UNIVERSITÁ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Economia e Statistica

Corso di laurea Magistrale in

BIOSTATISTICA

Effects of physical activity in the elderly on cognitive functions: a meta-analysis

Relatore: Antonella Zambon

Correlatore: Annalisa Biffi

Tesi di Laurea di:

Alessia Bacchiani

Matr. N. 889933

Anno Accademico 2023/2024

INDEX

1. INTRODUCTION

The aging of the population is one of the most significant challenges of the 21st century. Due to the increase in life expectancy and the decline in birth rates, adults aged 60 and older now represent the fastest-growing segment of the population. According to data from the United Nations Regional Information Center (UN data), in the last 20 years, the number of elderly people has increased by 70.3 million, and their percentage has risen from 13.4% to 17.5% in 2022. Moreover, it is estimated that by 2030, if this trend continues, people aged 65 and older will represent one-fifth of the total population in the UNECE region (United Nations Economic Commission for Europe), and those aged 80 and older will account for 5.4%.

Therefore, the significance of this phenomenon should not be underestimated, as some predictions highlight that many healthcare resources will likely be devoted to treating age-related conditions (cancer, fractures, dementia). [47]

Specifically, with aging, there is growing concern for cognitive health, as conditions like Mild Cognitive Impairment (MCI) and dementia pose an increasing challenge for healthcare and individual well-being. Mild Cognitive Impairment represents an intermediate state between normal aging and dementia, garnering increasing interest in scientific research due to its association with cognitive decline in the elderly. [9] The impact of MCI on the elderly population is multifactorial and can manifest in various aspects of daily life. Moreover, this condition, particularly the amnestic type, is associated with an increased risk of developing Alzheimer's disease (AD) or other forms of dementia. Studies indicate higher annual conversion rates for individuals with MCI compared to healthy controls, varying from 3% to 15% depending on the context (community or clinical) and the characteristics of the studied population. [34] Mild cognitive impairment, which exceeds what is expected in normal aging, significantly impacts the daily lives of those affected by MCI. Difficulties in performing household tasks, managing finances, and planning daily activities can make this population more vulnerable. [26] Additionally, for individuals aware of their MCI dysfunction and its high risk of progression to dementia, this can negatively influence their mental health; indeed, the presence of depressive symptoms is common, with a range varying from 36% to 63%; anxiety symptoms are also frequent, varying from 10% to 74%.[39-32]

What has been discussed so far aims to lead the reader to understand the fundamental importance of developing targeted diagnostic approaches and therapeutic interventions with the goal of improving the quality of life of elderly individuals, primarily by preventing the development of MCI and/or forms of dementia, and secondarily, by understanding their effects to delay progression.

However, at present, there are no FDA (Food and Drug Administration) approved specific treatments for the treatment or prevention of MCI. Some physicians prescribe medications approved for the treatment of Alzheimer's or other cognitive disorders to manage the symptoms of MCI, but there is no specific approved therapy for this condition or its prevention. [35] Current approaches primarily focus on symptom management and promoting a healthy lifestyle.

In this context, physical activity has increasingly gained prominence as a potential determinant of cognitive health in the elderly.

The World Health Organization (WHO) states in its 2020 physical activity guidelines that there is substantial evidence showing an inverse relationship between mortality and physical activity; regular practice of the latter can play a fundamental role in improving the functional abilities of the elderly and their quality of life. The Ministry of Health (2022) clearly reports the benefits of physical activity in elderly individuals: prevention of cardiovascular diseases, morbidity, and disability; significantly lower rates of various chronic disorders such as coronary artery disease, stroke, diabetes mellitus, and osteoporosis; reduction in blood pressure; and the fundamental importance of activity in maintaining muscle mass and balance, consequently reducing the risk of falls; positive effects on immune function, reducing the risk of respiratory infections. The favorable effects of physical exercise are amplified when elderly individuals are simultaneously involved in social and productive activities, leading to improvements in quality of life through social pathways as well. [29]

Although the benefits of physical activity in the elderly are well-established, to the point of having guidelines and recommendations from major global health authorities, the evidence that physical activity can delay cognitive decline and deterioration is more equivocal.

Indeed, the value of fitness in the elderly population, in the clinical context, has been critically addressed in various systematic reviews and meta-analytic syntheses. However, a lack of agreement has been found regarding the actual extent of cognitive function improvement associated with exercise interventions. The reasons for the current

lack of consensus are due to differences in the studies included in the analyses, the assessment of study methods and materials, the analytical approaches to the data, and the classification of the various cognitive measures used.

Cross-sectional studies have shown that physically active individuals tend to exhibit better neurocognitive function compared to inactive individuals. Prospective observational studies have reported similar results, demonstrating that individuals who maintain higher levels of physical activity show improvements in neurocognitive function compared to their sedentary counterparts. However, randomized studies have provided inconsistent results, with some reporting improvements in cognitive outcomes and others showing unclear results. Meta-analytic reviews of RCTs (randomized controlled trials) have also reported large variations in the extent of neurocognitive improvement associated with aerobic exercise, with some meta-analyses reporting modest cognitive gains and others more significant.

The present study proposes a systematic and quantitative analysis of the existing literature, aiming to accurately and objectively synthesize the evidence regarding the effect of physical activity on the cognitive health of the elderly. By analyzing aggregated data from previously conducted RCTs, it intends to provide as clear a perspective as possible on the extent of this association and identify any moderating factors that may influence the results. It should be noted that studies where the intervention was multifactorial, including not only physical activity but also cognitive training, social activities, and others, have also been considered.

The meta-analysis thus aims to (1) illustrate the specific interventions chosen in the RCTs, to give the reader a broad overview of the methodologies used, (2) describe the structural factors necessary to set up an effective intervention program (e.g., duration, frequency, etc.); and last but not least, (3) outline the cognitive changes (as the primary outcome) and functional abilities (as the secondary outcome) from baseline to post-intervention.

A systematic approach and rigorous analysis of the available evidence are essential to outline more effective and personalized guidelines, with the aim of preserving and improving cognitive health in the growing elderly population.

2. MATERIALS AND METHODS

2.1. Research strategy

The research process for this meta-analysis followed the guidelines outlined by the "Preferred Reporting Items for Systematic Reviews and Meta-Analysis" (PRISMA). The database used for searching studies aligned with the analysis purpose was PubMed. Additionally, further published studies were sought in the bibliographies of all included articles and additional meta-analyses found (but not included in this one). To enhance and facilitate the search, Key Medical Subject Headings (MeSH Terms) were utilized. Below is the search string used:

((((((((((((((((((((aged[MeSH Terms]) OR (frail elderly[MeSH Terms])) OR (aged, 80 and over[MeSH Terms])) OR (older adult, frail[MeSH Terms])) AND (geriatric assessment[MeSH Terms])) OR (inpatient[MeSH Terms])) OR (hospitalization[MeSH Terms])) AND (exercise therapy[MeSH Terms])) OR (motor activity[MeSH Terms])) OR (exercise[MeSH Terms])) OR (activity, motor[MeSH Terms])) OR (activity, physical[MeSH Terms])) OR (walking speed[MeSH Terms])) OR (tai chi[MeSH Terms])) OR (dance therapy[MeSH Terms])) OR (Exercise Movement Techniques[MeSH Terms])) OR (resistance training[MeSH Terms])) AND (Cognitive dysfunction[MeSH Terms])) OR (cognitive impairment[MeSH Terms])) OR (cognition disorders[MeSH Terms])) NOT (Alzheimer[MeSH Terms])) AND (treatment outcome[MeSH Terms])

The research methodology and protocol for this meta-analysis were not registered before conducting the review. However, all methods and procedures regarding the research and analysis are described in this document.

2.2. Inclusion and exclusion criteria

This paper used the PICO (Participants, Intervention, Controls, Outcome) approach in defining the specific research question and to include studies that are comparable. Only the following criteria were included (1) controlled and randomized studies (RCTs) were included.

(2) The population of interest consisted of elderly individuals aged over 60 years, residing in hospitals, communities, or geriatric clinics. (3) The included patients were overall healthy. Certainly, randomized controlled clinical trials involving patients with MCI were also included. (4) The examined intervention was physical activity, defined as any bodily movement that requires energy expenditure, performed in a programmed and periodic manner with sessions lasting from 20 to 90 minutes, one or more times per week. Studies were included where the intervention involved resistance exercises, strength training, flexibility exercises, balance training, as well as activities such as Tai-Chi, Yoga, dance, and walking. Additionally, multi-component interventions combining physical activity as described above with cognitive training were considered separately. However, they will not be included in the pooled estimate due to not reaching the minimum threshold of three studies for each outcome measure.

(5) The primary outcome was cognitive function measured by the MMSE (Mini-Mental State Examination), the MoCA (Montreal Cognitive Assessment) scale, or the Logical Memory Test.

It is noted that both the MMSE and MoCA scales range from a minimum of 0 to a maximum of 30 points. For the former, a score of 24 or higher is considered normal; for the latter, the threshold level is 26. Lower scores may indicate the presence of cognitive deficits, more or less mild depending on the obtained result. The MoCA scale is characterized by greater sensitivity compared to MMSE because it focuses on a broader range of cognitive functions; however, both are valid assessment tools for an individual's cognitive status.

The secondary outcome was geriatric depression, measured by the GDS (Geriatric Depression Scale), functional capacity assessed by the SPPB (Short Physical Performance Battery), and Gait Speed (m/sec). (6) For the control group, both active and passive controls were considered.

(7) No filter regarding a specific publication period was applied. (8) Only articles with full text available in European languages were included.

Consistently with the aforementioned, studies with inadequate design were excluded, specifically non-randomized clinical trials, prospective studies, observational studies, and generally studies lacking a control group, in order to reduce the risk of bias. Additionally, studies where the population was affected by dementia, such as Alzheimer's, were excluded to work with a population with similar baseline cognitive function. Patients with significant psychiatric or neurological disorders, severe

cardiovascular discomfort, or otherwise unable to engage in physical activity were excluded.

Furthermore, clinical trials where the intervention group underwent therapy with medication, aimed at alleviating MCI symptoms (such as dementia medications), were excluded. RCTs with missing data, such as standard deviation or even the mean difference of the outcome of interest, were excluded from the analysis, as well as studies for which the full text could not be retrieved.

To conclude, all duplicates were, obviously, excluded.

2.3.Statistical analysis

Before extracting the data, the included studies were assessed for risk of bias using PRISMA criteria, as will be further elaborated later on.

From the information reported in the included articles, the following data were extracted: sample size; size of intervention subgroups; mean age in the control group and intervention group(s) and their respective standard deviations; presence or absence of MCI (or other diseases); duration of the intervention; exercise protocol (duration and frequency of training sessions); characteristics of the control group; follow-up (Table 1 and 2).

To conduct statistical analysis of the data, R software version 4.2.2 was used. Treatment effects were calculated based on aggregated data for the same outcome from individual studies deemed clinically homogeneous. Cognitive measures are listed in **Table 3**. For the meta-analysis, cognitive tests for which at least three studies using the same instrument for calculating participants' cognitive level were not retrieved were excluded from the final pooled estimate.

The summary statistics required for each outcome were the number of participants in the intervention and control groups, the mean change from baseline, and the standard deviation of the mean change (SD). If variation in scores from baseline was not provided, these were calculated using baseline and post-test means and standard deviations. Where absent, variation SDs were calculated assuming a correlation of zero between baseline and follow-up measures (a conservative approach that may overestimate the standard deviation of change from baseline but is commonly used in meta-analyses).

The summary measure of treatment effect was the standardized mean difference (SMD) of the mean change from baseline to post-test between intervention (treated) and control (untreated) groups in the respective assessment scale or test.

Statistical heterogeneity was assessed using the I^2 heterogeneity index, which quantifies the variation among estimated effects in included studies. This measure represents the percentage of total variation among studies that is due to heterogeneity rather than chance. When the I^2 test shows a significant value, it indicates substantial heterogeneity among included studies, meaning that the observed differences in study results cannot be explained solely by chance. I^2 can range from 0% to 100%, and when the value is above approximately 70%, it indicates high heterogeneity that cannot be ignored in the final comprehensive estimate required in the meta-analysis. Heterogeneity in this study may stem from differences in study participants, intervention or control protocols.

The methodology chosen to address heterogeneity was twofold. Firstly, we opted to separate single-intervention RCTs from multi-intervention ones in the calculation of the final pooled estimate. Additionally, random-effects models were developed in cases where the I2 test yielded a significant result. The random-effects model allows for the presence of heterogeneity among studies. The estimated effect for each study is weighted based on the precision of its estimate and the total variance, which includes both within-study variance and between-study heterogeneity. Thus, the equation of the random-effects model incorporates variance between studies and assigns greater weight to more precise studies and lesser weight to less precise ones.

Furthermore, a sensitivity analysis was conducted to explore possible reasons for heterogeneity among studies, using confounding variables and constructing meta-regressions. The overall treatment difference estimates will be presented in Forest Plots.

3. RESULTS

3.1.Study selection

In order to identify all possible data for this meta-analysis, a comprehensive search method was conducted. A single reviewer screened the titles and abstracts of the 4,642 articles found according to the criteria described earlier (specifically, 4,635 identified through the search string entered on PubMed and 7 provided from the bibliographies of meta-analyses or reviews) and discarded irrelevant results based on a preliminary analysis. Subsequently, 4,466 articles were excluded, the full text of the remaining 176 articles was read, and those that met the inclusion criteria were selected. Upon further analysis, a total of 141 articles were removed, of which: 43 were not conducted using the methodology consistent with this analysis, namely, controlled and randomized clinical trials, as observational studies, prospective studies, systematic reviews, and meta-analyses, uncontrolled or non-randomized clinical trials were excluded; 8 studies were excluded due to the lack of a control group; 22 articles found a lack of suitability of the study participants, the most common reason being the participants' age being too low, but also those with specific diseases that did not meet the inclusion criteria already listed; 47 studies showed inappropriate intervention, such as exclusively cognitive therapy or mindfulness interventions, not combined with physical activity practices; 18 articles were discarded due to outcomes not conforming to our analysis. It is important to note that cognitive outcomes are currently measured by various instruments, therefore, studies that did not contain either cognitive measures or measures related to the functional capacity of individuals conforming to the purpose of this study were excluded; 3 studies were excluded because they were duplicates and finally one was excluded due to the absence of a full text in a translatable language by the reviewer.

The remaining 34 articles were examined in detail to extract the data on which to calculate the pooled estimate; 5 articles were excluded due to the inability to retrieve the mean difference between baseline and post-intervention or to calculate the standard deviation.

In conclusion, it is stated that, at the end of the inclusion/exclusion procedure, 29 controlled and randomized clinical trials were examined and included in the analysis.

It is noted that, of the 29 included RCTs, 21 of these are single-intervention studies, where the intervention group performed exclusively physical activity according to the specific study protocol. 8 RCTs are multi-intervention trials, meaning that in addition to physical activity performed in the intervention group, cognitive training was added, based on orientation exercises, cognitive stimulation, music therapy, or household activities. The process undertaken is shown in the flowchart in **Figure 1**.

