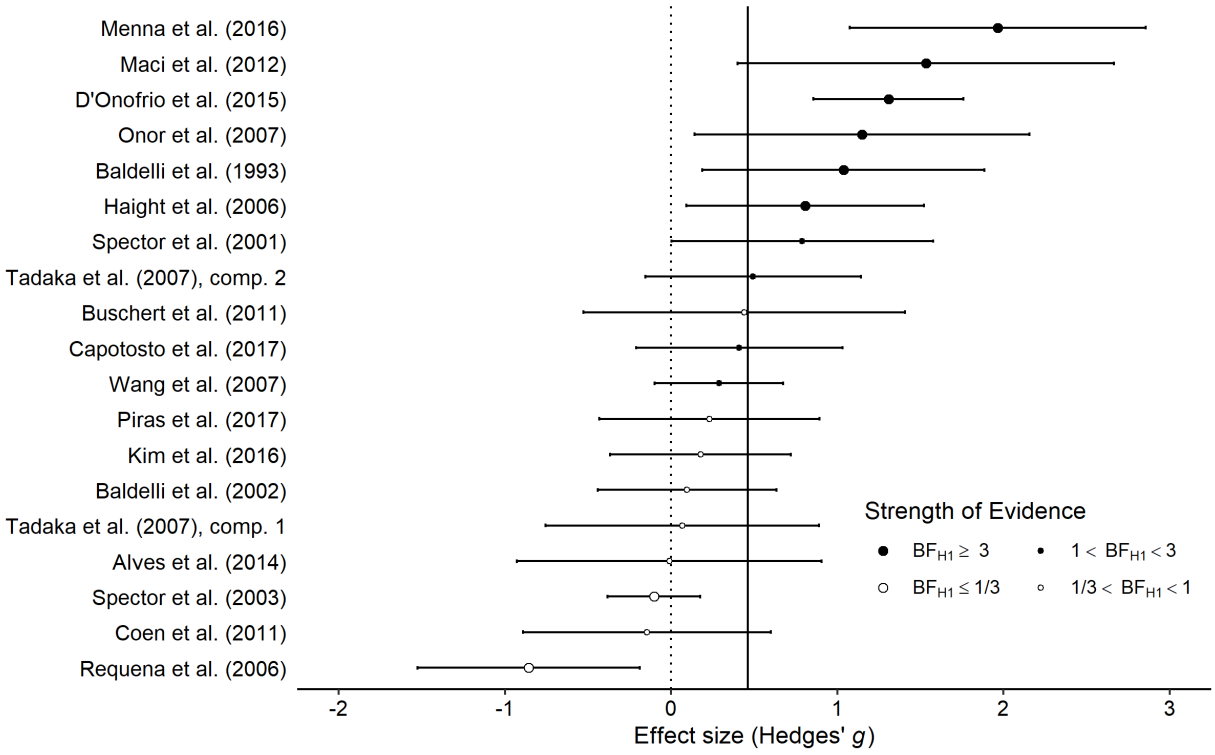
Average Effects of Cognitive Stimulation on Activities of Daily Living



Note. Effect sizes are weighted averages; error bars represent 95% confidence intervals. The filling of the circles reflects the direction of evidence (black: alternative hypothesis; white: null hypothesis) and their size the strength of evidence in terms of Bayes factors (BFs; small: ambiguous evidence with BFs between 1/3 and 3; large: at least substantial evidence with BFs ≥ 3 or $\leq 1/3$). The solid vertical line represents the weighted average effect size ($g = 0.17$), the dotted line represents the reference line at zero.

