Title:

# The effects of virtual reality training on cognition in older adults: a systematic review, meta-analysis, and meta-regression of randomized controlled trials.

Authors:

Mirjami Kantola, MSc [1], Outi Ilves, PhD [1, 2], Sari Honkanen, MSc [1], Harto Hakonen, MSc [3], Riku Yli-Ikkelä, MSc [1], Anna Köyhäjoki, MSc [4], Marjo-Riitta Anttila, PhD [1], Aki Rintala, PhD [5], Hilkka Korpi, PhD [1,6,7], Tuulikki Sjögren, PhD, Doc. [1], Juha Karvanen, DSc (Tech.), prof. [8], Eeva Aartolahti, PhD [9]

Affiliations:

[1] Faculty of Sport and Health Sciences, University of Jyväskylä, Finland

[2] Department of Sports and Rehabilitation, South-Eastern Finland University of Applied Sciences, Savonlinna, Finland

[3] Jamk University of Applied Sciences, LIKES, Jyväskylä, Finland

[4] Central Ostrobothnia wellbeing service county “Soite”, Kokkola, Finland

[5] Physical Activity and Functional Capacity Research Group, Faculty of Health Care and Social Services, LAB University of Applied Sciences, Lahti, Finland

[6] Social and Healthcare Unit, Vaasa University of Applied Sciences, Vaasa, Finland

[7] Well-being and Culture Unit, Oulu University of Applied Sciences, Oulu, Finland

[8] Faculty of Mathematics and Science, University of Jyväskylä, Jyväskylä, Finland

[9] Institute of Rehabilitation, Jamk University of Applied Sciences, Jyväskylä, Finland

Running head:

Effects of VRBT on cognition in older adults

Corresponding author:

Senior Researcher Dr. Eeva Aartolahti, Institute of Rehabilitation, JAMK University of Applied Sciences, Piippukatu 2, FI-40100 Jyväskylä, Finland. Email address: eeva.aartolahti@jamk.fi, Telephone number: +358 50 3237511.

# The effects of virtual reality training on cognition in older adults: a systematic review, meta-analysis, and meta-regression of randomized controlled trials

The aim of this systematic review, meta-analysis, and meta-regression was to examine the effects of virtual reality-based training (VRBT) on global cognition and executive function compared with conventional training or information-based treatment in older adults, regardless of cognitive level. A systematic literature search was conducted using four databases. A total of 31 randomized controlled trials (RCT) were identified. Pooled effect sizes were calculated, the risk of bias was assessed, and evidence was graded. The primary analyses showed a small but statistically significant effect of VRBT compared with control on global cognition (Hedges' g 0.42, 95% CI 0.17 to 0.68, $I^2$=70.1%, n=876, 20 RCTs, low evidence) and executive function (Hedges' g 0.35, 95% CI 0.06 to 0.65, $I^2$=68.4%, n=810, 16 RCTs, very low evidence). Meta-regression yielded inconclusive results. VRBT may be more effective than control in improving cognition in older adults; however, more high-quality studies are needed.

Keywords: Virtual Reality, Aged, Cognition, Meta-Analysis, Randomized Controlled Trial

## INTRODUCTION

Cognitive health is an important part of functional performance that enables independence and the management of daily living in older age (WHO, 2019). Biological aging is an important factor that causes changes in brain structure and, together with other risk factors, such as lifestyle and family history, may influence the onset of cognitive impairment (WHO, 2019). To date, more than 55 million people have dementia worldwide, and millions of new cases occur each year (WHO, 2023). Therefore, prevention and rehabilitation of cognitive impairment are essential.

In recent years, cognitive rehabilitation has shifted to the increasing use of virtual reality (VR) (Maggio et al., 2019), as it provides multisensory stimulation, increases adherence to training (Valenzuela et al., 2018), and often allows for simultaneous motor-cognitive or dual-task challenges (Monteiro-Junior et al., 2016; Pichierri et al., 2011), which have been shown to benefit cognition and brain function through neuroplasticity (Cramer et al., 2011).

VR is often described as a computer-simulated 3D environment that provides a sense of presence and real-time interaction in virtual space (Steuer, 1992; Wilson et al., 1997). VR can be divided into three different levels of immersion (non-immersive, semi-immersive, fully immersive) (Mujber et al., 2004), depending onthe hardware used, the experience and feeling of presence gained by replacing the physical environment with the virtual world, and the amount of sensory information available (Rizzo & Koenig, 2017; Steuer, 1992; Wilson et al., 1997). Different levels of virtual reality can also be identified through the concept of a virtuality continuum, which expresses a fusion of real and virtual environments with different levels of virtuality (Milgram & Kishino, 1994). However, a unified definition of virtual reality in rehabilitation research is lacking.

Recent meta-analyses have supported the beneficial effects of virtual reality-based training (VRBT) in improving cognitive function in older adults with MCI or dementia (Gómez-Cáceres et al., 2022; Kim et al., 2022; Papaioannou et al., 2022; Yan et al., 2022; Yu et al., 2022; Zhong et al., 2021; Zhu

et al., 2021), or over 60-year-olds with undefined cognitive level and other neurological disorders (Yen & Chiu, 2021). These studies included many different cognitive outcomes, of which overall cognitive performance, that is, global cognition, and executive function, are considered particularly important for this review, as they have been found to predict dependence (Gill et al., 1996), functional decline, and mortality (Johnson et al., 2007). With regard to these outcomes, the results of recent meta-analyses have varied. VRBT was found to improve global cognition and executive function in older persons with mild cognitive impairment (MCI) in one study (Gómez-Cáceres et al., 2022), induce no differences between intervention and control groups in another study (Kim et al., 2022), and have beneficial effects only on global cognition but not on executive function in a third study (Papaioannou et al., 2022). These findings are inconclusive. Only one meta-analysis with meta-regression analysis of VRBT in older adults has been published before; in that study the participants had other neurological disorders, and in study selection process, the primary outcome was depressive symptoms (Yen & Chiu 2021).

Grading of evidence was seen in only one recent meta-analysis, which found moderate evidence for global cognition and very low evidence for executive function in older adults with MCI or dementia (Papaioannou et al., 2022). However, Papaioannou et al. included other digitized forms of rehabilitation, such as computerized cognitive training, in the control group, making it difficult to separate the true effects of VRBT. In addition, none of the recent meta-analyses assessed the risk of bias separately for each outcome, which may have led to more biased assessments. Some adverse events have been reported with VRBT in older adults (Papaioannou et al., 2022; Yen & Chiu, 2021). These include dizziness, fatigue, mild symptoms of cybersickness, and delayed muscle soreness. However, reports on adverse events associated with VRBT are scarce.

This review aimed to address these limitations. The objectives of this systematic review, meta-analysis, and meta-regression were 1) to examine the effects of virtual reality-based training (VRBT) on global cognition and executive function compared with conventional training with physical and/or

cognitive exercises or information-based treatment in people aged 60 years and older (defined here as older adults); 2) to examine whether the covariates of the study factors were associated with these effects; 3) to grade the level of evidence; and 4) to assess the safety of VRBT by considering adverse events reported in the original studies. To reliably determine the effects of VRBT, this review was based on randomized controlled trials (RCTs). In this study, the focus was on medical rehabilitation, and interventions using interactive virtual reality or video game technology that simulated real-life situations or environments and provided a sense of presence were considered VR.

## METHODS

This meta-analysis and meta-regression review was conducted according to the guidelines of the Cochrane Handbook for Systematic Reviews of Interventions (Cochrane, 2021). The report corresponds with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-2020) (Page et al., 2021). The PRISMA checklist is provided in the Supplementary Material S1. This review was prospectively registered in the International Prospective Register of Systematic Reviews (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022319227).

### Search strategy

The data for this study were collected as part of a larger research project investigating the effectiveness and meanings of robotics, virtual reality, and augmented reality in medical rehabilitation (Ilves et al., 2022a), due to which a two-phase literature search and screening took place. A comprehensive systematic literature search was carried out using the Ovid MEDLINE, CINAHL, PsycINFO and ERIC databases. The first phase of the search was conducted from inception to

September 16, 2020. An updated literature search was conducted after the registration of this review using the same databases from September 2020 to March 25, 2022. The reference lists of previous systematic reviews were hand-searched to obtain all relevant literature. The search strategy consisted of a Cochrane filter for RCTs (Lefebvre et al., 2011), along with MeSH terms and keywords related to rehabilitation, virtual reality including exergaming and video gaming, and augmented reality. No language limitations were implemented during the database search. The search strategy for the Ovid MEDLINE database is presented in Supplementary Material S1.

**Eligibility criteria**

A two-phase screening process was used in this review, as described above. In the first phase, the records identified in the larger project (Ilves et al., 2022b) were screened according to the pre-specified PICOS strategy; participants (P) included both children and adults, regardless of age, in need of medical rehabilitation; intervention (I) included rehabilitation robots, virtual reality or augmented reality; and the control (C) group received different approaches or implementation of rehabilitation, usual care, or waiting list treatment. Outcomes (O) included any parameter related to the International Classification of Functioning, Disability, and Health (ICF) classification of bodily functions and structures, performance and participation, quality of life, physical and mental well-being, need for assistance, and ability to study and work. Eligible studies were limited to parallel and cross-over RCT designs (S).

In the second phase of screening, more specific PICOS criteria were applied. Original peer-reviewed studies were eligible for this review if they met the following PICOS criteria: participants (P) were at least 60 years of age regardless of cognitive level; and intervention (I) was delivered as VRBT involving either only VR training or VR training in addition to conventional training or information-based treatment. Interventions using interactive virtual reality or video game technology that

simulated real-life situations or environments and provided a sense of presence were considered as VR. Control (C) group received conventional training with physical and/or cognitive exercises, or information-based treatment, where participants received health education or written instructions. Outcomes (O) included global cognition or executive function, and the study design (S) was a parallel or cross-over RCT.

Conference abstracts without full text, letters to the editor, studies with participants having neurological disorders other than cognitive decline, or studies with a control group that received another type of VRBT, computerized training or no treatment, were excluded from this review. Studies published in English, German, Swedish, or Finnish were included because of the language skills of the research group.

**Study selection process**

Screening of abstracts and full texts was performed by two independent reviewers (MK, AK, RYI, EA, OI, SH) using the Covidence systematic review software (Covidence, 2022). All conflicting decisions of the two independent reviewers were discussed and resolved between these reviewers. If no consensus was reached, a third opinion was obtained (EA).

**Data extraction**

Data for this review were extracted independently by two researchers (MK, SH) using a pre-customized format in Covidence (Covidence, 2022). The extraction included the identification information of the included study, funding source, study aims and setting, population group, inclusion and exclusion criteria, baseline characteristics of the participants, details of the experimental intervention and control groups, and the outcome measures used. Quantitative data of mean values

with standard deviations or standard errors and the number of participants at baseline and post-intervention were extracted for statistical analyses according to the predefined priority list of measures. Priority lists for both global cognition and executive function are presented in Supplementary Material S1. In cases of insufficient quantitative data, the corresponding author was contacted three times by e-mail. If adequate data were not received after three requests, the study was excluded from meta-analyses. The results of these studies are reported narratively in Supplementary Material S2.

**Risk of bias assessment**

The included studies were appraised separately for each outcome by two independent assessors (MK, SH) using the Cochrane Risk of Bias 2 tool (RoB 2) for randomized and cluster-randomized (C-RCT) trials (Sterne et al., 2019). Potential conflicts between the assessors were resolved to reach a consensus. If no consensus was reached, a third opinion (EA) was obtained. The tool assesses five different domains: the process of randomization and concealment, deviations from intended interventions, missing outcome data, appropriate outcome measurements used, and possible selection of reported results (Higgins et al., 2019). A variant for C-RCT addresses additional considerations when assessing risk of bias in trials with groups of individuals randomized in clusters (Higgins et al., 2023). Overall risk of bias is estimated across five domains. The level of the risk of bias ranges from low (green) to unclear (yellow) or high (red). For the overall risk to be low, all domains must be rated as low. To gain understanding of the assessment tool and achieve reliable assessments, the use of the tool was practiced within the research team prior to implementation.

**Data analysis**

A meta-analysis was conducted to synthesize and pool data on global cognition and executive function. For the random-effects meta-analysis, R software with the Metafor package was used to calculate the effect sizes and variances (Viechtbauer, 2010). The post-intervention values of the mean, standard deviation (SD), and number of participants (n) were used in the analyses. When quantitative data were available as median, interquartile range (IQR), standard error (SE), or 95% confidence interval (CI), they were converted to mean and SD assuming a normal distribution. The direction of outcome measures (from low to high) was harmonized by multiplying the measures by -1 when necessary.

When there were multiple group comparisons within a study, all comparisons were analyzed simultaneously in the same meta-analysis (Htut et al., 2018; Karssemeijer et al., 2019; Manenti et al., 2020). This leads to a correlation within the meta-analysis, which is acknowledged here using a correlated effects model. Thus, the correlated effects model with robust variance estimation (RVE) was implemented using the Robumeta package in R (Fisher et al., 2017), as it considers the possible dependent effect of the studies used multiple times in the same meta-analysis. Cluster-randomized trials were included in the same meta-analysis with randomized controlled trials when cluster-randomization had been considered in analyses of the original studies (Higgins et al., 2023). If a correction due to cluster-randomization was not made in the original study or an intraclass correlation coefficient was not reported for the correction to be made for meta-analysis, the study was excluded from statistical analyses of this review (Higgins et al., 2023).

A user-specified within-study effect size correlation $\rho$ was set to 0.80 to conduct the analyses (Tanner-Smith et al., 2016). The intervention effect size (Hedges' g) and its' 95% CI were visualized using a forest plot. The Hedges' g scale was evaluated as follows: 0.20–0.49 was considered as a small effect, 0.50–0.79 as a medium effect, and 0.80 or more as a large effect (Cohen, 1992). Statistical heterogeneity ($I^2$) was assessed as follows: 0–40% may not represent important heterogeneity, 30–

60% may represent moderate heterogeneity, 50–90% may represent substantial heterogeneity, and 75–100% represents considerable heterogeneity (Deeks et al., 2021).

Subgroup analyses were performed according to the cognitive level for each outcome to reduce statistical heterogeneity. The level of cognition was categorized as normal, MCI, or dementia according to the characteristics or baseline cognitive scores of the participants by identifying the level when comparing the scores with the appropriate cut-off point of the measurement tests. Sensitivity analyses were carried out by excluding studies with a high risk of bias to assess the reliability of the results. Publication bias was assessed using funnel plots separately for both cognitive outcomes (Egger et al., 1997). Finally, despite the small number of participants and included studies, a meta-regression model was fitted to find out whether covariates of different study factors (age, baseline level of cognition, intervention duration in weeks, number of sessions per week, session duration in minutes per session, intervention volume, commercial versus customized technology, exercise- versus information-based treatment of control groups, supervision of intervention), and RoB 2 domains with a high risk of bias could influence the results.

**Certainty of evidence**

Evidence was graded separately for global cognition and executive function according to the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) (Guyatt et al., 2011; Schünemann et al., 2013). Five categories were assessed: risk of bias of the original studies, inconsistency, indirectness, imprecision, and publication bias (Schünemann et al., 2013). The certainty of evidence varies among very low, low, moderate, and high levels of evidence (Balshem et al., 2011). According to the GRADE guidelines, RCTs start with high-quality evidence and the level is lowered if serious limitations are identified in the categories described (Guyatt et al., 2011).

Formulated statements were used to describe the findings of this systematic review and meta-analysis (Santesso et al., 2020).

## RESULTS

### Study selection and characteristics

The literature search yielded a total of 4 921 records (Figure 1). After exclusion of duplicates, a first phase of study selection was applied by screening titles and abstracts of 3 621 studies and then skimming through the full text of 1 175 studies. A total of 802 studies were re-screened in the specified second phase using the PICOS criteria of this review, and 771 studies were excluded because their titles and abstracts did not meet the criteria. This left 61 full text articles for further evaluation. Finally, 31 original peer-reviewed studies published between 2010 and 2022 were included in this review. A list of the excluded full texts with justification (n=30) is presented in Supplementary Material S1.

All included studies were published in English, of which 26 were RCTs and five were C-RCTs. Two published articles were based on the same RCT (Liao et al., 2019; Liao et al., 2020). A total of 26 studies were included in the meta-analysis because four studies (Anderson-Hanley et al., 2012; Gunst et al., 2022; Optale et al., 2010; Padala et al., 2017) did not have sufficient data (see Supplementary Material S2) and one study (Ramnath et al. 2021) was unclear in taking cluster-randomization into account in the original analyses. The detailed characteristics of the selected studies that were included and excluded from the analyses are described in Table 1.

#### *Participants*

The mean age of the participants was 76 (SD 6) years (mean range 60–88 years), and 64% were female. Gender distribution was not reported in onestudy (Amjad et al., 2019). Cognitive levels varied between normal (five studies), mild cognitive impairment (16 studies), and dementia (five studies). The level was unclear in three studies (Bacha et al., 2018; Gomes et al., 2018; Stanmore et al., 2019), in which baseline cognitive scores were compared with the corresponding cut-off values of the measurement tests, revealing cognition as mildly impaired. Another study reported a normal level of cognition according to the Mini-Mental State Examination (MMSE), but the results from Montreal Cognitive Assessment (MoCA) showed values below normal limits (Htut et al., 2018). As these studies included participants with mild cognitive impairment, they were presented and analyzed according to MCI.

***Interventions***

The study setting was described in 22 studies (85%), varying from the home environment to the hospital, senior gymnasium, laboratories or research institutes, care homes, assisted living facilities or geriatric outpatient facilities, and welfare or other community centers.

The interventions included different VR technologies, of which 85% (22 studies) were described as non-immersive (e.g., Xbox 360 Kinect, Nintendo Wii) and 11% (3 studies) as immersive (VR headsets). One study (4%) classified the intervention as VR but did not specify the technology (Hwang & Lee, 2017). The length of an intervention session, number of training sessions per week, and length of the intervention period varied widely between studies. Virtual reality-based training was delivered an average of 48 (SD 20) minutes per session (range of variation 18–100 minutes) and 3 (SD 1) times per week (range 1–5 times). The mean length of the intervention period was 11 (SD 5) weeks (range 3–24 weeks). Follow-up was performed in 7 studies with post-intervention periods ranging from 4 weeks to 12 months. Where reported, the intensity and progression of experimental

interventions were generally based on the games played and, in some cases, Borg's scale, rate of perceived exertion (RPE), or heart rate. Interventions were performed either individually, in pairs, or in groups.

The reporting of adherence to the intervention varied, and the participation rate was reported either as a percentage, mean total exercise time, or mean number of sessions attended. According to studies that reported group-based adherence, participants were slightly more committed to VRBT than to the control treatment in five RCTs (Guimarães et al., 2018; Hughes et al., 2014; Karssemeijer et al., 2019; Moreira et al., 2021; Padala et al., 2017). The details of the study settings and adherence to the interventions are described in Table 1.

***Control groups***

The control groups received traditional physical and/or cognitive training or information-based treatment with health education or written instructions. In 16 studies, the amount of training and guidance received in the control group corresponded to that of the intervention group. In other studies, the training or guidance received by the control group was either less than that in the intervention group or was not described. Implementations in the control groups were executed individually, in pairs, or in groups.

***Outcomes***

The primary outcomes were global cognition and executive function, measured with various tests. For meta-analyses, results of global cognition were analyzed using the Montreal Cognitive Assessment (MoCA, n=9), Mini-Mental State Examination (MMSE, n=7), Addenbrooke's Cognitive Examination (ACE-III, n=1), Computerized Assessment of Mild Cognitive Impairment (CAMCI,

n=1), Cognitive Telephone Screening Instrument (COGTEL, n=1), and Loewenstein Occupational Therapy Cognitive Assessment –Geriatric (LOTCA-G, n=1). Executive function was measured with the Trail Making Test Part B (TMT-B, n=12), Stroop Color Word Test (SCWT, n=1), Frontal Assessment Battery (FAB, n=2), and CogState (n=1).

### *Adverse events*

Mild adverse events were reported in three (10%) studies (Anderson-Hanley et al., 2012; Bacha et al., 2018; Gomes et al., 2018) related to exercise-induced pain or muscle soreness, dizziness, fatigue, or frustration with the technology used (Table 1). Other adverse events unrelated to the study were also reported in one study, including acute upper respiratory illness, cancer diagnosis, back pain due to lifting, or a car accident (Anderson-Hanley et al., 2012). In 14 (45%) studies, experimental intervention did not cause any adverse events. The last 14 studies (45%) did not consider and report adverse events at all.

## Risk of bias

The overall risk of bias was assessed as unclear or high for each study, in terms of both global cognition and executive function. No study had a low overall risk of bias. The high risk originated mainly from deviations from the intended interventions, missing outcome data, and inappropriate outcome measures. Uncertainty was also found in the concealment of allocation in the randomization process and the lack of research protocols for statistical analyses. Detailed tables for the risk of bias are provided in the Supplementary Material S1.

## Meta-analysis of intervention effects

A total of 26 studies were included in the meta-analyses (Table 1). Meta-analyses were performed for global cognition (20 RCT, n=876) and executive function (16 RCT, n=810). Two studies were included twice in the meta-analysis for both global cognition (Htut et al., 2018; Manenti et al., 2020) and executive function (Karssemeijer et al., 2019; Manenti et al., 2020) because two appropriate control groups were compared with an intervention group and vice versa. The effect of the intervention was balanced using a correlated effects model, as described in the Methods section.

### *Global cognition*

The measures included in the meta-analysis of global cognition were MoCA (n=9), MMSE (n=7), ACE-III (n=1), CAMCI (n=1), COGTEL (n=1), and LOTCA-G (n=1). The higher the measured score, the better the outcome and performance in global cognition.

According to the meta-analysis, VRBT appeared to be more efficient than the control in improving global cognition, although the effect size was small (Hedges' g 0.42, 95% CI 0.17 to 0.68, 876 participants, 20 studies, low-quality evidence) (Figure 2). Statistical heterogeneity was high ($I^2$=70.1%). Subgroup analysis by cognitive level showed that VRBT was more effective in persons with normal cognition with a small effect size (Hedges' g 0.23, 95% CI 0.03 to 0.44, $I^2$=0%) and in MCI with a moderate effect size (Hedges' g 0.54, 95% CI 0.22 to 0.86, $I^2$=73%) compared with the control (Supplementary Material S2). No differences between the intervention and control groups were observed in older adults with dementia (Hedges' g 0.06, 95% CI -1.94 to 2.06, $I^2$=73%).

### *Executive function*

The measures included in the meta-analysis of executive function were the TMT-B (n=12), SCWT (n=1), CogState battery with an executive function test (n=1), and FAB (n=2). Most of the measures involved completion of time, meaning that the faster the performance, the better the result.

According to the meta-analysis, VRBT was more effective in improving executive function when compared with control, but the effect size was small (Hedges' g 0.35, 95% CI 0.06 to 0.65, 810 participants, 16 studies, very low-quality evidence) (Figure 3). Statistical heterogeneity was considered high ($I^2$=68.4%). Subgroup analysis by cognitive level showed no difference between the intervention and control groups in older adults with either normal cognition (Hedges' g 0.52, 95% CI -0.61 to 1.66, $I^2$=86%), MCI (Hedges' g 0.40, 95% CI –0.03 to 0.83, $I^2$=57%) or dementia (Hedges' g 0.01, 95% CI -0.59 to 0.60, $I^2$=29%).