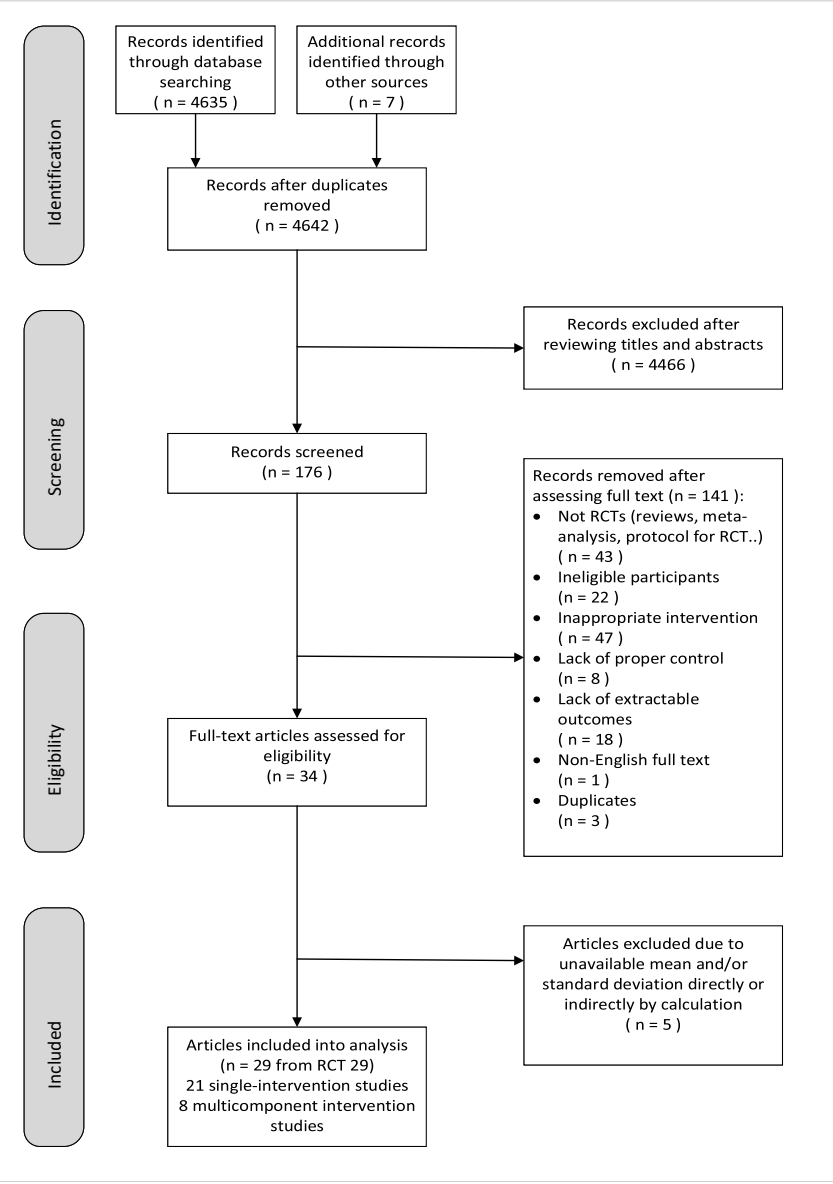

Figure 1. Flowchart of study selection according to the PRISMA criteria.

3.2. Description of included studies

The included studies come from around the world: Europe (n=6); America (n=5); Asia (n=18), with 8 from China.

In this meta-analysis, a total of N=4574 participants are included, of which 945 patients underwent a physical intervention associated with cognitive training (specifically, 8 RCTs).

The definition of "elderly" varies depending on the study considered (refer to Table 1 and Table 2 for details), with the minimum age considered being 60 years overall. The study by Zhu et al. [48] includes patients with an age range of 50 to 85 years, but in the actual population on which the RCT was conducted, there are no patients under the age of 60.

Approximately 40% of the included studies were conducted on a population with Mild Cognitive Impairment (MCI). The study by Martinez-Velilla et al.[27] included elderly patients in the ACE (Acute Care of the Elderly) ward but excluded those with severe cognitive decline, terminal illnesses, pulmonary embolisms, myocardial infarctions, uncontrolled arrhythmias, or bone fractures. Williamson et al.[44] enrolled elderly patients at high risk of disability but still able to walk 400 meters in less than 15 minutes. Lee et al.[20] included elderly individuals with physio-cognitive decline syndrome (PCDS). The remaining studies were conducted on a generally healthy elderly population.

Subsequently, the included studies in this meta-analysis will be further elaborated to clearly illustrate the main sources of heterogeneity among the RCTs. For this purpose, the articles are pre-divided into those where the planned intervention was only physical and those where physical activity was combined with cognitive training. The subsequent discussion will be presented in **Table 1** and **Table 2**.

3.2.1. Single-intervention studies

Of the twenty-nine studies included, twenty-one (approximately 70%) are RCTs where the treatment was solely physical activity, albeit in different forms.

It is worth noting that five of these twenty-one studies were designed for an elderly population with MCI (Zhu et al.; Lee et al.; Yoon et al.; Suzuki et al.; Hauer et al.).[48-20-45-41-11].

The study by Martinez-Velilla et al.[27] implemented in the intervention group two daily sessions of 20 minutes each with exercises for resistance, balance, and walking in the morning; followed by strength exercises with light weights and walking again in the afternoon; all performed consecutively for 5 days a week. The control group received the usual care, including physical rehabilitation, assigned to residents in the geriatric hospital. The follow-up considered was the length of hospital stay (the median was eight days). The study by Sink et al.[36] assigned similar exercises to those in the first study (based on strength, balance, flexibility, and walking), but with the difference that the sessions were scheduled 3-4 times a week, lasting 50 minutes each, and the follow-up length for the community-residing patients included in the study was 24 months. In this case, the control group participated in health education workshops, followed by flexibility and stretching exercises every week throughout the follow-up duration. Suzuki et al.[41] assigned the CG the same type of activities seen above and programmed a series of exercises in the IG to improve both aerobic performance and the posture and balance of the patients, conducted for 90 minutes twice a week, over a 12-month FU period. A similar intervention was implemented for the participants in the study by Hauer et al.[11], but with a follow-up period of 6 weeks; the control group received the usual care assigned to the residents of the geriatric community. The RCT conducted by Williamson et al.[44] also combined aerobic, strength, balance, and flexibility exercises, with the particularity of designing the intervention in three phases: adoption, transition, and finally maintenance of the acquired physical fitness, with a total FU of 25 weeks. In this case, the control group was assigned health education operations such as diet and personal care.

Some studies focused the intervention program on exercises aimed at increasing endurance through high/moderate-intensity aerobic sessions: Simley-Oyen et al.; Muscari et al.; Liu-Ambrose et al.; Cavalcante et al.[38-30-22-2]

The first study implemented cardio sessions for about 30 minutes, three times a week for 10 months; meanwhile, the control group performed flexibility and balance exercises. The second study involved one-hour resistance training sessions three times a week; the CG participated in meetings aimed at teaching a healthy lifestyle and promoting physical activity. The last two RCTs listed divided the intervention groups into two subgroups: Liu-Ambrose et al.[22] based on the number of weekly training sessions, 1 x RT

or 2 x RT. Cavalcante et al.[2] randomized the subset into RE (Resistance training) and REI (Resistance training with Instability). In both cases, the control group was considered "active" as it participated in health education seminars plus stretching and flexibility exercises. The study by Cassilhas et al.[1] worked with two teams in the IG, performing resistance exercises combined with press activities to develop strength in the elderly (three times a week for one hour), defined as "EMODERATE group" and "HIGH group" based on the lower or higher loads. This study had a follow-up of 24 weeks, and the control group performed warm-up exercises followed by stretching for the same period.

The studies by Lee et al.[20] and Yoon et al.[45] intervened in the interest group with activities based on high-intensity speed and power (HSPT training), conducted three times a week for 50 minutes in the first case; the control group received lectures to encourage physical activity, continuing for the eight weeks of intervention. In the second case, the intervention was divided into the subgroup practicing HSPT and the subgroup performing low-intensity strength and speed development activities (LSST), with sessions lasting one hour twice a week. In this second RCT, which lasted 12 weeks, both subgroups used elastic bands to perform the study exercises. The control group underwent dynamic stretching activities once a week for one hour.

The study by Venturelli et al.[43] chose a 12-week program based on 45-minute circuit training sessions three times a week, focusing on the upper body and with moderate intensity. The CG participated in physiotherapy sessions and electrostimulation, massages, and passive movements in bed.

Moving away from the traditional physical exercises seen so far, we will now briefly delve into studies that used Tai Chi or other forms of PA such as dancing, kayaking, and others.

The practice of Tai Chi is a form of Chinese martial art that involves slow, fluid movements coordinated with deep and mindful breathing; it improves balance, flexibility, and body coordination. Lam et al.[16] implemented Tai Chi techniques for 30 minutes a day, three days a week, with a control group performing relaxation and stretching exercises, for a twelve-month follow-up. Sungkart et al.[40] had the IG practice Tai Chi for 50 minutes, three days a week, with a follow-up of twelve weeks; the control group received educational material on fall prevention and cognitive decline. Both studies included instructors to supervise the correct execution of Tai Chi, aiming for the most effective intervention possible.

The study by Oken et al.[31] planned two intervention groups, where one subgroup practiced yoga for 90 minutes once a week, and the other performed aerobic exercises such as outdoor walking with the same frequency; the control group was defined as "passive" (no intervention was performed).

The RCT by Zhu et al.[48] involved group dance sessions for 35 minutes, three times a week; the patients in the CG continued to receive the usual care provided for the geriatric community. The follow-up was six months.

Choi et al.[5] designed a Virtual Kayak Paddling (VKP) activity, conducted by an instructor and an assistant, for one hour twice a week for six weeks. The control group performed exercises aimed at developing strength.

The study by Lü et al.[23] trained patients randomized to the intervention group in a dumbbell training (DTG) for one hour, three times a week for 12 weeks. Chen et al.[4] used elastic band exercises to increase the level of physical activity in wheelchair-bound patients (40 minutes, three times a week, for six months). Both of these last two studies chose a passive control group that continued to receive the usual care of the nursing home.

Finally, the RCT by Maki et al.[25] implemented a program of simple walking for 90 minutes once a week for three months, encouraging participants to continue walking daily with the goal of gradually increasing the number of steps. The control group received nutrition education lectures.

3.2.2. Multi-intervention studies

Considering the multi-intervention studies, it is noted that the majority (75%, or 6 RCTs) combine various physical exercises (such as resistance, balance, strength, and flexibility training) with cognitive training aimed at improving orientation, sensory use, memory, and attention (Train the Brain; Lee et al.; Jirayucharoensak et al.; Law et al.; Hagovská et al.; Kamegaya et al.)[42-19-13-18-10-14]. The remaining two articles by Li et al.[21] and Young et al.[46] combine cognitive training with physical training based on Tai Chi.

Almost all studies, except those by Kamegaya et al.[14] and Lee et al.[19], work with elderly patients with Mild Cognitive Impairment (MCI).

The study by Li et al.[21] randomized three groups: one assigned to cognitive training, another combining cognitive training with Tai Chi, and a control group that did not receive any intervention. The duration of this RCT is 12 months. Similarly, the RCT by Young et al.[46] used Tai Chi combined with physical games, orientation activities, and

sensory exercises (sounds, foods, etc.) for the intervention group to improve both physical and mental performance.

Train the Brain[42] program involved seven months of cognitive intervention combined with physical exercises and music therapy (three times a week for an hour) compared to controls who continued their routine activities.

Lee et al.[19] conducted 60 sessions of two hours each over a year, involving physical exercises, cognitive training, and nutritional education; the control group was passive.

Similar interventions were implemented by randomized controlled trials (Hagovská et al. and Kamegaya et al.)[10-14] that combined aerobic or balance exercises with competitive games, cooking, or sessions specifically designed for cognitive improvement (e.g., CogniPlus) for 10 and 12 weeks of follow-up, respectively.

The trial by Jirayucharoensak et al.[13] assigned the intervention group to two different types of activities: one to cognitive training plus physical training and the other to cognitive training only. Finally, the RCT by Law et al.[18] randomized more than one intervention group: FTG, a subgroup performing exercises to improve core muscles; ETG, aerobic training sessions; and CTG, cognitive training. The follow-up was eight weeks, and the control group engaged in activities provided by the community care center.

Ref. Author	Setting	Sample n (IG/CG)	Mean Age IG/CG	Inclusion criteria	Intervention	Control	Follow-up
Martínez-Velilla (2019)	Community-dwelling geriatric hospital - disease: ACE patients	370 (185/185)	87.1(5.2) / 87.6 (4.6)	Age 75 years or older, BI ≥ 60, able to ambulate, to comunicate and to collaborate	Two daily sessions of 20 minutes' duration during 5 to 7 consecutive days. Morning session: resistence, balance, walking exercises. Evening session: functional exercises using light loads + walking	Habitual hospital care	Measurement at baseline and at discharge (median 8 days)
Zhu (2018)	Community-dwelling dementia clinic - disease: MCI	60 (29/31)	70.3 (6.7) / 69.0 (7.3)	Age between 50 and 85, with MCI, MMSE ≥ 25 and MoCA ≤ 26	Patients danced togethere for 35-minutes dance session 3 times a week for 3 months. After 3 months, patients in the IG were encouraged to continue practing the dance routine	Habitual hospital care	Measurement at baseline, 3 months and 6 months
Sink (2015)	Community-living participants	1476 (735/741)		Aged 70-89 years, high risk for mobility disability based on SPPB but who could walk 400 m in 15 min without assistance	Physical activity intervention focused on walking, strength, flexibility and balance training, sessions during 50 minutes for 3-4 times x week	Health education workshop (presentation + flexibility and stretching)	Measurement at baseline and after 24 months
Liu-Ambrose (2010)	Community-dwelling women	n = 155 (1 x RT = 54, 2 x RT = 52, BAT = 49)	69.5 (2.7) / 69.4 (3.0) / 70.0 (3.3)	Women aged 65 to 75 years, living independently, MMSE ≥ 24, visual acuity of at least 20/40	Resistance Training session (RT) during 60 minutes. Progressive and high-intensity. Two groups: once-weekly RT or twice-week RT	Balance and tone exercise twice a week	Measurement at baseline, mid-point and trial completion
Cavalcante (2020)	Community-dwelling older adults - disease : MCI	n = 67 (REI = 22, RE = 23, CON = 22)	71 (6) / 71 (6) / 71 (4)	Aged 65 or older, subjective cognitive complaints or MoCA ≤ 26/30, no clinical disease that contraindicates training	Two type of intervention: resistance exercise (RE) and resistance exercise with instability (REI). 12 weeks of thrice-weekly sessions of seven exercises for three sets and 10-15 rep.	Health education seminars (lectures on prevention, healthy behaviors, stretching; 1 x week)	Measurment at baseline and at trail completion
Lee (2020)	Community-dwelling older adults - disease : MCI	40 (18 / 22)	74.2 (4.4) / 73.7 (4.6)	Aged 65 or older, with MCI, 0.5 score out of 3 on the CDR, frailty, 3 out of 5 on the Fried frail phenotype scale	HSPT (high-speed power training) 3 times a week, carried out for a total of 50 minutes per session	Lecture to encourage physical activity, in addition to exercise	Measurment at baseline and after 8 weeks of intervention
Choi (2019)	Community dwelling older adults	60 (30 / 30)	77.27 (4.37) / 75.37 (3.97)	Aged 65 years or older, MoCA < 26, able to communicate	Participants in the VKP (Virtual Kayak Paddling) group exercised for 6 weeks, twice a week for 60 min	Home exercises: flexion, curl-ups, leg lifts, hyperextensions	Measurment at baseline and at trail completion (6 weeks after)
Yoon (2017)	Community-dwelling women - disease: MCI	n = 30 (HSPT = 14, LSST = 9, CON = 7)	75.00 (3.46) / 76.00 (3.94) / 78.00 (2.77)	Aged 65 or older, MMSE scored 20-24 points, MoCA < 23, with ability to walk 10 m without a walking aid	Two types of exercise: HSPT (high speed power training) or LSST (low-speed strength training) groups. 1 h of exercises twice a week for 12 weeks with elastic band training for both groups	Usual care + stretching	Measurment at baseline and after 12-weeks (post test)
Suzuki (2012)	Community-dwelling elderly - disease: MCI	50 (25 / 25)	75.3 (7.5) / 76.8 (6.8)	Elderly 65 years and over dwelling in the community with MCI using the Peterson criteria	Multicomponent exercise for 90 min/d, 2 d/wk for total of 80 times over 12 months. The subjects practiced aerobic exercise, postural balance retraining and dual-task training	Health promotion	Measurment at baseline, after 6 months and after 12 months
Hauer (2017)	Community-dwelling geriatric hospital - disease: MCI	34 (17 / 17)	81.4 (6.6) / 83.3 (5.7)	Aged 65 or older, MMSE < 24, ability to stand or walk 5 m without support, no severe somatic or psychiatric disease	Patients in IG took part in a standardized 6 weeks home training: postural control and strength, interactive motor training and functional assessment	Habitual hospital care	Measurements at baseline, at the end of the 6-weeks training period and after 6-weeks FU period
Lam (2014)	Community-dwelling elderly	389 (171 / 218)	77.2 (6.3) / 78.3 (6.6)	Aged over 65 years old at risk of cognitive decline (MCI or amnestic MCI)	Intervention group received training on "24-style Tai Chi" . The frequency of the intervention was 30 minutes per day and 3 days per week	Stretching and relaxation exercise	Measurements were performed at baseline, 2, 6, and 12 months

