Psychology & Health

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gpsh20

Which cognitive tests are used to examine the acute effect of physical activity on cognition in healthy adults aged 50 and older? - A systematic review

Fien De Block, Delfien Van Dyck, Benedicte Deforche, Geert Crombez & Louise Poppe

To cite this article: Fien De Block, Delfien Van Dyck, Benedicte Deforche, Geert Crombez & Louise Poppe (08 May 2025): Which cognitive tests are used to examine the acute effect of physical activity on cognition in healthy adults aged 50 and older? - A systematic review, Psychology & Health, DOI: 10.1080/08870446.2025.2498577

To link to this article: https://doi.org/10.1080/08870446.2025.2498577

Published online: 08 May 2025.

PSYCHOLOGY & HEALTH
https://doi.org/10.1080/08870446.2025.2498577

Which cognitive tests are used to examine the acute effect of physical activity on cognition in healthy adults aged 50 and older? - A systematic review

Fien De Block[a,b], Delfien Van Dyck[a], Benedicte Deforche[b,c], Geert Crombez[d] and Louise Poppe[b]

[a]Department of Movement and Sport Sciences, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium; [b]Department of Public Health and Primary Care, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium; [c]Department of Movement and Sport Sciences, Faculty of Physical Education and Physical Therapy, Vrije Universiteit Brussel, Brussels, Belgium; [d]Department of Experimental Clinical and Health Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

ABSTRACT

Objectives: The review aims to create an overview of current evidence on the acute effect of physical activity (PA) on cognition in adults aged 50 and older, focusing on which cognitive (sub)domains, cognitive tests, and outcomes are used, and whether effects were demonstrated.
Methods: The MEDLINE, Embase, and Cochrane Central databases were searched for articles assessing the acute effect of PA on cognition in healthy adults aged 50 and older.
Results: Forty-two articles were included, using a variety of protocols. Executive functioning was the most frequently assessed cognitive domain. Overall, thirty-five different cognitive tests were administered, among which many variations and modifications were found. Furthermore, the reported outcomes varied greatly, even when using the same test. Across tests, 45.3% of the reported outcomes demonstrated an improvement in cognition shortly after PA. Time-based outcomes demonstrated an improvement more often than accuracy-based outcomes. However, because of the large variety among protocols and often insufficiently nuanced reporting, results should be interpreted carefully.
Conclusions: The acute effect of PA on executive functioning in older adults has been examined frequently, but research in other cognitive domains is limited. The variety among study protocols and test outcomes highlights the need for more rigorous research and reporting.

ARTICLE HISTORY
Received 8 March 2024
Accepted 14 April 2025

KEYWORDS
Cognitive functioning; cognitive domains; executive functioning; middle-aged adults; older adults; exercise

CONTACT Fien De Block fldblock.deblock@ugent.be; Delfien Van Dyck delfien.vandyck@ugent.be Department of Movement and Sport Sciences, Faculty of Medicine and Health Sciences, Ghent University, Watersportlaan 2, 9000 Ghent, Belgium.

1. Introduction

The aging process is inherently associated with changes in cognition. Although some cognitive functions, such as vocabulary knowledge, can remain stable or improve into late adulthood, most start declining around age 60, such as memory and reasoning (Salthouse, 2010). Beyond these long-term changes, cognition also fluctuates in the short term, i.e. from day to day or even within one day (McKinney et al., 2019; Sliwinski et al., 2006). Such short-term fluctuations appear to be larger in older adults compared to younger individuals (Hultsch et al., 2002), causing older adults to experience moments or days during which they are more forgetful and 'cognitively slower' than usual. Moreover, larger fluctuations in cognition among older adults have been associated with the development of dementia (Gamaldo et al., 2012; Hultsch et al., 2000).

Adopting a physically active lifestyle could be a promising approach to counter large short-term fluctuations in older adults' cognition, as physical activity (PA) has been shown to improve cognition across the lifespan and reduce the risk of cognitive decline later in life (Gallaway et al., 2017; Sofi et al., 2011). The majority of the existing literature has focused on the long-term effects of PA on cognition, with study durations ranging from a few weeks to several years (Bherer et al., 2013; Northey et al., 2018). Overall, moderate evidence has been found for a long-term effect of regular PA of moderate to vigorous intensity on cognition later in life (Erickson et al., 2019; Northey et al., 2018). In recent years, several studies have examined the effects of PA on cognition within the same day in middle-aged and older adults, hereafter referred to as acute effects (Chang et al., 2012b; Erickson et al., 2019; McSween et al., 2019). Overall, this research has revealed small to moderate evidence for an acute improvement in cognition after PA (Chang et al., 2012b; Erickson et al., 2019; McSween et al., 2019). The meta-analysis of Chang et al. (2012b) demonstrated that light to moderate-intensity PA appears to be the most beneficial and that the duration of the PA session should be at least 11 min to positively affect cognition. Furthermore, the delay between the PA session and cognitive assessment also appeared significant, with the strongest effects demonstrated when cognition was tested 11-20 min after finishing the PA session (Chang et al., 2012b).

Although previous studies have demonstrated certain acute effects of PA on cognition, primarily in the domains of executive functioning, attention, and memory (Chang et al., 2012b; Erickson et al., 2019; McSween et al., 2019), a frequently mentioned shortcoming is the large heterogeneity in the assessed cognitive domains and the cognitive tests that are used to measure them (Erickson et al., 2019; McSween et al., 2019). Furthermore, since the first signs of dementia can already start manifesting at age 50 (Haeger et al., 2020), more research on adults aged 50 and older is needed. However, there is currently no overview that summarizes which cognitive domains and tests are suitable to assess the acute effects of PA on cognition in this population. Consequently, this review aims to provide an overview of the existing literature on the acute effects of PA on cognition among healthy adults aged 50 and older, focusing on (1) which cognitive (sub)domains were assessed, (2) which cognitive tests and (3) outcomes were used, and (4) whether the reported results demonstrated acute effects of PA on cognition.

2. Methods

This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria (Page et al., 2021). See Appendix A for the PRISMA checklist. The review was preregistered at PROSPERO (ID: CRD42022301784) and Open Science Framework (De Block & Poppe, 2023).

2.1. Eligibility criteria

Only articles published in scientific journals or as conference papers were eligible for inclusion. Other publication types were not considered. Articles were required to be English-language with interventional or observational designs. To assess the acute effects of PA on cognition, at least one behavioral cognitive test had to be conducted after the PA and on the same day. All types, intensities, and durations of PA were considered. PA was defined as any bodily movement produced by skeletal muscles that requires energy expenditure (WHO guidelines on physical activity & sedentary behaviour, 2020).

Only articles in which all participants or a distinct group of participants were at least 50 years old were included. Furthermore, participants needed to be described as healthy to be eligible for inclusion, i.e. free of (neurological) conditions impeding daily life. Healthy controls in clinical trials were also eligible for inclusion. Participants residing in residential care centers or retirement homes were considered insufficiently healthy and were therefore excluded, since the National Institute on Aging describes residents of residential facilities as (temporarily) in need of help or care greater than what their environment can offer (*Residential Facilities, Assisted Living, and Nursing Homes | National Institute on Aging*, n.d.).

2.2. Information sources and search strategy

Three databases—MEDLINE (PubMed interface), Embase, and Cochrane CENTRAL—were searched. The search, which was conducted without filters and not restricted to a specific time period, followed the PICO framework (Population, Intervention, Comparison (not applicable to this review), and Outcome). The search strategy was developed in consultation with a center of expertise (Knowledge Centre for Health Ghent) and encompassed five concepts within the framework: 'physical activity', 'acute', 'cognition', 'adults aged 50 and older', and 'healthy'.

The search strategies can be found in Appendix B. In the Embase and Cochrane Central databases, the 'NEAR/n' operator was initially employed to ensure keeping the concepts 'acute' and 'cognition' near each other, within n words. However, because this search yielded no results in the Cochrane database, the 'NEAR/n' operator was replaced with an 'AND' operator in this database to broaden the search.

The first search was run on December 17th (MEDLINE and Embase) and December 20th, 2021 (Cochrane CENTRAL). To update the review, a second search of all three databases was performed on April 4th, 2023. This search was limited to articles published after the first search. In addition to articles retrieved directly from the search results, papers included in systematic reviews and meta-analyses identified by the search were also screened.

2.3. Selection process

References were imported into EndNote X9 and duplicates were deleted. Thereafter, the references were imported into Rayyan, an online screening tool (Ouzzani et al., 2016). After a second duplicate search was conducted in Rayyan, two reviewers (FDB and LP) independently screened the remaining articles' titles and abstracts. Articles were assigned 'include', 'exclude', or 'maybe'. After screening all articles, the reviewers' assignments were compared. When the reviewers disagreed or an article was assigned 'maybe', the reviewers screened these abstracts together to reach a consensus. Abstracts lacking sufficient evidence for inclusion or exclusion were included, so the full text could be reviewed. One reviewer (FDB) assessed the full text of the included articles to decide on final inclusion. Uncertainties were discussed with at least one other reviewer.

2.4. Data collection process and data items

Data from the included articles were collected in a file by one reviewer (FDB) and shared with all authors. Uncertainties were discussed with at least one other reviewer. The primary outcomes were (1) the assessed cognitive (sub)domain(s), (2) the used cognitive test(s), (3) the measured outcome(s), and (4) whether an acute change in cognition was detected after PA, compared to after a rest condition, rest group or baseline measurement. The presence of an acute effect of PA on cognition was determined based on the significance of the effects reported. In articles with a rest group or rest condition next to the PA group(s) or condition(s), a significant group-by-time interaction was necessary to conclude the presence of an effect. In articles only consisting of (a) PA group(s) or condition(s), a main effect of time or a group-by-time interaction effect was required.

Other extracted data included: year of publication, study design, conditions/groups (if applicable), sample size, sex distribution, participants' mean age or age range, PA condition (type, duration, and intensity), rest activity (type and duration; if applicable), administration method, the time between the PA/rest condition and cognitive assessment(s) (if applicable), and details on the test's conditions, version, etc. If any information regarding these variables was missing or unclear, the article's corresponding author was contacted. If the authors did not respond or could not provide the information, the data was labeled as 'not reported'.

2.5. Risk of bias assessment

One researcher (FDB) performed the risk of bias (ROB) assessment using the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies (*Quality Assessment Tool for Quantitative Studies*, n.d.). The total ROB score consisted of the following six component scores: (1) selection bias (the representativeness of the study sample, based on whether the participants were recruited through random sampling), (2) study design (the strength of the research design), (3) confounders (the presence of important differences between participant groups at baseline and whether these differences were controlled for), (4) blinding (whether the outcome assessors and participants were blinded during the data collection), (5) data collection methods (whether the tools to measure the outcomes were valid and reliable), and (6) withdrawals and drop-outs (whether the numbers of and reasons for withdrawals and drop-outs were reported). Each component received a 'strong' (low ROB), 'moderate', or 'weak' (high ROB) rating based on these criteria. Articles received a total 'strong' score if they had no weak components, a 'moderate' score if they had one weak component, and a 'weak' score if they had more than one weak component. To ensure accurate assessments, a second reviewer (LP) independently performed the ROB on five randomly selected included articles to verify the inter-rater agreement. Uncertainties were discussed with at least one other reviewer.