**Sensitivity analyses**

In global cognition, a statistically significant small effect in favor of the intervention group remained after excluding 9 studies from the meta-analysis owing to a high overall risk of bias (Hedges' g 0.44, 95% CI 0.14 to 0.75, $I^2$=71,3%). However, the levels of effect (confidence intervals) were seen to enlarge. Regarding executive function, the meta-analysis was no longer statistically significant between the groups when 10 studies were excluded owing to a high risk of bias (Hedges' g 0.28, 95% CI -0.05 to 0.62, $I^2$=0%). The sensitivity analyses are presented in Supplementary Material S2.

**Meta-regression**

Meta-regression models were fitted separately for age, baseline level of cognition, intervention duration in weeks, number of sessions per week, session duration in minutes per session, intervention volume, supervision of intervention, combined intervention vs. pure VRBT, customized vs.

commercial technology, fully immersive vs. non-immersive VR, type of the control group intervention, and RoB 2 domains with a high risk of bias. Supervision of the intervention increased the effect on executive function (SMD 0.51, 95% CI 0.18 to 0.84, p=0.005) whereas high risk of bias in the selection of reported results decreased the effect on executive function (SMD -0.48, 95% CI -0.79 to -0.17, p=0.005) (Table 2).

**Publication bias**

Some degree of publication bias is possible with smaller studies favoring intervention in both global cognition (Figure 4) and executive function (Figure 5); however no strong conclusions can be drawn due to the small number of included studies. There were more studies on MCI than on dementia or cognitively healthy older adults. However, studies with large sample sizes are scarce. It is also possible that high statistical heterogeneity is confounded with the interpretation of publication bias (Sterne et al., 2011).

**Certainty of evidence (GRADE)**

The GRADE of evidence was low for global cognition and very low for executive function. The level was lowered due to the high risk of bias in individual studies, inconsistency with high statistical heterogeneity and different directions in the results of individual studies, and imprecision with wide confidence intervals in executive function. The certainty of evidence is presented as a Summary of Findings in Table 3.

**DISCUSSION**

Our findings show that virtual reality-based training may be more effective in improving global cognition and executive function than conventional physical and/or cognitive exercise training or information-based treatment in older adults, regardless of cognitive level. This means that VRBT could also be a potential preventative option by increasing cognitive reserve, although its long-term effects should be studied. The certainty of evidence was graded as low for global cognition and very low for executive function indicating uncertainty in the results of this review. Nevertheless, the reliability of the results was increased by including all cognitive levels to ensure correct categorization of the studies in the analyses. We also aimed to reduce statistical heterogeneity and identify covariates behind the clinical heterogeneity affecting the results by performing subgroup analyses, sensitivity analyses, and meta-regression. These results showed that VRBT was particularly beneficial in improving global cognition in older adults with MCI when compared with the control group.

Similar results regarding the positive effects of VRBT on global cognition in older adults with MCI, have been reported in recent meta-analyses (Gómez-Cáceres et al., 2022; Papaioannou et al., 2022; Yan et al., 2022; Yu et al., 2022; Zhong et al., 2021; Zhu et al., 2021). However, differing results have also been published regarding the statistical significance of VRBT compared with active or passive controls, which may be due to a very small number of studies analyzed (Kim et al., 2022).

Our subgroup analysis showed the significant benefits of VRBT in improving global cognition in older adults with normal cognition and MCI. This may indicate the potential of VRBT as a preventative means in cognitive impairment to delay the clinical onset of cognitive decline (Chao & Chen, 2022) by increasing cognitive reserve (Song et al., 2022; Tucker et al., 2011) or as an early intervention to manage the increasing need for rehabilitation in primary care once the condition is diagnosed (Cieza et al., 2021). No subgroup reached statistically significant effect in executive function though main analysis favored VRBT. This is probably due to the smaller number of studies and participants analyzed per subgroup.

Mild adverse events were reported in three studies. As no serious adverse events were identified and the number of mild events was low, VRBT appears to be a safe training alternativefor use either alone or as part of conventional rehabilitation. Nevertheless, the use of virtual reality should always be considered individually according to the goals and preferences of the user.

Although this review found positive results for the effects of VRBT on cognition, the certainty of evidence was either low or very low according to GRADE. This differs from the gradings of a previous meta-analysis by Zhong et al., which assessed the evidence medium for both outcomes of global cognition and executive function, although the effect sizes of their analyses were smaller. This was because Zhong et al. rated the inconsistency of the results of individual studies and the level of statistical heterogeneity as not serious. This may in part be due to the fact that the meta-analyses were conducted per measurement tests. In addition, publication bias was not considered serious, although funnel plots were not performed because of the small number of studies included (Zhong et al., 2021).

The aim of this review was to ensure high-quality evidence of the effects of VRBT and to reduce the clinical heterogeneity associated with the studies by accepting only RCTs. The original studies were appraised by two independent reviewers for both global cognition and executive function to strengthen the quality assessment of this review. However, in this assessment, the included original studies mainly had an unclear or high risk of bias. Therefore, the Cochrane Risk of Bias 2 tool is often considered strict and more demanding than the previous version (Minozzi et al., 2020), which is why it is important to ensure that the interpretation of the appraised domains is consistent among assessors. In this review, the appraisal tool was practiced within the research team prior to implementation and was supported with written notes to ensure consistency.

High statistical heterogeneity was observed in the results, which may originate from a wide range of clinical heterogeneity, such as the age range of participants; different levels of education; and varying content, duration, intensity, or progression of training with different length of intervention periods. Although two moderators were found to influence the results according to the statistical significance

of the covariate, the small number of studies analyzed in our meta-regression (20 RCTs in global cognition, 16 RCTs in executive function) makes the interpretation difficult and unreliable. However, it is important to investigate the effects of different study factors and high risk of bias domains on the results obtained from the meta-analyses. Therefore, the results of this meta-regression should be confirmed in future reviews, as high-quality RCTs will be published.

As various tests are used to measure global cognition and executive function, an important question revolves around the sensitivity of a test in discriminating small improvements in cognitive performance after a short training period. For example, the MMSE was used as a global cognitive measure in two studies with cognitively healthy participants (Guimarães et al., 2018; Moreira et al., 2021) and in two studies with MCI (Manenti et al., 2020; Thapa et al., 2020). This may influence the direction or magnitude of the effect size calculated owing to its high ceiling effect and low sensitivity when used with cognitively healthy persons or with MCI (Hoops et al., 2009). Some researchers have suggested VR as a potential measurement tool through digital tests and games alongside training (e.g., Cabinio et al., 2020; Kourtesis et al., 2021; Liu et al., 2023) to contribute to this issue.

In this review, meta-analysis was justified by an adequate number of RCTs included, with a consistent PICO strategy used during screening. Although several similar meta-analyses of RCTs have been published recently, many discrepancies remain evident. Of the eight most recent meta-analyses published between 2021 and 2022, only one review (Papaioannou et al., 2022) observed most important methodological items when following the appraisal tool for systematic reviews (Shea et al., 2017). Our review also reported the funding sources of the included RCTs and evaluated the results using a clearly defined control group. In addition, the inclusion of older adults regardless of their cognitive level allowed a more accurate assessment of the cognitive category of participants by considering the cut-off values of the measurement tests, rather than analyzing participants in subgroups based only on the inclusion criteria of the original study. In three original studies, the cognitive level of participants remained unclear when cognition was evaluated according to inclusion

criteria (Bacha et al., 2018; Gomes et al., 2018; Stanmore et al., 2019). In another study, the cognitive level of participants was evaluated as normal with the MMSE at the inclusion stage (Htut et al., 2018). When the results were critically evaluated, the baseline cutoff values with the MoCA test revealed MCI in all participants in the described studies. Thus, the original RCTs were assessed according to the appropriate cognitive subgroups strengthening the reliability of the results of this review.

Although the results support the use of VR in training, clear evidence of the best intervention program remains unclear. That is, no optimal intensity, amount, or duration of VRBT has been identified, although in a previous meta-analysis, VRBT with a total duration of at least 15 h resulted in a greater efficacy in improving global cognition than interventions with less than 15 h (Gómez-Cáceres et al., 2022). However, this only shows the overall training time, rather than the optimal weekly or daily dose. There are also varying results regarding the preference for fully immersive, semi-immersive or non-immersive VR (Yu et al., 2022; Zhu et al., 2021), which may emphasize an individualized approach to VRBT. The possibility of improved accessibility and reduced loneliness through technology has long been suggested (White et al., 2002). Attitudes toward digitalization may vary among older adults (Hill et al., 2015), which is important to consider before implementing VRBT. In addition, cost-effectiveness should be examined when evaluating the overall benefits of VRBT as it could serve as an important implication for future rehabilitation, as learned from the COVID-19 pandemic.

**Strengths and limitations**

This review had several strengths. First, high-quality research methods were used with grading of evidence and reporting of adverse events, which were often lacking in previous meta-analyses. Second, a comprehensive review protocol was followed by conducting an extensive literature search and careful screening of records to provide sufficient data to justify this meta-analysis, by pre-

specifying the PICOS strategy and planning the outcome preference lists to obtain new and reliable information about the study phenomenon in question, and by including only RCTs that provided a basis for high-quality evidence. Third, the included original peer-reviewed studies were appraised with an updated tool independently by two assessors and separately for both examined outcomes, providing a reliable assessment of the risk of bias. Fourth, the conduction of sensitivity analyses and meta-regression, and the assessment of publication bias provided important information for interpreting the results as well as acted as an essential effort to observe statistical heterogeneity. Finally, all cognitive levels were included in the meta-analysis, and by subgrouping the effects, it was possible to identify whether VRBT could serve as a beneficial training alternative for postponing cognitive impairment in cognitively healthy older adults or as an effective treatment for already reduced cognitive performance.

Some limitations are also evident. The systematic literature search could have been more extensive when considering the guidelines of Cochrane Handbook (Lefebvre et al., 2022). However, compared with previous reviews, the literature search of RCTs was considered sufficient for the purpose of this review, and no new studies were found through reference and hand searches. The publication language in the eligibility criteria can be seen as a limitation of a review. However, no studies were excluded in the full-text phase of this review due to language limitations. In this review, medians and IQR were converted to means and SDs assuming a normal distribution. As medians are more likely to be reported when the data are skewed, this conversion may have resulted in biased pooled effects of outcomes. Furthermore, some individual studies included a small number of participants, which may have biased the effect size or the statistical significance of the results. Very little research has been conducted on the long-term effects of VRBT, which is why it was not studied in this review. Although subgroup analyses were conducted, this review could not determine an optimal dose of VRBT. Finally, despite the attempt to reveal moderators influencing the results of meta-analyses, the

meta-regression may have been unreliable because of the small number of studies included in the model.

## CONCLUSIONS

Virtual reality-based training may be more effective compared with conventional training or information-based treatment in improving global cognition in older adults. The evidence is uncertain. Similar results are obtained with executive function with very uncertain evidence. For global cognition the effect was mostly evident for older adults with normal cognition and MCI. Though two covariates influenced the effect on executive function, the reliability of the association remains inconclusive. The reported adverse events suggest that VR is a safe training alternative. However, individual assessments, goals, and preferences must be noticed when VRBT is considered as part of rehabilitation. There is a further need for high-quality studies to confirm this evidence, as new studies may influence the direction and magnitude of the effect of the current results. Meta-regression analyses should be repeated as more studies are being published. Implications for future research should also include investigating the long-term effects of VRBT using RCTs with longer intervention and follow-up periods.

## Conflict of interest

The authors declare no conflicts of interest.

**Acknowledgements**

The larger project was funded by the Social Insurance Institution of Finland (Kela) under the grant 50/26/2019. No financial support was received for this review. The authors would like to acknowledge and thank the research team members of the larger project for their contribution that helped in conducting this review.

## REFERENCES

Amjad, I., Toor, H., Niazi, I. K., Pervaiz, S., Jochumsen, M., Shafique, M., Haavik, H., & Ahmed, T. (2019). Xbox 360 Kinect cognitive games improve slowness, complexity of EEG, and cognitive functions in subjects with mild cognitive impairment: a randomized control trial. *Games for health journal*, *8*(2), 144–152. https://doi.org/10.1089/g4h.2018.0029

Anderson-Hanley, C., Arciero, P. J., Brickman, A. M., Nimon, J. P., Okuma, N., Westen, S. C., Merz, M. E., Pence, B. D., Woods, J. A., Kramer, A. F., & Zimmerman, E. A. (2012). Exergaming and older adult cognition: a cluster randomized clinical trial. *American journal of preventive medicine*, *42*(2), 109–119. https://doi.org/10.1016/j.amepre.2011.10.016

Bacha, J. M. R., Gomes, G. C. V., de Freitas, T. B., Viveiro, L. A. P., da Silva, K. G., Bueno, G. C., Varise, E. M., Torriani-Pasin, C., Alonso, A. C., Luna, N. M. S., D'Andrea Greve, J. M., & Pompeu, J. E. (2018). Effects of Kinect Adventures games versus conventional physical therapy on postural control in elderly people: a randomized controlled trial. *Games for health journal*, *7*(1), 24–36. https://doi.org/10.1089/g4h.2017.0065

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., Norris, S., & Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality

of evidence. *Journal of clinical epidemiology*, *64*(4), 401–406. https://doi.org/10.1016/j.jclinepi.2010.07.015

Cabinio, M., Rossetto, F., Isernia, S., Saibene, F. L., Di Cesare, M., Borgnis, F., Pazzi, S., Migliazza, T., Alberoni, M., Blasi, V., & Baglio, F. (2020). The use of a virtual reality platform for the assessment of the memory decline and the hippocampal neural injury in subjects with mild cognitive impairment: the validity of smart aging serious game (SASG). *Journal of clinical medicine*, *9*(5), 1355. https://doi.org/10.3390/jcm9051355

Cieza, A., Causey, K., Kamenov, K., Hanson, S. W., Chatterji, S., & Vos, T. (2021). Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet (London, England)*, *396*(10267), 2006–2017. https://doi.org/10.1016/S0140-6736(20)32340-0

Chao, G., & Chen, L. (2022). Meta-analysis of Virtual Reality Based on Delaying Mild Cognitive Impairment. *The Journal of nervous and mental disease*, *210*(3), 194–198. https://doi.org/10.1097/NMD.0000000000001426

Cochrane (2021). Cochrane handbook for systematic reviews of interventions, version 6.2. https://training.cochrane.org/handbook/current

Cohen J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Covidence systematic review software (2022), Veritas Health Innovation, Melbourne, Australia. https://www.covidence.org

Cramer, S. C., Sur, M., Dobkin, B. H., O'Brien, C., Sanger, T. D., Trojanowski, J. Q., Rumsey, J. M., Hicks, R., Cameron, J., Chen, D., Chen, W. G., Cohen, L. G., deCharms, C., Duffy, C. J., Eden, G. F., Fetz, E. E., Filart, R., Freund, M., Grant, S. J., Haber, S., … Vinogradov, S. (2011). Harnessing

neuroplasticity for clinical applications. *Brain: a journal of neurology*, *134*(6), 1591–1609. https://doi.org/10.1093/brain/awr039

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2021). Analysing data and undertaking meta-analyses: identifying and measuring heterogeneity. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. Welch (eds.). *Cochrane handbook for systematic reviews of interventions, version 6.2*. https://training.cochrane.org/handbook/current/chapter-10#section-10-10-2

Delbroek, T., Vermeylen, W., & Spildooren, J. (2017). The effect of cognitive-motor dual task training with the biorescue force platform on cognition, balance and dual task performance in institutionalized older adults: a randomized controlled trial. *Journal of physical therapy science*, *29*(7), 1137–1143. https://doi.org/10.1589/jpts.29.1137

Egger, M., Davey Smith, G., 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

Fisher, Z., Tipton, E., Zhipeng, H., & Fisher, M. Z. (2017). Package 'robumeta'. https://cran.r-project.org/web/packages/robumeta/robumeta.pdf

Gill, T. M., Williams, C. S., Richardson, E. D., & Tinetti, M. E. (1996). Impairments in physical performance and cognitive status as predisposing factors for functional dependence among nondisabled older persons. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *51*(6), M283–M288. https://doi.org/10.1093/gerona/51a.6.m283

Gomes, G. C. V., Simões, M. D. S., Lin, S. M., Bacha, J. M. R., Viveiro, L. A. P., Varise, E. M., Carvas Junior, N., Lange, B., Jacob Filho, W., & Pompeu, J. E. (2018). Feasibility, safety, acceptability, and functional outcomes of playing Nintendo Wii Fit Plus™ for frail older adults: A randomized feasibility clinical trial. *Maturitas*, *118*, 20–28. https://doi.org/10.1016/j.maturitas.2018.10.002

Gómez-Cáceres, B., Cano-López, I., Aliño, M., & Puig-Perez, S. (2022). Effectiveness of virtual reality-based neuropsychological interventions in improving cognitive functioning in patients with mild cognitive impairment: A systematic review and meta-analysis. *The Clinical neuropsychologist*, 1–34. https://doi.org/10.1080/13854046.2022.2148283

Gouveia, É. R., Smailagic, A., Ihle, A., Marques, A., Gouveia, B. R., Cameirão, M., Sousa, H., Kliegel, M., & Siewiorek, D. (2021). The efficacy of a multicomponent functional fitness program based on exergaming on cognitive functioning of healthy older adults: a randomized controlled trial. *Journal of aging and physical activity*, *29*(4), 586–594. https://doi.org/10.1123/japa.2020-0083

Gschwind, Y. J., Eichberg, S., Ejupi, A., de Rosario, H., Kroll, M., Marston, H. R., Drobics, M., Annegarn, J., Wieching, R., Lord, S. R., Aal, K., Vaziri, D., Woodbury, A., Fink, D., & Delbaere, K. (2015). ICT-based system to predict and prevent falls (iStoppFalls): results from an international multicenter randomized controlled trial. *European review of aging and physical activity: official journal of the European Group for Research into Elderly and Physical Activity*, *12*, 10. https://doi.org/10.1186/s11556-015-0155-6

Guimarães, A. V., Barbosa, A. R., & Meneghini, V. (2018). Active videogame-based physical activity vs. aerobic exercise and cognitive performance in older adults: a randomized controlled trial. *Journal of physical education and sport*, *18*(1), 203-209. https://doi.org/10.7752/jpes.2018.01026

Gunst, M., De Meyere, I., Willems, H., & Schoenmakers, B. (2022). Effect of exergaming on wellbeing of residents in a nursing home: a single blinded intervention study. *Aging clinical and experimental research*, *34*(1), 151–157. https://doi.org/10.1007/s40520-021-01903-1

Guyatt, G. H., Oxman, A. D., Schünemann, H. J., Tugwell, P., & Knottnerus, A. (2011). GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of clinical epidemiology*, *64*(4), 380–382. https://doi.org/10.1016/j.jclinepi.2010.09.011

Higgins, J. P. T., Savovic, J., Page, M. J., & Sterne, J. A. C. (2019). Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). Guidance document. https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/current-version-of-rob-2

Hill, R., Betts, L. R., & Gardner, S. E. (2015). Older adults' experiences and perceptions of digital technology: (Dis) empowerment, wellbeing, and inclusion. *Computers in Human Behavior*, *48*, 415–423. https://doi.org/10.1016/j.chb.2015.01.062

Higgins, J. P. T., Eldridge, S. & Li, T. (eds.). (2023). Chapter 23: Including variants on randomized trials. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, V. A. Welch (eds.). *Cochrane Handbook for systematic reviews of interventions*. Version 6.4. https://training.cochrane.org/handbook/current/chapter-23

Hoops, S., Nazem, S., Siderowf, A. D., Duda, J. E., Xie, S. X., Stern, M. B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*, *73*(21), 1738–1745. https://doi.org/10.1212/WNL.0b013e3181c34b47

Htut, T. Z. C., Hiengkaew, V., Jalayondeja, C., & Vongsirinavarat, M. (2018). Effects of physical, virtual reality-based, and brain exercise on physical, cognition, and preference in older persons: a randomized controlled trial. *European review of aging and physical activity: official journal of the European Group for Research into Elderly and Physical Activity*, *15*, 10. https://doi.org/10.1186/s11556-018-0199-5

Hughes, T. F., Flatt, J. D., Fu, B., Butters, M. A., Chang, C. C., & Ganguli, M. (2014). Interactive video gaming compared with health education in older adults with mild cognitive impairment: a feasibility study. *International journal of geriatric psychiatry*, *29*(9), 890–898. https://doi.org/10.1002/gps.4075

Hwang, J., & Lee, S. (2017). The effect of virtual reality program on the cognitive function and balance of the people with mild cognitive impairment. *Journal of physical therapy science*, *29*(8), 1283–1286. https://doi.org/10.1589/jpts.29.1283

Ilves, O., Korpi, H., Honkanen, S., & Aartolahti, E. (eds.). (2022a). *Robottien, virtuaalitodellisuuden ja lisätyn todellisuuden vaikuttavuus ja merkityksellisyys lääkinnällisessä kuntoutuksessa: Järjestelmälliset kirjallisuuskatsaukset*. Kela. http://urn.fi/URN:NBN:fi-fe2022052037517

Ilves, O., Korpi, H., Honkanen, S., Sjögren, T., & Aartolahti, E. (2022b). Järjestelmällisten kirjallisuuskatsausten toteutus. In O. Ilves, H. Korpi, S. Honkanen, & E. Aartolahti (eds.). *Robottien, virtuaalitodellisuuden ja lisätyn todellisuuden vaikuttavuus ja merkityksellisyys lääkinnällisessä kuntoutuksessa: Järjestelmälliset kirjallisuuskatsaukset* (pp. 32–33). Kela. http://urn.fi/URN:NBN:fi-fe2022052037517

Johnson, J. K., Lui, L. Y., & Yaffe, K. (2007). Executive function, more than global cognition, predicts functional decline and mortality in elderly women. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *62*(10), 1134–1141. https://doi.org/10.1093/gerona/62.10.1134

Karssemeijer, E. G. A., Aaronson, J. A., Bossers, W. J. R., Donders, R., Olde Rikkert, M. G. M., & Kessels, R. P. C. (2019). The quest for synergy between physical exercise and cognitive stimulation via exergaming in people with dementia: a randomized controlled trial. *Alzheimer's research & therapy*, *11*(1), 3. https://doi.org/10.1186/s13195-018-0454-z

Kela. The Social Insurance Institution of Finland. https://www.kela.fi/web/en

Kim, H., Jung, J., & Lee, S. (2022). Therapeutic application of virtual reality in the rehabilitation of mild cognitive impairment: a systematic review and meta-analysis. *Vision*, *6*(4), 68. https://doi.org/10.3390/vision6040068

Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2021). Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): an immersive virtual reality neuropsychological battery with enhanced ecological validity. *Journal of the International Neuropsychological Society: JINS, 27*(2), 181–196. https://doi.org/10.1017/S1355617720000764

Lefebvre, C., Manheimer, E., & Glanville, J. (2011). Searching for studies: designing search strategies. In J. P. T. Higgins, & S. Green (eds.). *Cochrane handbook for systematic reviews of interventions, version 5.1*. https://handbook-5-1.cochrane.org/chapter_6/6_4_11_search_filters.htm

Lefebvre, C., Glanville, J., Briscoe, S., Featherstone, R., Littlewood, A., Marshall, C., Metzendorf, M-I., Noel-Storr, A., Paynter, R., Rader, T., Thomas, J., Wieland, L. S. (2022). Searching for and selecting studies. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (eds.). *Cochrane Handbook for Systematic Reviews of Interventions, version 6.3*. https://www.training.cochrane.org/handbook