Table 1: continued

Lü (2015)	Community dwelling older adults	45 (22 / 23)	69.00 (3.83) / 70.43 (5.53)	Aged over 65 years older, MMSE ≥ 24, MoCA < 26, ADL < 26, not clinically demented	Momentum-based dumbbell training (DTG) three times per week over 12 weeks for 60 minutes	Habitual hospital care	Measurment at baseline and after intervention (after 13 weeks)
Sungkart (2017)	Community dwelling older adults	66 (33 / 33)	68.3 (6.7) / 67.5 (7.3)	Aged 60 and older, MMSE ≥ 24, MoCA < 26	Practiced Tai Chi 3 times per week for 12 weeks	Educational material related to cognitive impaitment and fall prevention	Measurement assessed at baseline and at the end of week 15
Maki (2012)	Community dwelling older adults	150 (75 / 75)	71.9 (4.1) / 72 (3.9)	Excluded patients with dementia, aged 80 years or over, with medical illness	Walking program was conduceted once a week for 90 minutes for 3 months + encouraged participants to walk on a regular basis and to increase their step per day gradually	Educational lectures on food, nutrition, oral care	Measurement assessed at baseline and after intervention (12 weeks)
Chen (2016)	Community dwelling older adults	138 (73 / 65)	80.79 (0.91) / 81.37 (0.84)	Aged 65 years or older with MCI, mobility by using wheelchair	Wheelchair-bound elastic band exercises (WSEB) 3 times per week, 40 minutes each time, for 6 months, in a group practice format: warm-up, aerobic motion and harmonic stretching.	Habitual hospital care	Measurment at baseline, at 3 months and at 6 months
Williamson (2009)	Community dwelling older adults - disease: risk of disability	102 (50 / 52)	76.80 (4.37) / 78.06 (4.11)	Age range of 70-89 years, able to walk 400 m within 15 min, SPPB ≤ 9	PA intervention consisted of a combination of aerobic, strength, balance and flexibility exercises in 3 phases: adoption (weeks 1-8), transition (weeks 9-24) and maintenance (week 25 to the end of the trial)	Health education intervention: health nutrition, medications, foot care	Measurment at baseline and after the end of the trial (12 months)
Oken (2006)	Community dwelling older adults	n = 135 (Y = 38, E = 38, CON = 42)	71.5 (4.9) / 73.6 (5.1) / 71.2 (4.4)	Healthy men and women aged 65-85 years, exclude subjects with medical illness that might impair cognition	Yoga classes praticed for 90 min, one a week, focused on specific poses, breathing and static strength. Aerobic exercises consited of walking on an outdoor 400 meter track for endurance training once time a week	Habitual hospital care	Measurment at baseline and after 6 months
Simley-Oyen (2008)	Community dwelling older adults	57 (28 / 29)	69.86 (4.59) / 70.52 (4.47)	Aged 65 years or older, without cancer, unstrable cardiovascular disease or autoimmune disorder	Cardio group exercised for 25-30 min on the aerobic exercise equipment of their choice (treadmills, stepper machine, arm ergometeres,..) for 3 times a week	Strength, flexibility and balance exercises	Measurment at baseline, after 1 month, 4-5 month and 10 month
Muscari (2010)	Community dwelling older adults	120 (60 / 60)	69.6 (2.8) / 68.8 (2.5)	Aged 65-74 years, MMSE ≥ 24, BMI > 18, without cardiovascular disease or respiratory insufficiency	Endurance exercise training (EET) consisted of sessions of 1 h, three times a week in a gym located: cycle ergometer, tradmill and free body activity	Educational materials: improve lifestyle and physical activity	Measurment at baseline and after 1 year (so at the end of the trial)
Cassilhas (2007)	Community dwelling older adults	n = 62 (M = 19, H = 20, CON = 23)	69.01 (1.10) / 68.4 (0.67) / 67.04 (0.54)	Aged 65-75 years, MMSE ≥ 23, without cardiovascular disease or psychiatric conditions	24 weeks of resistance exercise, six exercises like chest press, leg press, .. 3 session per week for 1 hour but there are two type of interventions: EMODERATE group and HIGH group	Warm-up and stretching following (same experimental group)	Measurment at baseline and at the end of the trial (after 24 weeks)
Venturelli (2010)	Community-dwelling geriatric institute	n =23 (12 / 11)	83.3 (6.7) / 84.1 (5.8)	Aged 65 years or older, 15 ≤ MMSE ≥ 25, mobility limitation	Training program during 12 weeks based on circuit-training exercise schedule, 3 times a week for 45 min. Focused on the upper body and moderate intensity	Electrostimulations, massage and passive leg movament on bed	Measurment at baseline and after 12 weeks

Table 1. Characteristics of Single-Intervention Studies (Physical Activity Only) Included in the Meta-Analysis.

Ref. Author	Setting	Sample n (IG/CG)	Mean Age IG/CG	Inclusion criteria	Intervention	Control	Method
Li (2023)	Patients from memory clinics - disease: MCI	n = 152 (CT=51, MixT=48, Control=53)	65.5 (7.2) / 65.6 (5.6) / 66.6 (7.1)	MCI diagnosis (detailed medical history, neurological examinations and MTA), not other causes of dementia	In the first 12 months, the cognitive training group (CT) had cognitive training, and the mixed group (MixT) had additional Tai Chi training. In the second 12 months, training was only provided for a subgroup of MixT. Both training during 120 minutes per week (60 minutes, twice a week)	Habitual hospital care	Measurements at baseline and every 6 months afterward (in months 6, 12, 18 and 24).
Young (2020)	Community-dwelling older adults - disease: MCI	80 (41 / 39)	80.05 (6.17) / 80.25 (6.33)	Age 60 or above, having never received a diagnosis of dementia/major neurocognitive disorder, MMSE ≥ 18, living in the community	14 group sessions with 2 sessions twice a week for 60 minutes and included: reality orientation, physical games, food, sounds, senses etc.. Tai chi was chosen for the physical exercise component (15-20 min at the end of each cognitive stimulation therapy)	Recreational activities, excluding cognitive training and Tai Chi	Measurment at baseline and post-treatment
Train the Brain Consortium (2017)	Older adults - disease: MCI	63 (38 / 25)	74.0 (4.8) / 74.9 (4.4)	Aged between 65 and 89 years with MCI, without neurological pathologies or severe/moderate dementia	The MCI-training group received a 7 months multidomain training, including cognitive, physical exercise and music therapy. 3 times a week of cognitive training (60 min), 3 times a week of physical training (60 min): muscle strength, flexibility, physical function. Once a week a session of 1	Habitual hospital care	Measurement at baseline and at the end of 7 months of training/usual life
Lee (2022)	Community-living older adults - disease: PCDS	n = 340 (PCDS = 96, MIND = 101, CIND = 50, Robust =	75.3 (6.6) / 72.8 (5.6) / 70.7 (5.4) / 70.5 (4.4)	Community-living people, aged ≥ 65 year-olds with ≥ 3 chronic medical conditions, not having terminal illness and/or sever disability	Sixteen 2-hour sessions per year were provided for participats, including physical exercise, cognitive training, dietician education and individualized integrated care for multimorbidity	Habitual hospital care	Measurments at baseline to 3rd, 6th, 9th and 12th month
Jirayucharoensak (2019)	Community-living older women - disease: aMCI	n = 119 (CAU + NFT = 58, CAU + exergame = 36, CAU = 25)	71.7 (6.5) / 73.9 (6.2) / 70.9 (5.1)	Women with CDR score of 0.5 and CDR memory component score of 0.5, MMSE score not lower than the dementia cutoff score	58 were treated with CAU + NFT, game-based neurofeedback training (20 sessions of 30 minutes each, 2-3 sessions per week), 36 with CAU + extragame-based training (20 sessions of 30 minutes each, 2-3 sessions per week)	Habitual hospital care	Measurment at baseline and post-treatment
Law (2019)	Community-dwelling older adults - disease: MCI	n = 59 (CTG = 15, ETG =16, FTG = 14, Con = 14)	76.93 (6.79) / 77.94 (7.43) / 71.57 (7.43) / 75.14 (8.53)	Age 60+ with MCI living in community without history of brain lesion, depression, medical conditions	Functional task group: 12 sessions of 1 h of core functional tasks exercise. Cognitive training group: 12 sessions for 1h center-based computer cognitive training of attention, memory… PA group: 12 sessions of 1h of moderate intensity aerobic exercise	Habitual hospital care	Measurment at baseline and post-treatment
Hagovská (2016)	Outpatient psychiatric clinic - disease: MCI	80 (40 / 40)	Outpatient psychiatric clinic	Age over 65 but less than 75, problems with memory, depression and anxiety, MMSE > 23	The experimental group was engaged in 20 cognitive training sessions twice per week, using CogniPlus together with balance training + dynamic balance training	Balance training programme	Measurment at baseline and after 10 weeks
Kamegaya (2014)	Community-dwelling elderly	52 (26 / 26)	73.6 (5.6) / 76.2 (6.1)	Aged 65 years old or older residing in Maebashi, without medical condition that made them unable to engage PA	Physical activity included muscle-stretching and aerobic exercise + leisure activities (cooking, handcrafts, competitive games): weekly 2-h programme at community center for 12 weeks	Habitual hospital care	Measurment at baseline and post-treatment

Table 2. Characteristics of Multi-Intervention Studies (Physical Activity + Cognitive Training/Other) Included in the Meta-Analysis.

3.3. Quality of the studies

The risk of bias was assessed for each study using PRISMA criteria. Specifically, the risk was categorized as low/high/unclear based on the type of bias, as follows:

- Selection Bias: This refers to the generation of the random sequence and the adequacy of the allocation process.
- Performance Bias: This is characterized by the adequacy and credibility of the control group, which should adopt training practices that enhance the representativeness of the data. It is noted that the use of a "passive" control group attributes a high risk of bias. Regarding the choice of an "active" control group, the credibility of the activity and its actual duration were evaluated, and a risk level of "unclear" or "low" was assigned based on the previously highlighted criteria.
- Detection Bias: The presence of blinding (single or double blind) among the evaluators and observers of the outcomes was analyzed to increase the objectivity and reliability of the results.
- Attrition Bias: This considered the details regarding participant loss (such as dropouts, lost follow-up, attrition rate) and whether the analysis was conducted as Intention-To-Treat (ITT) or not.
- Reporting Bias: This refers to the clarity in the presentation of the final results, with a precise definition of primary and secondary outcomes.

For each study, quality was assessed on a scale ranging from 1 to 3 where:

- High quality (Score 3): The study has a robust design and meets most or all the predefined methodological criteria. It has a low risk of bias and provides a reliable contribution to the evidence.
- Moderate quality (Score 2): The study has an acceptable design but has some methodological and reporting limitations. The risk of bias may be moderate, but the drawbacks do not seriously compromise the validity of the study.
- Low quality (Score 1): The study has serious methodological or reporting limitations. The risk of bias is high, and the study results may be less reliable.

	Random sequence generation	Allocation concealment	Blinding of participants personnel	Blinding of outcome assessment	Incomplete outcome data addressed	Selective reporting	Quality of studies
Martínez-Velilla (2019)	+	+	−	+	+	+	3
Zhu (2018)	+	+	−	+	−	?	2
Sink (2015)	+	+	+	+	−	+	3
Liu-Ambrose (2010)	+	+	−	?	+	+	2
Cavalcante (2020)	+	?	−	−	+	+	2
Lee (2020)	+	?	?	?	−	−	1
Choi (2019)	+	+	+	?	+	+	3
Yoon (2017)	+	?	−	?	−	?	1
Suzuki (2012)	+	?	?	?	+	?	2
Hauer (2017)	?	?	−	?	−	+	1
Lam (2014)	+	?	−	?	+	+	2
Lü (2015)	+	+	−	+	−	+	2
Sungkart (2017)	+	+	?	+	−	+	2
Maki (2012)	+	?	?	?	−	?	2
Chen (2016)	+	?	−	+	−	?	2
Williamson (2009)	+	?	+	+	−	+	2
Oken (2006)	+	+	−	+	−	+	2
Simley-Oyen (2008)	+	?	+	?	−	+	2
Muscari (2010)	+	?	?	?	−	+	2
Cassilhas (2007)	+	?	?	?	−	?	2
Venturelli (2010)	+	?	?	?	−	?	2

	Random sequence generation	Allocation concealment	Blinding of participants personnel	Blinding of outcome assessment	Incomplete outcome data addressed	Selective reporting	Quality of studies
Li (2023)	+	+	−	+	−	+	2
Young (2020)	+	?	−	?	+	?	2
Train the Brain (2017)	+	?	−	+	−	+	2
Lee (2022)	?	?	−	?	−	+	1
Jirayucharoensak (2019)	?	?	−	?	+	+	1
Law (2019)	+	?	−	+	+	+	2
Hagovská (2019)	+	+	−	−	−	?	2
Kamegaya (2014)	+	?	−	?	−	?	1

Table 3. Risk of Bias and Quality Assessment of Included Studies.
Note: "+" low risk; "-" high risk; "?" unclear risk. Regarding "blinding of participants," it is noted that in non-pharmacological interventions, such as in this case, it is impossible to blind participants. However, it was assessed as "low" when the control was active, engaging in activities that could be mistaken by participants for the intervention itself, and when attention was paid by the study creators to the credibility of the control group.

3.4. Outcome classification

In this meta-analysis, as predefined by the PICO criteria, cognitive function was designated as the primary outcome. The assessment of cognitive outcomes is conducted through various tests, which, on one hand, allows for a comprehensive and detailed evaluation of different aspects of cognition, but on the other hand, makes the analysis quite heterogeneous.

To obtain pooled estimates with a minimum of three studies per test/tool used, the following cognitive outcomes were selected: MMSE scale, MoCA scale, and Logical Memory (LM).

Regarding the secondary outcome, the following were chosen: the "Geriatric Depression Scale" (GDS), a tool used to detect the presence and severity of depression, where higher scores indicate a higher likelihood of depression; "Gait Speed" (GS) in m/sec, which refers to walking speed, with faster gait associated with better physical function and lower risk of mobility-related conditions, while slower GS may indicate health problems or increased risk of falls or other complications; and the "Short Physical Performance Battery" (SPPB), which evaluates physical performance in older adults or individuals with reduced mobility, composed of three tests measuring static balance, gait speed, and strength separately, with total scores ranging from 0 (low physical performance) to 12 (high physical performance).

The latter two outcomes are designed to assess an individual's physical functionality. The GDS does not specifically evaluate cognitive functions but is used to explore various aspects of mood, interest in daily activities, sleep, appetite, or other dimensions related to depression; therefore, it does not fall under either cognitive or functional assessment categories. Considering this, these three assessment tools (GDS, GS, SPPB) are deemed "partially cognitive," as they pertain to other spheres of the individual as well.

Table 4 provides a detailed breakdown of the included studies for each outcome domain (cognitive, partially cognitive, or other) and indicates the number of articles that will be included in the final pooled estimate. It also specifies the number of estimates for each outcome. It is clarified that different estimates can be obtained even using the same evaluation scale (as in our case); this phenomenon depends on factors such as the

condition of the participants at the time of testing (thus, the baseline data), the methodology used, the analyst's experience, and contingent factors.

Specifically, for the MMSE outcome, three versions are noted: the classic Mini-Mental State Examination, the 3MSE (modified MMSE version in Williamson et al.[44]), and the Korean Version of MMSE (Yoon et al.[45]); for the MoCA outcome, we have the classic version, the Korean Version of MoCA (Yoon et al. [45]), and the Beijing Version of MoCA (Zhu et al.[48]); the Logical Memory outcome was estimated in its Wechsler Memory Scale-Revised version for all RCTs; GDS was assessed in its classic version and the Spanish Version of Yesavage GDS; GS in m/sec; and SPPB was estimated in its classic version for all studies.

As stated in section "**2.3. Statistical Analysis**", to proactively manage at least some heterogeneity, the final pooled estimate was calculated by separating single-intervention RCTs from multi-intervention RCTs. It is noted that the latter (multifactorial studies) did not meet the minimum threshold of three studies evaluated with the same outcome measure, despite having chosen outcomes consistent with this meta-analysis.

To make the meta-analysis as comprehensive as possible, the subsequent subparagraph will delve into the results of the studies included in the "Other" column of **Table 4**, which are consistent with the present analysis but did not meet the minimum threshold of three studies.