This ROB assessment tool was originally developed for a wide range of health-related topics. However, it overlooks within-subject and longitudinal study designs and lacks information on how to score components in certain cases. Therefore, some criteria were slightly adapted to fit this review. Regarding the component 'selection bias', the original tool did not specify how to score articles that did not report their recruitment methods. In such cases, the component was scored as 'not reported'. Within the component 'study design', designs that were not mentioned by the ROB tool received the same scores as designs with similar strength. Randomized cross-over trials received a 'strong' score. Non-randomized cross-over trials and longitudinal designs received a 'moderate' score. For the 'confounders' component, the criteria were initially grounded in assessing differences between participant groups. However, as this review also included research designs that were not composed of different participant groups, 'n/a' was filled in for these articles.

Furthermore, within the component 'data collection methods', the tools used for the articles' outcome assessment(s) (i.e. the cognitive tests) were scored on validity and reliability. A test was considered valid and reliable if the validity and reliability of that specific test were published for healthy adults aged similar to the articles' participants or if the test was part of a validated test battery. For articles employing multiple tests, a weak score was assigned if less than half of the tests were valid and reliable, a moderate score if more than half but not all tests were valid and reliable, and a strong score if all tests were valid and reliable.

Within the component 'withdrawals and drop-outs', articles that did not explicitly report withdrawals or drop-outs, but reported a consistent participant number throughout all data-collection moments (meaning there were no drop-outs or withdrawals) were given a 'strong' rating. If participant numbers varied or were undisclosed across different data collection moments without explicitly mentioning drop-outs, the article received a 'weak' score.

2.6. Synthesis methods

The articles' characteristics were organized per study design and arranged alphabetically by the first author's name. The articles' results were structured based on the cognitive tests used, prioritized from most to least used, and further sorted alphabetically by the first author's name. The assignment of cognitive domains and subdomains to the tests was derived from the articles' descriptions. Many approaches exist to define and categorize cognitive (sub)domains. This review employed the six key domains of cognition and their respective subdomains, as described by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition: *language* (*object naming*, *word finding*, *fluency*, *grammar and syntax*, and *receptive language*), *learning and memory* (*free recall*, *cued* *recall*, *recognition memory*, *semantic and autobiographical long-term memory*, and *implicit learning*), *social cognition* (*recognition of emotions*, *theory of mind*, and *insight*), *complex attention* (*sustained attention*, *divided attention*, *selective attention*, and *processing speed*), *executive functioning* (*planning*, *decision-making*, *working memory*, *responding to feedback*, *inhibition*, and *flexibility*), and *perceptual-motor functioning* (*visual perception*, *visuoconstructional reasoning*, and *perceptual-motor coordination*) (Sachdev et al., 2014). If no subdomain was specified for a certain test, it was only assigned to the cognitive domain(s) described by the articles in which it was used.

3. Results

3.1. Study selection

Figure 1 demonstrates the selection process. The searches generated 2390 results, of which 1807 remained after having removed the duplicates. Following the abstract screening, 73 full-text articles were reviewed, of which 25 articles were included. Additionally, the search yielded five relevant reviews and meta-analyses (Cheng et al., 2022; Erlenbach et al., 2021; Griebler et al., 2022; McSween et al., 2019; Oberste et al., 2021). The 116 articles included in these reviews were screened for inclusion. After having removed the duplicates, 106 abstracts were reviewed, resulting in 20 full-text screenings, which led to 17 more articles being included in this systematic review. This brought the number of included articles to 42.

Authors of similar articles were contacted to verify whether the articles were written on the same study. Two of the included articles were confirmed to have been written on the same study (Won et al., 2019a, 2019b). Two articles were confirmed to be written on different studies (Tsai et al., 2021; Tsai & Pan, 2023). Other similar articles

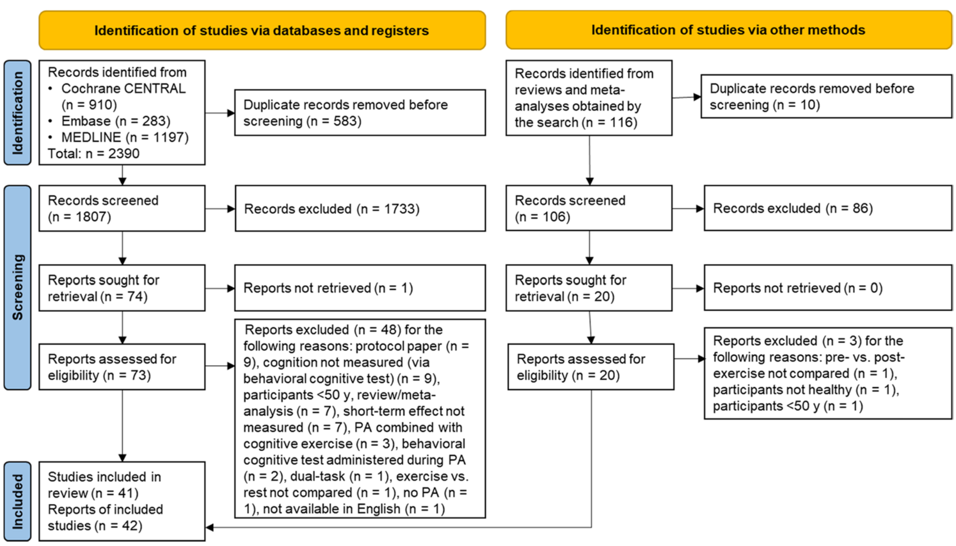

Figure 1. PRISMA flow diagram of the selection process. PRISMA definitions: report = a document supplying information about a particular study (e.g. journal article); record = the title and/or abstract of a report indexed in a database or website; study = an investigation that includes a defined group of participants and one or more interventions and outcomes (Page et al., 2021).

were not confirmed to be written on the same study or different studies, namely Hsieh et al. (2016) and Hsieh et al. (2016); and Chang et al. (2012a) and Chang et al. (2014b). These articles were therefore treated as different studies.

3.2. Study characteristics

Tables 1–4 provide an overview of the articles' characteristics. The included articles employed various designs: 24 used a cross-over design (Table 1), 11 were randomized controlled trials (Table 2), one article used a cross-over design for the first half of the participants, and a randomized controlled trial for the second half (Table 1), four employed pretest-posttest designs (no rest group) (Table 3), and two used a micro-longitudinal design (repeated measurements over several days) (Table 4).

The included articles were published between 1997 and 2023, with 15 (35.7%) being published in the last 5 years. The number of participants per article ranged from 11 to 150. Six articles included multiple age groups, of which only the groups with adults aged 50 and older were included in this review. Two articles consisted of patients and healthy controls, of which only the healthy controls were included. Thirty-four articles included men and women, four included only women, and four included only men.

Nine different PA types were used in the included articles, of which ergometer cycling was the most commonly used ($n = 18$). Other types of PA included walking (with or without a treadmill; $n = 8$), resistance exercises ($n = 5$), dancing ($n = 2$), combination exercise ($n = 1$), seated exercises ($n = 1$), and squeezing a latex ball ($n = 1$). Additionally, one article consisted of two PA groups with different PAs (one used a cycling ergometer and one performed resistance exercise); and one article had three PA groups (two used a cycling ergometer and one performed stretching exercises). Lastly, four articles measured the participants' PA levels in their daily lives or during their habitual PA sessions.

Of the 42 included articles, 29 implemented a rest condition or rest group. (Seated) reading ($n = 12$) was the most common, followed by rest ($n = 11$), watching videos ($n = 2$), providing picture quality ratings ($n = 1$), talking with a researcher (seated) ($n = 1$), and habitual sedentary activity ($n = 1$). Additionally, one article used a typical 'rest day' from the participants' daily lives as a rest condition. Lastly, 13 articles used baseline measurements or real-life sedentary measurements as a rest condition.

3.3. Risk of bias assessment

Appendix C shows the component ratings and total ratings for the ROB assessment. Four articles received a 'strong' total rating, 12 received a 'moderate' total rating and 26 received a 'weak' total rating. Although many articles received moderate or strong scores on the 'design', 'confounders', and 'blinding' components, the majority received a weak total rating. This was mainly because 26 of the included articles did not randomly select their participants, resulting in weak scores for the component 'selection bias'. Additionally, many articles did not use validated cognitive tests (for this age group) or did not clarify or refer to which version of a test was used, resulting in weak scores on the 'data collection methods' component. Lastly, several articles did not report withdrawals or drop-outs, leading to weak scores on the respective component.

Table 1. Characteristics of articles using cross-over designs.

Article	Conditions (and groups)	Sample size (#F)	Mean age ± sd	PA condition(s) (type, duration[a], intensity)	Rest condition (type, duration)	Time between PA and cognitive test(s)
Abe et al. (2018)	Conditions: 3 PA (different types of seated exercises) and 1 C	26 (18)	71.8 ± 4.7	Stepping, stretching, and finger movement for 10 min each	10 min rest	5 min after
Barella et al. (2010)b	Conditions: 1 PA and 1 C	40 (32)	69.5 ± 8.3	Walking on a treadmill for 20 min at 60% HHR	Sitting for 25 min, talking with experimenter	After 30-90 s, 5 min, 10 min, 15 min, 20 min, 30 min, 45 min, 60 min, 75 min, 90 min, 105 min, 120 min
Callow et al. (2023)	Conditions: 1 PA and 1 C	30 (23)	70.2 ± 6.1	Ergometer cycling for 20 min at moderate intensity	Resting for 30 min	5 min after
Chang et al. (2014a)	• • Conditions: 1 PA and 1 C • Groups: 1 HF and 1 LF	42 (0)	LF: 63.4 ± 3.3 HF: 62.8 ± 2.4	Ergometer cycling for 20 min at 50-60% HRR	Reading for 30 min	Within 15 min
Chang et al. (2012a)	Conditions: 1 PA and 1 C	30 (16)	57.2 ± 2.9	Resistance exercises for ±10 min at moderate intensity	Reading for 20 min	Within 3 min
Chang et al. (2014b)	Conditions: 1 PA and 1 C	30 (15)	58.1 ± 3.0	Resistance exercises for 20-25 min at 70% of 10 RM	Reading for 30 min	Not reported
Chen et al. (2018)	Conditions: 3 PA (3 durations) and 1 C	45 (26)	57.7 ± 5.1	Ergometer cycling for 10, 20, or 45 min at 65-70% HRR	Reading for 30 min (seated)	Not reported
Chu et al. (2015)	• • Conditions: 1 PA and 1 C • Groups: 1 HF and 1 LF	46 (22)	LF: 64.9 ± 4.0 HF: 63.8 ± 2.3	Ergometer cycling for 20 min at moderate intensity	Reading for 25 min (seated)	Within 5 min
Emery et al. (2001)	• • Conditions: 1 PA and 1 C • Groups: COPD patientsc and HC	29 (17)	68.7 ± 6.0	Ergometer cycling for 20 min at high intensity	Watching videos for ± 45 min (seated)	Immediately after
Faulkner et al. (2017)	• • Conditions: 4 PA (2 positions and 2 intensities) • Groups: TIA patientsc and HC	15 (2)	61.5 ± 7.1	Ergometer cycling in 2 positions: upright seated and supine, and at 2 intensities: moderate for 30 min and high until maximum was reached	No rest condition	Twice: 1.5 min and 15 min after
Hsieh et al. (2016)	• • Conditions: 1 PA and 1 C • Groups: youngerc and older adults	17 (0)	66.4 ± 1.2	Resistance exercises for 20-25 min at 70% of 10 RM	Reading for 30 min	10 min after
Hsieh et al. (2016)d	Conditions: 1 PA and 1 C	20 (0)	67.2 ± 1.8	Resistance exercises for 20-25 min at moderate intensity	Reading for 30 min (seated)	Not reported
Hyodo et al. (2012)	Conditions: 1 PA and 1 C	16 (3)	69.3 ± 3.5	Ergometer cycling for 10 min at moderate intensity	Resting for 25 min	15 min after
Hyodo et al. (2021)	Conditions: 2 PA (intermittent and continuous PA)	15 (8)	W: 70.0 ± 2.5 M: 71.1 ± 2.9	Continuous (exercises at 90 bpm, without breaks) or intermittent (exercises at 120 bpm with 30 s breaks) aerobic dancing for 10 min at light intensity	No rest condition	5 min after

(*Continued*)

Table 1. Continued.