Liao, Y. Y., Chen, I. H., Lin, Y. J., Chen, Y., & Hsu, W. C. (2019). Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. *Frontiers in aging neuroscience*, *11*, 162. https://doi.org/10.3389/fnagi.2019.00162

Liao, Y. Y., Tseng, H. Y., Lin, Y. J., Wang, C. J., & Hsu, W. C. (2020). Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. *European journal of physical and rehabilitation medicine*, *56*(1), 47–57. https://doi.org/10.23736/S1973-9087.19.05899-4

Liao, Y. Y., Chen, I. H., Hsu, W. C., Tseng, H. Y., & Wang, R. Y. (2021). Effect of exergaming versus combined exercise on cognitive function and brain activation in frail older adults: A randomised controlled trial. *Annals of physical and rehabilitation medicine*, *64*(5), 101492. https://doi.org/10.1016/j.rehab.2021.101492

Liu, C. L., Cheng, F. Y., Wei, M. J., & Liao, Y. Y. (2022). Effects of exergaming-based Tai Chi on cognitive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. *Frontiers in aging neuroscience*, *14*, 761053. https://doi.org/10.3389/fnagi.2022.761053

Liu, Q., Song, H., Yan, M., Ding, Y., Wang, Y., Chen, L., & Yin, H. (2023). Virtual reality technology in the detection of mild cognitive impairment: A systematic review and meta-analysis. *Ageing research reviews, 87,* 101889. Advance online publication. https://doi.org/10.1016/j.arr.2023.101889

Maggio, M. G., Maresca, G., De Luca, R., Stagnitti, M. C., Porcari, B., Ferrera, M. C., Galletti, F., Casella, C., Manuli, A., & Calabrò, R. S. (2019). The growing use of virtual reality in cognitive rehabilitation: fact, fake or vision? A scoping review. *Journal of the National Medical Association*, *111*(4), 457–463. https://doi.org/10.1016/j.jnma.2019.01.003

Manenti, R., Gobbi, E., Baglio, F., Macis, A., Ferrari, C., Pagnoni, I., Rossetto, F., Di Tella, S., Alemanno, F., Cimino, V., Binetti, G., Iannaccone, S., Bramanti, P., Cappa, S. F., & Cotelli, M. (2020). Effectiveness of an innovative cognitive treatment and telerehabilitation on subjects with mild cognitive impairment: a multicenter, randomized, active-controlled study. *Frontiers in aging neuroscience*, *12,* 585988. https://doi.org/10.3389/fnagi.2020.585988

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems 77 (12), 1321–1329.*

Minozzi, S., Cinquini, M., Gianola, S., Gonzalez-Lorenzo, M., & Banzi, R. (2020). The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. *Journal of clinical epidemiology*, *126*, 37–44. https://doi.org/10.1016/j.jclinepi.2020.06.015

Monteiro-Junior, R. S., Vaghetti, C. A., Nascimento, O. J., Laks, J., & Deslandes, A. C. (2016). Exergames: neuroplastic hypothesis about cognitive improvement and biological effects on physical

function of institutionalized older persons. *Neural regeneration research*, *11*(2), 201–204. https://doi.org/10.4103/1673-5374.177709

Moreira, N. B., Rodacki, A. L. F., Costa, S. N., Pitta, A., & Bento, P. C. B. (2021). Perceptive-cognitive and physical function in prefrail older adults: exergaming versus traditional multicomponent training. *Rejuvenation research*, *24*(1), 28–36. https://doi.org/10.1089/rej.2020.2302

Mujber, T. S., Szecsi, T., & Hashmi, M. S. (2004). Virtual reality applications in manufacturing process simulation. *Journal of materials processing technology*, *155*, 1834–1838. https://doi.org/10.1016/j.jmatprotec.2004.04.401

Oliveira, J., Gamito, P., Souto, T., Conde, R., Ferreira, M., Corotnean, T., Fernandes, A., Silva, H., & Neto, T. (2021). Virtual reality-based cognitive stimulation on people with mild to moderate dementia due to Alzheimer's disease: a pilot randomized controlled trial. *International journal of environmental research and public health*, *18*(10), 5290. https://doi.org/10.3390/ijerph18105290

Optale, G., Urgesi, C., Busato, V., Marin, S., Piron, L., Priftis, K., Gamberini, L., Capodieci, S., & Bordin, A. (2010). Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. *Neurorehabilitation and neural repair*, *24*(4), 348–357. https://doi.org/10.1177/1545968309353328

Padala, K. P., Padala, P. R., Malloy, T. R., Geske, J. A., Dubbert, P. M., Dennis, R. A., Garner, K. K., Bopp, M. M., Burke, W. J., & Sullivan, D. H. (2012). Wii-fit for improving gait and balance in an assisted living facility: a pilot study. *Journal of aging research*, *2012*, 597573. https://doi.org/10.1155/2012/597573

Padala, K. P., Padala, P. R., Lensing, S. Y., Dennis, R. A., Bopp, M. M., Roberson, P. K., & Sullivan, D. H. (2017). Home-based exercise program improves balance and fear of falling in community-

dwelling older adults with mild Alzheimer's disease: a pilot study. *Journal of Alzheimer's disease: JAD*, *59*(2), 565–574. https://doi.org/10.3233/JAD-170120

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Papaioannou, T., Voinescu, A., Petrini, K., & Stanton Fraser, D. (2022). Efficacy and moderators of virtual reality for cognitive training in people with dementia and mild cognitive impairment: a systematic review and meta-analysis. *Journal of Alzheimer's disease: JAD*, *88*(4), 1341–1370. https://doi.org/10.3233/JAD-210672

Park, J. S., Jung, Y. J., & Lee, G. (2020). Virtual reality-based cognitive-motor rehabilitation in older adults with mild cognitive impairment: a randomized controlled study on motivation and cognitive function. *Healthcare*, *8*(3), 335. https://doi.org/10.3390/healthcare8030335

Pichierri, G., Wolf, P., Murer, K., & de Bruin, E. D. (2011). Cognitive and cognitive-motor interventions affecting physical functioning: a systematic review. *BMC geriatrics*, *11*, 29. https://doi.org/10.1186/1471-2318-11-29

Ramnath, U., Rauch, L., Lambert, E. V., & Kolbe-Alexander, T. (2021). Efficacy of interactive video gaming in older adults with memory complaints: A cluster-randomized exercise intervention. *PloS one*, *16*(5), e0252016. https://doi.org/10.1371/journal.pone.0252016

Rizzo, A. S., & Koenig, S. T. (2017). Is clinical virtual reality ready for primetime? *Neuropsychology*, *31*(8), 877–899. https://doi.org/10.1037/neu0000405https://doi.org/10.1037/neu0000405

Santesso, N., Glenton, C., Dahm, P., Garner, P., Akl, E. A., Alper, B., Brignardello-Petersen, R., Carrasco-Labra, A., De Beer, H., Hultcrantz, M., Kuijpers, T., Meerpohl, J., Morgan, R., Mustafa, R., Skoetz, N., Sultan, S., Wiysonge, C., Guyatt, G., Schünemann, H. J., & GRADE Working Group (2020). GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *Journal of clinical epidemiology*, *119*, 126–135. https://doi.org/10.1016/j.jclinepi.2019.10.014

Schünemann, H., Brozek, J., Gyuatt, G., & Oxman, A. (2013). GRADE Handbook. Introduction to GRADE Handbook. GRADEpro. https://gdt.gradepro.org/app/handbook/handbook.html

Serino, S., Pedroli, E., Tuena, C., De Leo, G., Stramba-Badiale, M., Goulene, K., Mariotti, N. G., & Riva, G. (2017). A novel virtual reality-based training protocol for the enhancement of the "mental frame syncing" in individuals with Alzheimer's disease: a development-of-concept trial. *Frontiers in aging neuroscience, 9*, 240. https://doi.org/10.3389/fnagi.2017.00240

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical research ed.), 358*, j4008. https://doi.org/10.1136/bmj.j4008

Song, S., Stern, Y., & Gu, Y. (2022). Modifiable lifestyle factors and cognitive reserve: A systematic review of current evidence. *Ageing research reviews*, *74*, 101551. https://doi.org/10.1016/j.arr.2021.101551

Stanmore, E. K., Mavroeidi, A., de Jong, L. D., Skelton, D. A., Sutton, C. J., Benedetto, V., Munford, L. A., Meekes, W., Bell, V., & Todd, C. (2019). The effectiveness and cost-effectiveness of strength and balance Exergames to reduce falls risk for people aged 55 years and older in UK assisted living facilities: a multi-centre, cluster randomised controlled trial. *BMC medicine*, *17*(1), 49. https://doi.org/10.1186/s12916-019-1278-9

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*, d4002. https://doi.org/10.1136/bmj.d4002

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Steuer, J., Biocca, F., & Levy, M. R. (1992). Defining virtual reality: Dimensions determining telepresence. *Communication in the age of virtual reality*, *33*, 37–39. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: a tutorial in R. *Journal of developmental and life-course criminology*, *2*, 85-112. https://doi.org/10.1007/s40865-016-0026-5

Thapa, N., Park, H. J., Yang, J. G., Son, H., Jang, M., Lee, J., Kang, S. W., Park, K. W., & Park, H. (2020). The effect of a virtual reality-based intervention program on cognition in older adults with mild cognitive impairment: a randomized control trial. *Journal of clinical medicine*, *9*(5), 1283. https://doi.org/10.3390/jcm9051283

Torpil, B., Şahin, S., Pekçetin, S., & Uyanık, M. (2021). The effectiveness of a virtual reality-based intervention on cognitive functions in older adults with mild cognitive impairment: a single-blind, randomized controlled trial. *Games for health journal*, *10*(2), 109–114. https://doi.org/10.1089/g4h.2020.0086

Tucker, A. M, & Stern, Y. (2011). Cognitive reserve in aging. *Current Alzheimer Research*, *8*(4), 354-360. https://doi.org/10.2174/156720511795745320

Valenzuela, T., Okubo, Y., Woodbury, A., Lord, S. R., & Delbaere, K. (2018). Adherence to Technology-Based Exercise Programs in Older Adults: A Systematic Review. *Journal of geriatric physical therapy (2001)*, *41*(1), 49–61. https://doi.org/10.1519/JPT.0000000000000095

van Santen, J., Dröes, R. M., Twisk, J. W. R., Blanson Henkemans, O. A., van Straten, A., & Meiland, F. J. M. (2020). Effects of exergaming on cognitive and social functioning of people with dementia: a randomized controlled trial. *Journal of the American Medical Directors Association*, *21*(12), 1958–1967.e5. https://doi.org/10.1016/j.jamda.2020.04.018

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wang, R. Y., Huang, Y. C., Zhou, J. H., Cheng, S. J., & Yang, Y. R. (2021). Effects of exergame-based dual-task training on executive function and dual-task performance in community-dwelling older people: a randomized-controlled trial. *Games for health journal*, *10*(5), 347–354. https://doi.org/10.1089/g4h.2021.0057

White, H., McConnell, E., Clipp, E., Branch, L. G., Sloane, R., Pieper, C., & Box, T. L. (2002). A randomized controlled trial of the psychosocial impact of providing internet training and access to older adults. *Aging & mental health*, *6*(3), 213–221. https://doi.org/10.1080/13607860220142422

WHO (2019). Risk reduction of cognitive decline and dementia: WHO Guidelines. World Health Organization.

WHO (2023). Dementia. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/dementia

Wilson, P. N., Foreman, N., & Stanton, D. (1997). Virtual reality, disability and rehabilitation. *Disability and rehabilitation*, *19*(6), 213–220. https://doi.org/10.3109/09638289709166530

Yan, M., Zhao, Y., Meng, Q., Wang, S., Ding, Y., Liu, Q., Yin, H., & Chen, L. (2022). Effects of virtual reality combined cognitive and physical interventions on cognitive function in older adults with mild cognitive impairment: A systematic review and meta-analysis. *Ageing research reviews*, *81*, 101708. https://doi.org/10.1016/j.arr.2022.101708

Yen, H. Y., & Chiu, H. L. (2021). Virtual reality exergames for improving older adults' cognition and depression: a systematic review and meta-analysis of randomized control trials. *Journal of the American Medical Directors Association*, *22*(5), 995–1002. https://doi.org/10.1016/j.jamda.2021.03.009

Yu, D., Li, X., & Lai, F. H. (2022). The effect of virtual reality on executive function in older adults with mild cognitive impairment: a systematic review and meta-analysis. *Aging & mental health*, 1–11. Advance online publication. https://doi.org/10.1080/13607863.2022.2076202

Zhong, D., Chen, L., Feng, Y., Song, R., Huang, L., Liu, J., & Zhang, L. (2021). Effects of virtual reality cognitive training in individuals with mild cognitive impairment: A systematic review and meta-analysis. *International journal of geriatric psychiatry*, *36*(12), 1829–1847. https://doi.org/10.1002/gps.5603

Zhu, S., Sui, Y., Shen, Y., Zhu, Y., Ali, N., Guo, C., & Wang, T. (2021). Effects of virtual reality intervention on cognition and motor function in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. *Frontiers in aging neuroscience*, *13*, 586999. https://doi.org/10.3389/fnagi.2021.586999

## Table Legend

Table 1. Summary of included RCT studies

Table 2. Results of meta-regression analysis for global cognition and executive function on covariates of study factors and the high risk of bias domains

Table 3. Summary of Findings

## Figure Legend

Figure 1. PRISMA 2020 flow diagram for identification and selection of studies

Figure 2. Forest plot with estimated effect of VRBT compared with control on global cognition

Figure 3. Forest plot with estimated effect of VRBT compared with control on executive function

Figure 4. Publication bias assessed by funnel plots for global cognition

Figure 5. Publication bias assessed by funnel plots for executive function

1 Table 1. Summary of included RCT studies

| Study | Participants | Intervention (I) | Control (C) | Outcome[1,2] | Study setting<br>Follow-up<br>Adverse events | Main results<br>Risk of bias (RoB2) | Funding |
|---|---|---|---|---|---|---|---|
| INCLUDED IN META-ANALYSIS | | | | | | | |
| ***Normal cognition*** | | | | | | | |
| Gouveia et al. 2021<br>Portugal<br>RCT | Community-dwelling older adults.<br>n=31<br>(37 randomized, 31 analyzed)<br>I: 67.6 (SD 5) y<br>C: 69.1 (SD 4.2) y<br>22 F / 9 M (71% / 29%) | Multicomponent functional fitness with exergaming and traditional physical exercise.<br>Technology: Kinect V2.<br>2 x 45 min./week<br>12 weeks<br>Intensity: moderate to vigorous.<br>Progression: depending on participants' perceived exertion and performance achieved in games.<br>Adherence: No group-based adherence level. Total rate 79%–89% out of 24 sessions.<br>n=15 | Functional fitness exercise in groups.<br>2 x 45 min. / week<br>12 weeks<br>Intensity: moderate to vigorous.<br>Progression: NI<br>Adherence: No group-based adherence level. Total rate 79%–89% out of 24 sessions.<br>n=16 | Cognition: COGTEL[1] | Local senior gymnasium.<br>Follow-up at 4 weeks after intervention.<br>No adverse events. | Both groups improved in global cognition after intervention.<br>Memory was improved in intervention group only after follow-up.<br>Exergaming demonstrated a more beneficial effect on cognition compared with traditional physical exercise.<br>RoB2 (GC): Unclear. | Portuguese Foundation for Science and Technology and Swiss National Science Foundation. |
| Gschwind et al. 2015<br>Australia, Spain, Germany<br>RCT<br>(Registered: ACTRN12614000096651) | Community-dwelling older adults (MCI excluded).<br>n=136<br>(153 randomized, 136 analyzed)<br>Mean age 74.7 (SD 6.3) y.<br>93 F / 60 M (61% / 39%) | Exergaming with balance and strength exercise + evidence-based booklet.<br>Technology: TV + Kinect.<br>Balance: 3 x 40 min./week<br>Strength: 3 x 15–20 min./week<br>16 weeks<br>Intensity: NI<br>Progression: by reducing upper limb support, narrowing the base of support, adjusting speed of movement, increasing gaming duration and number of repetitions, sets and the difficulty level.<br>Adherence: iStoppFalls system used median 42 times (IQR 57) at a game level of median 2.1 (IQR 3.9) for a total duration of median 11.7 h (IQR 22.0) when including and median 7.0 h (IQR 12.8) when excluding instructions.<br>n=71 | Evidence-based booklet about general health and fall prevention.<br>Control group was encouraged to follow their habitual exercise routines if applicable.<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=65 | Executive function: TMT-B[2], VST, DSB, DSC | Home.<br>No follow-up.<br>No adverse events. | No difference was found between groups concerning cognitive performance.<br>RoB2 (EF): High. | European Union's Seventh Framework Program, Australian National Health and Medical Research Council (NHMRC), Margarete and Walter Lichtenstein Foundation. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Guimarães et al.<br>2018<br>Brazil<br>RCT<br>(Registered: RBR-9crpzc) | Older adults<br>n=27<br>(36 randomized, 27 analyzed)<br>I: 60 (SD 4) y<br>C: 60.7 (SD 3.6) y<br>16 F / 11 M (59% / 41%) | Active videogame-based physical activity program.<br>Technology: Xbox 360 Kinect.<br>3 x 60 min./week<br>12 weeks<br>Intensity: moderate.<br>Progression: gradually increased, not detailed.<br>Adherence: 90.6%, with 424 sessions completed out of 468 sessions (13 participants x 36 sessions).<br>n=13 | Aerobic exercise with treadmill and cycle ergometer.<br>3 x 60 min./week<br>12 weeks<br>Intensity: moderate.<br>Progression: gradually increased, not detailed.<br>Adherence: 86.9%, with 438 sessions completed out of 504 sessions (14 participants x 36 sessions).<br>n=14 | Cognition: MMSE[1]<br><br>Executive function: CogState Battery (Groton Maze Learning Test)[2] | Laboratory.<br>No follow-up.<br>Adverse events: NI. | Training effects were positive for both groups. No difference between groups was found.<br>RoB2 (GC): High.<br>RoB2 (EF): High. | Federal University of Santa Catarina. |
| Moreira et al.<br>2021<br>Brazil<br>RCT<br>(Registered: RBR-97jm74) | Pre-frail older adults (reduced cognitive capacity excluded).<br>n=66<br>(99 randomized, 66 analyzed)<br>I: 70.8 (SD 4.5) y<br>C: 70.8 (SD 5.6) y<br>66 F (100%) | Exergaming.<br>Technology: Xbox 360 Kinect.<br>3 x 50 min./week<br>12 weeks<br>Intensity: Borg's Scale 10 to 15.<br>Progression: as proposed within the game.<br>Adherence: Exercise attendance rate 83.1%.<br>n=32 | Strength, balance and cardiorespiratory exercises.<br>3 x 50 min./week<br>12 weeks<br>Intensity: Borg's Scale 10 to 15.<br>Progression: weight increment was 3% body mass on the first week, followed by gradual increases thereafter.<br>Adherence: Exercise attendance rate 81.2%.<br>n=34 | Cognition: MMSE[1]<br><br>Executive function: TMT-B[2] | Setting not reported.<br>No follow-up.<br>Adverse events: NI. | Intervention group improved more in MMSE, processing speed and TMT-B than control group. Both groups decreased the number of errors.<br>RoB2 (GC): High.<br>RoB2 (EF): High. | Federal University of Paraná, CAPES Foundation. |
| Wang et al.<br>2021<br>Taiwan<br>RCT<br>(Registered: ACTRN 12617000095369) | Community-dwelling older adults (MMSE at least 24 points).<br>n=20<br>I: 71.30 (SD 5.33) y<br>C: 73.50 (SD 5.66) y<br>13 F / 7 M (65% / 35%) | Exergaming.<br>Technology: Kinect sensor with TANO and PAPAMAMA software system with treadmill.<br>3 x 60 min./week<br>12 weeks<br>Intensity: walking self-selected speed on the treadmill.<br>Progression: by increasing the treadmill speed and the difficulty of the motor or cognitive tasks progressively.<br>Adherence: Attendance rate 100% out of 36 exercise sessions.<br>n=10 | Home-based multicomponent exercise training with aerobic, strengthening and balance exercise.<br>3 x 60 min./week<br>12 weeks<br>Intensity: RPE 12 to 14.<br>Progression: Record kept on the exercise sessions and physical therapist checking the progress once in 2-3 weeks.<br>Adherence: Attendance rate 100% out of 36 exercise sessions.<br>n=10 | Executive function: TMT-B[2], EXIT-25, SCWT | Setting not reported for intervention group (supervised). Home for control group.<br>No follow-up.<br>No adverse events. | Experimental group improved significantly in measures of general executive function and inhibitory control compared with control group.<br>RoB2 (EF): Unclear. | Ministry of Science and Technology of the Republic of China. |