Figure 5

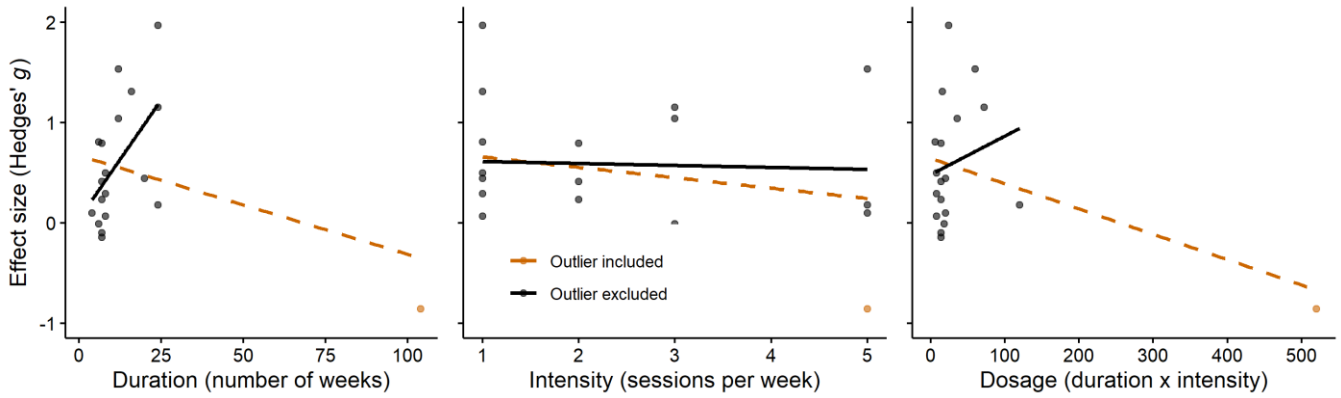
Average Effects of Cognitive Stimulation on Symptoms of Depression



Note. Effect sizes are weighted averages; error bars represent 95% confidence intervals. The filling of the circles reflects the direction of evidence (black: alternative hypothesis; white: null hypothesis) and their size the strength of evidence in terms of Bayes factors (BFs; small: ambiguous evidence with BFs between 1/3 and 3; large: at least substantial evidence with BFs ≥ 3 or $\leq 1/3$). The solid vertical line represents the weighted average effect size ($g = 0.46$), the dotted line represents the reference line at zero.

Figure 6

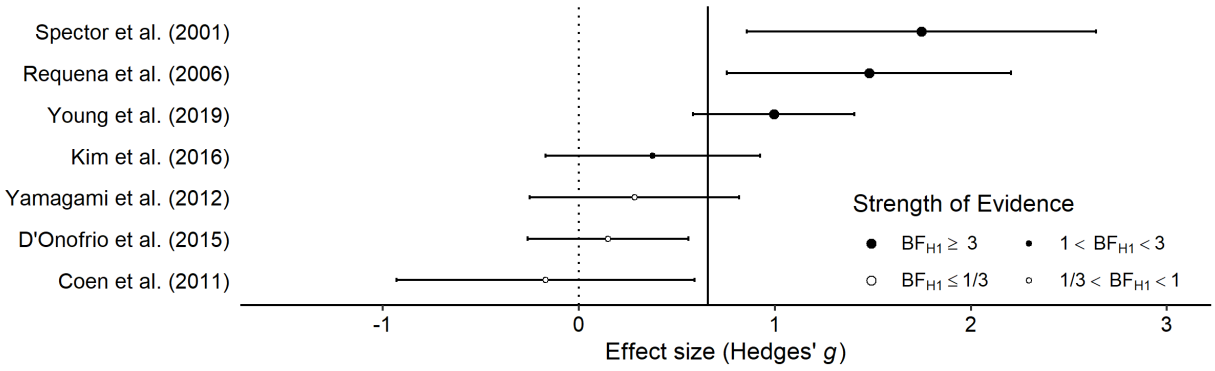
Relationship Between Intervention Duration, Intensity, and Dosage and Cognitive Stimulation Effect Size on Depression



Note. Scatterplots relating weighted average effect sizes of cognitive stimulation on depression to intervention duration (left), intensity (center), and their interaction (i.e., total number of sessions, right). The red dot represents the study by Requena et al. (2006), with the red regression line representing the trend when including this study, and the blue regression line representing the trend when excluding this study.