Article	Conditions (and groups)	Sample size (#F)	Mean age ± sd	PA condition(s) (type, duration[a], intensity)	Rest condition (type, duration)	Time between PA and cognitive test(s)
Ji et al. (2019)	Conditions: 1 PA, 1 PA + cognitive exercisec, 1 cognitive exercisec, 1 C	20 (9)	65.6 ± 1.3	Walking for 15 min at 65% heart rate returned	Reading, duration not reported	Immediately after (when HR had returned to rest level)
Kamijo et al. (2009)	• Conditions: 2 PA (2 intensities) • Groups: younger[c] and older adults	12 (0)	65.5 ± 1.5	Ergometer cycling for 20 min at 30% of VO_2max or 50% VO_2max	No rest condition	Less than 2 min after
Nielson et al. (2014)	• Conditions: 1 PA and 1 C • Groups: based on order of conditions	Group A: 25 (17) Group B: 22 (11)	A: 72.3 ± 4.9 B: 72.4 ± 5.4	Squeezing a sand-filled latex ball for 1 min, HR typically increases 10 beats per minute	Rest (seated) for 1 min	Immediately after (± 15 min) and delayed (2 wk after)c
Peiffer et al. (2015)	Conditions: 2 PA (2 intensities)	11 (11)	65.8 ± 3.8	Walking on treadmill for 20 min at 50% VO_2max or 75% VO_2max	No rest condition	Twice: immediately after (less than 2 min) and 30 min after
Pellegrini-Laplagne et al. (2022)	Conditions: 1 PA, 1 cognitive exercisec, 1 PA + cognitive exercisec	17 (12)	62.4 ± 3.9	Ergometer cycling for 30 min at moderate intensity	No rest condition	10 min after
Sturnieks et al. (2018)	Conditions: 1 PA and 1 C	50 (34)	72.5 ± 6.5	Participants scheduled as many activities as possible, based on their busiest days	Participants scheduled as few daily activities as possible, based on their least busy days	After 4 PM (when the participants' PA measurements ended)
Tsai et al. (2021)	Conditions: 2 PA (2 intensities) and 1 C	20 (10)	61.2 ± 4.4	Ergometer cycling for 24 min at moderate or high intensity	Reading (seated) for 35 min	3-5 min after
Tsai and Pan (2023)	Conditions: 2 PA (2 intensities) and 1 C	22 (11)	61.1 ± 5.4	Ergometer cycling for 24 min at moderate or high intensity	Reading (seated) for 30 min	Blood was drawn immediately after PA or rest, cognitive tests started 2-3 min after blood draw
Tsujii et al. (2013)	Conditions: 1 PA and 1 C	14 (9)e	65.9 ± 1.0	Ergometer cycling for 10 min at 40% of VO_2max	Resting for 10 min	10 min after
Won et al. (2019a)	Conditions: 1 PA and 1 C	32 (24)	66.2 ± 7.3	Ergometer cycling for 20 min at moderate intensity	Resting (seated) for 35 min	Immediately after
Won et al. (2019b)	Conditions: 1 PA and 1 C	26 (20)	65.9 ± 7.2	Ergometer cycling for 20 min at moderate intensity	Resting (seated) for 30 min	Twice: within 10 min after and 30–40 min after

[a]Warm-up and cool-down are not included.
[b]Because of carry-over effects, the study design was changed to RCT after the 1st half of the participants.
[c]Group, condition, or assessment was not included.
[d]Article consisted of 2 articles with identical designs: one with younger and one with older adults. Only older adults were included.
[e]Number of participants and sex distribution uncertain.
Abbreviations: F = female; PA = physical activity; C = control; W = women; M = men; HRR = heart rate reserve; HF = higher fitness level; LF = lower fitness level; RM = repetition maximum; COPD = chronic obstructive pulmonary disease; HC = healthy controls; TIA = transient ischemic attack; VO_2max = maximal aerobic capacity; HR = heart rate.

Table 2. Characteristics of articles using RCTs.

Article	Groups (and conditions)	Sample size (#F)	Mean age ± sd	PA group (s) (type, duration[a], intensity)	Rest group (type, duration)	Time between PA and cognitive test(s)
Córdova et al. (2009)	Groups: 3 PA (3 intensities) and 1 C	48 (48)	63.8 ± 4.6	Ergometer cycling for 20 min at 60%, 90% and 110% of AT	Reading for 25 min (seated)	8 min after
Hogan et al. (2013)	Groups: 1 PA and 1 C Participants were divided into 4 age-based quartilesb	• Total: 144 (73) • PA: 71 (37) • C: 73 (36)b	Q3 (EX: 50–72 and C: 54–69) Q4 (EX: 73–93 and C: 70–87)b	Ergometer cycling for 15 min at ± 50% HRR	Provide subjective picture quality ratings for 15 – 22.5 min	Immediately after (but participants completed affect assessment first)
Johnson et al. (2016)	• • Groups: 2 PA (2 types) • Conditions: 2 PA (2 durations)	31 (21)	71.7 ± 1.5	Ergometer cycling or resistance exercises, both for 10 or 30 min at moderate intensity	No rest group	Three times: immediately after, 30 min, and 60 min
Kimura and Hozumi (2012)	Groups: 2 PA (FS and CS)	34 (20)	FS: 71.1 ± 4.9 CS: 70.7 ± 4.3	Dance for 20-min at 40% HRR	No rest group	Immediately after
Martelli et al. (2021)	Groups: 2 PA	28 (22)	PA1: 70.7 ± 3.7 PA2: 71.6 ± 4.7	24 min walking with or without perturbations	No rest group	10 min after
McSween et al. (2021)	Groups: 3 PA (stretching, MICE, and HIIT)	60 (43)	66.4 ± 4.6	Stretching (38 min), MICE (ergometer cycling for 30 min at moderate intensity), or HIIT (ergometer cycling for 25 min at high intensity)	No rest group	Twice: immediately after (9.8 ± 5.3 min) and delayed (7.2 ± 0.8 days after 1st session)c
Naderi et al. (2019)	• • Groups: 1 PA and 1 C; each group is divided based on sex • Conditions: 2 PA (2 intensities)	40 (20)	PA W: 63.9 ± 2.9 PA M: 63.7 ± 2.6 C W: 63.6 ± 2.2 C M: 63.1 ± 2.1	Resistance exercises for 35-40 min at 40% of 10 RM and at 70% of 10 RM	Watching video for 45 min	15 min after
Nouchi et al. (2020)	Groups: 1 PA and 1 C; each group is divided into middle-agedc and older adults	30 (30)	PA: 69.7 ± 5.32 C: 70.1 ± 5.38	Combination exercise[3] for 30 min at 60% to 80% of max HR	Waiting for 30 min (seated)	5 min after
Olivo et al. (2021)	Groups: 1 PA and 1 C	49 (22)	PA: 69.6 ± 2.8 C: 70.7 ± 3.1	Ergometer cycling for 30 min at moderate intensity	Laying down and relaxing for 30 min	7 min after
Schramke and Bauer (1997)	• • Groups: 4 (order of conditions); each group is divided into youngerc and older adults • Conditions: 1 PA and 1 C	48d	PA - PA 68.1 ± 5.6 PA - C: 71.9 ± 5.3 C - C: 67.4 ± 4.8 C - PA: 70.0 ± 6.4	Walking for 5-7 min at a pace that participant found comfortable	Rest (seated) for 5 – 7 min	Immediately after
Wang et al. (2015)e	Groups: 1 PA and 1 C	24 (10)	PA: 56.4 ± 5.1 C: 55.0 ± 5.4	Ergometer cycling for 20 min at 65% of HRR	Reading for 30 min	Not reported

[a]Warm-up and cool-down not included.
[b]Only two oldest quartiles were included. Number of participants, sex distribution and mean age of each quartile were not reported.
[c]Group or assessment was not included.
[d]Sex distribution not reported.
[e]Article consisted of 2 articles with identical designs: one with younger and one with older adults. Only older adults were included.
Abbreviations: F = female; PA = physical activity; C = control; W = women; M = men; AT = anaerobic threshold; HRR = heart rate reserve; FS = freestyle (short, sequentially repeated routines); CS = combination style (long choreographic routine); MICE = moderate-intensity continuous exercise; HIIT = high-intensity interval exercise; RM = repetition maximum; HR = heart rate.

Table 3. Characteristics of articles using pretest-posttest designs.

Article	Groups	Sample size (#F)	Mean age	PA condition (type, duration[a], intensity)	Rest condition or group (type, duration)	Time between PA session and cognitive assessment
Hatta et al. (2013)	1 group	20 (10)	70.5 ± 3.4	Self-paced walk at moderate intensity for 80 – 120 min (target 7000 – 10 000 steps)	No rest condition	Immediately after
Lebeau et al. (2022)[b]	1 group	71 (46)	66.4 ± 4.7	Walking on a treadmill for 15 min at until RPE = 15	No rest condition	Not reported
Netz et al. (2009)	2 groups: youngerc and older adults	20 (20)	63.7 ± 3.6	Walking on a treadmill for 35 min at 60% of HRR	No rest condition	Twice: 5 min after and 1 h after
O'Brien et al. (2017)	3 groups: usual activity: OS, CS, or sedentary (C)	58 (37)	OS: 69.2 ± 5.1 CS: 69.2 ± 4.8 C: 70.5 ± 6.9	Usual PA type, duration, and intensity OS mean duration: 80 ± 20 min CS mean duration: 70 ± 20 min	Usual sedentary activity for 60 min (e.g. card games)	Within 10 min

[a]Warm-up and cool-down not included.
[b]Only part of artcile with pretest-posttest design was relevant for this review, so only this part was included.
[c]Group was not included.
Abbreviations: F = female; RPE = rate of perceived exertion; HRR = heart rate reserve; OS = open skill (dynamic and unpredictable environment, e.g. tennis); CS = closed skill (stable and predictable environment, e.g. swimming); C = control.

Table 4. Characteristics of articles using micro-longitudinal designs.