	Outcome domains													
	Cognition						**Partially cognition**						**Other**	
	MMSE		MoCA		LM		GDS		GS		SPPB		…	
	PA only	Multi-int.	PA only	Multi-int.	PA only	Multi-int.	PA only	Multi-int.	PA only	Multi-int.	PA only	Multi-int.	PA only	Multi-int.
Num. ARTICLES	**7**	2	**3**	1	**4**	0	**4**	0	**3**	0	**4**	0	6	5
Num. ESTIMATES	3		3		1		2		1		1		/	

Table 4. Outcome Domains, Number of Articles, and Number of Estimates.
Note: "PA only" = single intervention (physical activity only); "multi-int." = multi-intervention studies.

3.5. In-depth Analysis: Other outcomes

The study by Sink et al.[36] uses the Digit Symbol Coding task (DSC) to measure participants' cognitive functions. This test primarily evaluates information processing abilities, working memory, and visuomotor processing speed. Additionally, it uses the Hopkins Verbal Learning Test-Revised (HVLT-R), a neuropsychological assessment tool to evaluate verbal memory and learning. Post-follow-up scores for DSC and HVLT-R (adjusted for clinic site, sex, and baseline values) show no significant differences between groups, with mean differences of -0.014 points (p-value = 0.97) for DSC and -0.03 (p-value = 0.84) for HVLT-R.

To measure cognitive and executive functions in the trial by Lee et al.[20], the frontal lobe test (FAB-K) was conducted. This test comprises six parts: mental flexibility, conceptualization, motor programming, behavioral self-control, inhibitory control, and primitive reflex. The mean difference between the intervention and control groups is 0.7 points, with a non-significant p-value.

The RCTs by Lü et al.[23] and Train the Brain[42] show a statistically significant improvement in cognitive functions for subjects with MCI, measured by the ADAS-Cognitive subscale, with scores of MD: -5.02 points, p-value = 0.012, and -2.17 points, p-value < 0.0001, respectively.

Chen et al.[4] evaluate changes in functional and cognitive abilities using the Barthel Index, which consists of 10 items related to the ability to manage daily life independently. Results show a statistically significant p-value, with a mean difference between groups of 7.98 points.

In their RCT, American researchers Oken et al.[31] consider a range of cognitive measures focused on attention and vigilance. Relevant to this analysis are performance on the Stroop Test, a psychological tool to estimate the speed of information processing and decision-making, and vigilance assessed via quantitative electroencephalogram (EEG). The p-values are 0.72 and 0.46, indicating no statistically significant differences in neuro-cognitive performance between the study groups. Similarly, Simley-Oyen et al.[38] find no significant differences using the Wisconsin Card Sorting Test (WCST), a test for executive brain functions involved in planning, organization, behavioral inhibition, and cognitive flexibility, with a p-value of 0.104. Other tests analyzed include Stroop Word, Stroop Color, and Stroop Word-Color conflict, with only the latter showing significance (p-value = 0.031).

Lee et al.[19] use alternative outcomes to measure global cognitive performance pre- and post-multifactorial intervention, with significant mean differences in concentration (p-value = 0.011), language (p-value = 0.006), abstract thinking (p-value = 0.027), and orientation (p-value = 0.013). These domains are also covered by the MoCA scale.

Chinese researchers Law et al.[18] employ the Trail Making Test (TMT) in versions A and B to measure attention, cognitive flexibility, visuospatial processing ability, and processing speed. Mean differences and p-values are 0.79 and 0.085 (TMT-A) and 1.07 and 0.095 (TMT-B), respectively.

Hagovská et al.[10] evaluate cognitive function differences using the Addenbrooke's Cognitive Examination (ACE), finding significant differences favoring the experimental group (p-value < 0.05-0.0001).

Kamegaya et al.[14] submit participants to the Five-Cog test, which evaluates cognitive domains of attention, memory, spatial function, language, and reasoning, and estimates depressive symptoms through a questionnaire. Results show a significant improvement (p-value = 0.046) in cognitive performance favoring the intervention group with physical and cognitive programs compared to the control group.

3.6. Pooled estimates

At this point in the analysis, the Forest Plots created using R software version 4.2.2 (as already explained in the "METHODS AND MODELS" section) are presented.

The process by which the pooled estimates were generated is described subsequently. The R package useful for conducting the meta-analysis was installed using "*install.packages: ("meta")*". A data frame was entered for each outcome with the respective studies, year, number of participants in the control group (N_CON) and in the intervention group (N_TRT), the mean differences from baseline to follow-up (MU_CON and MU_TRT), and the standard deviations (SD_CON and SD_TRT) for each intervention group. Standardized mean differences (SMD) between the intervention (treated) and control (untreated) groups were used to calculate the final result, along with the respective standard errors. Using the "*metacont*" command, applied here for continuous outcomes, the pooled estimates with their respective confidence intervals were calculated. Finally, the "*forest*" command allowed for the easy creation of the Forest Plots that follow. Naturally, these are as many as the outcomes chosen for this meta-analysis.

The Forest Plots presented below show the point estimates of the effects of each study, along with the confidence intervals, in the form of horizontal lines. Each line represents an RCT, and the horizontal position of the line indicates the effect estimate. The confidence interval (CI) is represented as a horizontal bar around each estimate. The graphs are useful for visualizing the variability between studies and for identifying the consistency or inconsistency of the estimates among them. Additionally, they can help identify outlier studies or those with particularly narrow or wide confidence intervals. The size of the green squares corresponds to the relative weight of each study in the overall estimate.

The calculation of each study's weight in the fixed-effects model follows the inverse variance principle, namely:

$$w_i = \frac{1}{estimated\ variance\ of\ the\ effect\ in\ study\ i}$$

Regarding the random effects model:

$$w_i = \frac{1}{estimated\ variance\ of\ the\ effect\ in\ study\ i + \tau^2}$$

where τ^2 = variance between studies, heterogeneity among studies.

The weight of a study influences its respective impact on the calculations of the overall estimate of the weighted mean effect. More precise studies will contribute more to the overall estimate, while less precise studies will have a smaller impact. The fixed-effects approach assumes that the true effects are all equal and that differences are mainly due to random errors or variability; the random-effects approach assumes that the weight assigned to each study is determined both by the precision of its effect estimate (fixed-effects model) and by the total variance, which includes both the estimated effect variance within the study (within-study) and the heterogeneity among studies (between-study).

The final pooled estimate is presented using a black diamond, with the precise value and its corresponding 95% CI to the left of it.

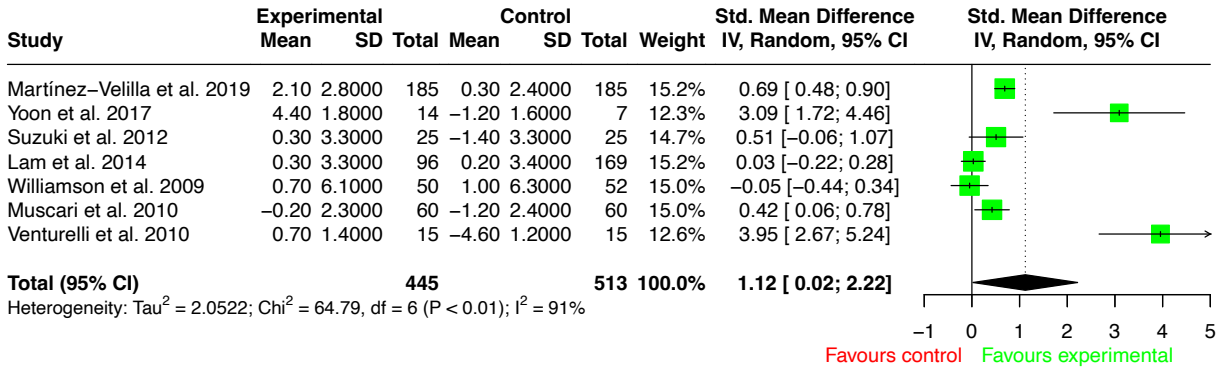

Figure 2: Forest Plot, single intervention: physical activity, outcome: MMSE.

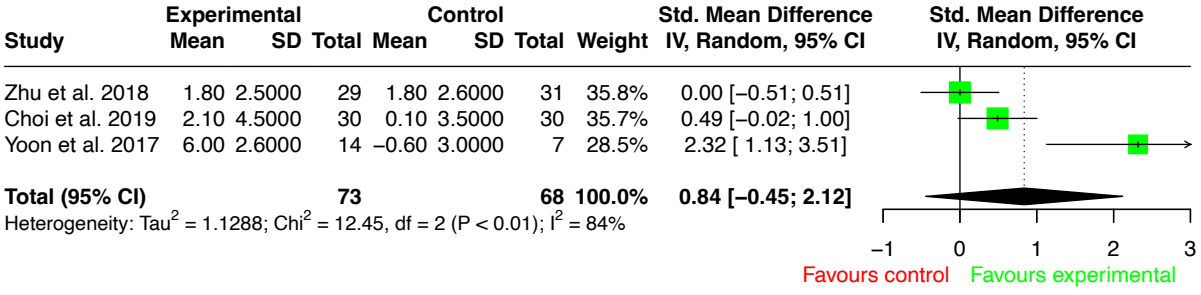

Figure 3: Forest Plot, single intervention: physical activity, outcome: MoCA.

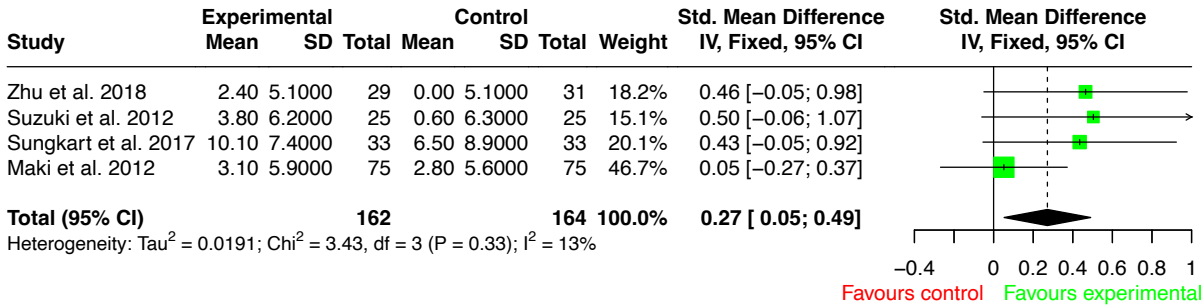

Figure 4: Forest Plot, single intervention: physical activity, outcome: LM.

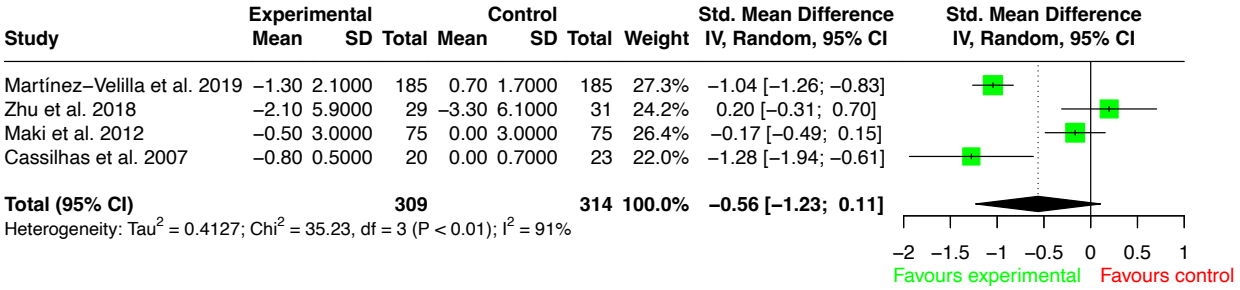

Figure 5: Forest Plot, single intervention: physical activity, outcome: GDS.

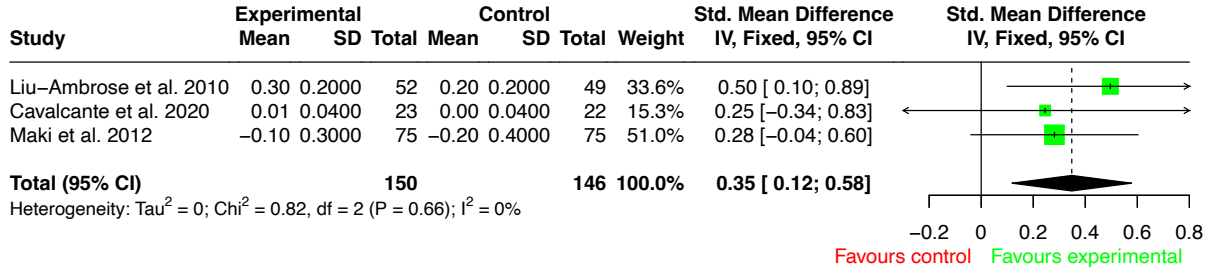

Figure 6: Forest Plot, single intervention: physical activity, outcome: GS.

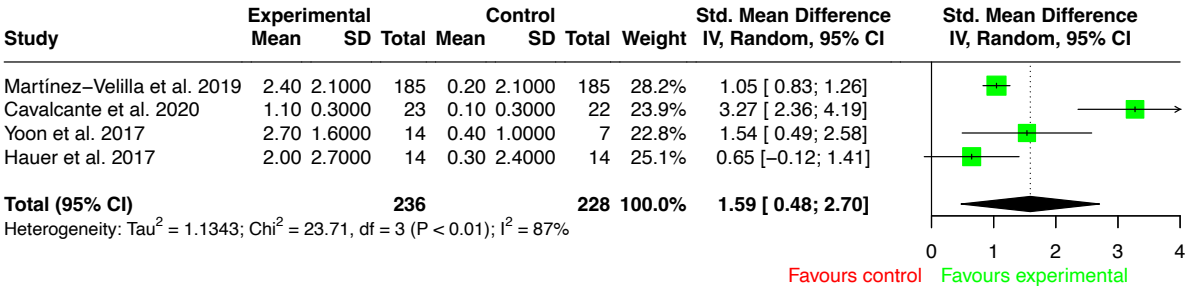

Figure 7: Forest Plot, single intervention: physical activity, outcome: SPPB.

3.7. Results commentary

The results depicted in the above Forest Plots are summarized. Looking at the I2, it is noted, first, that a random-effects model was employed to account for the considerable variability among the studies for all outcome measures, except for Logical Memory (**Figure 4**) and Gait Speed (**Figure 6**), with respective I^2 values of 13% and 0%, for which a fixed-effects model was used. Indeed, examining the studies included for the LM outcome, it can be observed that, except for the study by Maki et al.[25] with a larger sample size and consequently a narrower confidence interval (CI), the others included a similar number of participants, yielding very similar results for the final effect measure (SMD), both in point estimate and in the CI. For the GS outcome, it is noted, to investigate the homogeneity among the three included RCTs, that all utilized an active control group and the highest percentage of participants in each trial were female.

The pooled estimates resulting for each outcome measure are then analyzed: two of these were not statistically significant, particularly the one for the MoCA outcome, 0.84 [-0.45; 2.12] (p-value = 0.2009) (**Figure 3**); and that for the GDS outcome, -0.56 [-1.23; 0.11] (p-value = 0.0986) (**Figure 5**). In-depth analysis reveals, for the first mentioned outcome measure, that only the study by Yoon et al.[43] of the three included trials was significant in favor of the intervention group, with a low sample size and consequently a wide CI, indicating low precision in the final estimate. Regarding the GDS, it is observed that two of the four included RCTs have a CI of the SMD effect measure that includes 0, while of the remaining two studies, that of Martínez-Velilla et al.[27] with a large sample size and a narrow and statistically significant CI stands out; the trial by Cassilhas et al.[1], although also significant, shows a wide CI, caused by the low sample size. From the Forest Plot, heterogeneity in the point and interval estimates of each included study is noticeable, justifying the choice of a model that accounts for variability among RCTs, as well as within studies.

Moving on to examine the other comprehensive effect measures showing superiority of the treated group over the untreated one. The overall treatment effect estimate for physical activity on the MMSE outcome (**Figure 2**) is 1.12 with a confidence interval [0.02; 2.22]. This result indicates that, based on the data available from this meta-analysis, significant differences are observed between the group assigned to physical activity and the control group on the MMSE measure (p-value = 0.0454). Of the seven included studies, it is seen, by checking each CI, that three of them were not significant;

these used an active control group, and the follow-up period was 12 months for all three. The lower weights (around 12%) were assigned to the studies by Yoon et al.[45] and Venturelli et al.[43], both with wide CIs and low precision of the final effect. The CI of the final pooled estimate, as expected, shows moderate precision, being quite wide.