Figure 7

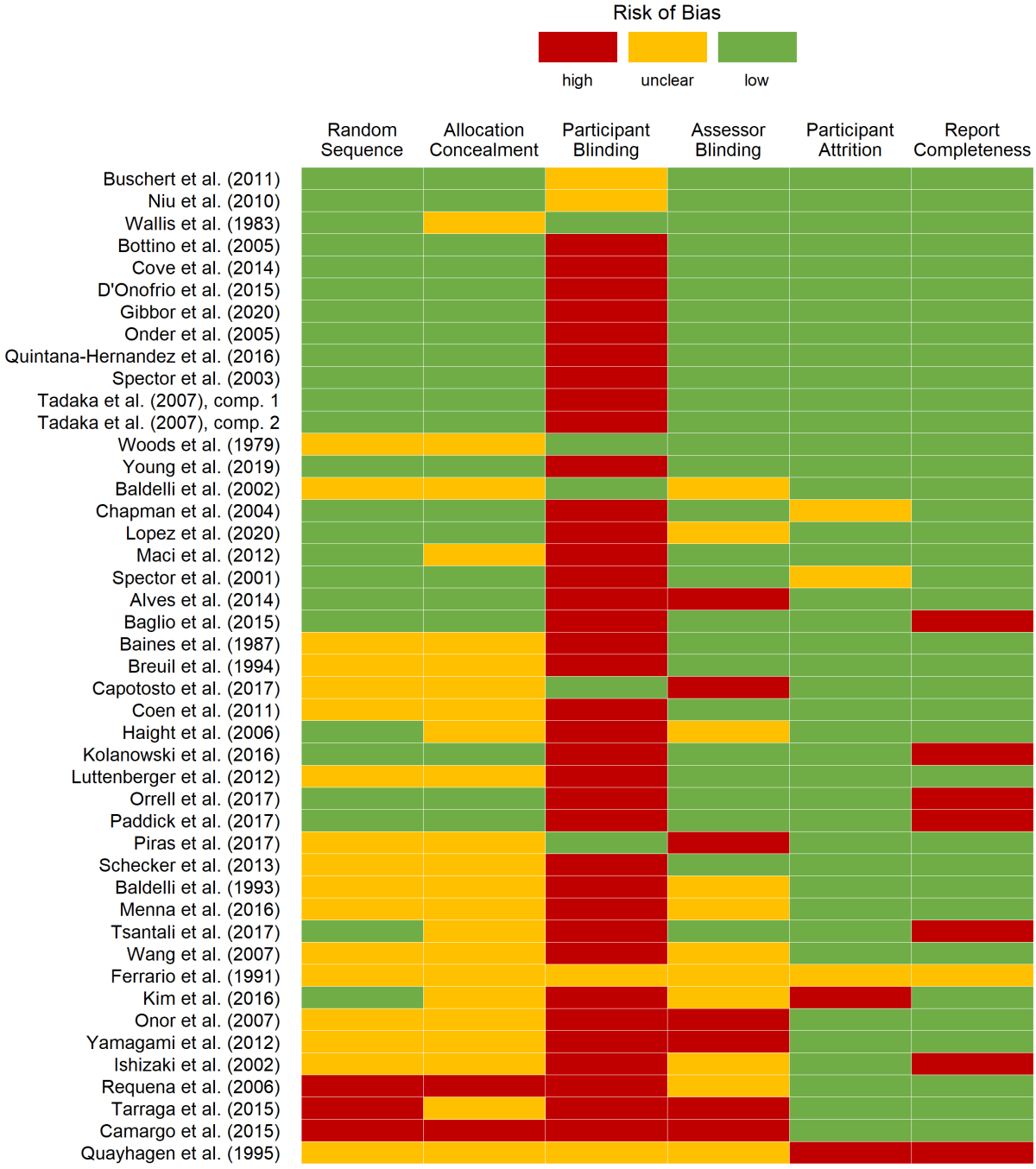
Average Effects of Cognitive Stimulation on Ratings of Dementia Symptoms



Note. Effect sizes are weighted averages; error bars represent 95% confidence intervals. The filling of the circles reflects the direction of evidence (black: alternative hypothesis; white: null hypothesis) and their size the strength of evidence in terms of Bayes factors (BFs; small: ambiguous evidence with BFs between 1/3 and 3; large: at least substantial evidence with BFs ≥ 3 or $\leq 1/3$). The solid vertical line represents the weighted average effect size ($g = 0.66$), the dotted line represents the reference line at zero.