Article	Sample size (#F)	Mean age	Duration of assessment period/ number and timing of assessments every day	PA/rest assessments
Luo et al. (2023)	150 (74)	73.5 ± 5.6	15 days/7 times per day ± 2 h apart	Ecological momentary assessment: participants were asked to indicate their current activity (socio-cognitive activities[a], passive leisure activities[a], physical activities, and self-maintenance activities[a]) and perform a cognitive test.
Phillips et al. (2016)	51 (31)	70.1 ± 7.0	4 days/once per day at a self-selected time	Participants continuously wore an activity monitor and performed a cognitive assessment battery daily.

[a]Assessments for these activities were not included.
Abbreviations: F = female..

3.4. Study results

All study results are presented in Table 5. Frequencies of tests, assessments, and outcomes per (sub)domain are shown in Table 6. It is important to note that several cognitive tests were assigned to multiple cognitive domains and subdomains, based

Table 5. Study results (terms and concepts that require further explanation are clarified in Appendix D).

Article	Design	Administration method	Condition, part, score	Basis of outcome	Results per outcome
Stroop Test - executive functioning (inhibition, flexibility); complex attention (processing speed, sustained attention, selective attention)					
Abe et al. (2018)	Cross-over	Computer		Accuracy	Error rate: no evidence for an effect found.
				Time	RT: no evidence for an effect found.
					Interference RT[a]: better scores post-PA compared to pre-PA and post-rest. No evidence for an effect found in rest condition.
Barella et al. (2010)	1st half of participants: cross-over, 2nd half: RCT	Verbal + computer	Color condition[b]	Time	Mean RT across all trials: faster RT post-PA compared to pre-PA up to 60 min after PA. No evidence for an effect found in rest condition/group.
			Interference condition[c]	Time	Mean RT across all trials: no evidence for an effect found.
			Inhibition condition[d]	Time	Mean RT across all trials: no evidence for an effect found.
Chang et al. (2014a)	Cross-over	Computer		Time	RT: faster RT in both PA groups after PA. No evidence for an effect found in rest condition.
Chang et al. (2014b)	Cross-over	Verbal + stopwatch	Congruent condition[e]	Time	Mean RT: faster RT post-PA compared to post-rest.
			Word condition[f]	Time	Mean RT: faster RT post-PA compared to post-rest.
			Color condition[g]	Time	Mean RT: faster RT post-PA compared to post-rest.
			Neutral condition[h]	Time	Mean RT: faster RT post-PA compared to post-rest.
			Incongruent condition[i]	Time	Mean RT: faster RT post-PA compared to post-rest.
Chu et al. (2015)	Cross-over	Computer		Accuracy	Accuracy: higher accuracy in both PA groups after PA. No evidence for an effect found in rest condition.
				Time	RT: faster RT in both PA groups after PA. No evidence for an effect found in rest condition.
Faulkner et al. (2017)	Cross-over	Computer	Interference condition[j]	Accuracy	Number of correct answers: no evidence for an effect found.
				Time	Completion time: faster completion time 1.5 min post-PA compared to pre-PA in both PA conditions.
Hyodo et al. (2012)	Cross-over	Computer		Accuracy	Interference accuracy rate[a]: no evidence for an effect found.
				Time	Interference RT[a]: faster RT post-PA compared to pre-PA in PA condition. No evidence for an effect found in rest condition.
Hyodo et al. (2021)	Cross-over	Computer		Time	Interference RT[a]: better interference score post-PA compared to pre-PA for both PA conditions.
Ji et al. (2019)	Cross-over	Computer	Naming condition[k]	Accuracy	Error rate: no evidence for an effect found.
				Time	RT: no evidence for an effect found.
			Executive condition[l]	Accuracy	Error rate: no evidence for an effect found.
				Time	RT: faster RT post-PA compared to pre-PA in PA condition. No evidence for an effect found in rest condition.

(*Continued*)

Table 5. Continued.

Article	Design	Administration method	Condition, part, score	Basis of outcome	Results per outcome
Johnson et al. (2016)	RCT	Verbal + computer	Color condition[m]	Time	RT: faster RT immediately post-PA compared to pre-PA in both PA groups. RT: faster RT 60 min post-PA compared to pre-PA after a 10 min PA bout.
			Word interference condition[n]	Time	RT: no evidence for an effect found.
			Inhibition condition[o]	Time	RT: faster RT immediately post-PA compared to pre-PA in both PA groups and both PA bout durations.
Lebeau et al. (2022)	Pretest-posttest	Not reported		Accuracy	Accuracy: no evidence for an effect found.
				Time	Mean RT: faster RT immediately post-PA compared to pre-PA.
Nouchi et al. (2020)	RCT	Pen and paper	Stroop condition[p]	Accuracy	Number of correct items: improved performance post-PA compared to pre-PA in PA group. No evidence for an effect found in rest group.
			Reverse Stroop condition[q]	Accuracy	Number of correct items: improved performance post-PA compared to pre-PA in PA group. No evidence for an effect found in rest group.
Pellegrini-Laplagne et al. (2022)	Cross-over	Computer	Naming condition[r]	Accuracy	Number of errors: no evidence for an effect found.
				Time	RT: no evidence for an effect found.
			Inhibition condition[s]	Accuracy	Number of errors: no evidence for an effect found.
				Time	RT: faster RT post-PA compared to pre-PA.
			Inhibition score[t]	Accuracy	Number of errors: no evidence for an effect found.
				Time	RT: no evidence for an effect found.
			Switching/flexibility condition[u]	Accuracy	Number of errors: fewer errors post-PA compared to pre-PA.
				Time	RT: faster RT post-PA compared to pre-PA.
			Switching/flexibility score[v]	Accuracy	Number of errors: no evidence for an effect found.
				Time	RT: faster RT post-PA compared to pre-PA.
Trail Making Test - executive functioning (flexibility, working memory); complex attention (processing speed); perceptual-motor functioning (visual perception)					
Córdova et al. (2009)	RCT	Pen and paper	Parts A and B	Time	Completion time part A: no evidence for an effect found. Completion time part B: faster completion time in post-PA in PA group exercising at 90% of anaerobic threshold compared to rest group.
Emery et al. (2001)	Cross-over	Pen and paper	Parts A and B	Time	Total score[w]: no evidence for an effect found.
Lebeau et al. (2022)	Pretest-posttest	Pen and paper	Parts A and B	Time	Total score[w]: faster RT post-PA compared to pre-PA.
Martelli et al. (2021)	RCT	Pen and paper	Parts A and B	Time	Completion time part A: no evidence for an effect found. Completion time part B: no evidence for an effect found.
Sturnieks et al. (2018)	Cross-over	Pen and paper	Parts A and B	Time	Total score[w]: no evidence for an effect found.

(*Continued*)

Table 5. Continued.

Article	Design	Administration method	Condition, part, score	Basis of outcome	Results per outcome
Eriksen Flanker Test - executive functioning (inhibition)					
Kamijo et al. (2009)	Cross-over	Computer		Accuracy	Accuracy rate: no evidence for an effect found.
				Time	Mean RT: no evidence for an effect found.
Naderi et al. (2019)	RCT	Computer		Accuracy	Accuracy: better accuracy 15 min post-PA compared to pre-PA in PA groups in both PA conditions. No evidence for an effect found in rest groups.
				Time	RT: faster RT 15 min post-PA compared to pre-PA in PA groups in both PA conditions. No evidence for an effect found in rest groups.
Peiffer et al. (2015)	Cross-over	Computer		Time	Total RT: faster RT immediately post-PA compared to pre-PA in both PA conditions. Congruent[x] RT: faster RT immediately post-PA compared to pre-PA in both PA conditions. Incongruent[y] RT: faster RT immediately post-PA and 30 min post-PA compared to pre-PA in both PA conditions.
Won et al. (2019a)	Cross-over	Computer	Congruent[x]	Accuracy	Accuracy rate: better accuracy post-PA compared to post-rest.
				Time	RT: no evidence for an effect found. Interference RT[z]: no evidence for an effect found.
			Incongruent[aa]	Accuracy	Accuracy rate: better accuracy post-PA compared to post-rest.
				Time	RT: no evidence for an effect found. Interference score[z]: no evidence for an effect found.
Digit Symbol Substitution Test - executive functioning (working memory); complex attention (processing speed); perceptual-motor functioning (perceptual-motor coordination)					
Emery et al. (2001)	Cross-over	Pen and paper		Accuracy	Number of correct items: no evidence for an effect found.
Nouchi et al. (2020)	RCT	Pen and paper		Accuracy	Number of correct items: no evidence for an effect found.
Phillips et al. (2016)	Micro-longitudinal	Pen and paper		Accuracy	Number of correct items: no evidence for an effect found.
Sturnieks et al. (2018)	Cross-over	Pen and paper		Accuracy	Number of correct items: no evidence for an effect found.
Verbal Fluency Test - executive functioning (flexibility); language					
Córdova et al. (2009)	RCT	Verbal		Accuracy	Number of correct items: better performance post-PA compared to pre-PA in PA group exercising at 90% of anaerobic threshold. No evidence for an effect found in rest group.
Emery et al. (2001)	Cross-over	Verbal		Accuracy	Number of correct items: no evidence for an effect found.
Nouchi et al. (2020)	RCT	Verbal		Accuracy	Number of correct items: no evidence for an effect found.
Sternberg memory test - executive functioning (working memory)					
Hsieh et al. (2016)	Cross-over	Computer		Accuracy	Accuracy: no evidence for an effect found.
				Time	RT: faster RT post-PA compared to pre-PA and post-rest.

(*Continued*)

Table 5. Continued.

Article	Design	Administration method	Condition, part, score	Basis of outcome	Results per outcome
Naderi et al. (2019)	RCT	Computer		Accuracy	RT: faster RT post-PA compared to pre-PA in PA groups. No evidence for an effect found in rest groups.
				Time	Accuracy: better accuracy post-PA compared to pre-PA in PA groups. No evidence for an effect found in rest groups.
n-back test - executive functioning (working memory)					
Hogan et al. (2013)	RCT	Computer	2-back	Accuracy	Mean accuracy rate: no evidence for an effect found.
				Time	Mean RT of correct responses: greater reduction in RT in PA group compared to rest group.
Olivo et al. (2021)	RCT	Computer	Figurative n-back in 3 blocks: 1-, 2-, and 3-back	Accuracy	Mean accuracy: no evidence for an effect found. Performance score[bb]: no evidence for an effect found.
				Time	RT: no evidence for an effect found.
Wisconsin Card Sorting Test - executive functioning					
Hatta et al. (2013)	Pretest-posttest	Computer		Accuracy	Number of completed categories: no evidence for an effect found. Number of errors: no evidence for an effect found. Perseverative errors of Nelson[cc]: no evidence for an effect found.
Wang et al. (2015)	RCT	Manual		Accuracy	Number of completed categories: no evidence for an effect found. Perseverative errors[cc]: no evidence for an effect found. Perseverative responses[dd]: no evidence for an effect found. Non-perseverative errors[ee]: no evidence for an effect found.
Digit Span Test - executive functioning (working memory); complex attention					
Emery et al. (2001)	Cross-over	Verbal		Accuracy	Accuracy: no evidence for an effect found.
O'Brien et al. (2017)	Pretest-posttest	Verbal	Forward digit span[ff]	Accuracy	Digit span score[gg]: no evidence for an effect found. Digit span product score[hh]: better score post-PA compared to pre-PA for open skill group. No evidence for an effect found in other groups.
Mnemonic similarity task - learning and memory (recognition memory)					
Callow et al. (2023)	Cross-over	Computer		Accuracy	Object recognition memory[ii]: no evidence for an effect found. Mnemonic discrimination[jj]: similar score post-PA compared to pre-PA for PA condition, but worse score post-PA compared to pre-PA in for rest condition.
Tower of London Test - executive functioning (planning)					
Chang et al. (2012a)	Cross-over	Manual		Time	Time violation score[nn]: no evidence for an effect found. Total initial time[oo]: faster performance post-PA compared to pre-PA, slower performance post-rest compared to pre-rest. Total execution time[pp]: no evidence for an effect found. Total planning-solving time[qq]: no evidence for an effect found.
				Accuracy	Total move score[kk]: better score post-PA compared to pre-PA and post-rest. Total correct score[ll]: better score post-PA compared to pre-PA and post-rest. Rule violation score[mm]: no evidence for an effect found.