***Mild cognitive impairment (MCI)***

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Amjad et al.<br>2019<br>Pakistan<br>RCT | MCI<br>n=38<br>(44 randomized, 38 analyzed)<br>I: 62.8 (SD 5.1) y<br>C: 65.6 (SD 5.,0) y<br>Gender distribution not reported. | Exergaming.<br>Technology: Xbox 360 Kinect.<br>5 x 25–30 min./week<br>6 weeks<br>Intensity: NI<br>Progression: according to game difficulty.<br>Adherence: NI<br>n=20 | Range of motion (ROM) exercises.<br>5 x 25–30 min./week<br>6 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=18 | Cognition: MoCA[1], MMSE<br><br>Executive function: TMT-B[2] | Hospital, under the supervision of a therapist.<br>No follow-up.<br>Adverse events: NI. | MMSE, MoCA and TMT improved significantly. These changes were not observed in the control group.<br>RoB2 (GC): High.<br>RoB2 (EF): High. | Possible funders were Riphah College of Rehabilitation Sciences, Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, and Railway General Hospital. |
| Bacha et al.<br>2018<br>Brazil<br>RCT<br>(Registered: RBR-4z4f48) | Community dwelling older adults (MoCA baseline < 26 points).<br>n=46<br>Mean age 69.3 (SD 5.3) y.<br>34 F / 12 M (74% / 26%) | Virtual reality -based training.<br>Technology: Xbox 360 Kinect.<br>2 x 60 min./week<br>7 weeks<br>Intensity: moderate.<br>Progression: according to performance and game difficulty.<br>Adherence: 91.3% (assessed by frequency of the number of elderly individuals who completed the interventions and safety through the presence of adverse effects).<br>n=23 | Conventional physical therapy.<br>2 x 60 min./week<br>7 weeks<br>Intensity: moderate.<br>Progression: NI<br>Adherence: 91.3% (assessed by frequency of the number of elderly individuals who completed the interventions and safety through the presence of adverse effects).<br>n=23 | Cognition: MoCA[1] | Clinics Hospital of the School of Medicine of the University of São Paulo.<br>Follow-up at 4 weeks after intervention.<br>Adverse events: Delayed muscle pain in the lower limbs after the first session, 34% in VR group and 26% in comparison group. | Both groups presented a significant improvement in cognition (MoCA) posttreatment that was maintained at fourth week after treatment. There were no significant differences between groups.<br>RoB2 (GC): Unclear. | According to register: Faculty of Medicine at University of São Paulo and CAPES Foundation. |
| Delbroek et al.<br>2017<br>Belgium<br>RCT | MCI<br>n=17<br>(20 randomized, 17 analyzed)<br>I: 86.9 (SD 5.6) y<br>C: 87.5 (SD 6.6) y<br>13 F / 7 M (65% / 35%) | Virtual reality -based training + usual care.<br>Technology: BioRescue.<br>2 x 18–30 min./week<br>6 weeks<br>Intensity: A training session consisted of several 3-minute exercises. The level of exertion not described.<br>Progression: gradually increasing training time from 18 min. to 30 min.<br>Adherence: NI<br>n=8 | Standard of usual care including physical therapy.<br>6 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=9 | Cognition: MoCA (Dutch version)[1] | Nursing home, supervised by a therapist.<br>No follow-up.<br>No adverse events. | No changes were detected over time for either group with regards to MoCA.<br>RoB2 (GC): Unclear. | NI |
| Gomes et al.<br>2018<br>Brazil<br>RCT<br>(Registered: RBR-823rst) | Frail and pre-frail older adults (MoCA baseline < 26 points).<br>n=30<br>84 (SD 6.0) y<br>28 F / 2 M (93% / 7%) | Interactive videogaming.<br>Technology: Nintendo Wii Fit Plus.<br>2 x 50 min./week<br>7 weeks | Booklet of physical activity.<br>Intensity: NI<br>Progression: NI<br>Adherence: No group-based adherence reported. Out of all 30 | Cognition: MoCA[1] | Geriatric outpatient facility.<br>Follow-up 30 days after intervention.<br>Adverse events: Four participants (33%) | No significant effect on MoCA was found. No difference between groups.<br>RoB2 (GC): Unclear. | No external funding. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Intensity: motor-cognitive demand varied in games (blocks A and B). The level of exertion not described.<br>Progression: NI<br>Adherence: No group-based adherence reported. Out of all 30 participants enrolled in the study, 26 completed the full protocol (14 sessions).<br>n=15 | participants enrolled in the study, 26 completed the full protocol.<br>n=15 | | reported fatigue and one participant (8%) reported muscle soreness in the legs after the first session. | | |
| Htut et al.<br>2018<br>Thailand<br>RCT<br>(Registered: NCT03118414) | Older adults living at homes for the aged<br>(normal cognition according to MMSE, but MoCA scores reveal MCI).<br>n=84,<br>(n=63 when passive control group excluded)<br>Mean age 75.8 (SD 5.2) y.<br>37 F / 47 M (44% / 56%) | Virtual reality -based exercise (VRE).<br>Technology: Xbox 360.<br>3 x 30 min./week<br>8 weeks<br>Intensity: In 30 min of playing, participants chose 6 games involving upper and lower limb movements, and balance training. The level of exertion not described.<br>Progression: according to game performance and difficulty.<br>Adherence: NI<br>n=21 | Physical exercise (PE).<br>3 x 30 min./week<br>8 weeks<br>Intensity: NI<br>Progression: increasing intensity and resistance with elastic TheraBand® or decreasing support after 4 weeks.<br>Adherence: NI<br>n=21<br><br>Brain exercise (BE) with board and card games.<br>3 x 30 min./week<br>8 weeks<br>Intensity: 10 min./game.<br>Progression: from 2-player games to 4-player games.<br>Adherence: NI<br>n=21<br><br>Passive control with no exercise and conducting their lives as usual.<br>n=21 | Cognition:<br>MoCA[1],<br>TUG-Cog | Homes for the aged.<br>No follow-up.<br>Adverse events: NI. | VRE and BE improved MoCA more compared with PE or passive control group.<br>RoB2 (GC): Unclear. | Norway Government and Mahidol University. |
| Hughes et al.<br>2014*<br>USA<br>RCT | MCI<br>n=20<br>(For CAMCI-test, n=19)<br>Mean age 77.4 (SD 5.8) y.<br>14 F / 6 M (70% / 30%) | Interactive videogaming.<br>Technology: Nintendo Wii.<br>1 x 90 min./week<br>24 weeks<br>Intensity: Players used their arms and/or bodies to simulate actions required for each game. Brief health education at the beginning of each session (10–15 min), followed by interactive videogaming (75–80 | Health education.<br>1 x 90 min./week<br>24 weeks<br>Intensity: NA<br>Progression: On week 10 and 20 a Jeopardy® style tournament was held in small groups to encourage retention of the health information and to match the | Cognition:<br>CAMCI[1],<br>CSRQ | Centrally located church within the study area.<br>1-year follow-up.<br>Adverse events: NI. | No statistically significant improvement was found concerning cognition, although medium effect sizes were found for cognitive functioning in favor of interactive | The National Institute on Aging at NIH (The National Institute of Health). |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | min). The level of exertion not described.<br>Progression: The first 6 weeks focused on training and developing competence with the Wii system. On week 7 new games were introduced. On week 10 and 20 a Wii tournament was held in small groups.<br>Adherence: Participants attended an average of 23 (SD 1.1; range 21–24) sessions out of 24 sessions.<br>n=10 | level of friendly competition in the Wii group.<br>Adherence: Participants attended on average 22 (SD 3.3; range 14–24) sessions out of 24 sessions.<br>n=10 | | | video gaming compared with health education.<br>RoB2 (GC): Unclear. | |
| Hwang & Lee 2017<br>South Korea<br>RCT | MCI<br>n=24<br>I: 74.2 (SD 6.1) y<br>C: 70.2 (SD 5.4) y<br>17 F / 7 M (71% / 29%) | Virtual reality (VR) program.<br>Technology: NI.<br>5 x 30 min./week<br>4 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=12 | Traditional occupational therapy.<br>4 weeks (not detailed)<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=12 | Executive function / attention: (Stroop) Word Color Test (WCT)[2] | Welfare center.<br>No follow-up.<br>Adverse events: NI. | VR-training was significantly more effective than traditional occupational therapy in improving performance in WCT.<br>RoB2 (EF): High. | NI |
| Liao et al. 2019**<br>Taiwan<br>RCT<br>(Registered: TCTR20180531001) | MCI<br>n=34<br>(42 randomized, 34 analyzed)<br>I: 75.5 (SD 5.2) y<br>C: 73.1 (SD 6.8) y<br>23 F / 11 M (68% / 32%) | Virtual reality -based physical and cognitive training.<br>Technology: Microsoft Kinect + HTC VIVE VR-glasses.<br>3 x 60 min./week<br>12 weeks<br>Intensity: physical training for 40 min. and cognitive training for 20 min. The level of exertion not described.<br>Progression: according to game performance and difficulty.<br>Adherence: NI<br>n=18 | Combined physical and cognitive (CPC) training meeting the ACSM standards for seniors.<br>3 x 60 min./week<br>12 weeks<br>Intensity: MHR 50–75%, Borg's Scale 13 to 14.<br>Progression: by adding more weights, increasing the number of repetitions or the difficulty of the performance.<br>Adherence: NI<br>n=16 | Executive function: TMT-B[2], SCWT | Setting not reported.<br>No follow-up.<br>No adverse events. | VR-group showed more improvement in TMT-B than CPC-group. Both groups improved in SCWT.<br>RoB2 (EF): Unclear. | Ministry of Science and Technology. |
| Liao et al. 2020**<br>Taiwan<br>RCT<br>(Registered: TCTR20181001001) | MCI<br>n=34<br>(42 randomized, 34 analyzed)<br>I: 75.5 (SD 5.2) y<br>C: 73.1 (SD 6.8) y<br>23 F / 11 M (68% / 32%) | Virtual reality -based physical and cognitive training.<br>Technology: Microsoft Kinect + HTC VIVE VR-glasses.<br>3 x 60 min./week<br>12 weeks<br>Intensity: physical training for 40 min. and cognitive training for 20 | Combined physical and cognitive (CPC) training meeting the ACSM standards for seniors.<br>3 x 60 min./week<br>12 weeks<br>Intensity: MHR 50–75%, Borg's Scale 13 to 14.<br>Progression: by adding more weights, increasing the number of | Cognition: MoCA[1]<br><br>Executive function: EXIT-25 | Setting not reported.<br>No follow-up.<br>No adverse events. | Both groups showed improvement in cognitive performance and executive function but no differences between groups were observed.<br>RoB2 (GC): Unclear. | Ministry of Science and Technology. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | min. The level of exertion not described.<br>Progression: according to game performance and difficulty.<br>Adherence: NI<br>n=18 | repetitions or the difficulty of the performance.<br>Adherence: NI<br>n=16 | | | | |
| Liao et al.<br>2021<br>Taiwan<br>RCT<br>(Registered: TCTR20180531001) | Frail older adults (MoCA baseline < 26 points).<br>n=46<br>(61 randomized, 46 analyzed)<br>I: 79.6 (SD 9.0) y<br>C: 83.8 (SD 5.1) y<br>31 F / 15 M (67% / 33%) | Exergaming with resistance, aerobic, balance and Tai Chi exercise.<br>Technology: Kinect system.<br>3 x 60 min.<br>12 weeks<br>Intensity: HR and RPE monitored during training to ensure consistency between intervention and control groups.<br>Progression: resistance applied by TheraBand from 1.1 to 3.1 kg gradually.<br>Adherence: 27 out of 31 participants randomized completed all exercise sessions.<br>n=25 | Combined physical training including resistance, aerobic and balance exercises, meeting ACSM standards for older population.<br>3 x 60 min.<br>12 weeks<br>Intensity: intensity of aerobic exercise was 50-75% of participants MHR and targeted RPE 13 to 14.<br>Progression: resistance was applied by TheraBand from 1.1 to 3.1 kg gradually.<br>Adherence: 25 out of 30 participants randomized completed all exercise sessions.<br>n=21 | Cognition: MoCA[1]<br><br>Executive function: TMT-B[2], SCWT, EXIT-25 | Care centers.<br>No follow-up.<br>No adverse events. | Both groups improved in global cognition and executive function. Intervention group significantly enhanced global cognition more than the control group.<br>RoB2 (GC): Unclear.<br>RoB2 (EF): Unclear. | Ministry of Science and Technology. |
| Liu et al.<br>2022<br>Taiwan<br>RCT<br>(Registered: TCTRTCTR20210530003) | MCI<br>n=50<br>(54 randomized, 50 analyzed, n=33 when passive control group excluded)<br>I: 74.6 (SD 6.1) y<br>C1: 73.2 (6.3) y<br>C2: 73.4 (6.5) y<br>35 F / 15 M (70% / 30%) | Exergaming-based Tai Chi<br>Technology: Kinect system.<br>3 x 50 min.<br>12 weeks<br>Intensity: RPE 12 to 14<br>Progression: NI<br>Adherence: 16 out of 18 randomized participants completed all intervention.<br>n=16 | Traditional Tai Chi<br>3 x 50 min.<br>12 weeks<br>Intensity: RPE 12 to 14<br>Progression: NI<br>Adherence: 17 out of 18 randomized participants completed the whole intervention.<br>n=17<br><br>Passive control with no exercise and maintaining their usual daily physical activities.<br>n=17 | Cognition: MoCA[1], MMSE<br><br>Executive function: TMT-B[2], SCWT | Setting not reported.<br>No follow-up.<br>No adverse events. | Both experimental and active comparison groups performed significantly better in executive function than passive control group. Only the experimental group gained significant benefits in MoCA.<br>RoB2 (GC): High.<br>RoB2 (EF): High. | Taipei City Hospital and Cheng Hsin General Hospital. |
| Manenti et al.<br>2020<br>Italy<br>RCT<br>(Registered: NCT03486704) | MCI with memory complaints but absence of dementia<br>n=49<br>I1: 75.3 (SD 3.3) y<br>I2: 76.3 (SD 4.9) y | Face-to-face cognitive virtual reality rehabilitation (VRRS) + home-based virtual rehabilitation<br>Technology: VRRS by Khymeia.<br>3 x 60 min./week<br>4 weeks + 12 weeks | Face-to-face cognitive treatment as usual<br>60 min./session<br>12 sessions<br>Intensity: NI<br>Progression: NI | Cognition: MMSE[1], CDT (in the study categorized as a test for | Research institutes and participants' home.<br>No follow-up.<br>Adverse events: NI. | No significant differences were found in global cognition or executive function between the groups. | Italian Ministry of Health. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | C: 78.1 (SD 4.1) y<br>25 F / 24 M (51% / 49%) | Intensity: starting level and number of trials adjusted according to the participant's performance level using an adaptive staircase procedure.<br>Progression: task difficulty adaptively progressed. In home-based treatment, individualized cognitive training exercises adjusted by the therapist once a week.<br>Adherence: All 18 participants completed the 12 sessions of face-to-face VRRS treatment, 6 participants completed all 36 sessions of at-home VRRS training, whereas all other subjects completed more than 70% of the telerehabilitation sessions.<br>n=18<br><br>Face-to-face VRRS + home-based unstructured cognitive stimulation<br>Technology: VRRS by Khymeia.<br>3 x 60 min./week<br>4 weeks + 12 weeks<br>Intensity: starting level and number of trials adjusted according to the participant's performance level using an adaptive staircase procedure<br>Progression: task difficulty adaptively progressed.<br>Adherence: All 14 participants completed the 12 sessions of face-to-face VRRS treatment, 7 subjects completed the 36 sessions of at-home unstructured cognitive stimulation, and the other subjects completed more than 70% of the at-home unstructured cognitive stimulation sessions.<br>n=14 | Adherence: 8 participants completed the 12 sessions, whereas all the other subjects completed more than 70% of the usual treatment sessions.<br>n=17 | visuo-constructional abilities)<br><br>Executive function: TMT-B[2] | | RoB2 (GC): High.<br>RoB2 (EF): Unclear. | |
| Park et al. 2020<br>South Korea | MCI<br>n=35 | Virtual reality -based cognitive-motor rehabilitation.<br>Technology: MOTOCOG®. | Conventional cognitive rehabilitation (CCR) | Cognition: MoCA[1] | Setting not reported.<br>No follow-up.<br>Adverse events: NI. | VR-intervention showed a significantly greater improvement | The National Research Foundation of Korea (NRF) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| RCT | (40 randomized, 35 analyzed)<br>I: 75.8 (SD 8.5) y<br>C: 77.2 (SD 7.2) y<br>18 F / 17 M (51% / 49%) | 5 x 30 min./week<br>6 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=18 | (e.g. puzzles, card games, paper-pencil table activities).<br>5 x 30 min./week<br>6 weeks<br>Intensity: NI<br>Progression: Chosen by experienced occupational therapists to match the patient's cognitive function.<br>Adherence: NI<br>n=17 | Executive function: TMT-B[2], DST | | in MoCA, TMT and DST compared with CCR. In addition, subjects in VR-group had significantly higher interest and motivation compared with CCR-group.<br>RoB2 (GC): High.<br>RoB2 (EF): High. | funded by the Ministry of Education. |
| Stanmore et al. 2019<br>UK<br>C-RCT<br>(Registered: NCT02634736) | Older adults in assisted living facilities<br>(ACE-III baseline < 88 points).<br>n=92<br>(106 randomized, 92 analyzed),<br>18 living facilities included.<br>I: 77.8 (SD 10.2) y<br>C: 77.9 (SD 8.9) y<br>83 F / 23 M (78% / 22%) | Exergaming + Standard care.<br>Technology: Microsoft Kinect.<br>3 x 30 min./week<br>12 weeks<br>Intensity: individually prescribed program of standardized exergames suiting the participant's starting level of ability. Level of exertion not described.<br>Progression: tailored progression with more exergames within a session, greater challenge and longer duration.<br>Adherence: Participants attended a mean number of 25 (SD 8.5) out of 36 sessions offered over the 12-week study period. Attendance rate at 12 weeks was 87.5% for the intervention group. The mean total exercise time at the end of the 12 weeks was 359 (SD 151.2) min.<br>n=49 | Standard care with a community fall prevention advice comprising the Age UK Staying Steady leaflet and the OTAGO strength and balance home exercise program leaflet.<br>3 times per week<br>12 weeks<br>Intensity: 3 preselected exercises from the OTAGO list over the 12-week period. Details of exertion not described.<br>Progression: NI<br>Adherence: NI<br>n=43 | Cognition: ACE-III[1] | Assisted living facilities.<br>No follow-up.<br>No adverse events. | No statistically significant result concerning cognitive function.<br>RoB2 (GC): Unclear. | Innovate UK and Medical Reasearch Council (MRC). |
| Thapa et al. 2020<br>South Korea<br>RCT<br>(Registered: UMIN000040107) | MCI<br>n=66<br>(68 randomized, 66 analyzed)<br>Mean age 72.5 (SD 5.3) y.<br>52 F / 16 M (76% / 24%) | Virtual reality -based intervention + health education.<br>Technology: VR-glasses Oculus HMD.<br>VR: 3 x 100 min./week<br>Education: 1 x 30–50 min. per week<br>8 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: Participants strongly adhered to the intervention. Further | Health education.<br>1 x 30–50 min./week<br>8 weeks<br>Intensity: NA<br>Progression: NA<br>Adherence: NI<br>n=33 | Cognition: MMSE-DS[1]<br><br>Executive function: TMT-B[2], SDST | Setting not reported.<br>No follow-up.<br>No adverse events. | TMT-B improved significantly in the intervention group compared with the control group. Positive changes were observed in MMSE and SDST that were not significant.<br>RoB2 (GC): High.<br>RoB2 (EF): Unclear. | Dong-A University research grant. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | information on the rate not reported.<br>n=33 | | | | | |
| Torpil et al. 2021<br>Turkey<br>RCT | MCI<br>n=61<br>(64 randomized, 61 analyzed)<br>I: 70.12 (SD 2.57) y<br>C: 70.30 (SD 2.73) y<br>36 F / 25 M (59% / 41%) | Virtual reality -based rehabilitation + conventional cognitive rehabilitation.<br>Technology: Microsoft Kinect.<br>2 x 45 min./week<br>12 weeks<br>Intensity: NI<br>Progression: according to game challenge levels.<br>Adherence: NI<br>n=30 | Conventional cognitive rehabilitation.<br>2 x 45 min./week<br>12 weeks<br>Intensity: NI<br>Progression: the first 8 weeks featured interventions for one of the cognitive domains of the LOTCA-G test. The last 4 weeks included interventions involving all cognitive functions.<br>Adherence: NI<br>n=31 | Cognition: LOTCA-G[1] | Setting not reported.<br>No follow-up.<br>No adverse events. | Experimental group gained significantly greater improvements in several domains of LOTCA-G test compared with conventional group.<br>RoB2 (GC): Unclear. | No funding. |
| ***Dementia*** | | | | | | | |
| Karssemeijer et al. 2019<br>Netherlands<br>RCT<br>(Registered: NTR5581) | Mild or moderate dementia (AD n=59, vascular n=11, mixed n=24, not specified n=21)<br>MMSE baseline 22.4 (SD 3.2) points.<br>n=115<br>Mean age 79.9 (SD 6.5) y.<br>53 F / 62 M (46% / 54%) | Exergaming<br>Technology: Bike Labyrinth.<br>3 x 30–50 min./week<br>12 weeks<br>Intensity: 65–75% HRR or RPE 12-15.<br>Progression: Gradually increasing the level of exertion in addition to training duration. Start of training with<br>50–60% HRR or RPE 12–15 and 20 min. duration.<br>Adherence: 87.3% (SD 13.6) by dividing the attended sessions with total offered sessions.<br>n=38 | Aerobic exercise with a stationary bike.<br>3 x 30–50 min./week<br>12 weeks<br>Intensity: 65–75% HRR or RPE 12-15.<br>Progression: Gradually increasing the level of exertion in addition to training duration. Start of training with<br>50–60% HRR or RPE 12–15 and 20 min. duration.<br>Adherence: 81.1% (SD 13.7).<br>n=38<br><br>Active control with relaxation and flexibility exercises.<br>3 x 30–50 min./week<br>12 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: 85.4% (SD 12.9)<br>n=39 | Executive function: TMT-B[2], SCWT, LF, RSCT | Community centers.<br>Follow-up 3 months after intervention.<br>No adverse events. | No significant differences between the intervention and control groups were found for executive function.<br>RoB2 (EF): High. | Netherlands Organisation for Health Research and Development, (ZonMw). |
| Oliveira et al. 2021<br>Portugal | Mild to moderate dementia | Virtual reality -based cognitive stimulation. | Traditional cognitive stimulation using paper-and-pencil materials. | Cognition: MMSE[1], CDT | Residential care homes.<br>No follow-up. | A marginally significant effect found for global | No external funding, but a research unit supporting the |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Pilot RCT | (AD, questionable dementia n= 2, mild dementia n=7, moderate dementia n=8).<br>n=17<br>(18 randomized, 17 analyzed)<br>I: 82.60 (SD 5.42) y<br>C: 84.14 (SD 6.30) y<br>12 F / 5 M (71% / 29%) | Technology: Systemic Lisbon Battery virtual reality platform.<br>2 x 45 min./week<br>6 weeks<br>Intensity: NI<br>Progression: The exercise sessions presented different difficulty levels for progression throughout the intervention. Progression not detailed.<br>Adherence: NI<br>n=10 | Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=7 | Executive function: FAB[2], TMT-B (no data available for TMT) | Adverse events: NI. | cognition (MMSE) favoring experimental group, yet no differences found for improving executive function between groups.<br>RoB2 (GC): Unclear.<br>RoB2 (EF): High. | project was funded by the Foundation for Science and Technology (FCT) of Portugal. |
| Padala et al. 2012<br>USA<br>Pilot RCT | Mild dementia (AD, MMSE ≥ 17 points).<br>n=22<br>I: 79.3 (SD 9.8) y<br>C: 81.6 (SD 5.2) y<br>16 F / 6 M (73% / 27%) | Exergaming.<br>Technology: Nintendo Wii Fit.<br>5 x 30 min./week<br>8 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: Mean exercise time 11.1 (SD 3.5) hours of maximum 20 hours.<br>n=11 | Walking.<br>5 x 30 min./week<br>8 weeks<br>Intensity: self-paced indoor walking.<br>Progression: NI<br>Adherence: Mean exercise time 13.1 (SD 4.3) hours of maximum 20 hours.<br>n=11 | Cognition: MMSE[1] | Assisted living facility.<br>No follow-up.<br>No adverse events. | No statistically significant difference between groups in MMSE.<br>RoB2 (GC): High. | AMDA Foundation / Pfizer Quality Improvement Award and Alzheimer's Association New Investigator Award. |
| Serino et al. 2017<br>Italy<br>RCT | Older adults with AD and probable dementia (MODS < 85,5 points).<br>n=20<br>I: 86.6 (SD 6.1) y<br>C: 88.7 (SD 3.6) y<br>17 F / 3 M (85% / 15%) | Virtual reality -based training with navigation and recalling.<br>Technology: Computer + NeuroVirtual 3D software.<br>3 x 20 min./week<br>3–4 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=10 | Traditional cognitive rehabilitative activities (e.g. cards games, naming, fluency, and music listening).<br>3 times per week<br>3–4 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=10 | Executive function: FAB[2] | Social senior center.<br>No follow-up.<br>Adverse events: NI. | No significant difference between groups in improving executive function (FAB).<br>RoB2 (EF): Unclear. | Supported by several Italian research projects and the Cariplo Foundation. |
| van Santen et al. 2020<br>Netherlands<br>C-RCT<br>(Registered: NTR5537/NL5420) | Mild or moderate dementia (AD n=37, vascular n=9, mixed n=7, other n=13, unknown n=46).<br>n=112<br>(112 randomized, 84 analyzed)<br>23 day-care centers.<br>I: 79 (SD 6.0) y<br>C: 79 (SD 7.0) y<br>52 F / 60 M (46% / 54%) | Exergaming + regular activity program.<br>Technology: interactive cycling with e.g. speed-syncing SilverFit Mile -system.<br>2–5 x 20–30 min./week<br>6 months<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=52 | Activity program (e.g. arts and crafts, music and physical exercise such as walking outdoors).<br>5 times per week<br>6 months<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=32 | Cognition: MMSE[1]<br><br>Executive function: TMT-B[2] | Psychogeriatric day-care centers.<br>No follow-up.<br>No adverse events. | Exergaming showed positive effects on MMSE at 6 months compared with control group. Effect size was small but clinically relevant.<br>RoB2 (GC): High.<br>RoB2 (EF): High. | ZonMw-Memorabel programme/ Alzheimer Nederlands, Stichting Dioraphte and the EU. |