The Forest Plot for the LM outcome (**Figure 4**) presents a final pooled estimate of 0.27 [0.05; 0.49] (p-value = 0.0150); it is noted that four studies are included for this outcome, of which three; those by Zhu et al.[48], Suzuki et al.[41], and Sungkart et al.[40]; have a similar point estimate value of the SMD, wide CIs, and are not statistically significant; the study with the most weight (46.7%), due to greater precision, is that of Maki et al.[25]; none of the studies are European, and, except for that of Suzuki et al.[41], all have a higher percentage of female participants than males.

The analysis continues with the examination of the pooled estimate of gait speed, Gait Speed (**Figure 6**), with a value of 0.35 m/sec and a confidence interval [0.12; 0.58] (p-value = 0.0030). From the RCTs considered in this meta-analysis, it is interpreted that there is a statistically significant difference regarding the change in the gait of the elderly between the group treated with physical activity and the control group. The CI is moderately narrow, thus it can be stated that the estimate is moderately precise.

The Forest Plot for the SPPB measure (**Figure 7**) highlights a final estimate of 1.59 [0.48; 2.70] ($p < 0.0052$), therefore, the treated group shows a significant improvement for this partially cognitive outcome compared to the untreated one. It is noted that only one trial, among the four included studies, resulted in non-significance, that of Hauer et al.[11]. The latter was conducted on a population with MCI and a training period of six weeks, the chosen physical activity mainly comprised postural and strength exercises, and less focused on aerobic training; the control group did not engage in any activity. The study by Cavalcante et al.[2] shows a marked point estimate in favor of the intervention group (of 3.27) and chose resistance activity, aimed at increasing aerobic capacity, for the intervention group.

3.8. Heterogeneity Analysis: Meta-Regression

Meta-regression is a statistical analysis used in meta-analyses to explore sources of heterogeneity among included studies, assessing how the characteristics of RCTs may influence effect estimates. Study heterogeneity can arise from multiple sources, such as differences in intervention methods, study designs, follow-up duration, or patient characteristics (e.g., RCTs on patients with or without MCI).

Considering what has just been discussed, meta-regressions constructed using R software (version 4.2.2.) will be illustrated below, using the "*metareg*" command available in the "*meta*" package, previously mentioned. Simple multivariate models were created considering some variables selected as effect modifiers; that is, factors that may influence the strength or direction of the association between the independent and dependent variables. Meta-regressions were conducted for outcomes with four or more studies and for which the I^2 index was high enough to reject the hypothesis of homogeneity for that specific measure considered.

For a complete visualization of the output generated in R, resulting from the meta-regressions, please refer to the "**APPENDICES**" collection at the end of the document. Below is the commentary, for each outcome, on the estimates ("*estimates*") of each modifying variable, paying attention to their respective confidence intervals ("*ci.lb*" and "*ci.ub*") and p-values ("*pval*"), to inform the reader about the significance or lack thereof of the inserted covariates. Furthermore, in order to take into account residual heterogeneity after performing the meta-regression and introducing new variables, the change in the I^2 index will be observed.

The outcome of the meta-regression for the MMSE outcome, using the REML method, is presented in **Appendix 1**. Five variables were included as potential effect modifiers, four of which are categorical: "MCI," which classifies studies conducted on subjects without MCI (noMCI, reference category) or with MCI; "control," if the RCTs chose a passive (reference category) or active control group; "region," studies conducted outside Europe ("nonEurope," reference) or European ("Europe"); "gender," divided into the "male" reference category, indicating that the majority of individuals in the trial were male, and "female" if the participants were mainly female. Continuing, a quantitative effect modifier, "FU," indicating the months of follow-up, is included; readers interested in retrieving this information can consult **Table 1** seen earlier.

Proceeding with the analysis of the estimates and the respective p-values of the inserted variables, the significance (with a p-value < 0.05) of all modifiers is highlighted, except for "control"; consistently, the "Test of Moderators" shows a p-value < 0.0001. For example, the value of the "MCI" estimate suggests that studies conducted on subjects without this condition show lower effect values on the MMSE outcome (due to the negative value of the coefficient in question) compared to trials conducted on subjects with MCI; a comment that can be repeated for the other categorical variables. Regarding the coefficient associated with the quantitative variable, "FU," it can be interpreted as the amount of variation in the effect estimate (0.7725 MMSE points) for each additional month of follow-up. The I^2 index is now 0%, and consistently, the QE test (Cochran's Q Test) shows a non-significant p-value, leading to the conclusion that, by including effect modifiers, a large part of the heterogeneity has been controlled.

In order to have evidence regarding the robustness and consistency of the results for the MMSE outcome, a multivariate meta-regression model is produced using the Sidik-Jonkman option. This method is used to address the problem of heteroscedasticity and proposes a correction of the variance in the estimates, taking into account the correlation between the mean and variance of the studies included in the meta-analysis; see **Appendix 1a** for its visualization. From the comparison between the outcomes obtained with the first and second methods (REML vs. SJ), it can be concluded that there is consistency in the results: the changes in the upper and lower limits of the CIs are almost negligible; the standard error in the second case is slightly higher, consequently, the value of the residual heterogeneity is almost 10%, but still the QE test is non-significant.

Moving on to the commentary of the meta-regression performed for the GDS outcome, contained in **Appendix 2** and **Appendix 2a**. It is anticipated that, to avoid a problem of model overfitting, the number of parameters plus the intercept (k + 1) must be less than the number of observations (studies) included in the meta-analysis. For the GDS outcome, only one modifying variable, "region," was considered significant.

The interpretation of the "region" variable can be made by stating that studies from non-European origins have a significantly higher effect value for the GDS outcome (because the coefficient is 1.0049) compared to European ones. The introduction of this covariate brings the I^2 heterogeneity index to a value of 0%. It should be noted that, for the calculation of confidence intervals, both the classical method (**Appendix 2**) and the Hartung-Knapp method (**Appendix 2a**) were used; the latter approach is designed to

correct variance estimates that may be influenced by heterogeneity or by a limited number of studies included in the meta-analysis (as is the case for the outcome in question with only four RCTs). With the Hartung-Knapp method, the CIs are corrected to account for variability among studies; indeed, they are wider than those calculated with the classical method; however, the "region" modifier remains significant, with a p-value of 0.0263.

Moving on to the meta-regression conducted for the SPPB outcome. The preliminary considerations are equivalent to those made for the GDS outcome, as well as the choice to present, in order to show the robustness of the analysis, both the classical method (**Appendix 3**) and the Hartung-Knapp method (**Appendix 3a**) for the calculation of confidence intervals. In this case, the modifying variable chosen is "control," as before, inserted into the model because it was the only one found to be significant.

It can be concluded, based on the analysis conducted, that studies with a passive control group have a lower effect value for the SPPB outcome compared to those choosing an active control; this modifier is significant, although, correcting the width of the CIs with the Hartung-Knapp method, a p-value close to the acceptance threshold of 0.05 is observed. In this case as well, Cochran's Q test leads to the acceptance of the test, and the heterogeneity index is 0.01%.

3.9. Influence Analysis

Influence analysis is a procedure carried out to assess the impact of individual studies on the overall outcome of the meta-analysis. The main objective here is not to identify the "truth" of the effects (already extensively analyzed so far), but rather to evaluate the robustness of the estimates and identify potential influences of individual studies that, if excluded, significantly change the overall estimates. Therefore, one study will be excluded at a time, and the exclusion's effect on the overall result for that outcome measure will be evaluated. This analysis will be carried out only for outcomes with a number of studies of at least four.

Starting with the MMSE outcome, the decision is made to exclude the study by Yoon et al.[45]. The reasons for this choice are as follows: the sample size is only 21 participants (14 in the intervention group and 7 in the control group), which is much smaller than the higher sample sizes of the other RCTs (except for Venturelli et al.[43]); the quality is poor, as evidenced by the high risk of bias (refer to **Table 3**); the RCT by Yoon et al.[45] can be seen as an outlier because its result deviates from that of the other articles (see **Figure 2**). Furthermore, for almost the same reasons as those just mentioned, the study by Venturelli et al.[43] will also be excluded. Thus, a multivariate influence analysis will be conducted, involving the simultaneous exclusion of multiple studies. This approach is used because we want to assess the combined effect of these two studies, with similar characteristics, on the overall effect estimate. The resulting Forest Plots will be presented to provide a clear graphical representation of the impact of the excluded RCTs on the overall effect estimate.

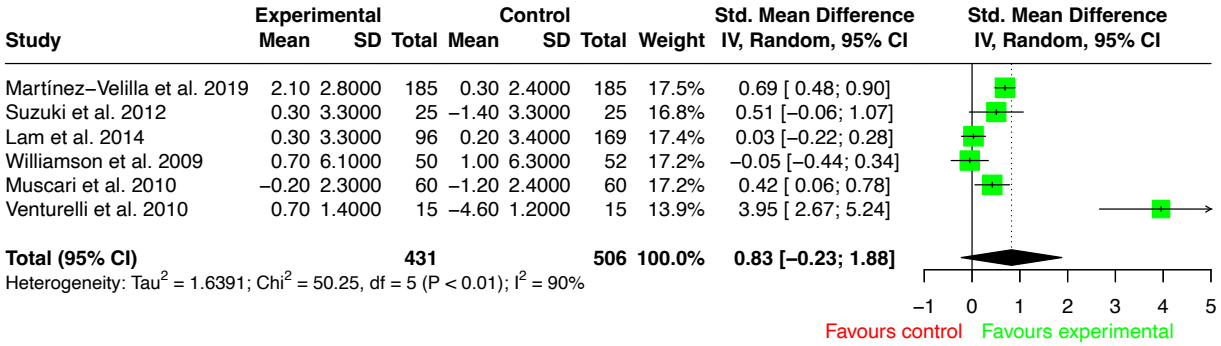

Figure 11. Forest Plot for MMSE outcome, excluding RCT by Yoon et al. 2017.

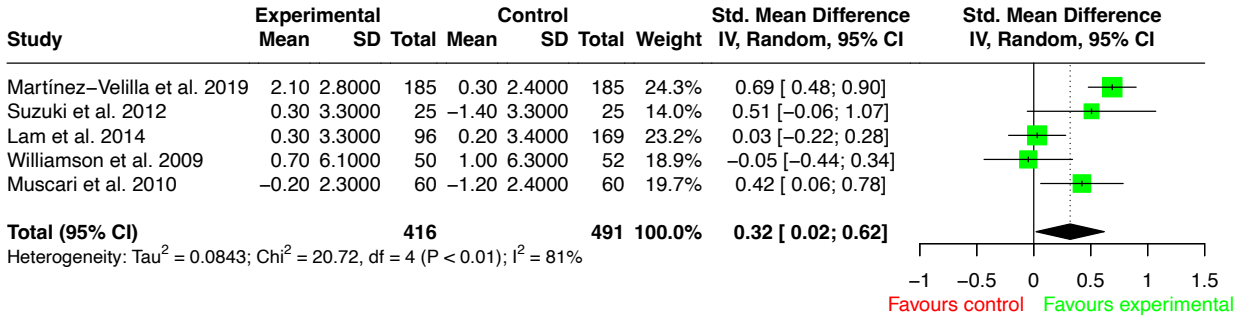

Figure 11.1. Forest Plot for MMSE outcome, excluding RCTs by Yoon et al. 2017 and Venturelli et al. 2010.

Figure 11 shows the Forest Plot conducted without the study by Yoon et al.[45], indicating a total pooled estimate of 0.83 points for the MMSE outcome. The confidence interval (CI) is [-0.23; 1.88], with a corresponding p-value of 0.1240. Therefore, excluding this RCT from the analysis impacts the final pooled estimate, rendering the

treatment outcome no longer statistically significant. Continuing, the study by Venturelli et al.[43] is also removed, and the results are observed (**Figure 11.1**). The diamond is positioned at 0.32 [0.02; 0.62], with a p-value of 0.0364. The CI width is narrower compared to that of the analysis conducted on all selected studies (**Figure 2**), indicating greater precision in the effect measurement. In summary, excluding the two articles characterized by outlier results, low sample size, and poor quality, it can be concluded that the point estimate of the effect on the MMSE outcome remains significant but decreases from 1.12 to 0.32. The I^2 heterogeneity index is now reduced from 91% in the initial case to 81%.

To further enhance understanding of the variability among the included studies and provide a more comprehensive analysis of the influence on the MMSE outcome, we proceed by removing the study by Suzuki et al. [41]. At this point, this RCT stands out due to its wider CI compared to the others included, along with a smaller sample size.

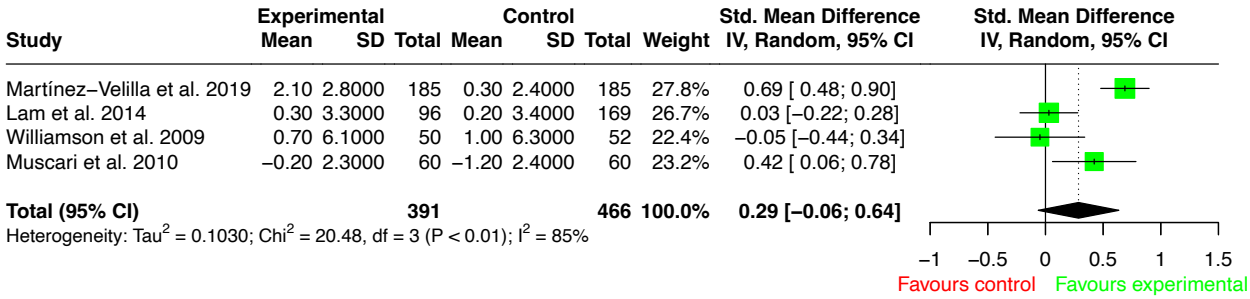

Figure 11.2 displays the Forest Plot for the MMSE outcome, excluding the RCTs by Yoon et al. 2017, Venturelli et al. 2010, and Suzuki et al. 2012.

The pooled estimate of the effect for the MMSE outcome is now 0.29 with a p-value of 0.1079 and CI [-0.06; 0.64], indicating the lack of significance of the intervention on this outcome measure. The weight of each study included in the analysis is approximately similar, hovering around 25% (**Figure 11.2**).

At this point, with the intention of further exploring the robustness of the meta-analysis and identifying sources of bias, we choose to exclude, in addition to the RCTs by Yoon et al.[45] and Venturelli et al.[43], first that of Martínez-Velilla et al.[27], and then that of Lam et al.[16], as seen in **Figure 11.1.**, they were the ones with the highest weight in the analysis, respectively 24.3% and 23.2%.

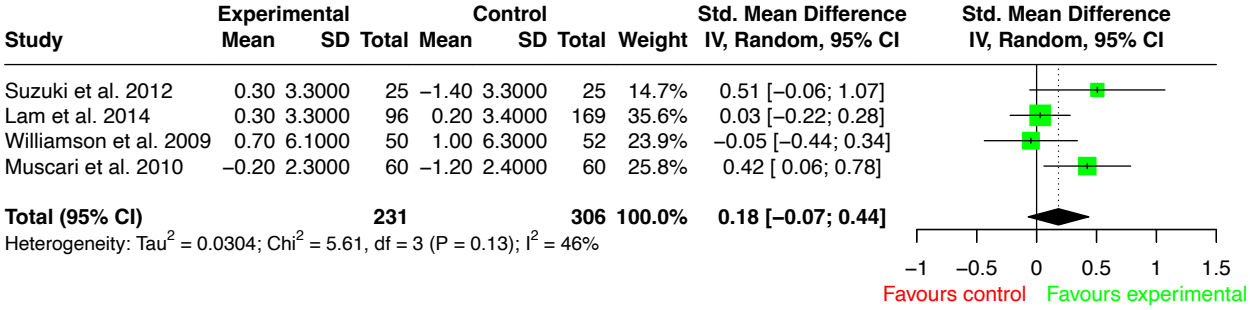

Figure 11.3. Forest Plot for MMSE outcome, without RCTs by Yoon et al. 2017, Venturelli et al. 2010, Martínez-Velilla et al. 2019.