(*Continued*)

Table 5. Continued.

Article	Design	Administration method	Condition, part, score	Basis of outcome	Results per outcome
Task Switching Paradigm - executive functioning					
Chen et al. (2018)	Cross-over	Computer	Homogeneous condition[rr]	Accuracy	Global switching[ss] accuracy: no evidence for an effect found.
				Time	Global switching[ss] RT: no evidence for an effect found.
			Heterogeneous condition[tt]	Accuracy	Global switching[ss] accuracy: no evidence for an effect found. Local switching[uu] accuracy: no evidence for an effect found.
				Time	Global switching[ss] RT: faster RT post-PA compared to post-rest in 20 min PA condition. No evidence for an effect found for other PA durations. Local switching[uu] RT: no evidence for an effect found.
Simple Response Time Test - complex attention					
Córdova et al. (2009)	RCT	Computer		Time	Mean RT: faster RT post PA compared to post-rest in all PA groups.
Tower of Hanoi - executive functioning (planning)					
Córdova et al. (2009)	RCT	Manual + stopwatch		Accuracy	Number of moves: improved performance in post-PA compared to pre-PA in PA group exercising at 90% of anaerobic threshold. No evidence for an effect found in rest group.
				Time	Completion time: no evidence for an effect found.
Finger Tapping Test - executive functioning; perceptual-motor functioning					
Emery et al. (2001)	Cross-over	Manual		Motor	Mean number of finger taps: no evidence for an effect found.
Go/No-Go Sustained Attention to Response Test - complex attention					
Hsieh et al. (2016)	Cross-over	Computer		Accuracy	Rate of commission errors[ww]: no evidence for an effect found. Rate of omission errors[xx]: no evidence for an effect found.
				Time	RT in go trials[vv]: faster RT post-PA compared to post-rest. SD of RT in go trials: smaller SD post-PA compared to post-rest. Coefficient of variation of RT in go trials: smaller coefficient of variation of RT post-PA compared to post-rest.
Task-switching Reaction Time Test - executive functioning					
Kimura and Hozumi (2012)	RCT	Computer		Accuracy	Accuracy rate: better accuracy post-PA compared to pre-PA in both PA groups.
				Time	Mean RT: faster RT post-PA compared to pre-PA in both PA groups. Switch cost[yy]: better performance post-PA compared to pre-PA in combination style group. No evidence for an effect found in free style group.
Numerical memory updating task - executive functioning (working memory)					
Luo et al. (2023)	Micro-longitudinal	Smartphone		Accuracy	Accuracy rate across the 2 administered trials: no evidence for an effect found.

(*Continued*)

Table 5. Continued.

Article	Design	Administration method	Condition, part, score	Basis of outcome	Results per outcome
Symbol Digit Modalities Test - complex attention (processing speed, selective attention)					
Martelli et al. (2021)	RCT	Pen and paper		Accuracy	Number of correct items: more correct answers post-PA compared to pre-PA in PA group walking with perturbations. No evidence for an effect found for PA group walking without perturbations.
				Time	Completion time: faster completion time post-PA compared to pre-PA in PA group walking with perturbations. No evidence for an effect found for PA group walking without perturbations.
Associative novel word learning task - learning and memory					
McSween et al. (2021)	RCT	Computer		Accuracy	Immediate recall accuracy rate: better accuracy post-PA in moderate intensity cycling group compared to stretching group, but only in participants with lower baseline learning abilities. Immediate recognition accuracy rate: no evidence for an effect found.
				Time	Immediate recognition RT: no evidence for an effect found.
More-odd - executive functioning (flexibility)					
Naderi et al. (2019)	RCT	Computer		Accuracy	Accuracy: better accuracy post-PA compared to pre-PA in PA groups. No evidence for an effect found in rest groups.
				Time	RT: faster RT post-PA compared to pre-PA in PA groups. No evidence for an effect found in rest groups. Switch cost[zz]: better performance post-PA compared to pre-PA in PA groups. No evidence for an effect found in rest groups.
Alternative Uses Test - executive functioning (flexibility)					
Netz et al. (2009)	Pretest-posttest	Verbal		Accuracy	Number of correct items: better scores post-PA compared to pre-PA. Repetition of same use for one object: no evidence for an effect found. Repetition of same use for different objects: no evidence for an effect found. Number of responses not complying with instructions: no evidence for an effect found.
Logical Memory Test - learning and memory					
Nielson et al. (2014)	Cross-over	Verbal		Accuracy	Number of principal events correctly recalled: no evidence for an effect found. Number of correctly identified themes: no evidence for an effect found.
Benton Visual Retention Test - learning and memory					
Nielson et al. (2014)	Cross-over	Pen and paper		Accuracy	Number of correctly drawn figures: no evidence for an effect found.
Working Memory Updating Test - executive functioning					
Nouchi et al. (2020)	RCT	Pen and paper		Accuracy	Number of correct items: no evidence for an effect found.
Digit Cancellation Test - complex attention					
Nouchi et al. (2020)	RCT	Pen and paper		Accuracy	Number of correct items: no evidence for an effect found.

(*Continued*)

Table 5. Continued.

Article	Design	Administration method	Condition, part, score	Basis of outcome	Results per outcome
Sound-Induced Flash Illusion Test - perceptual-motor functioning					
O'Brien et al. (2017)	Pretest-posttest	Computer		Accuracy	Accuracy rate: better accuracy post-PA compared to pre-PA in open skill group. No evidence for an effect found in closed skill group or rest group.
d2 Test - executive functioning; complex attention (sustained attention, selective attention)					
Peiffer et al. (2015)	Cross-over	Computer		Accuracy	Error rate: lower error rate 30 min post-PA, compared to pre-PA in both PA conditions. Working speed[aaa]: faster working speed immediately and 30 min post-PA, compared to pre-PA in both PA conditions. Attention span[bbb]: better attention span 30 min post-PA, compared to pre-PA in both PA conditions.
Daily Everyday Cognitive Assessment - executive functioning					
Phillips et al. (2016)	Micro-longitudinal	Pen and paper		Accuracy	Number of correct items: no evidence for an effect found.
Letter Series Test - executive functioning					
Phillips et al. (2016)	Micro-longitudinal	Pen and paper		Accuracy	Number of correctly solved problems: no evidence for an effect found.
California Verbal Learning Test - learning and memory					
Schramke and Bauer (1997)	RCT	Not reported		Accuracy	Number of correctly recalled items: no evidence for an effect found. Number of targets correctly identified: no evidence for an effect found. Distracters correctly identified: no evidence for an effect found.
Saccadic Paradigm - executive functioning (inhibition)					
Tsai et al. (2021)	Cross-over	Computer	Prosaccade task[ccc]	Time	RT: no evidence for an effect found.
			Antisaccade task[ddd]	Time	RT: faster RT in post-PA compared to pre-PA in both PA conditions. No evidence for an effect found in rest condition.
Delayed match to sample - executive functioning (working memory)					
Tsai and Pan (2023)	Cross-over	Computer		Accuracy	Accuracy rate: better accuracy post-PA compared to pre-PA in moderate intensity condition. No evidence for an effect found in rest condition and high-intensity condition.
				Time	RT: faster RT post-PA compared to pre-PA in both PA conditions. No evidence for an effect found in rest condition.
Unspecified working memory test - executive functioning (working memory)					
Tsujii et al. (2013)	Cross-over	Computer		Time	Mean RT: faster RT post-PA compared to post-rest.
Famous Names Test - learning and memory (semantic and autobiographical memory)					
Won et al. (2019b)	Cross-over	Computer		Accuracy Time	RT: no evidence for an effect found. Accuracy rate: no evidence for an effect found.

Table 6. The number of tests, assessments, and outcomes per (sub)domain. Abbreviations: LT = long term.

Cognitive domains and subdomains	Number of times assessed	Number of tests	Number of outcomes	Number of significant outcomes	Proportion of significant outcomes (%)
Executive functioning	52	24	125	58	46.4
Inhibition	18	3	56	32	57.1
Working memory	18	8	27	10	37.0
Flexibility	23	5	58	31	53.4
Planning	2	2	9	4	44.4
No subdomain(s) mentioned	n/a	8	n/a	n/a	n/a
Complex attention	29	9	66	35	53.0
Processing speed	23	4	54	28	51.9
Selective attention	15	3	46	29	63.0
Sustained attention	14	2	44	27	61.4
No subdomain(s) mentioned	n/a	4	n/a	n/a	n/a
Perceptual-motor functioning	11	4	13	3	23.1
Visual perception	5	1	7	2	28.6
Perceptual-motor coordination	4	1	4	0	0
No subdomain(s) mentioned	n/a	2	n/a	n/a	n/a
Learning and memory	6	6	13	2	15.4
Recognition memory	1	1	2	1	50.0
Semantic and autobiographical LT memory	1	1	2	0	0
No subdomain(s) mentioned	n/a	4	n/a	n/a	n/a
Language	3	1	3	1	33.3
No subdomain(s) mentioned	n/a	1	n/a	n/a	n/a

on how the tests were described by the articles in which they were used (see Table 5 for the (sub)domains each test was assigned to). Therefore, the results of a domain are not equal to the sum of their subdomains' results.

Executive functioning was the most frequently measured cognitive domain (effect assessed 52 times by 24 tests), followed by complex attention (effect assessed 29 times by 9 tests), learning and memory (effect assessed 6 times by 6 tests), perceptual-motor functioning (effect assessed 11 times by 4 tests), and language (effect assessed 3 times by 1 test). When examining the specific outcomes per domain and subdomain, it becomes apparent that the highest proportions of acute effects of PA on cognition were found in the domains and subdomains of complex attention and executive functioning.

In total, 35 different cognitive tests were employed. The Stroop Test (executive functioning, complex attention) was used most frequently, in 13 articles. Furthermore, the Trail Making Test (executive functioning, complex attention) was used in five articles, and the Eriksen Flanker Test (executive functioning, complex attention) and Digit Symbol Substitution Test (complex attention, perceptual-motor functioning, learning and memory) were each used in four articles. The Verbal Fluency Test (executive functioning, language) was used in three articles, and the Sternberg Memory Test (executive functioning), n-back test (executive functioning), Wisconsin Card Sorting

Test (executive functioning), and Digit Span Test (executive functioning) were each used in two articles. The remaining 26 tests were each used in one article.