EXCLUDED FROM META-ANALYSIS

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Anderson-Hanley et al.<br>2012<br>USA<br>C-RCT<br>(Registered: NCT01167400) | MCI and healthy older adults.<br>n=79<br>I: 75.7 (SD 9.9) y<br>C: 81.6 (SD 6.2) y<br>62 F / 17 M (78% / 22%) | Cybercycling.<br>Technology: Stationary bikes + virtual reality display for 3D-tour.<br>Gradually 5 x 45 min./week<br>12 weeks<br>Intensity: mid-intervention adjustments made to maintain HRR of 60%.<br>Progression: gradually increasing exercise time and frequency.<br>Adherence: The frequency of rides (n) was 51.3 (SD 3.32).<br>n=38 | Stationary cycling.<br>Gradually 5 x 45 min. /week,<br>12 weeks<br>Intensity: mid-intervention adjustments made to maintain HRR of 60%.<br>Progression: gradually increasing exercise time and frequency.<br>Adherence: The frequency of rides (n) was 53.3 (SD 3.14).<br>n=41 | Executive function: Stroop C, DSB, CTT | Independent living facilities.<br>No follow-up.<br>Adverse events:<br>In VR group, a total of 7 adverse events reported: Knee or sciatica pain while cycling n=2, other injuries (hurt back lifting, car accident) n=1, cancer diagnosis and treatment n=2, frustrated interacting with bike computer n=1, vertigo while cycling n=1.<br>In comparison group, a total of 6 adverse events reported: Knee or sciatica pain while cycling n=2, acute illness (upper respiratory) n=1, other injuries (hurt back lifting, car accident) n=1, frustrated interacting with bike computer n=2. | Cybercycling was more effective on improving cognitive function than stationary cycling.<br>RoB2 (EF): High. | Robert Wood Johnson Foundation, Union College and Skidmore College. |
| Gunst et al.<br>2022<br>Belgium<br>C-RCT | Residents in a nursing home (MMSE > 20 points).<br>n=32<br>3 residential care centers included.<br>Age varied between 76-80 years in both groups.<br>22 F / 10 M (69% / 31%) | Exergaming + other planned activities.<br>Technology: Xbox 360 Kinect Sport.<br>2 x 60 min./week<br>13 weeks<br>Intensity: Borg's Scale at baseline was on average 10.9.<br>Progression: Borg's Scale at 3 months was 11.6.<br>Adherence: Mean 23.3 exergame sessions were organized in the three residential care centers and an average participation rate was 88%. | Planned activities (usual care).<br>13 weeks<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=17 | Cognition: CDT<br><br>Executive function: SCWT | Residential care centers.<br>Follow-up at 3 months after intervention.<br>Adverse events: NI. | The median time difference for SCWT slightly increased at 3 months for intervention group but decreased with the control group.<br>Results remained inconclusive due to unavailable comparative data. In addition, the direction of the CDT test remained unclear. | NI |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | n=15 | | | | RoB2 (GC): High.<br>RoB2 (EF): High. | |
| Optale et al.<br>2010<br>Italy<br>Pilot RCT | Older adults with memory deficits<br>n=31<br>(36 randomized, 31 analyzed)<br>Median age 80 y.<br>24 F / 12 M (67% / 33%) | Virtual reality memory training (VRMT) + auditory sessions + recreational activities (e.g. reading, painting).<br>Technology: VR-glasses HMD V6.<br>Initial phase (IP) 3 months:<br>3 x 30 min./week<br>Booster phase (BP) 3 months:<br>2 x 30 min./week<br>Total of 6 months<br>Intensity: IP with 3 auditory sections alternating with 3 VR sessions every 2 weeks. BP with 1 auditory and 1 VR session every week. The level of exertion not described.<br>Progression: gradually increasing the complexity of the stimuli.<br>Adherence: NI<br>n=15 | Music therapy + recreational activities (e.g. reading, painting).<br>Initial phase 3 months:<br>3 x 30 min./week<br>Booster phase 3 months:<br>2 x 30 min./week<br>Total of 6 months<br>Intensity: NI<br>Progression: NI<br>Adherence: NI<br>n=16 | Cognition:<br>MMSE,<br>MSN<br><br>Executive function:<br>CET, DTP, PVF | Rest-care home.<br>No follow-up.<br>Adverse events: NI. | VRMT-group showed significant improvements in several aspects of cognition. Control group showed a progressive decline.<br>RoB2 (GC): High.<br>RoB2 (EF): High. | Consorzio Sociale CPS and The Scientific Institute (IRCCS) Eugenio Medea. |
| Padala et al.<br>2017<br>USA<br>Pilot RCT<br>(Registered: NCT01002586) | Mild dementia<br>(AD, MMSE ≥ 18 points).<br>n=30<br>Mean age 73 (SD 6.2) y.<br>11 F / 19 M (37% / 63%) | Exergaming.<br>Technology: Nintendo Wii Fit.<br>5 x 30 min./week<br>8 weeks<br>Intensity: based on prior studies with respect to ability and safety.<br>Progression: according to performance in games.<br>Adherence: Total number of exercise sessions was mean 38 (SD 2) out of maximum 40 sessions (adherence 95%).<br>n=15 | Walking.<br>5 x 30 min./week<br>8 weeks<br>Intensity: self-paced indoor or outdoor walking.<br>Progression: NI<br>Adherence: Total number of exercise sessions was mean 37 (SD 6) out of maximum 40 sessions (adherence 93%).<br>n=15 | Cognition:<br>3MS,<br>MMSE | Home-based, supervised by caregivers.<br>Follow-up 8 weeks after intervention.<br>No study-related adverse events. | No difference between groups found concerning cognition.<br>RoB2 (GC): High. | New Investigator grant from Alzheimer's Association. |
| Ramnath et al.<br>2021<br>South Africa<br>C-RCT<br>(Registered: PACTR202008547335106) | Older adults with memory complaints<br>(Adapted Petersen criteria).<br>n=45<br>6 different retirement homes included.<br>I: 70.8 (SD 4.52) y<br>C: 74.14 (SD 5.8) y | Interactive video gaming.<br>Technology: Xbox Kinect Sport.<br>2 x 60 min./week<br>12 weeks<br>Intensity: Borg's RPE Scale from 1 to 10. Researchers ensured that the individuals participated within their limit.<br>Progression: NI | Conventional multimodal exercise.<br>2 x 60 min./week<br>12 weeks<br>Intensity: Borg's RPE Scale from 1 to 10. Researchers ensured that the individuals participated within their limit. | Cognition:<br>MMSE[1]<br><br>Executive function:<br>Stroop (modified)[2] | Retirement homes.<br>No follow-up.<br>No adverse events. | Experimental group showed significant improvement in the total number of correct responses and reaction time of correct color-words on the Stroop test | No funding. |

| | | | | |
|---|---|---|---|---|
| Gender distribution not reported. | Adherence: 100% for both groups with all 45 participants completing 24 one-hour sessions over the 12-week intervention period.<br>n=23 | Progression: body weight used during first 6 weeks. Intensity of upper body strength training for weeks 7-12 increased by using external weights.<br>Adherence: 100% for both groups with all 45 participants completing 24 one-hour sessions over the 12-week intervention period.<br>n=22 | | compared with conventional group.<br>RoB2 (GC): High.<br>RoB2 (EF): High. |

1 Outcome measures considered in quantitative analyses for global cognition are expressed as upper index number one (1).
2 Outcome measures considered in quantitative analyses for executive function are expressed as upper index number two (2).

*In Hughes et al. 2014 all participants were analyzed except for CAMCI, in which one participant was unable to complete the test. Thus, the number of analyzed participants (n=20) reported differs from the number of participants included in meta-analysis for CAMCI-test (VR n=9, Control n=10).
**Liao et al. 2019 and Liao et al. 2020 are different publications from the same RCT. Thus, the overall number of participants reported in this review are counted once for this RCT.

Abbreviations: ACE-III = Addenbrooke's Cognitive Examination III, ACSM = American College of Sports Medicine, AD = Alzheimer's Disease, C = Control group, CAMCI = The Computerized Assessment of Mild Cognitive Impairment, CDT = The Clock Drawing Test, CET = Cognitive Estimation Test, COGTEL = Cognitive Telephone Screening Instrument, C-RCT = Cluster-Randomized Controlled Trial, CSRQ = The Cognitive Self-Report Questionnaire 25, CTT = Color Trails Test, DSB = Digit Span Backwards, DSC = Digit Symbol Coding Test, DST = Digit Span Test, DTP = Dual Task Performance, EF = Executive function, e.g. = for example, EXIT-25 = The Executive Interview 25, F = Female, FAB = Frontal Assessment Battery, GC = Global cognition, HMD = Head-Mounted Display, HR = Heart Rate, HRR = Heart Rate Reserve, I = Intervention group, LF = Letter Fluency, LOTCA-G = Loewenstein Occupational Therapy Cognitive Assessment -Geriatric, M = Male, MCI = Mild Cognitive Impairment, MHR = Maximum Heart Rate, MMSE = Mini-Mental State Examination, MMSE-DS = Mini-Mental State Examination-Dementia screening test, MoCA = Montreal Cognitive Assessment, MODS = Milan Overall Dementia Scale, MSN = Mental Status in Neurology, n = number of participants, NA = Not Applicable, NI = No Information, PVF = Phonemic Verbal Fluency, RCT = Randomized Controlled Trial, RPE = Rate of Perceived Exertion, RSCT = Rule Shift Cards Test, SCWT = Stroop Color Word Test, SD = Standard Deviation, SDST = Symbol Digit Substitution Test, TMT = Trail Making Test, TMT-B = Trail Making Test Part B, TUG-Cog = Timed Up and Go Cognition, VST = Victoria Stroop Test, y = age in years, 3MS = Modified Mini-Mental State Examination.

1

Table 2. Results of meta-regression analysis for global cognition and executive function on covariates of study factors and the high risk of bias domains

| Covariates | Global cognition | | | | | | Executive function | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimated Effect Size | SE | Lower CI | Upper CI | *p* | sig. | Estimated Effect Size | SE | Lower CI | Upper CI | *p* | sig. |
| *Study factors* | | | | | | | | | | | | |
| Age (years) | -0.029 | 0.026 | -0.091 | 0.033 | 0.305 | | -0.032 | 0.030 | -0.112 | 0.048 | 0.343 | |
| Level of cognition at baseline | | | | | | | | | | | | |
| Normal | -0.230 | 0.154 | -0.748 | 0.287 | 0.240 | | 0.206 | 0.400 | -0.822 | 1.234 | 0.629 | |
| MCI | 0.392 | 0.239 | -0.148 | 0.932 | 0.135 | | 0.104 | 0.283 | -0.504 | 0.712 | 0.718 | |
| Dementia | -0.430 | 0.441 | -1.990 | 1.130 | 0.414 | | -0.357 | 0.247 | -1.004 | 0.291 | 0.213 | |
| Intervention duration (weeks) | -0.018 | 0.020 | -0.076 | 0.040 | 0.414 | | -0.049 | 0.018 | -0.103 | 0.004 | 0.062 | * |
| Number of sessions per week | -0.009 | 0.225 | -0.566 | 0.548 | 0.969 | | 0.201 | 0.214 | -0.423 | 0.826 | 0.406 | |
| Session duration (min/session) | -0.003 | 0.005 | -0.017 | 0.011 | 0.588 | | -0.002 | 0.006 | -0.022 | 0.017 | 0.745 | |
| Intervention volume (min/week) | -0.002 | 0.001 | -0.006 | 0.002 | 0.209 | | 0.000 | 0.001 | -0.005 | 0.006 | 0.798 | |
| Supervision of intervention | 0.086 | 0.125 | -0.176 | 0.349 | 0.500 | | 0.507 | 0.151 | 0.179 | 0.835 | 0.005 | *** |
| Combined intervention (ref. pure VRBT) | 0.002 | 0.253 | -0.545 | 0.550 | 0.993 | | -0.435 | 0.231 | -1.006 | 0.136 | 0.110 | |
| Customized technology (ref. commercial) | -0.195 | 0.232 | -0.683 | 0.294 | 0.412 | | -0.378 | 0.399 | -1.307 | 0.551 | 0.372 | |
| Fully immersive (ref. non-immersive) | -0.287 | 0.142 | -1.412 | 0.838 | 0.251 | | -0.119 | 0.271 | -2.011 | 1.773 | 0.720 | |
| Control group intervention (ref. all other types) | | | | | | | | | | | | |
| Physical exercise | 0.025 | 0.292 | -0.590 | 0.641 | 0.933 | | 0.364 | 0.294 | -0.272 | 1.001 | 0.237 | |
| Cognitive exercise | 0.366 | 0.378 | -0.565 | 1.296 | 0.372 | | 0.130 | 0.254 | -0.539 | 0.800 | 0.632 | |
| Combined exercise (physical+cognitive) | -0.209 | 0.178 | -0.679 | 0.261 | 0.298 | | -0.211 | 0.320 | -2.336 | 1.914 | 0.601 | |
| Information | -0.334 | 0.135 | -1.394 | 0.726 | 0.200 | | -0.535 | 0.177 | -1.105 | 0.035 | 0.058 | * |
| *High Risk of Bias* | | | | | | | | | | | | |
| Overall | -0.044 | 0.259 | -0.590 | 0.502 | 0.867 | | 0.077 | 0.244 | -0.466 | 0.621 | 0.759 | |
| Randomization process | NA | | | | | | NA | | | | | |
| Deviations from intended interventions | 0.172 | 0.426 | -0.947 | 1.292 | 0.704 | | 0.095 | 0.345 | -0.667 | 0.856 | 0.789 | |
| Missing outcome data | 0.261 | 0.293 | -0.393 | 0.915 | 0.395 | | 0.566 | 0.361 | -0.269 | 1.400 | 0.156 | |
| Measurement of the outcome | -0.494 | 0.283 | -1.247 | 0.259 | 0.148 | | 0.093 | 0.144 | -0.217 | 0.403 | 0.528 | |
| Selection of the reported results | -0.008 | 0.129 | -0.278 | 0.263 | 0.952 | | -0.479 | 0.145 | -0.790 | -0.168 | 0.005 | *** |

SE: Standard Error; CI: 95 % Confidence Interval; VRBT: virtual reality-based training; NA: not applicable; ****p* ≤ *0.01; * p* ≤ *0.10*

Table 3. Summary of Findings

**Virtual reality-based training (VRBT) compared with conventional training or information-based treatment**

**Patient or population:** Older adults at least 60 years of age regardless of the cognitive level

**Setting:** Varied between home, hospital, laboratory and research institutes, senior gymnasium, recreational or community settings, and different living facilities

**Intervention:** Virtual reality-based training

**Comparison:** Conventional training or information-based treatment

| Outcomes | Hedges' g with 95% CI* | No. of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|
| Global cognition | Hedges' g **0.42 higher** (CI 0.17 to 0.68) | 876 (20 RCTs) | ⊕⊕◯◯ Low [1,2] | Global cognition was assessed with valid or otherwise widely used measures of Montreal Cognitive Assessment, Addenbrooke's Cognitive Examination, Computerized Assessment of Mild Cognitive Impairment, Cognitive Telephone Screening Instrument, Loewenstein Occupational Therapy Cognitive Assessment –Geriatric and Mini-Mental State Examination. |
| Executive function | Hedges' g **0.35 higher** (CI 0.06 to 0.65) | 810 (16 RCTs) | ⊕◯◯◯ Very low [1,2,3] | Executive function was assessed with valid or otherwise widely used measures of Trail Making Test Part B, Stroop Color Word Test, CogState and Frontal Assessment Battery. |

*__CI:__ confidence interval

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

**Explanations**

1. Study limitations: Downgraded by one level due to high risk of bias in individual studies. Results of meta-regression were inconclusive due to a small number of studies analyzed, and there was a high number of unclear and high risk of bias found in all the domains assessed. For global cognition, high risk of bias was identified in randomization process (1), deviation from intended intervention (5), missing outcome data (7), measurement of the outcome (9) and selection of the reported results (1). For executive function, high risk of bias was identified in randomization process (2), deviation from intended intervention (8), missing outcome data (7), measurement of the outcome (1) and selection of the reported results (1).

2. Inconsistency of results: Downgraded by one level due to inconsistency in individual studies and high heterogeneity ($I^2$).

3. Imprecision: Downgrading by one level due to imprecision. The number of participants both in intervention and control groups are over 200 per group (intervention n=433, control n=422) but the number of studies is small and confidence interval wide.

4. Indirectness: No downgrading due to indirectness, as the review follows the specified PICOS criteria well.

5. Publication bias: No downgrading due to publication bias. Slight asymmetry indicated in funnel plots, but publication bias was not seen serious.

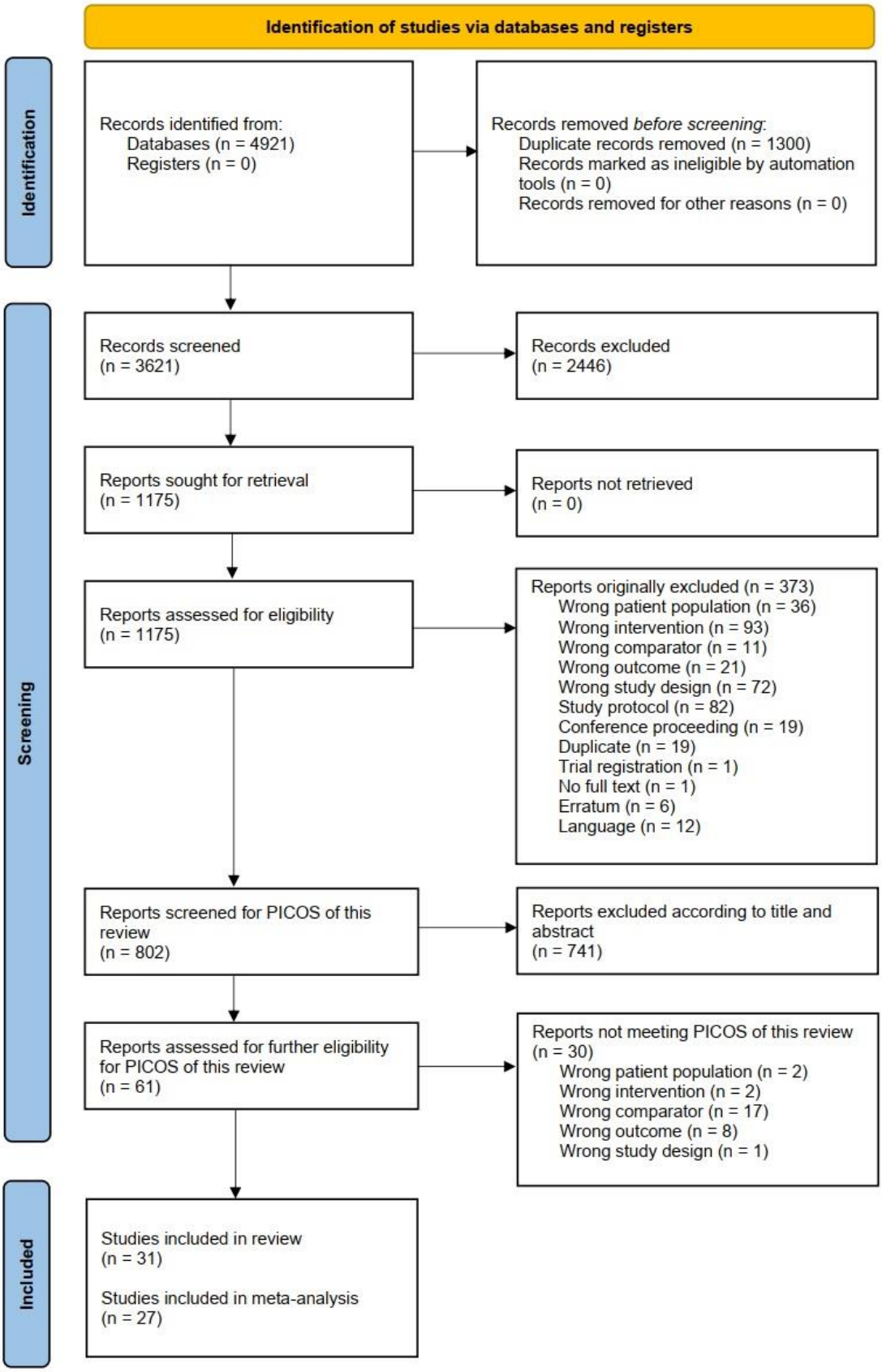



Figure 1. PRISMA 2020 flow diagram for identification and selection of studies.

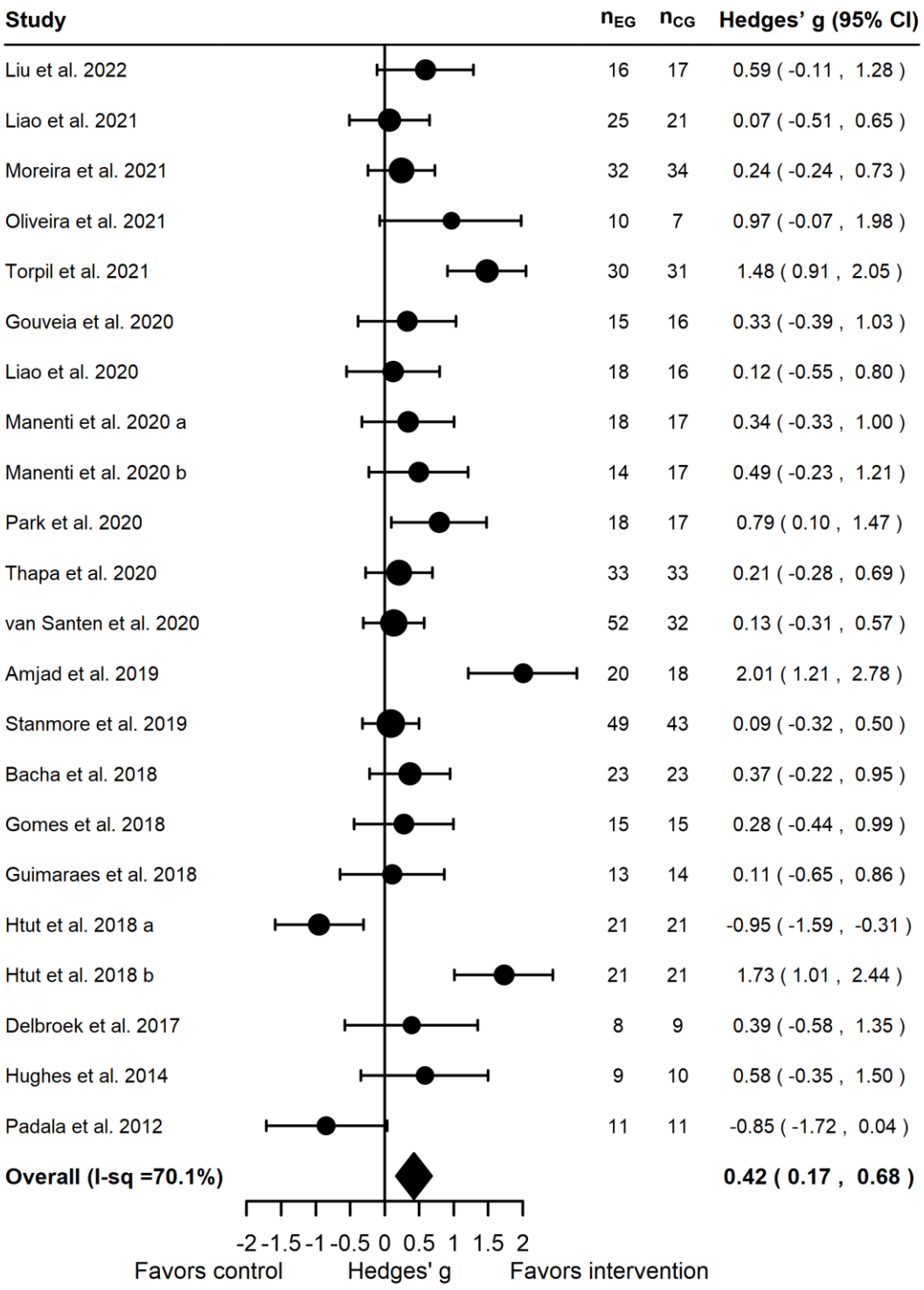



Figure 2. Forest plot with estimated effect of VRBT compared with control on global cognition. The size of the circle in Hedges' g value shows the emphasis of an individual study in the pooled effect.