Figure 8

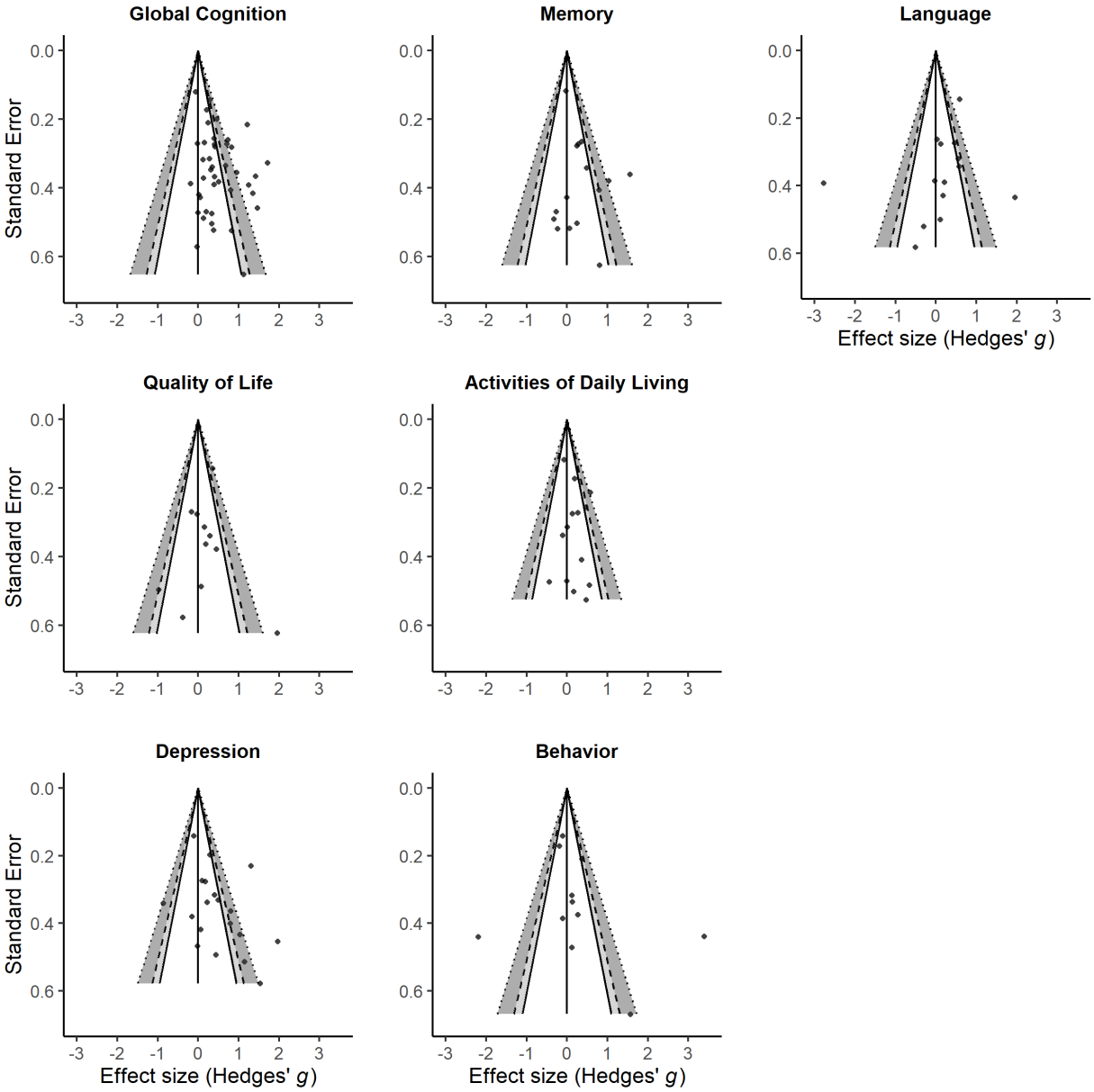
Risk of Bias Ratings



Note. Risk of bias was rated using the Cochrane Risk of Bias Tool (Higgins et al., 2011).

Figure 9

Contour-Enhanced Funnel Plots of Effect Sizes of Cognitive Stimulation on Primary and Secondary Outcomes



Note. Primary outcomes are presented in the top row, secondary outcomes in the middle and bottom row. The solid vertical line represents the null hypothesis of $d = 0$. The tilted solid, dashed, and dotted lines show levels of significance for the effect sizes at $\alpha = .05$, $\alpha = .01$, and $\alpha = .001$, respectively. Thus, the areas between the tilted solid and dashed lines, the dashed and dotted lines, and beyond the dotted lines represent $p < .05$, $p < .01$, and $p < .001$, respectively. Effect sizes within the inner white area are non-significant.