Across all conducted tests, 24 reported both accuracy- and time-based outcomes, 24 used only accuracy-based outcomes, 14 used only time-based outcomes, and one used a motor-based outcome (number of finger taps in 10 s). In total, all test administrations produced 148 outcomes, of which 77 were accuracy-based, 70 were time-based, and one was motor-based. Across all outcomes, 67 (45.3%) showed an effect of PA. Across the 77 accuracy-based outcomes, 24 (31.2%) demonstrated an improvement in accuracy scores after PA. Across the 70 time-based outcomes, 43 (61.4%) showed faster response times after PA. The one motor-based outcome did not demonstrate a better performance after PA. Furthermore, a large variety of outcomes was used, sometimes specific to the cognitive test. The most frequently used accuracy-based outcomes were accuracy rate, error rate, number of correct items, and number of errors. The most frequently used time-based outcomes were response time and completion time.

When comparing the number of significant outcomes of articles with strong designs to articles with moderate designs (no articles employed weak designs according to the ROB assessment), the proportion of significant outcomes was similar (44.6% and 47.1%, respectively).

4. Discussion

This systematic review aimed to provide an overview of the current body of literature on the acute effects of PA on cognition in adults aged 50 and older, focusing on (1) the assessed cognitive domains, (2) the employed cognitive tests, (3) the reported outcomes, and (4) whether reported results provided evidence for acute changes in cognition. The review of Pontifex et al. (2019) already offered a comprehensive overview of this effect in all age groups. However, because of the aging population and the consequences it entails, a more detailed overview focusing on middle-aged and older adults was warranted. Furthermore, as the review of Pontifex et al. (2019) identified articles that were published before 2018, many new studies have been performed since. The current systematic review included 18 articles that were published in 2018 or later, which demonstrates the growing interest in the acute effects of PA on cognition. Lastly, the current systematic review further complements the review of Pontifex et al. (2019) by including a wider range of study designs, including micro-longitudinal designs.

4.1. Cognitive (sub)domains

The most frequently assessed cognitive domain was executive functioning. This reflects previous research, as executive functioning is historically the most well-researched domain when assessing the acute effect of PA on cognition (Erickson et al., 2019; Pontifex et al., 2019). Across the executive functioning tests for which at least one subdomain was specified, cognitive flexibility, working memory, and inhibition were assessed most frequently. The subdomain of planning was assessed less often. Next to executive functioning, complex attention was also assessed by a fair amount of studies. The subdomain processing speed was measured most often, but selective and sustained attention were also assessed frequently. The domains and subdomains of learning and memory, perceptual-motor functioning, and language were tested less often. The domain of social cognition was not assessed in any article. These results are similar to the findings of Pontifex et al. (2019), as these authors also found that executive functioning and complex attention were assessed most often, and learning and memory and perceptual-motor functioning were assessed less frequently. Contrary to this review, Pontifex et al. (2019) identified more inhibition assessments compared to other subdomains, whereas this review found that inhibition was assessed a similar amount of times as working memory, flexibility, processing speed, selective attention, and sustained attention.

Furthermore, the current review found higher proportions of significant effects in specific (sub)domains: within executive functioning, the subdomains of inhibition and flexibility showed the highest proportion of significant effects (57.1% and 53.4%, respectively). Planning and working memory outcomes demonstrated a lower proportion of significant effects (44.4% and 37.0%, respectively). All assessed subdomains within complex attention (i.e. processing speed, selective attention, and sustained attention) showed significant effects in more than half of the outcomes (51.9%, 63.0%, and 61.4%, respectively). It has already been suggested that the acute effects of PA on cognition may be most pronounced on functions of the prefrontal cortex, such as executive functioning and complex attention, as this area is known to be affected by short bouts of PA (Basso & Suzuki, 2017).

This review found, similar to the existing literature, that the other domains were assessed considerably less frequently (Basso & Suzuki, 2017; Pontifex et al., 2019). However, these domains should not be overlooked, since certain skills in these domains are known to decline throughout the normal aging process (Grainger et al., 2023; Harada et al., 2013) and several studies have already demonstrated acute effects of PA on these cognitive domains (Basso & Suzuki, 2017). Therefore, further research is warranted to further clarify whether the other cognitive domains can also benefit from an acute bout of PA.

4.2. Cognitive tests

In contrast to the limited variation among the reported cognitive (sub)domains, many different cognitive tests were used in the included articles. Among the 42 included articles, 35 different cognitive tests were mentioned, with the Stroop test being the most frequently administered. Among the tests used in multiple articles, various versions, and administration methods were reported. For instance, the 13 articles that administered the Stroop test each used a different version. Moreover, certain articles did not refer to or provide detailed descriptions of the cognitive testing protocol, such as unexplained modifications to a test. Therefore, comparing test results across articles is not straightforward, even when the same test is used.

Furthermore, the ROB component score data collection methods revealed that the majority of the articles used cognitive tests that have not been shown to be valid or reliable for the specific population. Important to note is that determining the reliability of repeated cognitive testing is different from the standard reliability analyses (e.g. test-retest reliability). Most traditional cognitive tests are meant to be administered once and are used to discern between-person differences, e.g. identifying neurodiversity, intellectual giftedness, dementia, etc. (Sliwinski et al., 2018). Therefore, many cognitive tests may not offer reliable assessments of intra-individual changes in cognition in healthy populations (Pontifex et al., 2019). In addition, as most tests are not designed to be administered more than once within a short time span, improvements might be affected by practice effects, particularly in studies without a control group or control condition. Furthermore, many of the cognitive tests were not validated in middle-aged and older adults. This is important since research has shown that one cognitive test may assess different cognitive functions in different age groups, likely caused by an age-related change in processing strategies (Gajewski et al., 2018). Some examples of tests that have been shown to be valid and reliable in assessing acute changes in cognition in older adults with repeated measures designs are the 2-back test (Sliwinski et al., 2018), dot memory test (Sliwinski et al., 2018), and symbol search test (Hernandez et al., 2023; Sliwinski et al., 2018), with their specific protocols and outcomes. Lastly, as demonstrated in Table 5, several studies performed their timekeeping manually with a stopwatch or did not report their timekeeping or administration methods. However, cognitive tests that intend to identify small intra-individual differences require consistent procedures and precise measurements. Consequently, it could be argued that computerized administration methods are preferred over procedures requiring an administrator to keep time or present cues.

4.3. Reported outcomes

Among the administered cognitive tests, various outcomes were reported, of which 45.3% found an acute effect of PA on cognition. In general, outcomes were mostly accuracy- or time-based. Time-based outcomes appeared more sensitive in detecting acute effects of PA on cognition, with more than half demonstrating an effect, compared to one-third of the accuracy-based outcomes. However, drawing such a conclusion would be insufficiently nuanced, since the majority of the reported cognitive tests did not use both accuracy- and time-based outcomes, despite this being highly recommended when measuring cognition (Liesefeld & Janczyk, 2019; Wickelgren, 1977). Both accuracy- and time-based outcomes should be measured and reported, as this respects the speed-accuracy tradeoff, which is inherent in decision-making processes. This tradeoff is defined as ‘the complex relationship between an individual’s willingness to respond slowly and make relatively fewer errors compared to their willingness to respond quickly and make relatively more errors’ (Zimmerman, 2011). Hence, accuracy- and time-based outcomes are needed to conclude whether an effect was found. This could explain why conclusions on the acute effect of PA on cognition in the existing literature vary so strongly.

Moreover, it might not be realistic to expect all outcomes across all tests to improve, as many tests often have high accuracy scores (ceiling effect) (Faria et al., 2024). This reduces variation across assessments and decreases the likelihood that changes in cognition will be reflected in accuracy scores. Additionally, some outcomes were not reported in sufficient detail. For instance, response time was frequently used as outcome but it was often unclear whether it was measured across all trials or only correct trials. Furthermore, various outcomes were reported across all tests, even among articles that administered the same test. Moreover, articles sometimes used different names for the same outcome or the same name for different outcomes. This ambiguous variation of outcomes complicates study reproducibility and the comparison of findings across articles.

4.4. Study designs and protocols

Pontifex et al. (2019) already emphasized the importance of selecting a strong research design that is appropriate for the specific research question. In this review, none of the included articles received a weak rating on the 'study design' component of the ROB analysis. When comparing the articles published before 2018 with the articles that were published later, a higher percentage of recent articles utilized strong research designs (37.5% before 2018 vs. 55.6% in 2018 or later), suggesting a potential shift towards stronger research designs in recent years. In addition to the five types of study designs Pontifex et al. (2019) defined, the current review also identified two articles with micro-longitudinal designs, in which participants were observed multiple times over several days in their daily lives. One of these articles used ecological momentary assessment (EMA) to repeatedly test participants' cognition in real-time in their natural environment (Luo et al., 2023). Therefore, EMA minimizes recall bias and maximizes ecological validity, making it suitable to assess whether lab-based findings can be translated to real-life settings. Important to note is that the cognitive tests used in such designs must be suitable for repeated assessments and cannot be overly complicated, since the participants perform the tests on their own.

Despite all included articles having moderate or strong designs, the ROB analysis revealed that many articles had an overall low quality, due to low component scores for 'selection bias', 'data collection methods', and 'withdrawals and drop-outs'. First, the majority of the included articles recruited volunteers instead of randomly selecting participants, which could have resulted in samples that were already interested in PA and cognition, and therefore could have caused selection bias (Smith & Vanderweele, 2019). Second, as discussed earlier, many of the reported cognitive tests have not been demonstrated valid or reliable for repeatedly assessing cognition in this specific population. Third, withdrawals and drop-outs were often not reported or could not be determined. Since withdrawals and drop-outs can affect a study's validity, it is crucial to clearly report the number of participants at each stage of a study and analyze the data accordingly (Bell et al., 2013).

Furthermore, the PA and rest procedures in the included articles varied greatly. Various types, durations, and intensities of PA were reported, and cognition was assessed at different post-PA intervals (e.g. immediately after vs. 30 min after), which complicates the comparison of studies. These parameters can also influence the acute effects of PA on cognition. Chang et al. (2012b) demonstrated that light to moderate-intensity PA appeared to be the most beneficial, and higher intensity resulted in negligible effects. Regarding the duration of the PA session, sessions of 11 min or longer resulted in positive effects, while shorter duration demonstrated no significant effects on cognition (Chang et al., 2012b). The delay between a PA session and cognitive assessment also appeared significant (Chang et al., 2012b). The strongest effects were demonstrated when cognition was tested 11-20 min after finishing the exercise. When the delay was more than 20 min, the overall effect was also positive but smaller. If a test was conducted within 10 min after finishing the session, exercise appeared to have a negative effect on cognition (Chang et al., 2012b).

Additionally, various rest conditions were used across the included articles. However, some rest activities may affect cognition differently than others, e.g. lying down vs. playing cards. Pontifex et al. (2019) suggested using a control condition in which the only difference from the PA condition is the absence of PA. This large variety of procedures challenges drawing conclusions regarding which specific types of PA affect cognition and when and how long this effect takes place. Therefore, detailed reporting of PA sessions and cognitive test administration is recommended, e.g. whether a warming-up was included or what the duration of the cognitive assessment was. This will allow future meta-analyses to assess which types, durations, and intensities of PA affect which cognitive (sub)domains and when these effects take place.