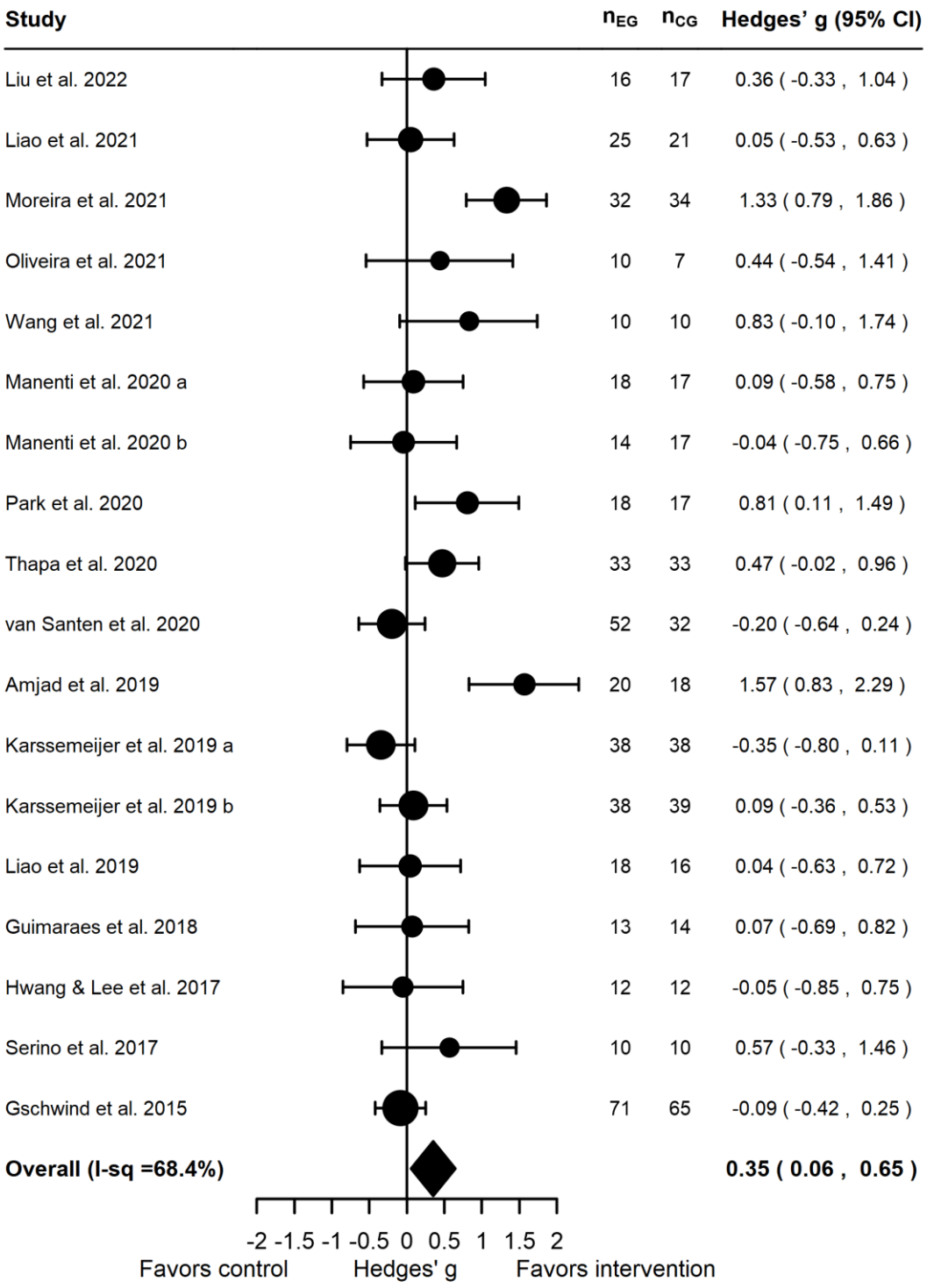



Figure 3. Forest plot with estimated effect of VRBT compared with control on executive function. The size of the circle in Hedges' g value shows the emphasis of an individual study in the pooled effect.

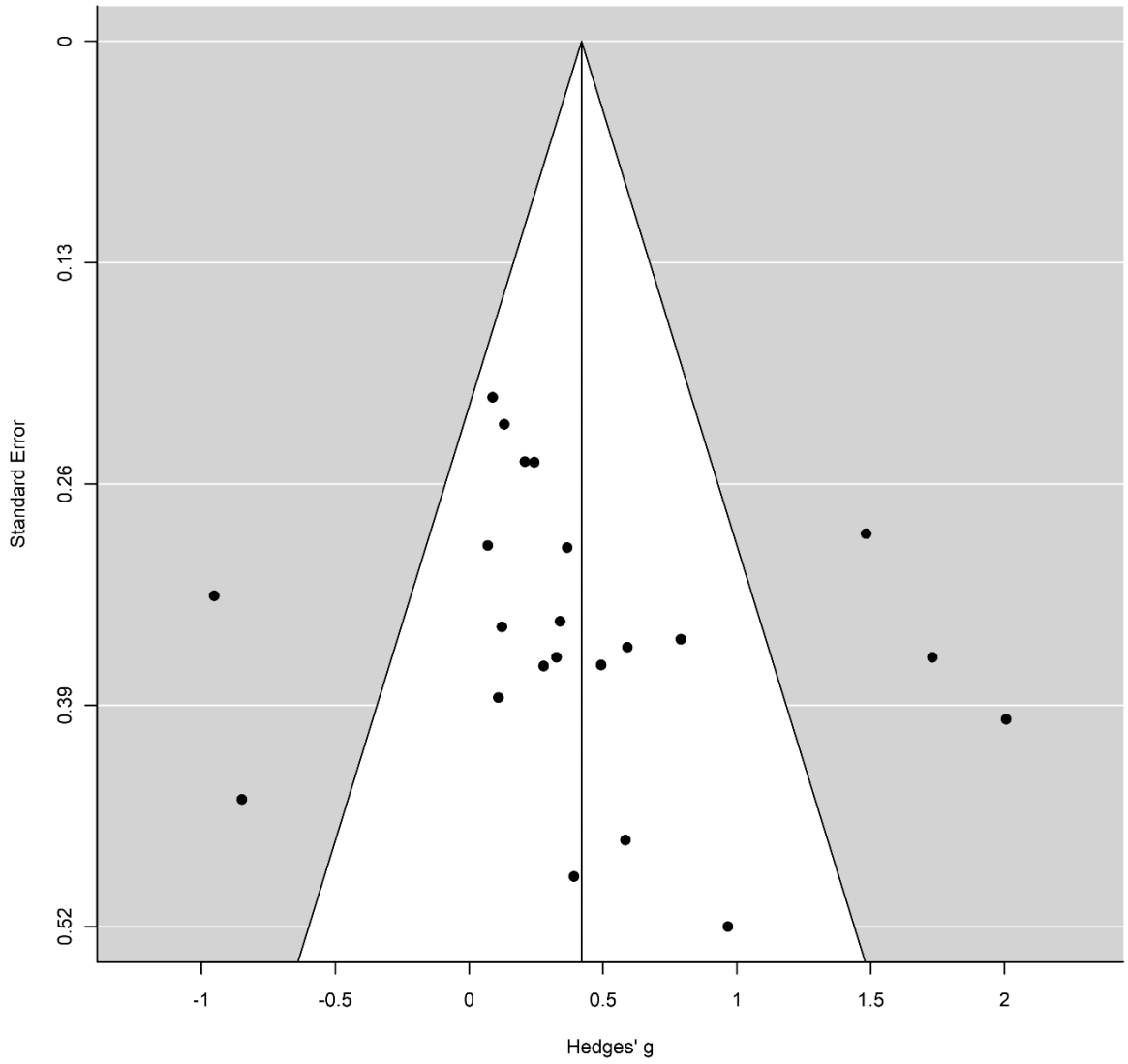



Figure 4. Publication bias assessed by funnel plot for global cognition.

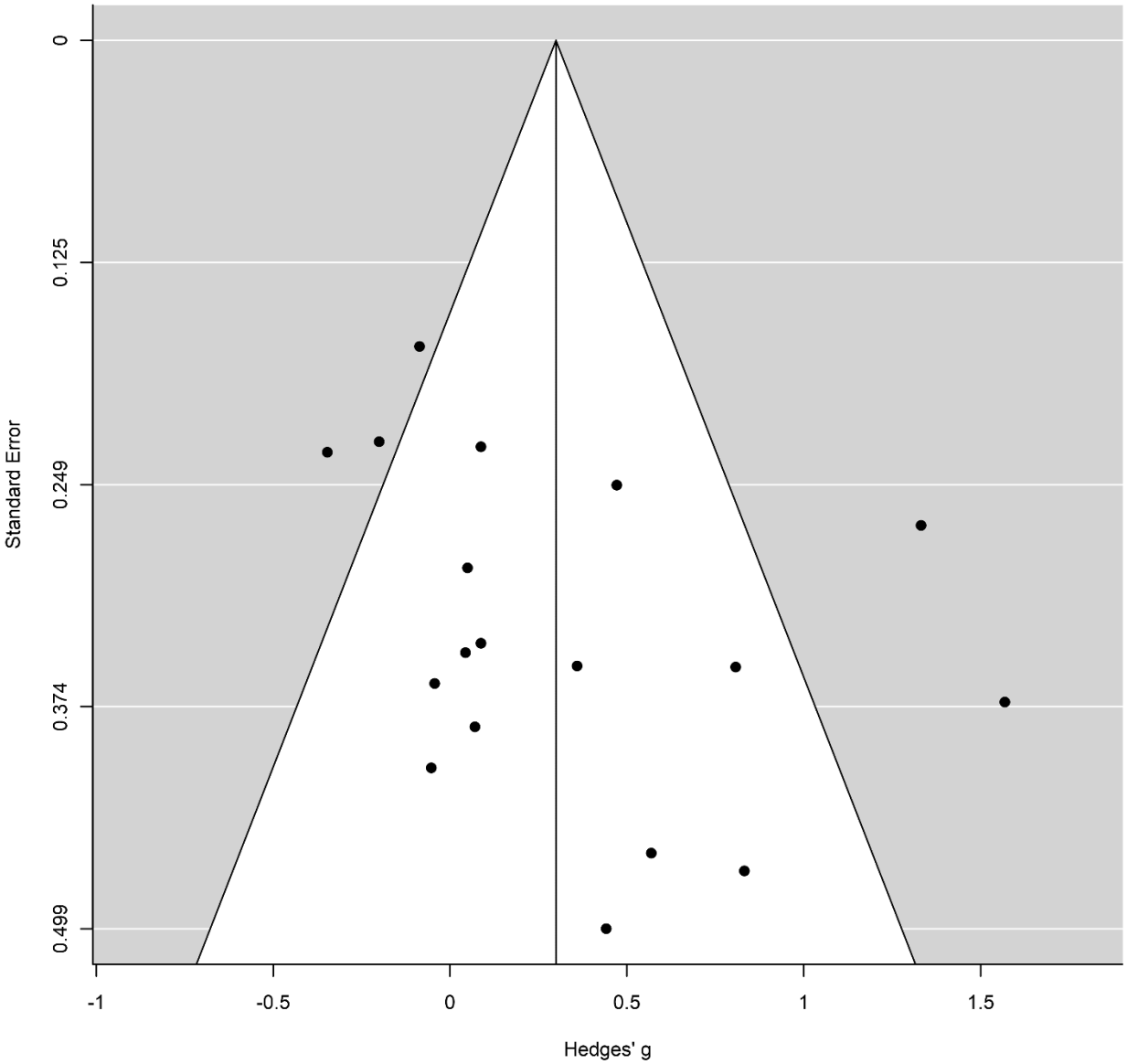



Figure 5. Publication bias assessed by funnel plot for executive function.

Supplementary Material S1

Appendix 1: PRISMA 2020 Checklist

Appendix 2: Search strategy for Ovid MEDLINE database

Appendix 3: Priority lists of measurements for global cognition and executive function

Appendix 4: A list of excluded studies (n=30) with justification for exclusion

Appendix 5: Tables of risk of bias for global cognition and executive function

## Appendix 1.

PRISMA 2020 Checklist

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review. | Title |
| **ABSTRACT** | | | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | JAPA abstract |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | Introduction |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Introduction |
| **METHODS** | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Methods |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Methods |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Supplementary Material S1 (Appendix 2) |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Methods |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Methods |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Methods |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | Methods, Table 1 |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Methods |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Methods |

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Methods |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Methods |
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Methods |
| | 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Methods |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | Methods |
| | 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | Methods |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Methods |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | Methods |
| **RESULTS** | | | |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | Results (Figure 1) |
| | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | Results, Supplementary Material S1 (Appendix 4) |
| Study characteristics | 17 | Cite each included study and present its characteristics. | Results (Table 1) |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | Supplementary Material S1 (Appendix 5) |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | Results |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Results |
| | 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | Results |
| | 20c | Present results of all investigations of possible causes of heterogeneity among study results. | Results |
| | 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | Results, Supplementary Material S2 (Appendix 2) |

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | Supplementary Material S1 (Appendix 5) |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | Results (Table 2) |
| **DISCUSSION** | | | |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | Discussion |
| | 23b | Discuss any limitations of the evidence included in the review. | Discussion |
| | 23c | Discuss any limitations of the review processes used. | Discussion |
| | 23d | Discuss implications of the results for practice, policy, and future research. | Discussion, Conclusions |
| **OTHER INFORMATION** | | | |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | Methods |
| | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | / |
| | 24c | Describe and explain any amendments to information provided at registration or in the protocol. | / |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | Acknowledgements |
| Competing interests | 26 | Declare any competing interests of review authors. | Conflict of interest |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | / |

PRISMA 2020 Checklist from: Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71

## Appendix 2.

Search strategy for Ovid MEDLINE database

1 Randomized Controlled trial.pt.
2 Controlled clinical trial.pt.
3 (Randomized or Randomised).ab.
4 Placebo.ab.
5 clinical trials as topic.sh.
6 randomly.ab.
7 trial.ti.
8 OR/1-7
9 rehabilitee*.mp.
10 Therapist/ or therapist*.mp.
11 exp Disabled Persons/ or disabled person.mp.
12 Caregivers/ or caregiver*.mp.
13 exp Rehabilitation/ or rehab*.mp.
14 exp Exercise/ or exercise.mp.
15 exp Exercise therapy/ or exercise therapy.mp.
16 therapeutic exercise.mp.
17 (Physical therap* or Physiotherap*).mp.
18 exp Physical Therapy Modalities/ or physical therapy modalities.mp.
19 physical rehabilitation.mp.
20 exp Occupational Therapy/ or occupational therap*.mp.
21 exp "Rehabilitation of Speech and Language Disorders"/
22 exp Speech Therapy/ or speech therap*.mp.
23 speech-language therap*.mp.
24 logoped*.mp.
25 exp Audiology/ or audiolog*.mp.
26 exp Sign Language/
27 exp Psychotherapy/ or psychotherap*.mp.
28 Neuropsychotherap*.mp.
29 exp Neuropsycology/ or neuropsychol*.mp.
30 (riding therap* or equine facilitated therap* or hippotherap* or horse riding therap* or horse back riding therap*).mp.
31 exp Dance Therapy/ or dance therap*.mp.
32 exp Music Therapy/ or music therap*.mp.
33 exp Art Therapy/ or art therap*.mp.
34 exp Optometry/ or optomet*.mp.
35 exp Orthoptics/ or orthoptic*.mp.
36 orthotic*.mp.
37 orthopedic techn*.mp.
38 exp Podiatry/ or podiat*.mp.
39 exp "Physical Education and Training"/ or physical education*.mp.
40 mobility special*.mp.
41 Rehabilitation Nursing/ or rehabilitation nurs*.mp.
42 (practical nurs* or practice nurs*).mp.
43 (asthma nurs* or respiratory nurs*).mp.
44 (diabetes nurs* or diabetes specialist nurs*).mp.
45 (geriatric nurs* or gerontological nurs* or gerontology nurs*).mp.
46 (sexual health therap* or sexual therap*).mp.
47 exp Sexology/
48 Exp Nutritionists/ or nutritionist.mp.
49 leisure activit*.mp.
50 play therap*.mp.
51 (drama therap* or psychodrama therap*).mp.
52 psychodram*.mp.
53 creative art therap*.mp.
54 (expression skills or expressive art therap*).mp.
55 (youth counselor* or youth leader*).mp.
56 OR/9-55
57 exp Augmented Reality/ or augmented realit*.mp.
58 exp Virtual Reality/ or virtual realit*.mp.
59 exp Video Games/
60 (video gam* or videogam*).mp.
61 serious gam*.mp.
62 exergam*.mp.
63 kinect*.mp.
64 nintendo*.mp.
65 (play station* or playstation*).mp.
66 wii*.mp.
67 xbox*.mp.
68 avatar*.mp.
69 OR/57-68
70 8 AND 56 AND 69
71 animal/
72 human/
73 71 NOT (71 AND 72)
74 70 NOT 73

## Appendix 3.

Priority lists of measurements at outcome level

**Global cognition**

1. Montreal Cognitive Assessment (MoCA)
   - high sensitivity (90 %) and specificity (87 %) for detecting MCI [1]
   - lower ceiling effect compared to MMSE [2]
2. The Addenbrooke Cognitive Examination (ACE)
   - valid and reliable (alpha coefficient 0.8) dementia screening test that is sensitive to early cognitive dysfunction [3]
3. The Computer Assessment of Mild Cognitive Impairment (CAMCI)
   - highly sensitive (86 %) and specific (94 %) for the identification of MCI among community-dwelling older adults [4]
4. Cognitive Telephone Screening Instrument (COGTEL)
   - reliable (test-retest reliability 0.85) and valid instrument for assessing cognitive function with inter-individual differences and performance level among healthy older adults [5], MCI and dementia [6]
5. The Loewenstein Occupational Therapy Cognitive Assessment – Geriatric (LOTCA-G)
   - valid and reliable version of a LOTCA-test for geriatric population distinguishing differences in cognitive performance between mild or moderate dementia and people with normal level of cognition among older adults 70-90 years old [7]
6. The Cognitive Self-Report Questionnaire 25 (CSRQ)
   - a self-report questionnaire [8]
   - found to be an appropriate measure for assessing hearing, cognition, and auditory processing among older adults with or without probable MCI [9]
7. Modified Mini-Mental State Examination (3MS)
   - developed to acknowledge the limitations of MMSE in recognizing mild dementia [10*]
   - found to be superior to MMSE with higher alpha internal consistency (0.87) though not sensitive for detecting MCI [10]
8. The Clock-Drawing Test (CDT)
   - quick and easy test for assessing global cognition [11], however scoring and direction of the test can vary [12, 13]
   - suites better for screening moderate or severe dementia as the test is not sensitive for detecting mild dementia [14]
   - validity acceptable compared with MMSE [15]
9. Mini-Mental State Examination (MMSE)
   - high ceiling effect lowering the sensitivity to detect differences in cognitive performance among healthy older adults and ones with MCI [16, 17]

**Executive function**

1. Trail Making Test, Part B (TMT-B)
    - widely used measure for executive function that is included in the Halstead-Reitan Neuropsychological Battery [18, 19]
2. Stroop Color Word Test (SCWT, WCT, Stroop C, VST)
    - commonly known measure of attention and inhibition [20] that is also suitable for detecting mild dementia [21]
3. Digit Span Backward (DSB)
    - usable measure especially with major cognitive impairment [22]
    - performance in DSB strongly related to gray matter volume of the brain among healhty older adults [23]
4. Symbol Digit Substitution Test (SDST, DSST, DSC)
    - valid, reliable and sensitive measure for cognitive dysfunction with low impact of language, culture and education on test performance [24]
    - apparently measures several cognitive domains such as executive function
5. Color Trails Test (CTT)
    - developed as equivalent version of TMT test with universal sign language symbols instead of cultural-based language [25], but some distinction has been found between CTT-2 and TMT-B [26]
    - age and education level seem to influence on CTT results [25]
6. Executive Interview (EXIT-25)
    - a measure with adequate reliability that is often used for measuring executive function, although studies show its usability also for global cognition [27]
    - correlates well with other measurements of executive function [28]
7. CogState Battery
    - game-like computerized test battery with several cognitive tests [29]
    - valid and reliable test in detecting cognitive decline among healthy older adults, adults with aMCI or Alzheimer's disease [30]
    - learning is shown to affect results in some tests [16, 29, 31, 32]
8. Cognitive Estimation Test (CET)
    - includes both quantitative and qualitative questions rated from 0 to 3 [33**]
    - sensitivity uncertain
9. The Frontal Assessment Battery (FAB)
    - test-retest and interrater reliabilities found to be satisfactory in screening of global executive dysfunction [34]
10. Rule Shift Cards Test (RSCT)
    - a promising measure of executive function, though test-retest reliability has proven to be weak [35]
11. The Dual Task Performance (DTP)
    - measures performance of two tasks simultaneously and division of attention [36, 37]
    - apparently no standardized DTP-measure available
    - paper-pencil assessment of DTP identifies impairment in Alzheimer's disease but sensitivity to recognize changes in cognitive performance between healthy and MCI is not clear [38]

* Original reference according to McDowell et al. 1997: Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *The Journal of clinical psychiatry*, *48*(8), 314–318.