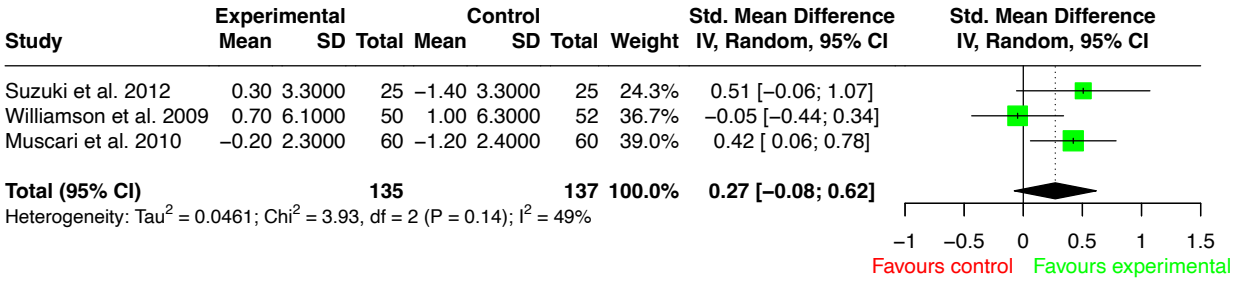

Figure 11.4. Forest Plot for MMSE outcome, without RCTs by Yoon et al. 2017, Venturelli et al. 2010, Martínez-Velilla et al. 2019, Lam et al. 2014.

As observed from the two Forest Plots (**Figure 11.3.** and **Figure 11.4**.), the I^2 index is respectively 46% and 49%, considerably lower measures compared to the previous ones, leading us to accept the hypothesis of homogeneity of the included studies in both cases. The study by Martínez-Velilla et al.[27] met high-quality standards and had a high sample size, as did that of Lam et al.[16], thus their removal likely resulted in studies that were less heterogeneous. The overall effect estimate on the MMSE outcome is not statistically significant with respective p-values of 0.1568 and 0.1263 and, consistently, narrow confidence intervals, but including zero.

We move on to the influence analysis conducted for the Logical Memory (LM) outcome. In this case, we exclude the study by Maki et al.[25], justifying this choice as, looking at **Figure 4**, its weight in creating the final effect estimate is just under 50%; furthermore, it shows a fairly high sample size, a value that differs from the other three included RCTs.

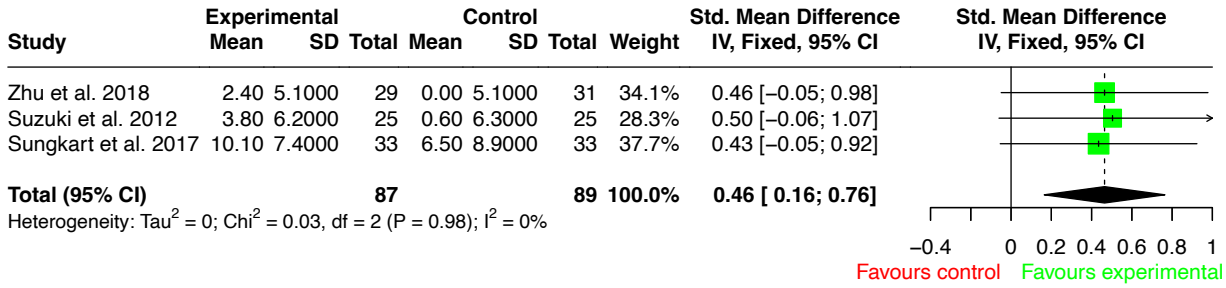

Figure 12. Forest Plot for LM outcome, without RCT: Maki et al. 2012.

The Forest Plot just above (**Figure 12**) shows the overall effect estimate on the LM outcome, after excluding the study by Maki et al.[25], of 0.46, CI [0.16; 0.76], with a p-value = 0.0024. Therefore, if the study with the highest weight is removed, the difference in effect between treated and control groups, regarding the LM outcome, remains significant, indicating the robustness of the analysis. Another observation concerns the I^2 heterogeneity index, which is now 0%.

Moving on to the influence analysis on the GDS outcome. Since the total number of included RCTs for this evaluative tool is four, we use the univariate sensitivity analytical approach, thus, one study is excluded at a time, and the overall effect of the meta-analysis is evaluated after each exclusion. It will be shown how the pooled estimate changes by removing: the study by Martínez-Velilla et al.[27], because it has a sample size that differs from the others and is the one with the highest weight percentage (27.3%); the RCT by Cassilhas et al.[1] because it reports a result that differs the most from the others (outlier).

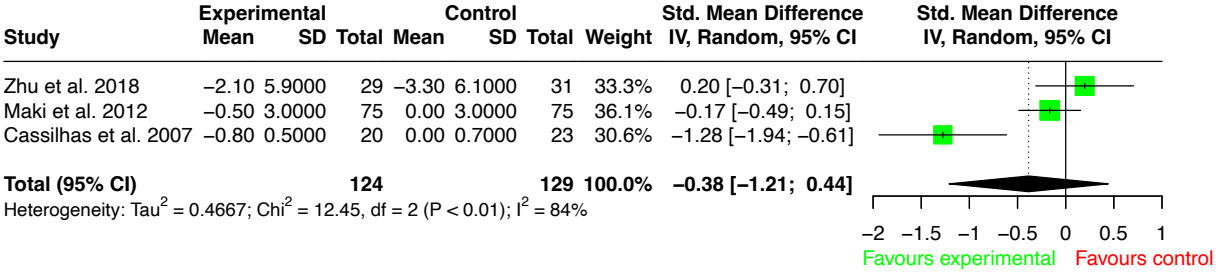

Figure 13. Forest Plot for GDS outcome, without RCT: Martínez-Velilla et al. 2019.

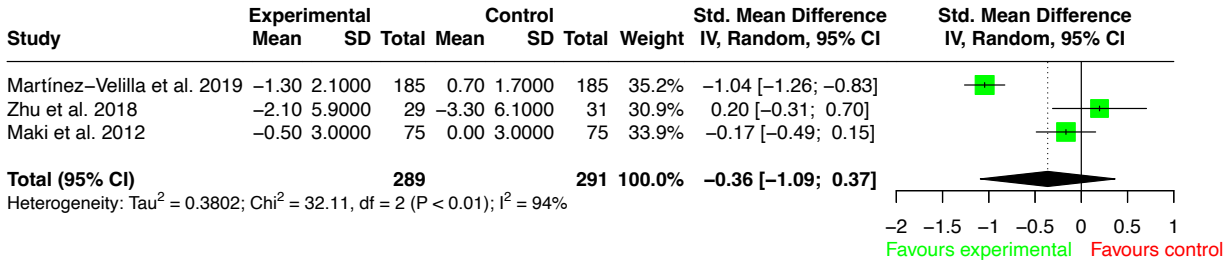

Figure 13.1. Forest Plot for GDS outcome, without RCT: Cassilhas et al. 2007.

Examining **Figure 13** and **Figure 13.1**, one can observe the lack of statistical significance of the studied effect on the GDS outcome for both resulting analyses after the removal of the two previously mentioned studies. The I^2 index decreases slightly in the case of excluding the study by Martínez-Velilla et al.[27], although showing a test that rejects the homogeneity of the studies included in the analysis. Therefore, as can be readily seen from the graphical representation of the results, the exclusion of the two RCTs does not lead to significant differences when compared with the analysis previously performed on the GDS outcome, which included all studies (**Figure 5**).

Finally, the impact of some studies on the overall effect estimate for the SPPB outcome is evaluated. The RCTs removed, one at a time, are those of: Martínez-Velilla et al.[27], because it is the most weighted in the analysis (28.2% as shown in **Figure 6**); that of Yoon et al.[45], and that of Hauer et al.[11], because both are considered of poor quality. Therefore, we want to see how the analysis varies by removing RCTs with a higher likelihood of bias (**Figure 14**, **Figure 14.1**, **Figure 14.2**).

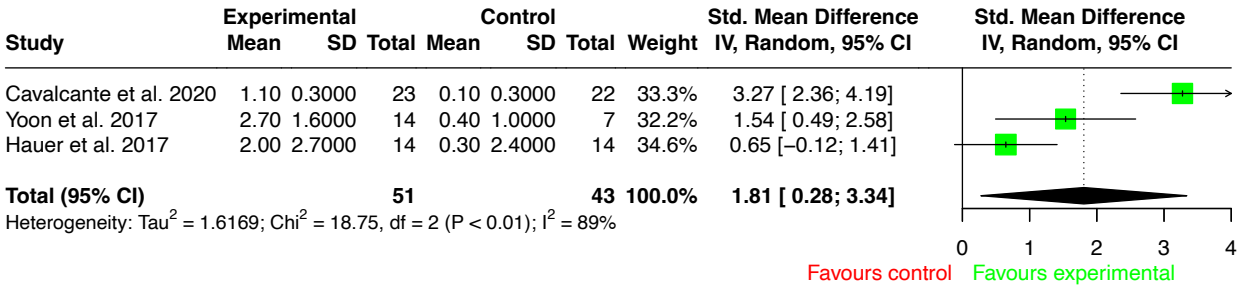

Figure 14. Forest Plot for SPPB outcome, without RCT: Martínez-Velilla et al. 2019.

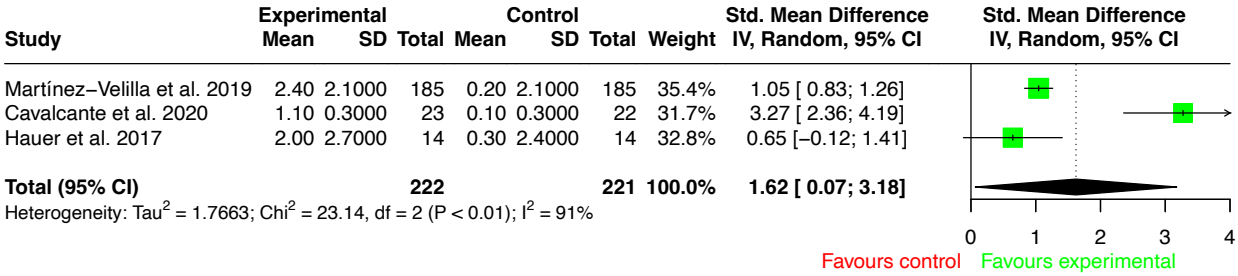

Figure 14.1. Forest Plot for SPPB outcome, without RCT: Yoon et al. 2017.

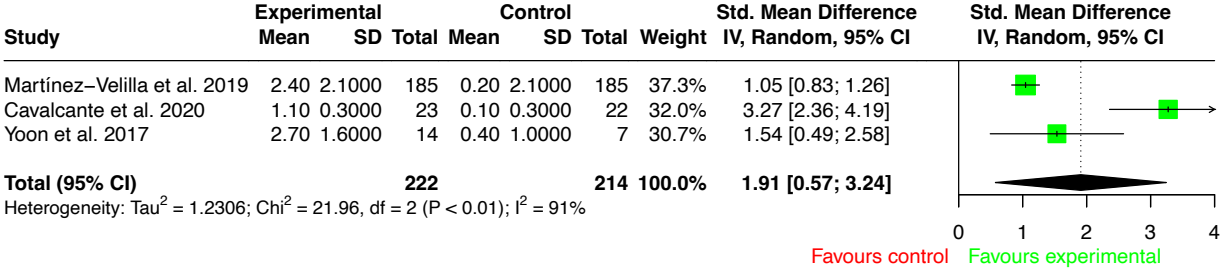

Figure 14.2. Forest Plot for SPPB outcome, without RCT: Hauer et al. 2017.

The above Forest Plots show situations very similar to that already seen (**Figure 7**) in the case of including all RCTs in the analysis, indicating the robustness of the overall effect estimate on the SPPB outcome. The p-values, in the order shown by the figures, are respectively: 0.0208, 0.0410, and 0.0051; it is interpreted that there is a statistically significant difference between the intervention group and the control group on the SPPB outcome, in favor of the treated.

3.10. Funnel Plot

When conducting a meta-analysis, there's a risk that studies with negative or non-significant results are less likely to be published compared to those with positive or significant results. This phenomenon, called publication bias, can influence the overall estimates, leading to an overestimation of treatment effects. To examine publication bias and visually display the dispersion of results from included studies, Funnel Plots are used. This graphical tool shows, on the x-axis, the treatment effect measures of the studies (standardized mean difference, SMD) and, on the y-axis, a measure of study precision, in this case, the Standard Error (SE). In the ideal case where there is no publication bias or heterogeneity, the Funnel Plot will have a funnel-shaped distribution of RCTs symmetrically around the overall effect estimate. Alternatively, it may not be symmetric

(e.g., if there is an excess of studies with negative/positive results) or indicate outlier studies, points that significantly deviate from the rest of the Funnel Plot.

In the present meta-analysis, the Funnel Plot was created using R software (version 4.2.2.) only for the MMSE outcome because a minimum threshold of five studies was used for this graphical representation. It should be noted that the interpretation of the Funnel Plot becomes less reliable with few studies. This is because with a limited number of studies, the plot may not provide an accurate representation of the expected dispersion of results around the effect estimate, showing an unstable representation subject to random variation, which may not reflect true publication bias or heterogeneity problems.

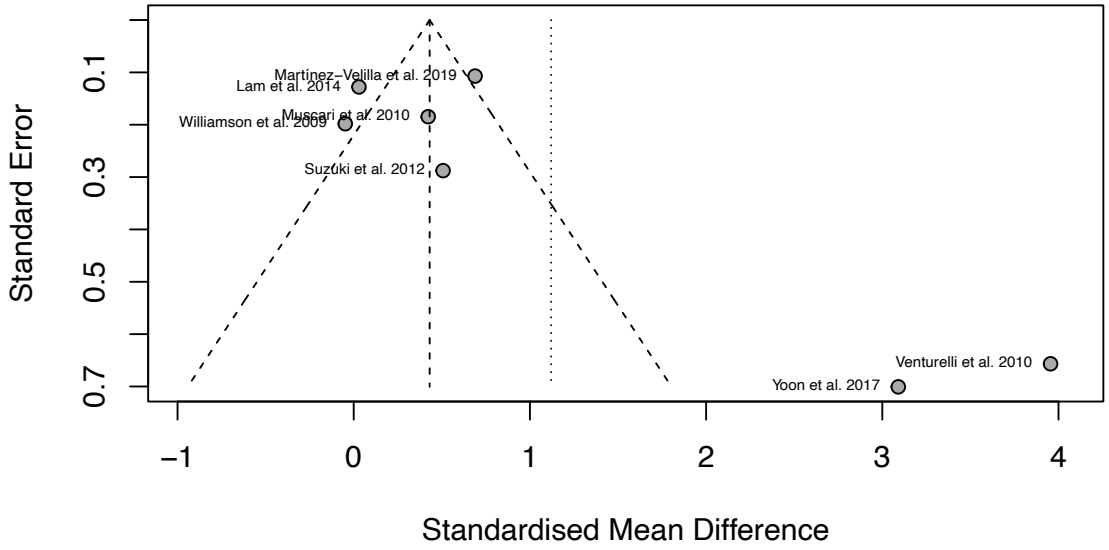

Figure 15. Funnel Plot, MMSE outcome.

From the Funnel Plot for the MMSE outcome (**Figure 15.**), the presence of two outlier studies is noticeable, with a treatment effect measure clearly in favor of the treatment group and high standard errors (indicating low precision), namely, Yoon et al.[45] and Venturelli et al.[43]. Looking at the distribution of the other points, a slight leftward asymmetry is observed, with the studies of Lam et al.[16] and Williamson et al.[44], which were not significant. Two statistical tests are used to assess the presence of asymmetry in the Funnel Plot: the Egger's test and Begg and Mazumdar's test. The p-values are respectively: 0.1631 and 0.1765. It is interpreted by accepting the null hypothesis, concluding that there is no significant evidence of asymmetry in the Funnel Plot. To complete the analysis and critically evaluate the results obtained from the tests, contextualizing them in the shown Funnel Plot, the "*Trim and Fill*" procedure is chosen.

This statistical method is implemented to address publication bias in meta-analyses, correcting for any overestimation of treatment effects by assessing how many studies might be missing from the Funnel Plot. The process involves two stages: (1) Trimming, where studies that appear to create asymmetry in the Funnel Plot are temporarily removed; (2) Filling, where "missing studies" are added to the truncated part of the plot, assuming they were omitted due to publication bias; these are speculative studies used to correct the asymmetry. A new corrected Funnel Plot is then produced.