4.5. Strengths and limitations of the review processes

This systematic review has several strengths. First, it is the first that includes such a broad overview of the acute effect of PA on cognition in healthy adults aged 50 and older. The definitions of PA and cognition were kept broad and the search was not limited to specific research designs to give an extensive overview of the existing research on this topic. Second, the search strategy was created together with experts from Knowledge Centre for Health Ghent, ensuring a thorough literature search.

Despite its strengths, this review has some limitations. First, the ROB assessment tool chosen for this review, while clear, overlooked certain study designs like cross-over and (micro)longitudinal designs. This required some components to be slightly adapted for specific designs and made certain components not applicable to certain cases. Second, several articles consisted of multiple PA conditions/groups or administered the cognitive tests at multiple post-PA time points. However, this review did not describe all possible comparisons between groups and time points, as this would have made the results section too extensive. For instance, when an article reported an improvement from baseline to immediately post-PA, but not from baseline to 30 min post-PA, this review drew positive conclusions for that outcome, despite not finding an effect at both time points. Therefore, the results of this review should be interpreted carefully, as they overestimate the acute effect of PA on cognition. Third, we recognize that different approaches for classifying cognitive (sub)domains exist (McCaffrey & Wright, 2022). We chose an existing classification of cognitive domains (i.e. the classification used in the DSM-5), because of its comprehensiveness. Using any other system could slightly change the domains and results, but the overall conclusions would likely remain the same. Fourth, linking the tests to cognitive (sub) domains was based on how the tests were described by the articles. Therefore, the terminology used by the articles to describe the tests could be inconsistent with the DSM-5 classification.

4.6. Implications for practice, policy, and future research

This review demonstrated that the most evidence for an acute effect of PA on cognition was found in the domains of complex attention and executive functioning. Nevertheless, since these domains were also the most frequently assessed, further research on the acute effect of PA on the other cognitive domains (e.g. learning and memory, language, etc.) is warranted. Furthermore, researchers should select valid and reliable cognitive tests for the target population and study protocol. In addition, the tests' properties, administration methods, and outcomes should be reported as clearly and in as much detail as possible. Preferably, cognitive tests should include both accuracy- and time-based outcomes, as this respects the speed-accuracy trade-off and provides more nuanced conclusions.

5. Conclusion

This systematic review investigated the acute effect of PA on cognition in healthy adults aged 50 and older. Executive functioning was the most frequently assessed cognitive domains, with slightly less than half of the outcomes indicating improvements following PA. However, research on this effect in other domains remains limited. Despite numerous articles suggesting cognitive benefits, the large variety and insufficient transparency regarding cognitive testing and PA protocols underscore the need for more high-quality research in this area, with clear and detailed reporting.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This work was supported by the Special Research Fund of Ghent University under Grant BOF/24J/2021/065; and the Research Foundation Flanders under Grant 12ZF122N and published with the support of the University Foundation of Belgium.

ORCID

Fien De Block http://orcid.org/0000-0003-0422-7897
Delfien Van Dyck http://orcid.org/0000-0003-1783-9075
Benedicte Deforche http://orcid.org/0000-0002-1064-0525
Geert Crombez http://orcid.org/0000-0002-4744-8561
Louise Poppe http://orcid.org/0000-0001-9285-6962

References

Abe, T., Fujii, K., Hyodo, K., Kitano, N., & Okura, T. (2018). Effects of acute exercise in the sitting position on executive function evaluated by the Stroop task in healthy older adults. *Journal of Physical Therapy Science*, *30*(4), 609–613. https://doi.org/10.1589/JPTS.30.609

Barella, L. A., Etnier, J. L., & Chang, Y. K. (2010). The immediate and delayed effects of an acute bout of exercise on cognitive performance of healthy older adults. *Journal of Aging and Physical Activity*, *18*(1), 87–98. https://doi.org/10.1123/JAPA.18.1.87

Basso, J. C., & Suzuki, W. A. (2017). The effects of acute exercise on mood, cognition, neurophysiology, and neurochemical pathways: A review. *Brain Plasticity (Amsterdam, Netherlands)*, *2*(2), 127–152. https://doi.org/10.3233/BPL-160040

Bell, M. L., Kenward, M. G., Fairclough, D. L., & Horton, N. J. (2013). Differential dropout and bias in randomised controlled trials: When it matters and when it may not. *BMJ (Clinical Research Ed.)*, *346*(jan21 1), e8668–e8668. https://doi.org/10.1136/BMJ.E8668

Bherer, L., Erickson, K. I., & Liu-Ambrose, T. (2013). A review of the effects of physical activity and exercise on cognitive and brain functions in older adults. *Journal of Aging Research*, *2013*, 657508. https://doi.org/10.1155/2013/657508

Callow, D. D., Pena, G. S., Stark, C. E. L., & Smith, J. C. (2023). Effects of acute aerobic exercise on mnemonic discrimination performance in older adults. *Journal of the International Neuropsychological Society: JINS*, *29*(6), 519–528. https://doi.org/10.1017/S1355617722000492

Chang, Y. K., Chu, C. H., Wang, C. C., Song, T. F., & Wei, G. X. (2014a). Effect of acute exercise and cardiovascular fitness on cognitive function: An event-related cortical desynchronization study. *Psychophysiology*, *52*(3), 342–351. https://doi.org/10.1111/PSYP.12364

Chang, Y. K., Ku, P. W., Tomporowski, P. D., Chen, F. T., & Huang, C. C. (2012a). Effects of acute resistance exercise on late-middle-age adults' goal planning. *Medicine and Science in Sports and Exercise*, *44*(9), 1773–1779. https://doi.org/10.1249/MSS.0B013E3182574E0B

Chang, Y. K., Labban, J. D., Gapin, J. I., & Etnier, J. L. (2012b). The effects of acute exercise on cognitive performance: A meta-analysis. *Brain Research*, *1453*, 87–101. https://doi.org/10.1016/J.BRAINRES.2012.02.068

Chang, Y. K., Tsai, C. L., Huang, C. C., Wang, C. C., & Chu, I. H. (2014b). Effects of acute resistance exercise on cognition in late middle-aged adults: General or specific cognitive improvement? *Journal of Science and Medicine in Sport*, *17*(1), 51–55. https://doi.org/10.1016/J.JSAMS.2013.02.007

Chen, F. T., Etnier, J. L., Wu, C. H., Cho, Y. M., Hung, T. M., & Chang, Y. K. (2018). Dose-response relationship between exercise duration and executive function in older adults. *Journal of Clinical Medicine*, *7*(9), 279. https://doi.org/10.3390/JCM7090279

Cheng, A., Zhao, Z., Liu, H., Yang, J., & Luo, J. (2022). The physiological mechanism and effect of resistance exercise on cognitive function in the elderly people. *Frontiers in Public Health*, *10*, 1013734. https://doi.org/10.3389/FPUBH.2022.1013734/BIBTEX

Chu, C. H., Chen, A. G., Hung, T. M., Wang, C. C., & Chang, Y. K. (2015). Exercise and fitness modulate cognitive function in older adults. *Psychology and Aging*, *30*(4), 842–848. https://doi.org/10.1037/PAG0000047

Córdova, C., Silva, V. C., Moraes, C. F., Simões, H. G., & Nóbrega, O. T. (2009). Acute exercise performed close to the anaerobic threshold improves cognitive performance in elderly females. *Brazilian Journal of Medical and Biological Research = Revista Brasileira De Pesquisas Medicas e Biologicas*, *42*(5), 458–464. https://doi.org/10.1590/S0100-879X2009000500010

De Block, F., & Poppe, L. (2023, November 24). *Which cognitive tests are used to examine the short-term effect of physical activity on cognitive functioning in healthy adults aged 50 and older? - A systematic review*.

Emery, C. F., Honn, V. J., Frid, D. J., Lebowitz, K. R., & Diaz, P. T. (2001). Acute effects of exercise on cognition in patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, *164*(9), 1624–1627. https://doi.org/10.1164/AJRCCM.164.9.2104137

Erickson, K. I., Hillman, C., Stillman, C. M., Ballard, R. M., Bloodgood, B., Conroy, D. E., Macko, R., Marquez, D. X., Petruzzello, S. J., & Powell, K. E., FOR 2018 PHYSICAL ACTIVITY GUIDELINES ADVISORY COMMITTEE*. (2019). Physical activity, cognition, and brain outcomes: A review of the 2018 physical activity guidelines. *Medicine and Science in Sports and Exercise*, *51*(6), 1242–1251. https://doi.org/10.1249/MSS.0000000000001936

Erlenbach, E., Mcauley, E., & Gothe, N. P. (2021). The association between light physical activity and cognition among adults: A scoping review. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *76*(4), 716–724. https://doi.org/10.1093/GERONA/GLAB013

Faria, L. O., Frois, T., Fortes, L., de, S., Bertola, L., & Albuquerque, M. R. (2024). Evaluating the stroop test with older adults: Construct validity, short term test-retest reliability, and sensitivity to mental fatigue. *Perceptual and Motor Skills*, *131*(4), 1120–1144. https://doi.org/10.1177/00315125241253425

Faulkner, J., Stoner, L., Grigg, R., Fryer, S., Stone, K., & Lambrick, D. (2017). Acute effects of exercise posture on executive function in transient ischemic attack patients. *Psychophysiology*, *54*(8), 1239–1248. https://doi.org/10.1111/PSYP.12868

Gajewski, P. D., Hanisch, E., Falkenstein, M., Thönes, S., & Wascher, E. (2018). What Does the n-Back Task Measure as We Get Older? Relations Between Working-Memory Measures and Other Cognitive Functions Across the Lifespan. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/FPSYG.2018.02208

Gallaway, P. J., Miyake, H., Buchowski, M. S., Shimada, M., Yoshitake, Y., Kim, A. S., & Hongu, N. (2017). Physical activity: A viable way to reduce the risks of mild cognitive impairment, alzheimer's disease, and vascular dementia in older adults. *Brain Sciences*, *7*(2), 22. https://doi.org/10.3390/BRAINSCI7020022

Gamaldo, A. A., An, Y., Allaire, J. C., Kitner-Triolo, M. H., & Zonderman, A. B. (2012). Variability in performance: Identifying early signs of future cognitive impairment. *Neuropsychology*, *26*(4), 534–540. https://doi.org/10.1037/A0028686

Grainger, S. A., Crawford, J. D., Riches, J. C., Kochan, N. A., Chander, R. J., Mather, K. A., Sachdev, P. S., & Henry, J. D. (2023). Aging is associated with multidirectional changes in social cognition: Findings from an adult life-span sample ranging from 18 to 101 years. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *78*(1), 62–72. https://doi.org/10.1093/GERONB/GBAC110

Griebler, N., Schröder, N., Artifon, M., Frigotto, M., & Pietta-Dias, C. (2022). The effects of acute exercise on memory of cognitively healthy seniors: A systematic review. *Archives of Gerontology and Geriatrics*, *99*, 104583. https://doi.org/10.1016/J.ARCHGER.2021.104583