** Original reference according to Wagner et al. 2011: Shallice, T., & Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex; a journal devoted to the study of the nervous system and behavior*, *14*(2), 294–303. https://doi.org/10.1016/s0010-9452(78)80055-0

References in the outcome lists


1. Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x
2. Trzepacz, P. T., Hochstetler, H., Wang, S., Walker, B., Saykin, A. J., & Alzheimer's Disease Neuroimaging Initiative (2015). Relationship between the Montreal Cognitive Assessment and Mini-mental State Examination for assessment of mild cognitive impairment in older adults. *BMC geriatrics*, *15*, 107. https://doi.org/10.1186/s12877-015-0103-3
3. Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International journal of geriatric psychiatry*, *21*(11), 1078–1085. https://doi.org/10.1002/gps.1610
4. Saxton, J., Morrow, L., Eschman, A., Archer, G., Luther, J., & Zuccolotto, A. (2009). Computer assessment of mild cognitive impairment. *Postgraduate medicine*, *121*(2), 177–185. https://doi.org/10.3810/pgm.2009.03.1990
5. Ihle, A., Gouveia, É. R., Gouveia, B. R., & Kliegel, M. (2017). The Cognitive Telephone Screening Instrument (COGTEL): A Brief, Reliable, and Valid Tool for Capturing Interindividual Differences in Cognitive Functioning in Epidemiological and Aging Studies. *Dementia and geriatric cognitive disorders extra*, *7*(3), 339–345. https://doi.org/10.1159/000479680
6. Alexopoulos, P., Skondra, M., Charalampopoulou, M., Aligianni, S., Kontogianni, E., Lentzari, I., Vratsista, A., Kliegel, M., & Politis, A. (2020). Validation of the Cognitive Telephone Screening Instrument (COGTEL) for detecting mild cognitive impairment and dementia due to Alzheimer's disease (AD). Neuropsychology/early detection of cognitive decline with neuropsychological tests. Poster presentation in Alzheimer's & Dementia, 16 (Suppl. 6), e041461. https://doi.org/10.1002/alz.041461.
7. Erez, A. B., & Katz N. (2004). Cognitive Profiles of Individuals with Dementia and Healthy Elderly: The Loewenstein Occupational Therapy Cognitive Assessment (LOTCA-G). *Physical & Occupational Therapy In Geriatrics, 22*(2), 29-42. https://doi.org/10.1080/J148v22n02_03
8. O'Brien, J. L., Lister, J. J., Fausto, B. A., Clifton, G. K., & Edwards, J. D. (2017). Cognitive Training Enhances Auditory Attention Efficiency in Older Adults. *Frontiers in aging neuroscience*, *9*, 322. https://doi.org/10.3389/fnagi.2017.00322
9. Fausto, B. A., Badana, A. N. S., Arnold, M. L., Lister, J. J., & Edwards, J. D. (2018). Comparison of Subjective and Objective Measures of Hearing, Auditory Processing, and Cognition Among Older Adults With and Without Mild Cognitive Impairment. *Journal of speech, language, and hearing research : JSLHR*, *61*(4), 945–956. https://doi.org/10.1044/2017_JSLHR-H-17-0263
10. McDowell, I., Kristjansson, B., Hill, G. B., & Hébert, R. (1997). Community screening for dementia: the Mini Mental State Exam (MMSE) and Modified Mini-Mental State Exam (3MS) compared. *Journal of clinical epidemiology*, *50*(4), 377–383. https://doi.org/10.1016/s0895-4356(97)00060-7
11. Ehreke, L., Luppa, M., König, H. H., & Riedel-Heller, S. G. (2010). Is the Clock Drawing Test a screening tool for the diagnosis of mild cognitive impairment? A systematic review. *International psychogeriatrics*, *22*(1), 56–63. https://doi.org/10.1017/S1041610209990676
12. Agrell, B., & Dehlin, O. (1998). The clock-drawing test. *Age and ageing*, *27*(3), 399-404.
13. Shulman K. I. (2000). Clock-drawing: is it the ideal cognitive screening test? *International journal of geriatric psychiatry*, *15*(6), 548–561. https://doi.org/10.1002/1099-1166(200006)15:6<548::aid-gps242>3.0.co;2-u
14. Pinto, E., & Peters, R. (2009). Literature review of the Clock Drawing Test as a tool for cognitive screening. *Dementia and geriatric cognitive disorders*, *27*(3), 201–213. https://doi.org/10.1159/000203344
15. Nishiwaki, Y., Breeze, E., Smeeth, L., Bulpitt, C. J., Peters, R., & Fletcher, A. E. (2004). Validity of the Clock-Drawing Test as a screening tool for cognitive impairment in the elderly. *American journal of epidemiology*, *160*(8), 797–807. https://doi.org/10.1093/aje/kwh288
16. de Jager, C. A., Schrijnemaekers, A. C., Honey, T. E., & Budge, M. M. (2009). Detection of MCI in the clinic: evaluation of the sensitivity and specificity of a computerised test battery, the Hopkins Verbal Learning Test and the MMSE. *Age and ageing*, *38*(4), 455–460. https://doi.org/10.1093/ageing/afp068
17. Hoops, S., Nazem, S., Siderowf, A. D., Duda, J. E., Xie, S. X., Stern, M. B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*, *73*(21), 1738–1745. https://doi.org/10.1212/WNL.0b013e3181c34b47
18. Horton, A. M. (2008). The Halstead-Reitan neuropsychological test battery: past, present, and future. In A. M. Horton & D. Wedding (edit.), The neuropsychology handbook, 3rd version. New York: Springer Publishing Company, 251–278.
19. Llinàs-Reglà, J., Vilalta-Franch, J., López-Pousa, S., Calvó-Perxas, L., Torrents Rodas, D., & Garre-Olmo, J. (2017). The Trail Making Test: Association with other neuropsychological measures and normative values for adults aged 55

years and older from a Spanish-speaking population-based sample. *Assessment*, *24*(2), 183–196. https://doi.org/10.1177/1073191115602552
20. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643. https://doi.org/10.1037/h0054651
21. Bondi, M. W., Serody, A. B., Chan, A. S., Eberson-Shumate, S. C., Delis, D. C., Hansen, L. A., & Salmon, D. P. (2002). Cognitive and neuropathologic correlates of Stroop Color-Word Test performance in Alzheimer's disease. *Neuropsychology, 16*(3), 335–343. https://doi.org/10.1037/0894-4105.16.3.335
22. Leung, J. L., Lee, G. T., Lam, Y. H., Chan, R. C., & Wu, J. Y. (2011). The use of the Digit Span Test in screening for cognitive impairment in acute medical inpatients. *International psychogeriatrics*, *23*(10), 1569–1574. https://doi.org/10.1017/S1041610211000792
23. Ruscheweyh, R., Deppe, M., Lohmann, H., Wersching, H., Korsukewitz, C., Duning, T., Bluhm, S., Stehling, C., Keller, S. S., & Knecht, S. (2013). Executive performance is related to regional gray matter volume in healthy older individuals. *Human brain mapping*, *34*(12), 3333–3346. https://doi.org/10.1002/hbm.22146
24. Jaeger J. (2018). Digit Symbol Substitution Test: The Case for Sensitivity Over Specificity in Neuropsychological Testing. *Journal of clinical psychopharmacology*, *38*(5), 513–519. https://doi.org/10.1097/JCP.0000000000000941
25. Lee, T. M., & Chan, C. C. (2000). Are trail making and color trails tests of equivalent constructs? *Journal of clinical and experimental neuropsychology*, *22*(4), 529–534. https://doi.org/10.1076/1380-3395(200008)22:4;1-0;FT529
26. Dugbartey, A. T., Townes, B. D., & Mahurin, R. K. (2000). Equivalence of the Color Trails Test and Trail Making Test in nonnative English-speakers. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, *15*(5), 425–431. https://doi.org/10.1093/arclin/15.5.425
27. Campbell, G. B., Whyte, E. M., Sereika, S. M., Dew, M. A., Reynolds, C. F., 3rd, & Butters, M. A. (2014). Reliability and validity of the Executive Interview (EXIT) and Quick EXIT among community dwelling older adults. *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry*, *22*(12), 1444–1451. https://doi.org/10.1016/j.jagp.2013.08.009
28. Matioli, M. N. P. S., Caramelli, P., Marques, B. D., da Rocha, F. D., de Castro, M. C. C., Yamashita, S. R., & Soares, A. M. (2008). EXIT25 - Executive interview applied to a cognitively healthy elderly population with heterogeneous educational background. *Dementia & neuropsychologia*, *2*(4), 305–309. https://doi.org/10.1590/S1980-57642009DN20400013
29. Darby, D. G., Fredrickson, J., Pietrzak, R. H., Maruff, P., Woodward, M., & Brodtmann, A. (2014). Reliability and usability of an internet-based computerized cognitive testing battery in community-dwelling older people. *Computers in Human Behavior*, *30*, 199-205. https://doi.org/10.1016/j.chb.2013.08.009
30. Lim, Y. Y., Jaeger, J., Harrington, K., Ashwood, T., Ellis, K. A., Stöffler, A., Szoeke, C., Lachovitzki, R., Martins, R. N., Villemagne, V. L., Bush, A., Masters, C. L., Rowe, C. C., Ames, D., Darby, D. & Maruff, P. 2013. Lim, Y. Y., Jaeger, J., Harrington, K., Ashwood, T., Ellis, K. A., Stöffler, A., Szoeke, C., Lachovitzki, R., Martins, R. N., Villemagne, V. L., Bush, A., Masters, C. L., Rowe, C. C., Ames, D., Darby, D., & Maruff, P. (2013). Three-month stability of the CogState brief battery in healthy older adults, mild cognitive impairment, and Alzheimer's disease: results from the Australian Imaging, Biomarkers, and Lifestyle-rate of change substudy (AIBL-ROCS). *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, *28*(4), 320–330. https://doi.org/10.1093/arclin/act021
31. Fredrickson, J., Maruff, P., Woodward, M., Moore, L., Fredrickson, A., Sach, J., & Darby, D. (2010). Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Neuroepidemiology*, *34*(2), 65–75. https://doi.org/10.1159/000264823
32. Maruff, P., Lim, Y. Y., Darby, D., Ellis, K. A., Pietrzak, R. H., Snyder, P. J., Bush, A. I., Szoeke, C., Schembri, A., Ames, D., Masters, C. L., & AIBL Research Group (2013). Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC psychology*, *1*(1), 30. https://doi.org/10.1186/2050-7283-1-30
33. Wagner, G. P., MacPherson, S. E., Parente, M. A., & Trentini, C. M. (2011). Cognitive estimation abilities in healthy and clinical populations: the use of the Cognitive Estimation Test. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *32*(2), 203–210. https://doi.org/10.1007/s10072-010-0449-4
34. Appollonio, I., Leone, M., Isella, V., Piamarta, F., Consoli, T., Villa, M. L., Forapani, E., Russo, A., & Nichelli, P. (2005). The Frontal Assessment Battery (FAB): normative values in an Italian population sample. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *26*(2), 108–116. https://doi.org/10.1007/s10072-005-0443-4
35. Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological rehabilitation*, *8*(3), 213-228. https://doi.org/10.1080/713755570
36. Della Sala, S., Logie, R. H., & Spinnler, H. (1992). Is primary memory deficit of Alzheimer patients due to a “central executive” impairment? *Journal of neurolinguistics*, *7*(4), 325-346. https://doi.org/10.1016/0911-6044(92)90022-O

37. Della Sala, S., Baddeley, A., Papagno, C., & Spinnler, H. (1995). Dual-task paradigm: a means to examine the central executive. *Annals of the New York Academy of Sciences*, *769*, 161–171. https://doi.org/10.1111/j.1749-6632.1995.tb38137.x
38. Foley, J. A., Kaschel, R., Logie, R. H., & Della Sala, S. (2011). Dual-task performance in Alzheimer's disease, mild cognitive impairment, and normal ageing. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, *26*(4), 340–348. https://doi.org/10.1093/arclin/acr032

## Appendix 4.

A list of excluded studies (n=30) with justification for exclusion.

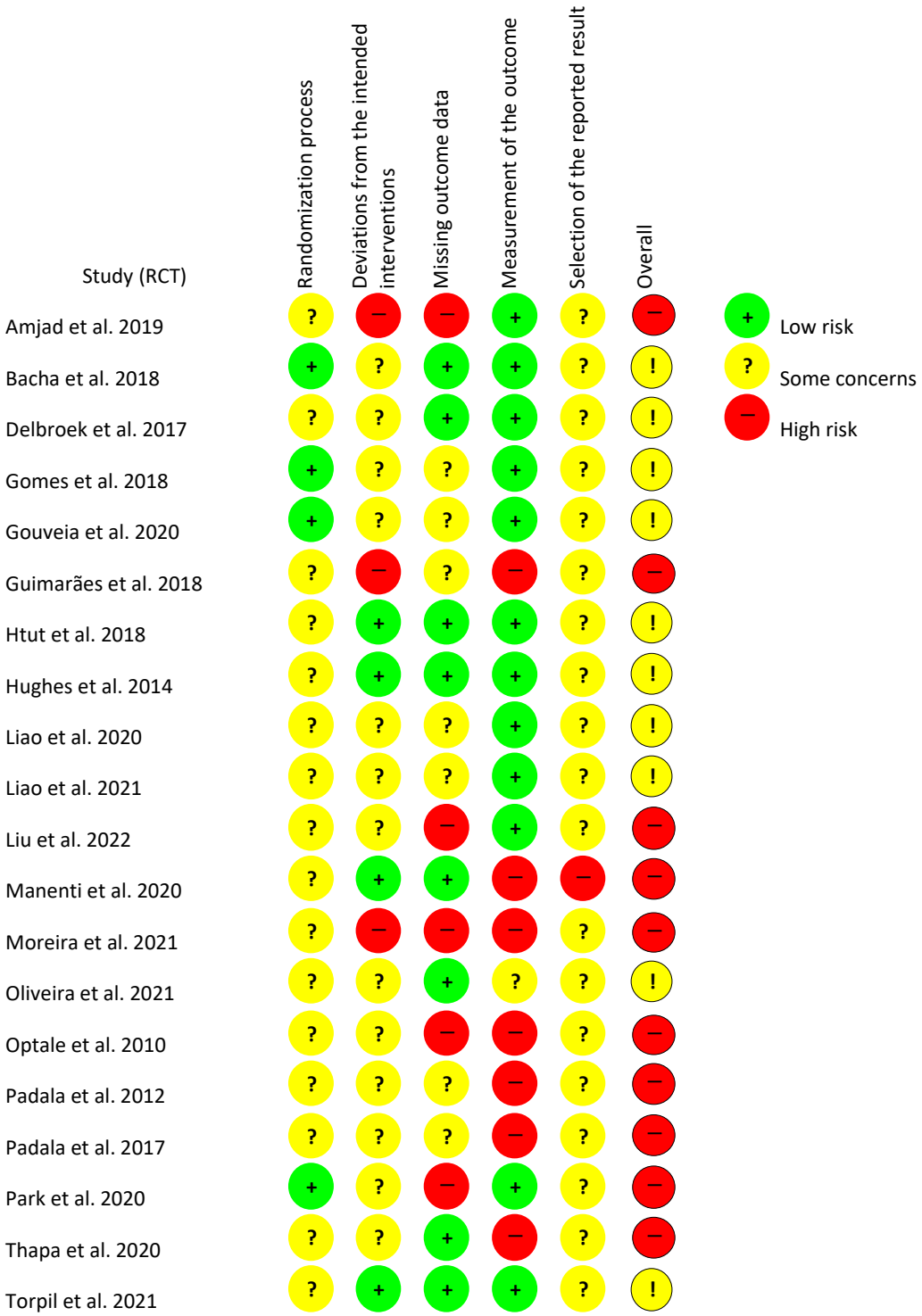
| Study | Title | Reason for exclusion |
|---|---|---|
| Adcock et al. 2020 | Effects of an in-home multicomponent exergame training on physical functions, cognition, and brain volume of older adults: a randomized controlled trial. | Wrong comparator |
| Anderson-Hanley et al. 2017 | Neuropsychological benefits of neuro-exergaming for older adults: a pilot study of an interactive physical and cognitive exercise system (iPACES). | Wrong comparator |
| Anderson-Hanley et al. 2018 | The aerobic and cognitive exercise study (ACES) for community-dwelling older adults with or at-Risk for mild cognitive impairment (MCI): neuropsychological, neurobiological and neuroimaging outcomes of a randomized clinical trial. | Wrong comparator |
| Barcelos et al. 2015 | Aerobic and cognitive exercise (ACE) pilot study for older adults: executive function improves with cognitive challenge while exergaming. | Wrong comparator |
| Belchior et al. 2019 | Computer and videogame interventions for older adults' cognitive and everyday functioning. | Wrong intervention |
| Hsieh et al. 2014 | Virtual reality system based on Kinect for the elderly in fall prevention. | Wrong outcome |
| Huang et al. 2020 | Exergaming executive functions: an immersive virtual reality-based cognitive training for adults aged 50 and older. | Wrong comparator |
| Kang et al. 2021 | Effect of Cognitive Training in Fully Immersive Virtual Reality on Visuospatial Function and Frontal-Occipital Functional Connectivity in Predementia: Randomized Controlled Trial. | Wrong comparator |
| Karssemeijer et al. 2019 | Exergaming as a physical exercise strategy reduces frailty in people with dementia: a randomized controlled trial. | Wrong outcome |
| Kim et al. 2021 | Effects of ICT-Based Multicomponent Program on Body Composition and Cognitive Function in Older Adults: A Randomized Controlled Clinical Study. | Wrong comparator |
| Liu et al. 2021 | Study on Adjuvant Medication for Patients with Mild Cognitive Impairment Based on VR Technology and Health Education. | Wrong patient group |
| Maillot et al. 2012 | Effects of interactive physical-activity video-game training on physical and cognitive function in older adults. | Wrong comparator |
| Man et al.2012 | Evaluation of a virtual reality-based memory training programme for Hong Kong Chinese older adults with questionable dementia: A pilot study. | Wrong outcome |
| McCord et al. 2020 | Short video game play improves executive function in the oldest old living in residential care. | Wrong intervention |
| Micarelli et al. 2019 | Vestibular rehabilitation in older adults with and without mild cognitive impairment: Effects of virtual reality using a head-mounted display. | Wrong outcome |
| Mirelman et al. 2016 | Addition of a non-immersive virtual reality component to treadmill training to reduce fall risk in older adults (V-TIME): a randomised controlled trial. | Wrong outcome |
| Monteiro-Junior et al. 2017 | Virtual reality-based physical exercise with exergames (PhysEx) improves mental and physical health of institutionalized older adults. | Wrong study design |

| | | |
|---|---|---|
| Mrakic-Sposta et al. 2018 | Effects of combined physical and cognitive virtual reality-based training on cognitive impairment and oxidative stress in MCI patients: A pilot study. | Wrong comparator |
| Ordnung et al. 2017 | No overt effects of a 6-week exergame training on sensorimotor and cognitive function in older adults. A preliminary investigation. | Wrong comparator |
| Padala et al. 2017 | Efficacy of Wii-Fit on static and dynamic balance in community dwelling older veterans: a randomized controlled pilot trial. | Wrong comparator |
| Park & Park 2018 | Does cognition-specific computer training have better clinical outcomes than non-specific computer training? A single-blind, randomized controlled trial. | Wrong comparator |
| Park Eunhee et al. 2019 | Effects of a mixed reality-based cognitive training system compared to a conventional computer-assisted cognitive training system on mild cognitive impairment: A pilot study. | Wrong comparator |
| Park Jin-Hyuck 2020 | Effects of virtual reality-based spatial cognitive training on hippocampal function of older adults with mild cognitive impairment. | Wrong comparator |
| Park Jin-Hyuck 2022 | Does the virtual shopping training improve executive function and instrumental activities of daily living of patients with mild cognitive impairment? | Wrong comparator |
| Park Jong-Hwan et al. 2020 | Feasibility and tolerability of a culture-based virtual reality (VR) training program in patients with mild cognitive impairment: a randomized controlled pilot study. | Wrong comparator |
| Rica et al. 2020 | Effects of a Kinect-based physical training program on body composition, functional fitness and depression in institutionalized older adults. | Wrong outcome |
| van Santen et al. 2021 | Cost-effectiveness of exergaming compared to regular day-care activities in dementia: results of a randomised controlled trial in the Netherlands. | Wrong outcome |
| Schwenk et al. 2016 | Sensor-based balance training with motion feedback in people with mild cognitive impairment. | Wrong comparator |
| Taylor et al. 2018 | Exergames to improve the mobility of long-term care residents: a cluster randomized controlled trial. | Wrong outcome |
| Wittelsberger et al. 2013 | The influence of Nintendo-Wii® bowling upon residents of retirement homes. | Wrong patient group |

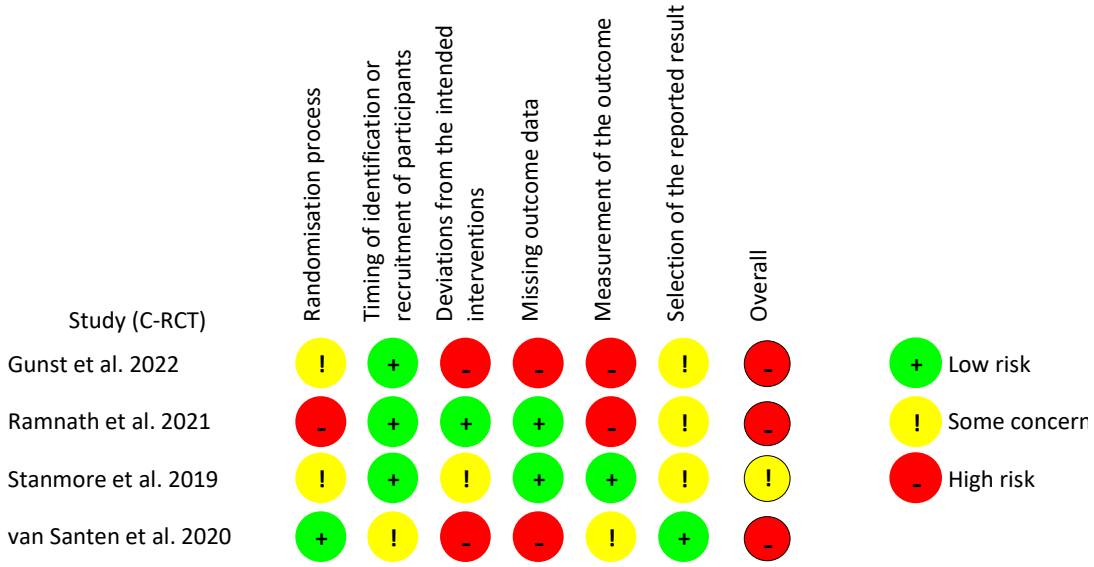
**Appendix 5.**

Tables of risk of bias for RCT and C-RCT studies in (a) global cognition and (b) executive function.