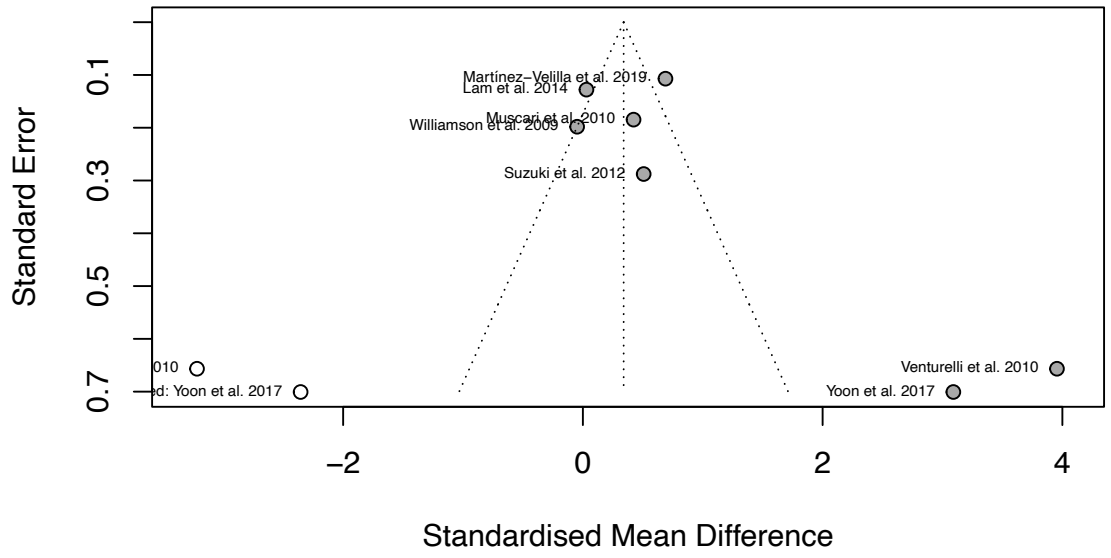

Figure 15.1. Funnel Plot, MMSE outcome with "Trim and Fill" method.

Figure 15.1. shows the result of the "Trim and Fill" procedure, with two speculative studies added in the bottom left, mirroring those of Venturelli et al.[43] and Yoon et al.[45], marked with two white circles. These two "imaginary" RCTs are inserted to account for potential publication bias, resulting in a more balanced and symmetric Funnel Plot. At this point, the corrected final effect estimate is calculated, taking into account the initial overestimation due to bias: the standardized mean difference on the MMSE outcome resulting from the "Trim and Fill" procedure is now 0.3401 with a 95% confidence interval of [-1.0681; 1.7483], with a p-value = 0.6359. Therefore, the overall effect estimate and its statistical significance change when considering publication bias. However, it's important to note that the "Trim and Fill" method is an approximate procedure and has been included for completeness of the analysis; it should still be interpreted with caution.

4. DISCUSSION

In this meta-analysis, the effect of physical activity, in its broadest sense, including it combined with cognitive training (multi-intervention case), on the cognitive and functional performance of hospitalized elderly individuals was examined.

Twenty-nine studies enrolling a total of 4574 patients were identified. Just under half of the RCTs were conducted on patients with Mild Cognitive Impairment (MCI), a condition associated with an increased risk of developing dementia and mortality.

Analyzing the results, it can be concluded that physical activity leads to a significant improvement in some of the considered outcomes: on the MMSE scale, on "Logical Memory," on "Gait Speed," and on the "Short Physical Performance Battery" (SPPB). Conversely, no statistically significant differences were found in the MoCA outcome and on the "Geriatric Depression Scale" (GDS).

Therefore, physical exercise shows a potential beneficial effect in hospitalized elderly patients regarding cognitive functions (MMSE and LM) and partially cognitive functions (GS and SPPB).

The MoCA scale does not show a statistically significant change; however, this should not be seen as a symptom of inconsistency in the results. It should be noted that the two cognitive assessment scales (MMSE and MoCA) have different specific requirements: the former is preferred for its practicality and simplicity, while the latter is more sensitive and detailed. Thus, the MMSE scale evaluates: orientation, registration, attention and calculation, repetition, language, and copying a geometric figure; while the MoCA scale assesses: memory, executive function, attention, language, visuospatial abilities, and temporal and spatial orientation.

Simultaneously, a comparison with the current landscape of scientific literature is proposed. The systematic review by Kelly et al.[15] examines the impact of aerobic exercise, resistance training, and Tai Chi on the cognitive function of elderly individuals without known cognitive impairment. This study included twenty-five randomized controlled trials and created as many pooled estimates as intervention types (the three listed above). Significant improvements were observed for resistance training-treated individuals compared to controls on measures of reasoning (p-value < 0.005); furthermore, for the Tai Chi group, significant increases were found in attention measures (p-value < 0.001) and processing speed (p-value < 0.00001). Moving forward, the

systematic review by Özbe et al.[33] examines studies using standardized multicomponent group interventions (physical activity + cognitive training) for people with MCI or dementia. Nine studies were included, with a total sample size of 513 participants; three studies had a positive effect on cognitive outcome measures (MMSE, ADAS-Cog, and "other"). Finally, Farhang et al.[8] studied the impact of mind-body interventions (multifactorial: physical activity and mindfulness) and reported improvement in cognitive function, daily activities, memory, resilience, and awareness in elderly individuals with MCI. It should be noted that this last review did not result in pooled estimates but rather a systematic comparison of the retrieved studies.

Now, let's emphasize the strengths of the present study. This meta-analysis combined the results of multiple studies, allowing for increased statistical power and a consequent reduction in the risk of random error, generating a broader and clearer view of the topic at hand. Due to the significant heterogeneity among the included RCTs, techniques were used to control it, such as creating random-effects models and designing meta-regressions, searching for possible effect modifiers, which were found to be significant. A sensitivity analysis was proposed to disclose the effect of one or more trials on the pooled estimate. The possible presence of publication bias was considered, as evidenced by the Funnel Plot.

The reader must also consider some limitations. Firstly, systematic searches were not conducted in different databases but rather a targeted search using Mesh Terms on PubMed. Furthermore, despite the fairly promising results of this meta-analysis, some of the available studies involved samples ranging from small to moderate sizes, only one study was multicentric, and problems were encountered regarding the comparison of the measures used to assess cognition. In fact, **Table 4**, in the "Other" section, shows that some of the studies did not enter the final pooled estimate due to the presence of compliant but non-comparable outcomes with the other RCTs. Furthermore, it is worth emphasizing the strong presence of heterogeneity among the studies that participated in this meta-analysis. The causes of this phenomenon are found in patient characteristics; intervention characteristics; follow-up characteristics; specific study designs, and so forth. Therefore, the issue of the generalizability of these results cannot be fully resolved. Moreover, long follow-up periods would be necessary, as well as studies assessing whether cognitive improvement is sustained after the conclusion of clinical experimentation. Without these studies, it is unclear whether physical exercise treatment has a long-term effect on

cognition or functional abilities in hospitalized elderly individuals or whether the effect is limited to the context of a clinical trial.

Therefore, for future research, two strategies should be implemented. Firstly, there is a need to standardize the measures used for cognitive function to make scientific research on cognitive outcomes less scattered and more homogeneous, currently assessed through a large number of scales and tools. Secondly, exercise and physical activity programs, combined with social interaction, should be implemented in the clinical routine of hospitalized and outpatient elderly patients. This approach will improve cognitive functioning in elderly patients, also contributing to better performance in partially cognitive domains, thereby reducing the likelihood of rehospitalization and the risk of developing dementia, which are positively correlated with the risk of early mortality. Therefore, it is believed that physical exercise in elderly patients should not be limited to clinical studies. On the other hand, the awareness of the difficulty in implementing such programs is clear. The World Health Organization's Global Recommendations on Physical Activity for Health address all age groups, promoting policies to increase physical activity levels and discouraging physical inactivity, which is becoming increasingly common worldwide. According to the WHO guidelines of 2020, individuals over 65 should engage in at least 150-300 minutes of moderate-intensity physical activity every week, or at least 75-150 minutes of vigorous-intensity physical activity, or an equivalent combination of the two. It is also recommended to include muscle-strengthening exercises for major muscle groups, 2 or more times a week, as well as multicomponent physical activity (aerobic activity, muscle strengthening, and balance training) performed in a single session at least 3 days a week to increase functional capacity and reduce the risk of falls.

The WHO guidelines were adopted in Italy in 2021. The Ministry of Health has developed "Guidelines on Physical Activity: Review of recommendations for different age groups and physiological situations and new recommendations for specific pathologies". In the "Elderly" section, what was stated by the WHO is adopted, specifying that individuals in this age group who cannot reach the recommended levels due to their health conditions should still adopt an active lifestyle and engage in low-intensity activities, within their capabilities and conditions. Furthermore, a section "For the healthcare setting" is reserved, sensitizing healthcare professionals and others to encourage their sedentary counterparts to change their lifestyle, introducing a sufficient amount of physical activity.

In summary, physical exercise has a positive impact on cognitive functions (MMSE and LM) and partially cognitive functions (SPPB and GS); however, further research is still needed. Efforts should be made to standardize cognitive outcomes to clarify and make the overall results more precise and to increase physical activity in the daily lives of hospitalized elderly patients.

BIBLIOGRAPHY

1. Cassilhas RC, Viana VA, Grassmann V, Santos RT, Santos RF, Tufik S, Mello MT. The impact of resistance exercise on the cognitive function of the elderly. *Med Sci Sports Exerc*. 2007 Aug;39(8):1401-7. doi: 10.1249/mss.0b013e318060111f. PMID: 17762374.

2. Cavalcante BR, de Souza MF, Falck RS, Liu-Ambrose T, Behm DG, Pitangui ACR, de Araújo RC. Effects of Resistance Exercise with Instability on Cognitive Function (REI Study): A Proof-Of-Concept Randomized Controlled Trial in Older Adults with Cognitive Complaints. *J Alzheimers Dis.* 2020;77(1):227-239. doi: 10.3233/JAD-200349. PMID: 32804132.

3. Chan WC, Yeung JW, Wong CS, Lam LC, Chung KF, Luk JK, Lee JS, Law AC. Efficacy of physical exercise in preventing falls in older adults with cognitive impairment: a systematic review and meta-analysis. *J Am Med Dir Assoc.* 2015 Feb;16(2):149-54. doi: 10.1016/j.jamda.2014.08.007. Epub 2014 Oct 7. PMID: 25304179.

4. Chen MC, Chen KM, Chang CL, Chang YH, Cheng YY, Huang HT. Elastic Band Exercises Improved Activities of Daily Living and Functional Fitness of Wheelchair-bound Older Adults with Cognitive Impairment: A Cluster Randomized Controlled Trial. *Am J Phys Med Rehabil.* 2016 Nov;95(11):789-799. doi: 10.1097/PHM.0000000000000518. PMID: 27149585.

5. Choi W, Lee S. The Effects of Virtual Kayak Paddling Exercise on Postural Balance, Muscle Performance, and Cognitive Function in Older Adults with Mild Cognitive Impairment: A Randomized Controlled Trial. *J Aging Phys Act.* 2019 Dec 1;27(4):861-870. doi: 10.1123/japa.2018-0020. PMID: 31185775.

6. Denkinger MD, Nikolaus T, Denkinger C, Lukas A. Physical activity for the prevention of cognitive decline: current evidence from observational and controlled studies. *Z Gerontol Geriatr.* 2012 Jan;45(1):11-6. doi: 10.1007/s00391-011-0262-6. PMID: 22278001.

7. Devenney KE, Sanders ML, Lawlor B, Olde Rikkert MGM, Schneider S; NeuroExercise Study Group. The effects of an extensive exercise programme on the

progression of Mild Cognitive Impairment (MCI): study protocol for a randomised controlled trial. *BMC Geriatr.* 2017 Mar 22;17(1):75. doi: 10.1186/s12877-017-0457-9. Erratum in: BMC Geriatr. 2017 May 19;17 (1):112. PMID: 28330458; PMCID: PMC5361785.

8. Farhang M, Miranda-Castillo C, Rubio M, Furtado G. Impact of mind-body interventions in older adults with mild cognitive impairment: a systematic review. *Int Psychogeriatr.* 2019 May;31(5):643-666. doi: 10.1017/S1041610218002302. Epub 2019 Feb 4. PMID: 30712518.

9. Geda YE. Mild cognitive impairment in older adults. *Curr Psychiatry Rep.* 2012 Aug;14(4):320-7. doi: 10.1007/s11920-012-0291-x. PMID: 22773365; PMCID: PMC3963488.

10. Hagovská M, Takáč P, Dzvoník O. Effect of a combining cognitive and balanced training on the cognitive, postural and functional status of seniors with a mild cognitive deficit in a randomized, controlled trial. *Eur J Phys Rehabil Med.* 2016 Feb;52(1):101-9. PMID: 26325026.

11. Hauer K, Ullrich P, Dutzi I, Beurskens R, Kern S, Bauer J, Schwenk M. Effects of Standardized Home Training in Patients with Cognitive Impairment following Geriatric Rehabilitation: A Randomized Controlled Pilot Study. *Gerontology.* 2017;63(6):495-506. doi: 10.1159/000478263. Epub 2017 Aug 17. PMID: 28813696.

12. Hong YJ, Jang EH, Hwang J, Roh JH, Lee JH. The Efficacy of Cognitive Intervention Programs for Mild Cognitive Impairment: A Systematic Review. Curr Alzheimer Res. 2015;12(6):527-42. doi: 10.2174/1567205012666150530201636. PMID: 26027815.

13. Jirayucharoensak S, Israsena P, Pan-Ngum S, Hemrungrojn S, Maes M. A game-based neurofeedback training system to enhance cognitive performance in healthy elderly subjects and in patients with amnestic mild cognitive impairment. *Clin Interv Aging.* 2019 Feb 19;14:347-360. doi: 10.2147/CIA.S189047. PMID: 30863028; PMCID: PMC6388796.

14. Kamegaya T, Araki Y, Kigure H; Long-Term-Care Prevention Team of Maebashi City; Yamaguchi H. Twelve-week physical and leisure activity programme improved cognitive

function in community-dwelling elderly subjects: a randomized controlled trial. Psychogeriatrics. 2014 Mar;14(1):47-54. doi: 10.1111/psyg.12038. Epub 2014 Feb 16. PMID: 24528600.

15. Kelly ME, Loughrey D, Lawlor BA, Robertson IH, Walsh C, Brennan S. The impact of exercise on the cognitive functioning of healthy older adults: a systematic review and meta-analysis. *Ageing Res Rev.* 2014 Jul; 16:12-31. doi: 10.1016/j.arr.2014.05.002. Epub 2014 May 23. PMID: 24862109.

16. Lam LC, Chan WM, Kwok TC, Chiu HF. Effectiveness of Tai Chi in maintenance of cognitive and functional abilities in mild cognitive impairment: a randomised controlled trial. *Hong Kong Med J.* 2014 Jun;20(3 Suppl 3):20-3. PMID: 25001031.

17. Langlois F, Vu TT, Chassé K, Dupuis G, Kergoat MJ, Bherer L. Benefits of physical exercise training on cognition and quality of life in frail older adults. *J Gerontol B Psychol Sci Soc Sci.* 2013 May;68(3):400-4. doi: 10.1093/geronb/gbs069. Epub 2012 Aug 28. PMID: 22929394.

18. Law LLF, Mok VCT, Yau MMK. Effects of functional tasks exercise on cognitive functions of older adults with mild cognitive impairment: a randomized controlled pilot trial. *Alzheimers Res Ther.* 2019 Dec 4;11(1):98. doi: 10.1186/s13195-019-0548-2. PMID: 31801630; PMCID: PMC6894271.

19. Lee WJ, Peng LN, Lin MH, Lin CH, Chen LK. Clinical Efficacy of Multidomain Interventions among Multimorbid Older People Stratified by the Status of Physio-Cognitive Declines: A Secondary Analysis from the Randomized Controlled Trial for *Healthy Aging. J Nutr Health Aging.* 2022;26(10):909-917. doi: 10.1007/s12603-022-1843-3. PMID: 36259579.

20. Lee DW, Yoon DH, Lee JY, Panday SB, Park J, Song W. Effects of High-Speed Power Training on Neuromuscular and Gait Functions in Frail Elderly with Mild Cognitive Impairment Despite Blunted Executive Functions: A Randomized Controlled Trial. *J Frailty Aging.* 2020;9(3):179-184. doi: 10.14283/jfa.2020.23. PMID: 32588034.

21. Li B, Tang H, He G, Jin Z, He Y, Huang P, He N, Chen S. Tai Chi enhances cognitive training effects on delaying cognitive decline in mild cognitive impairment. *Alzheimers*

Dement. 2023 Jan;19(1):136-149. doi: 10.1002/alz.12658. Epub 2022 Mar 15. PMID: 35290704.