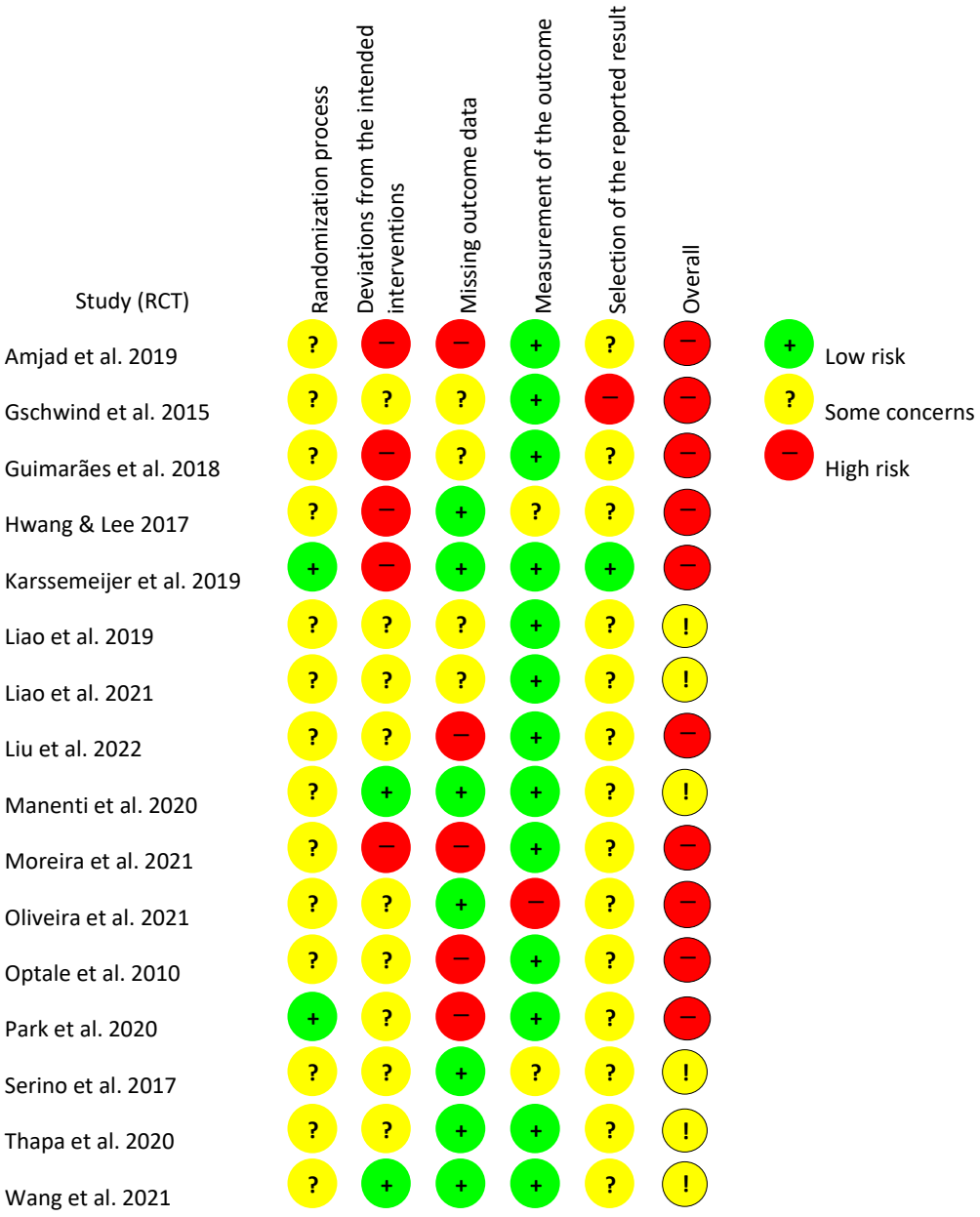
(a) Global cognition with RCT studies:



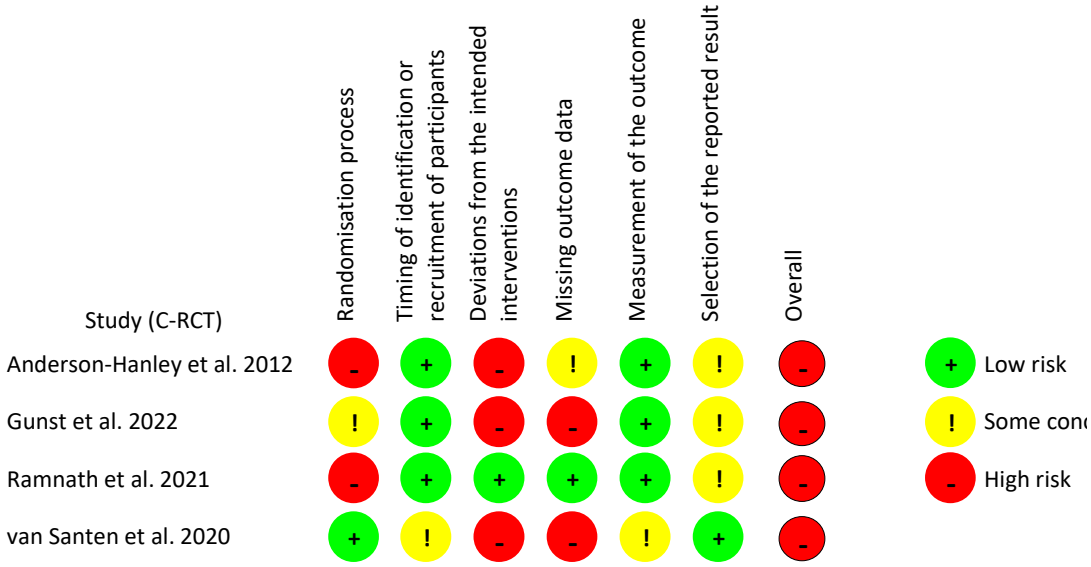
Global cognition with C-RCT studies:



(b) Executive function with RCT studies:



Executive function with C-RCT studies:



Supplementary Material S2

Appendix 1: Subgroup analyses

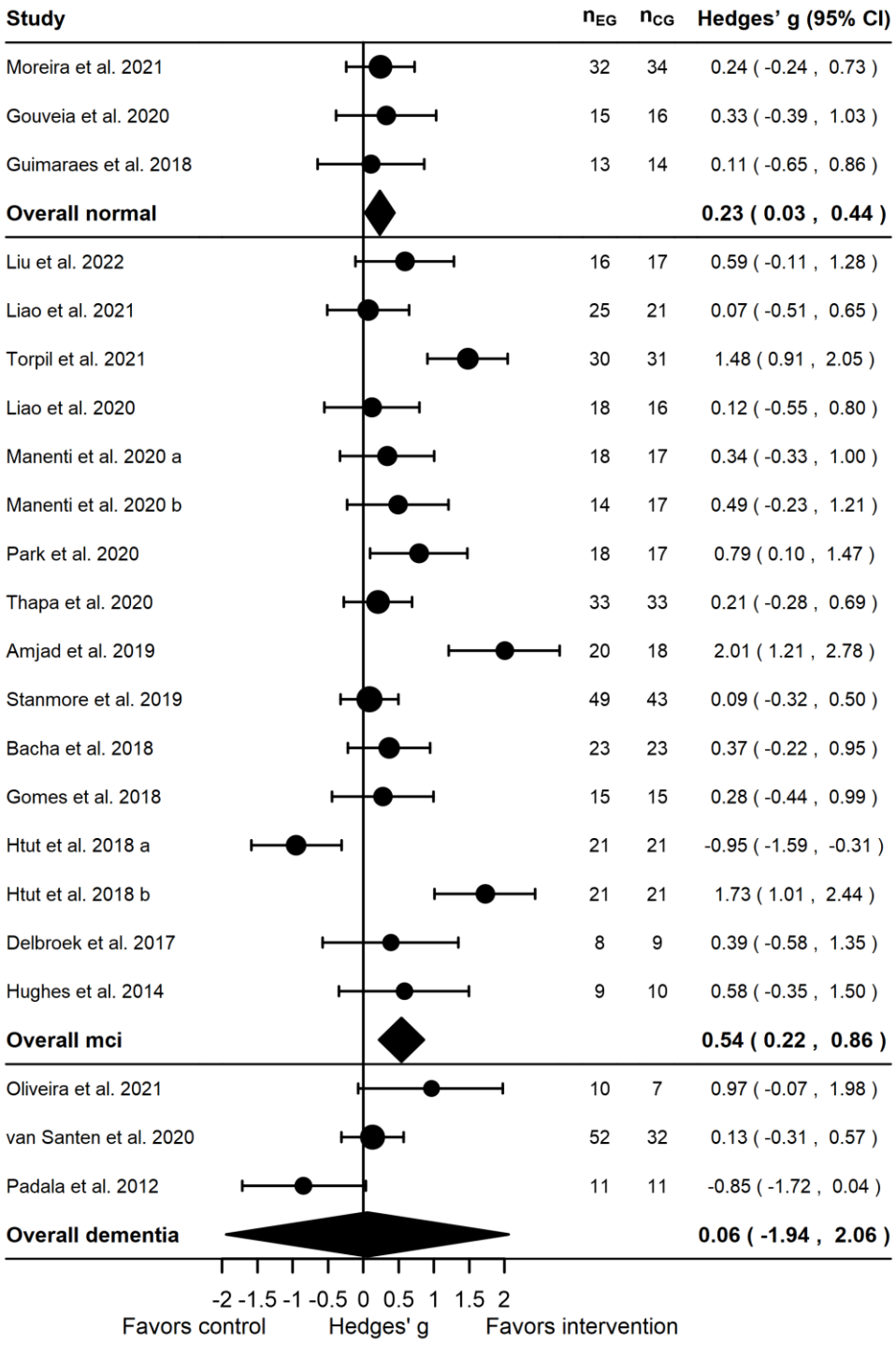
Appendix 2: Sensitivity analyses

Appendix 3: Narrative synthesis of results

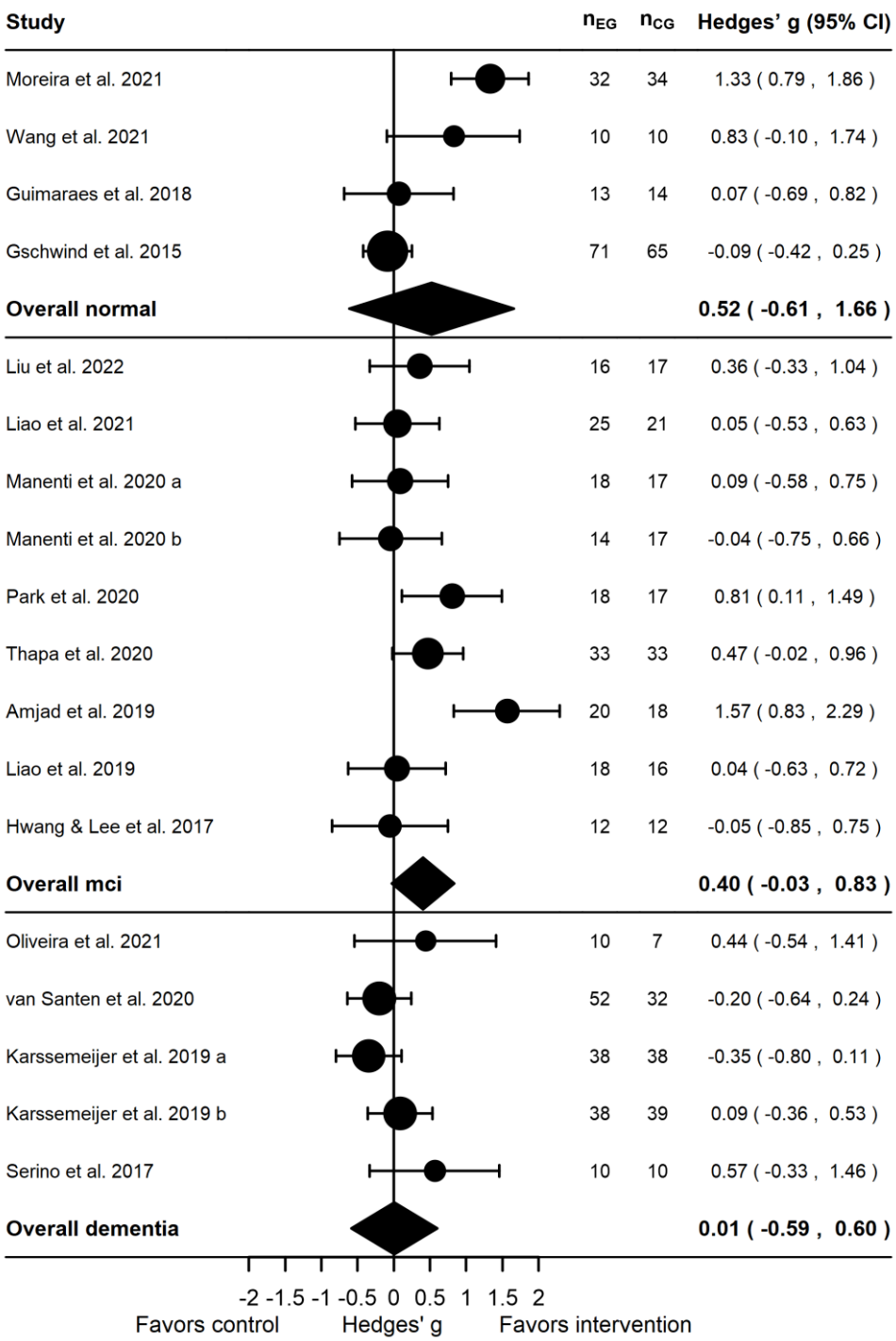
## Appendix 1

Subgroup analyses for (a) global cognition and (b) executive function. The size of the circle in Hedges' g value shows the emphasis of an individual study in the pooled estimated effect.

(a)



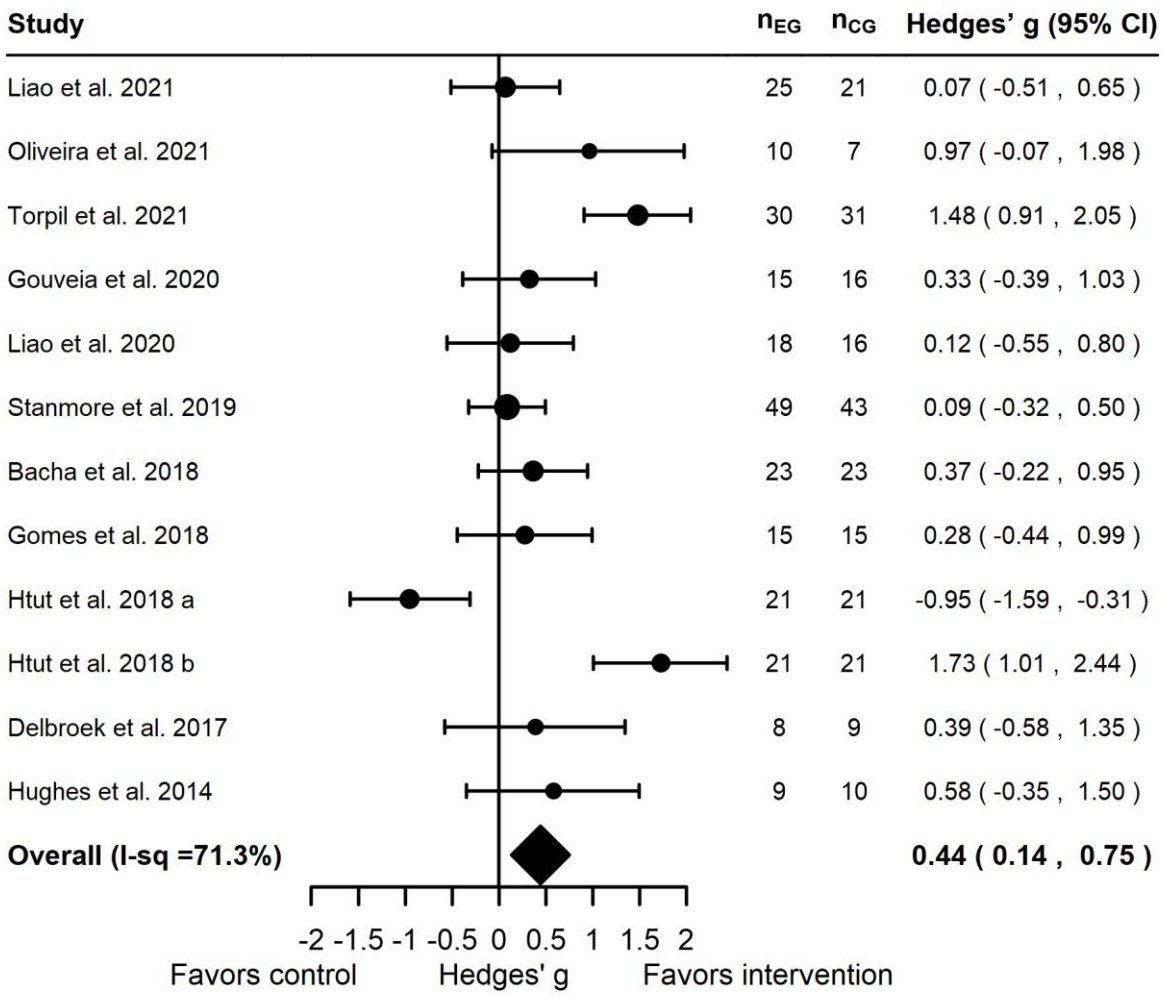
(b)


Study
$n_{EG}$
$n_{CG}$
Hedges' g (95% CI)
Moreira et al. 2021 32 34 1.33 ( 0.79 , 1.86 )
Wang et al. 2021 10 10 0.83 ( -0.10 , 1.74 )
Guimaraes et al. 2018 13 14 0.07 ( -0.69 , 0.82 )
Gschwind et al. 2015 71 65 -0.09 ( -0.42 , 0.25 )
Overall normal 0.52 ( -0.61 , 1.66 )
Liu et al. 2022 16 17 0.36 ( -0.33 , 1.04 )
Liao et al. 2021 25 21 0.05 ( -0.53 , 0.63 )
Manenti et al. 2020 a 18 17 0.09 ( -0.58 , 0.75 )
Manenti et al. 2020 b 14 17 -0.04 ( -0.75 , 0.66 )
Park et al. 2020 18 17 0.81 ( 0.11 , 1.49 )
Thapa et al. 2020 33 33 0.47 ( -0.02 , 0.96 )
Amjad et al. 2019 20 18 1.57 ( 0.83 , 2.29 )
Liao et al. 2019 18 16 0.04 ( -0.63 , 0.72 )
Hwang & Lee et al. 2017 12 12 -0.05 ( -0.85 , 0.75 )
Overall mci 0.40 ( -0.03 , 0.83 )
Oliveira et al. 2021 10 7 0.44 ( -0.54 , 1.41 )
van Santen et al. 2020 52 32 -0.20 ( -0.64 , 0.24 )
Karssemeijer et al. 2019 a 38 38 -0.35 ( -0.80 , 0.11 )
Karssemeijer et al. 2019 b 38 39 0.09 ( -0.36 , 0.53 )
Serino et al. 2017 10 10 0.57 ( -0.33 , 1.46 )
Overall dementia 0.01 ( -0.59 , 0.60 )
-2 -1.5 -1 -0.5 0 0.5 1 1.5 2
Favors control
Hedges' g
Favors intervention

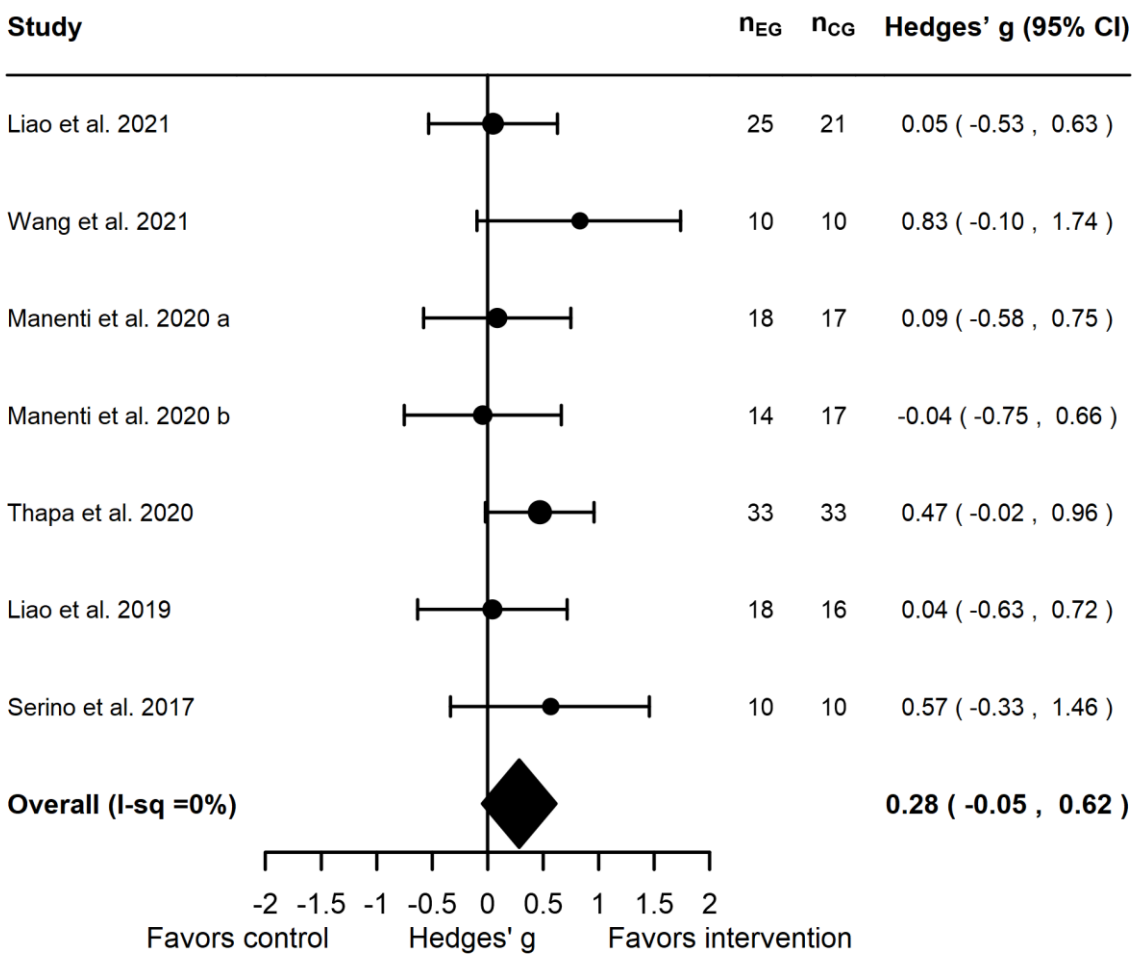
## Appendix 2

Sensitivity analyses for (a) global cognition and (b) executive function

(a)



(b)


Study
$n_{EG}$
$n_{CG}$
Hedges' g (95% CI)
Liao et al. 2021
25
21
0.05 ( -0.53 , 0.63 )
Wang et al. 2021
10
10
0.83 ( -0.10 , 1.74 )
Manenti et al. 2020 a
18
17
0.09 ( -0.58 , 0.75 )
Manenti et al. 2020 b
14
17
-0.04 ( -0.75 , 0.66 )
Thapa et al. 2020
33
33
0.47 ( -0.02 , 0.96 )
Liao et al. 2019
18
16
0.04 ( -0.63 , 0.72 )
Serino et al. 2017
10
10
0.57 ( -0.33 , 1.46 )
Overall (I-sq =0%)
0.28 ( -0.05 , 0.62 )
-2 -1.5 -1 -0.5 0 0.5 1 1.5 2
Favors control
Hedges' g
Favors intervention

## Appendix 3

Narrative synthesis of results

Five studies were excluded from meta-analysis due to insufficient numerical data (Anderson-Hanley et al. 2012, Gunst et al. 2022, Optale et al. 2010, Padala et al. 2017) and unclarity of cluster-randomization taken into account in the original analyses (Ramnath et al. 2021). Among MCI, intervention group with head-mounted display was observed to be more efficient in improving global cognition than music therapy and recreational activities (Optale et al. 2010). VR-training implemented as cyber cycling seemed to be more effective in improving executive function than stationary cycling (Anderson-Hanley et al. 2012). Interactive video gaming with Xbox Kinect Sport induced statistically significant improvement in global cognition and executive function compared with conventional multimodal physical exercise (Ramnath et al. 2021). Also some improvements were found in cognition by Xbox 360 Kinect Sport training compared with usual care of planned activities among older residents living in a residential care center (Gunst et al. 2022). However, the effect remained otherwise unclear and cognition level of participants was not reported nor was it able to be estimated from results as the direction of the outcome measure was not defined (Gunst et al. 2022).

Among people with mild dementia of Alzheimer's disease, no difference was found between VR-training and walking in global cognition (Padala et al. 2017).