22. Liu-Ambrose T, Nagamatsu LS, Graf P, Beattie BL, Ashe MC, Handy TC. Resistance training and executive functions: a 12-month randomized controlled trial. *Arch Intern Med.* 2010 Jan 25;170(2):170-8. doi: 10.1001/archinternmed.2009.494. PMID: 20101012; PMCID: PMC3448565.

23. Lü J, Sun M, Liang L, Feng Y, Pan X, Liu Y. Effects of momentum-based dumbbell training on cognitive function in older adults with mild cognitive impairment: a pilot randomized controlled trial. *Clin Interv Aging.* 2015 Dec 22;11:9-16. doi: 10.2147/CIA.S96042. PMID: 26766905; PMCID: PMC4699540.

24. Maci T, Pira FL, Quattrocchi G, Nuovo SD, Perciavalle V, Zappia M. Physical and cognitive stimulation in Alzheimer Disease. the GAIA Project: a pilot study. *Am J Alzheimers Dis Other Demen.* 2012 Mar;27(2):107-13. doi: 10.1177/1533317512440493. PMID: 22495338.

25. Maki Y, Ura C, Yamaguchi T, Murai T, Isahai M, Kaiho A, Yamagami T, Tanaka S, Miyamae F, Sugiyama M, Awata S, Takahashi R, Yamaguchi H. Effects of intervention using a community-based walking program for prevention of mental decline: a randomized controlled trial. *J Am Geriatr Soc.* 2012 Mar;60(3):505-10. doi: 10.1111/j.1532-5415.2011.03838.x. Epub 2012 Jan 30. PMID: 22288578.

26. Mallya, S., Fiocco, A.J. Effects of Mindfulness Training on Cognition and Well-Being in Healthy Older Adults. *Mindfulness* 7, 453–465 (2016). https://doi.org/10.1007/s12671-015-0468-6.

27. Martínez-Velilla N, Casas-Herrero A, Zambom-Ferraresi F, Sáez de Asteasu ML, Lucia A, Galbete A, García-Baztán A, Alonso-Renedo J, González-Glaría B, Gonzalo-Lázaro M, Apezteguía Iráizoz I, Gutiérrez-Valencia M, Rodríguez-Mañas L, Izquierdo M. Effect of Exercise Intervention on Functional Decline in Very Elderly Patients During Acute Hospitalization: A Randomized Clinical Trial. *JAMA Intern Med.* 2019 Jan 1;179(1):28-36. doi: 10.1001/jamainternmed.2018.4869. Erratum in: JAMA Intern Med. 2019 Jan 1;179(1):127. PMID: 30419096; PMCID: PMC6583412.

28. Maurus I, Röh A, Falkai P, Malchow B, Schmitt A, Hasan A. Nonpharmacological treatment of dyscognition in schizophrenia: effects of aerobic exercise. *Dialogues Clin Neurosci.* 2019 Sep;21(3):261-269. doi: 10.31887/DCNS.2019.21.3/aschmitt. PMID: 31749650; PMCID: PMC6829165.

29. MINISTERO DELLA SALUTE (2022), Normativa dell'area "Stili di Vita – Guadagnare salute".

30. Muscari A, Giannoni C, Pierpaoli L, Berzigotti A, Maietta P, Foschi E, Ravaioli C, Poggiopollini G, Bianchi G, Magalotti D, Tentoni C, Zoli M. Chronic endurance exercise training prevents aging-related cognitive decline in healthy older adults: a randomized controlled trial. *Int J Geriatr Psychiatry.* 2010 Oct;25(10):1055-64. doi: 10.1002/gps.2462. PMID: 20033904.

31. Oken BS, Zajdel D, Kishiyama S, Flegal K, Dehen C, Haas M, Kraemer DF, Lawrence J, Leyva J. Randomized, controlled, six-month trial of yoga in healthy seniors: effects on cognition and quality of life. *Altern Ther Health Med.* 2006 Jan-Feb;12(1):40-7. PMID: 16454146; PMCID: PMC1457100.

32. Orgeta V, Qazi A, Spector A, Orrell M. Psychological treatments for depression and anxiety in dementia and mild cognitive impairment: systematic review and meta-analysis. *Br J Psychiatry.* 2015 Oct;207(4):293-8. doi: 10.1192/bjp.bp.114.148130. PMID: 26429684; PMCID: PMC4589662.

33. Özbe D, Graessel E, Donath C, Pendergrass A. Immediate Intervention Effects of Standardized Multicomponent Group Interventions on People with Cognitive Impairment: A Systematic Review. *J Alzheimers Dis.* 2019;67(2):653-670. doi: 10.3233/JAD-180980. PMID: 30689588; PMCID: PMC6398841.

34. Rosenberg PB, Mielke MM, Appleby BS, Oh ES, Geda YE, Lyketsos CG. The association of neuropsychiatric symptoms in MCI with incident dementia and Alzheimer disease. *Am J Geriatr Psychiatry.* 2013 Jul;21(7):685-95. doi: 10.1016/j.jagp.2013.01.006. Epub 2013 Feb 6. PMID: 23567400; PMCID: PMC3428504.

35. Sanford AM. Mild Cognitive Impairment. *Clin Geriatr Med.* 2017 Aug;33(3):325-337. doi: 10.1016/j.cger.2017.02.005. Epub 2017 May 17. PMID: 28689566.

36. Sink KM, Espeland MA, Castro CM, Church T, Cohen R, Dodson JA, Guralnik J, Hendrie HC, Jennings J, Katula J, Lopez OL, McDermott MM, Pahor M, Reid KF, Rushing J, Verghese J, Rapp S, Williamson JD; LIFE Study Investigators. Effect of a 24-Month Physical Activity Intervention vs Health Education on Cognitive Outcomes in Sedentary Older Adults: The LIFE Randomized Trial. *JAMA.* 2015 Aug 25;314(8):781-90. doi: 10.1001/jama.2015.9617. PMID: 26305648; PMCID: PMC4698980.

37. Siu MY, Lee DTF. Effects of tai chi on cognition and instrumental activities of daily living in community dwelling older people with mild cognitive impairment. *BMC Geriatr.* 2018 Feb 2;18(1):37. doi: 10.1186/s12877-018-0720-8. PMID: 29394884; PMCID: PMC5797349.

38. Smiley-Oyen AL, Lowry KA, Francois SJ, Kohut ML, Ekkekakis P. Exercise, fitness, and neurocognitive function in older adults: the "selective improvement" and "cardiovascular fitness" hypotheses. *Ann Behav Med*. 2008 Dec;36(3):280-91. doi: 10.1007/s12160-008-9064-5. Epub 2008 Sep 30. PMID: 18825471; PMCID: PMC2748860.

39. Solfrizzi V, D'Introno A, Colacicco AM, Capurso C, Del Parigi A, Caselli RJ, Scapicchio PL, Scafato E, Gandin C, Capurso A, Panza F; Italian Longitudinal Study on Aging Working Group. Incident occurrence of depressive symptoms among patients with mild cognitive impairment - the Italian longitudinal study on aging. *Dement Geriatr Cogn Disord.* 2007;24(1):55-64. doi: 10.1159/000103632. Epub 2007 Jun 11. PMID: 17565214.

40. Sungkarat S, Boripuntakul S, Chattipakorn N, Watcharasaksilp K, Lord SR. Effects of Tai Chi on Cognition and Fall Risk in Older Adults with Mild Cognitive Impairment: A Randomized Controlled Trial. *J Am Geriatr Soc*. 2017 Apr;65(4):721-727. doi: 10.1111/jgs.14594. Epub 2016 Nov 22. PMID: 27874176.

41. Suzuki T, Shimada H, Makizako H, Doi T, Yoshida D, Tsutsumimoto K, Anan Y, Uemura K, Lee S, Park H. Effects of multicomponent exercise on cognitive function in older adults with amnestic mild cognitive impairment: a randomized controlled trial.

BMC Neurol. 2012 Oct 31;12:128. doi: 10.1186/1471-2377-12-128. PMID: 23113898; PMCID: PMC3534485.

42. Train the Brain Consortium. Randomized trial on the effects of a combined physical/cognitive training in aged MCI subjects: the Train the Brain study. *Sci Rep.* 2017 Jan 3;7:39471. doi: 10.1038/srep39471. PMID: 28045051; PMCID: PMC5206718.

43. Venturelli M, Lanza M, Muti E, Schena F. Positive effects of physical training in activity of daily living-dependent older adults. *Exp Aging Res.* 2010 Apr;36(2):190-205. doi: 10.1080/03610731003613771. PMID: 20209421.

44. Williamson JD, Espeland M, Kritchevsky SB, Newman AB, King AC, Pahor M, Guralnik JM, Pruitt LA, Miller ME; LIFE Study Investigators. Changes in cognitive function in a randomized trial of physical activity: results of the lifestyle interventions and independence for elders pilot study. *J Gerontol A Biol Sci Med Sci*. 2009 Jun;64(6):688-94. doi: 10.1093/gerona/glp014. Epub 2009 Feb 24. PMID: 19244157; PMCID: PMC2679423.

45. Yoon DH, Kang D, Kim HJ, Kim JS, Song HS, Song W. Effect of elastic band-based high-speed power training on cognitive function, physical performance and muscle strength in older women with mild cognitive impairment. *Geriatr Gerontol Int.* 2017 May;17(5):765-772. doi: 10.1111/ggi.12784. Epub 2016 Jul 10. PMID: 27396580.

46. Young DK. Multicomponent intervention combining a cognitive stimulation group and tai chi to reduce cognitive decline among community-dwelling older adults with probable dementia: A multi-center, randomized controlled trial. *Dementia (London)*. 2020 Aug;19(6):2073-2089. doi: 10.1177/1471301218814637. Epub 2018 Nov 28. PMID: 30486656.

47. Zaccaria D., NNA, *Network Non Autosufficienza.* "L'invecchiamento della popolazione in Europa: trend e sfide per il futuro.", 2015.

48. Zhu Y, Wu H, Qi M, Wang S, Zhang Q, Zhou L, Wang S, Wang W, Wu T, Xiao M, Yang S, Chen H, Zhang L, Zhang KC, Ma J, Wang T. Effects of a specially designed aerobic dance routine on mild cognitive impairment. *Clin Interv Aging.* 2018 Sep

11;13:1691-1700. doi: 10.2147/CIA.S163067. PMID: 30237705; PMCID: PMC6138969.

APPENDICES

Appendices 1. Multivariate model for MMSE outcome, Method: "REML".

```
Mixed-Effects Model (k = 7; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0 (SE = 0.0393)
tau (square root of estimated tau^2 value):             0
I^2 (residual heterogeneity / unaccounted variability): 0.00%
H^2 (unaccounted variability / sampling variability):   1.00
R^2 (amount of heterogeneity accounted for):            100.00%

Test for Residual Heterogeneity:
QE(df = 1) = 0.1085, p-val = 0.7419

Test of Moderators (coefficients 2:6):
QM(df = 5) = 64.6808, p-val < .0001

Model Results:

                estimate      se     zval    pval     ci.lb    ci.ub
intrcpt          12.6220  1.9654   6.4221  <.0001    8.7699  16.4741  ***
MCInoMCI        -10.9845  2.3307  -4.7130  <.0001  -15.5526  -6.4165  ***
controlpassive   -0.9487  1.0193  -0.9307  0.3520   -2.9465   1.0492
regionnonEurope -10.9002  2.4528  -4.4439  <.0001  -15.7077  -6.0927  ***
FU                0.7725  0.2623   2.9447  0.0032    0.2583   1.2866   **
gendermale      -10.4843  2.4155  -4.3405  <.0001  -15.2185  -5.7501  ***
```

Appendices 1a. Multivariate model for MMSE outcome, Method: "Sidik Jonkman".

```
Mixed-Effects Model (k = 7; tau^2 estimator: SJ)

tau^2 (estimated amount of residual heterogeneity):     0.0030 (SE = 0.0430)
tau (square root of estimated tau^2 value):             0.0546
I^2 (residual heterogeneity / unaccounted variability): 9.67%
H^2 (unaccounted variability / sampling variability):   1.11
R^2 (amount of heterogeneity accounted for):            99.87%

Test for Residual Heterogeneity:
QE(df = 1) = 0.1085, p-val = 0.7419

Test of Moderators (coefficients 2:6):
QM(df = 5) = 61.2757, p-val < .0001

Model Results:

                estimate      se     zval    pval     ci.lb    ci.ub
intrcpt          12.6236  1.9764   6.3871  <.0001    8.7499  16.4973  ***
MCInoMCI        -10.9861  2.3482  -4.6785  <.0001  -15.5885  -6.3836  ***
controlpassive   -0.9487  1.0251  -0.9254  0.3548   -2.9579   1.0606
regionnonEurope -10.9017  2.4755  -4.4038  <.0001  -15.7537  -6.0498  ***
FU                0.7725  0.2648   2.9168  0.0035    0.2534   1.2915   **
gendermale      -10.4843  2.4351  -4.3055  <.0001  -15.2570  -5.7116  ***
```

Appendices 2. Multivariate model for GDS outcome.

```
Mixed-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0 (SE = 0.0533)
tau (square root of estimated tau^2 value):             0
I^2 (residual heterogeneity / unaccounted variability): 0.00%
H^2 (unaccounted variability / sampling variability):   1.00
R^2 (amount of heterogeneity accounted for):            100.00%

Test for Residual Heterogeneity:
QE(df = 2) = 1.8272, p-val = 0.4011

Test of Moderators (coefficient 2):
QM(df = 1) = 33.3998, p-val < .0001

Model Results:

                estimate      se      zval    pval    ci.lb    ci.ub
intrcpt          -1.0672  0.1054  -10.1271  <.0001  -1.2737  -0.8607  ***
regionnonEurope   1.0049  0.1739    5.7793  <.0001   0.6641   1.3457  ***
```

Appendices 2a. Multivariate model for GDS outcome, IC: “Hartung-Knapp”.

```
tau^2 (estimated amount of residual heterogeneity):     0 (SE = 0.0533)
tau (square root of estimated tau^2 value):             0
I^2 (residual heterogeneity / unaccounted variability): 0.00%
H^2 (unaccounted variability / sampling variability):   1.00
R^2 (amount of heterogeneity accounted for):            100.00%

Test for Residual Heterogeneity:
QE(df = 2) = 1.8272, p-val = 0.4011

Test of Moderators (coefficient 2):
F(df1 = 1, df2 = 2) = 36.5585, p-val = 0.0263

Model Results:

                estimate      se      tval  df    pval    ci.lb    ci.ub
intrcpt          -1.0672  0.1007  -10.5952   2  0.0088  -1.5006  -0.6338  **
regionnonEurope   1.0049  0.1662    6.0464   2  0.0263   0.2898   1.7200   *
```

Appendices 3. Multivariate model for SPPB outcome.

```
Mixed-Effects Model (k = 4; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0.0000 (SE = 0.0955)
tau (square root of estimated tau^2 value):             0.0025
I^2 (residual heterogeneity / unaccounted variability): 0.01%
H^2 (unaccounted variability / sampling variability):   1.00
R^2 (amount of heterogeneity accounted for):            100.00%

Test for Residual Heterogeneity:
QE(df = 2) = 1.8939, p-val = 0.3879

Test of Moderators (coefficient 2):
QM(df = 1) = 21.8133, p-val < .0001

Model Results:

                estimate      se     zval    pval    ci.lb    ci.ub
intrcpt           3.2748  0.4679   6.9992  <.0001   2.3578   4.1918  ***
controlpassive   -2.2392  0.4794  -4.6705  <.0001  -3.1789  -1.2995  ***
```

Appendices 3a. Multivariate model for SPPB outcome, IC: “Hartung-Knapp”.

```
tau^2 (estimated amount of residual heterogeneity):     0.0000 (SE = 0.0955)
tau (square root of estimated tau^2 value):             0.0025
I^2 (residual heterogeneity / unaccounted variability): 0.01%
H^2 (unaccounted variability / sampling variability):   1.00
R^2 (amount of heterogeneity accounted for):            100.00%

Test for Residual Heterogeneity:
QE(df = 2) = 1.8939, p-val = 0.3879

Test of Moderators (coefficient 2):
F(df1 = 1, df2 = 2) = 23.0359, p-val = 0.0408

Model Results:

                estimate      se     tval  df    pval    ci.lb    ci.ub
intrcpt           3.2748  0.4553   7.1926   2  0.0188   1.3158   5.2338  *
controlpassive   -2.2392  0.4665  -4.7996   2  0.0408  -4.2466  -0.2318  *
```