Haeger, A., Mangin, J. F., Vignaud, A., Poupon, C., Grigis, A., Boumezbeur, F., Frouin, V., Deverre, J. R., Sarazin, M., Hertz-Pannier, L., Bottlaender, M., Baron, C., Berland, V., Blancho, N., Desmidt, S., Doublé, C., Ginisty, C., Joly-Testault, V., Laurier, L., … Vuillemard, C. (2020). Imaging the aging brain: Study design and baseline findings of the SENIOR cohort. *Alzheimer's Research & Therapy*, *12*(1), 77. https://doi.org/10.1186/S13195-020-00642-1

Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine*, *29*(4), 737–752. https://doi.org/10.1016/J.CGER.2013.07.002

Hatta, A., Nishihira, Y., & Higashiura, T. (2013). Effects of a single bout of walking on psychophysiologic responses and executive function in elderly adults: A pilot study. *Clinical Interventions in Aging*, *8*, 945–952. https://doi.org/10.2147/CIA.S46405

Hernandez, R., Hoogendoorn, C., Gonzalez, J. S., Jin, H., Pyatak, E. A., Spruijt-Metz, D., Junghaenel, D. U., Lee, P. J., & Schneider, S. (2023). Reliability and validity of noncognitive ecological momentary assessment survey response times as an indicator of cognitive processing speed in people's natural environment: Intensive longitudinal study. *JMIR mHealth and uHealth*, *11*(1), e45203. https://doi.org/10.2196/45203

Hogan, C. L., Mata, J., & Carstensen, L. L. (2013). Exercise holds immediate benefits for affect and cognition in younger and older adults. *Psychology and Aging*, *28*(2), 587–594. https://doi.org/10.1037/A0032634

Hsieh, S. S., Chang, Y. K., Fang, C. L., & Hung, T. M. (2016). Acute resistance exercise facilitates attention control in adult males without an age-moderating effect. *Journal of Sport & Exercise Psychology*, *38*(3), 247–254. https://doi.org/10.1123/JSEP.2015-0282

Hsieh, S. S., Chang, Y. K., Hung, T. M., & Fang, C. L. (2016). The effects of acute resistance exercise on young and older males' working memory. *Psychology of Sport and Exercise*, *22*, 286–293. https://doi.org/10.1016/j.psychsport.2015.09.004

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *57*(2), P101–P115. https://doi.org/10.1093/GERONB/57.2.P101

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology*, *14*(4), 588–598. https://doi.org/10.1037//0894-4105.14.4.588

Hyodo, K., Dan, I., Suwabe, K., Kyutoku, Y., Yamada, Y., Akahori, M., Byun, K., Kato, M., & Soya, H. (2012). Acute moderate exercise enhances compensatory brain activation in older adults. *Neurobiology of Aging*, *33*(11), 2621–2632. https://doi.org/10.1016/J.NEUROBIOLAGING.2011.12.022

Hyodo, K., Suwabe, K., Yamaguchi, D., Soya, H., & Arao, T. (2021). Comparison between the effects of continuous and intermittent light-intensity aerobic dance exercise on mood and executive functions in older adults. *Frontiers in Aging Neuroscience*, *13*, 723243. https://doi.org/10.3389/FNAGI.2021.723243

Ji, Z., Feng, T., Mei, L., Li, A., & Zhang, C. (2019). Influence of acute combined physical and cognitive exercise on cognitive function: An NIRS study. *PeerJ.*, *7*(8), e7418. https://doi.org/10.7717/PEERJ.7418

Johnson, L., Addamo, P. K., Raj, I. S., Borkoles, E., Wyckelsma, V., Cyarto, E., & Polman, R. C. (2016). An acute bout of exercise improves the cognitive performance of older adults. *Journal of Aging and Physical Activity*, *24*(4), 591–598. https://doi.org/10.1123/JAPA.2015-0097

Kamijo, K., Hayashi, Y., Sakai, T., Yahiro, T., Tanaka, K., & Nishihira, Y. (2009). Acute effects of aerobic exercise on cognitive function in older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *64*(3), 356–363. https://doi.org/10.1093/GERONB/GBP030

Kimura, K., & Hozumi, N. (2012). Investigating the acute effect of an aerobic dance exercise program on neuro-cognitive function in the elderly. *Psychology of Sport and Exercise*, *13*(5), 623–629. https://doi.org/10.1016/j.psychsport.2012.04.001

Lebeau, J. C., Mason, J., Roque, N., & Tenenbaum, G. (2022). The effects of acute exercise on driving and executive functions in healthy older adults. *International Journal of Sport and Exercise Psychology*, *20*(1), 283–301. https://doi.org/10.1080/1612197X.2020.1849353

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy trade-offs. *Behavior Research Methods*, *51*(1), 40–60. https://doi.org/10.3758/S13428-018-1076-X/TABLES/3

Luo, M., Moulder, R. G., & Röcke, C. (2023). The short-term effects of activity engagement on working memory performance in older age. *Psychology and Aging*, *38*(2), 117–131. https://doi.org/10.1037/PAG0000727

Martelli, D., Kang, J., Aprigliano, F., Staudinger, U. M., & Agrawal, S. K. (2021). Acute effects of a perturbation-based balance training on cognitive performance in healthy older adults: A pilot study. *Frontiers in Sports and Active Living*, *3*, 688519. https://doi.org/10.3389/FSPOR.2021.688519/BIBTEX

McCaffrey, J., & Wright, J. (2022). Neuroscience and cognitive ontology: A case for pluralism. In F. De Brigard & W. Sinnott-Armstrong (Eds.), *Neuroscience and philosophy* (pp. 427–466). The MIT Press.

McKinney, T. L., Euler, M. J., & Butner, J. E. (2019). It's about time: The role of temporal variability in improving assessment of executive functioning. *The Clinical Neuropsychologist*, *34*(4), 619–642. https://doi.org/10.1080/13854046.2019.1704434 https://doi.org/10.1080/13854046.2019.1704434

McSween, M. P., Coombes, J. S., MacKay, C. P., Rodriguez, A. D., Erickson, K. I., Copland, D. A., & McMahon, K. L. (2019). The immediate effects of acute aerobic exercise on cognition in healthy older adults: A systematic review. *Sports Medicine (Auckland, N.Z.)*, *49*(1), 67–82. https://doi.org/10.1007/S40279-018-01039-9

McSween, M. P., McMahon, K. L., Maguire, K., Coombes, J. S., Rodriguez, A. D., Erickson, K. I., & Copland, D. A. (2021). The acute effects of different exercise intensities on associative novel word learning in healthy older adults: A randomized controlled trial. *Journal of Aging and Physical Activity*, *29*(5), 793–806. https://doi.org/10.1123/JAPA.2020-0093

Naderi, A., Shaabani, F., Esmaeili, A., Salman, Z., Borella, E., & Degens, H. (2019). Effects of low and moderate acute resistance exercise on executive function in community-living older adults. *Sport, Exercise, and Performance Psychology*, *8*(1), 106–122. https://doi.org/10.1037/spy0000135

Netz, Y., Argov, E., & Inbar, O. (2009). Fitness's moderation of the facilitative effect of acute exercise on cognitive flexibility in older women. *Journal of Aging and Physical Activity*, *17*(2), 154–166. https://doi.org/10.1123/JAPA.17.2.154

Nielson, K. A., Wulff, L. L., & Arentsen, T. J. (2014). Muscle tension induced after learning enhances long-term narrative and visual memory in healthy older adults. *Neurobiology of Learning and Memory*, *109*, 144–150. https://doi.org/10.1016/J.NLM.2014.01.008

Northey, J. M., Cherbuin, N., Pumpa, K. L., Smee, D. J., & Rattray, B. (2018). Exercise interventions for cognitive function in adults older than 50: A systematic review with meta-analysis. *British Journal of Sports Medicine*, *52*(3), 154–160. https://doi.org/10.1136/BJSPORTS-2016-096587

Nouchi, R., Nouchi, H., & Kawashima, R. (2020). A single 30 minutes bout of combination physical exercises improved inhibition and vigor-mood in middle-aged and older females: Evidence from a randomized controlled trial. *Frontiers in Aging Neuroscience*, *12*, 179. https://doi.org/10.3389/FNAGI.2020.00179

Oberste, M., Sharma, S., Bloch, W., & Zimmer, P. (2021). Acute exercise-induced set shifting benefits in healthy adults and its moderators: A systematic review and meta-analysis. *Frontiers in Psychology*, *12*, 528352. https://doi.org/10.3389/FPSYG.2021.528352

O'Brien, J., Ottoboni, G., Tessari, A., & Setti, A. (2017). One bout of open skill exercise improves cross-modal perception and immediate memory in healthy older adults who habitually exercise. *PLOS One*, *12*(6), e0178739. https://doi.org/10.1371/JOURNAL.PONE.0178739

Olivo, G., Nilsson, J., Garzón, B., Lebedev, A., Wåhlin, A., Tarassova, O., Ekblom, M., & Lövdén, M. (2021). Immediate effects of a single session of physical exercise on cognition and cerebral blood flow: A randomized controlled study of older adults. *NeuroImage*, *225*, 117500. https://doi.org/10.1016/J.NEUROIMAGE.2020.117500

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ (Clinical Research Ed.)*, *372*, n160. https://doi.org/10.1136/BMJ.N160

Peiffer, R., Darby, L. A., Fullenkamp, A., & Morgan, A. L. (2015). Effects of acute aerobic exercise on executive function in older women. *Journal of Sports Science & Medicine*, *14*(3), 574–583.

Pellegrini-Laplagne, M., Dupuy, O., Sosner, P., & Bosquet, L. (2022). Acute effect of a simultaneous exercise and cognitive task on executive functions and prefrontal cortex oxygenation in healthy older adults. *Brain Sciences*, *12*(4), 455. https://doi.org/10.3390/BRAINSCI12040455

Phillips, C. B., Edwards, J. D., Andel, R., & Kilpatrick, M. (2016). Daily physical activity and cognitive function variability in older adults. *Journal of Aging and Physical Activity*, *24*(2), 256–267. https://doi.org/10.1123/JAPA.2014-0222

Pontifex, M. B., McGowan, A. L., Chandler, M. C., Gwizdala, K. L., Parks, A. C., Fenn, K., & Kamijo, K. (2019). A primer on investigating the after effects of acute bouts of physical activity on cognition. *Psychology of Sport and Exercise*, *40*, 1–22. https://doi.org/10.1016/j.psychsport.2018.08.015

Quality Assessment Tool for Quantitative Studies. (n.d). Retrieved. October 26, 2023, from https://www.ephpp.ca/quality-assessment-tool-for-quantitative-studies/

Residential Facilities, Assisted Living, and Nursing Homes | National Institute on Aging. (n.d). Retrieved December 20, 2022, from https://www.nia.nih.gov/health/residential-facilities-assisted-living-and-nursing-homes

Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., & Petersen, R. C. (2014). Classifying neurocognitive disorders: The DSM-5 approach. *Nature Reviews. Neurology*, *10*(11), 634–642. https://doi.org/10.1038/NRNEUROL.2014.181

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society: JINS*, *16*(5), 754–760. https://doi.org/10.1017/S1355617710000706
Schramke, C. J., & Bauer, R. M. (1997). State-dependent learning in older and younger adults. *Psychology and Aging*, *12*(2), 255–262. https://doi.org/10.1037//0882-7974.12.2.255
Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and validity of ambulatory cognitive assessments. *Assessment*, *25*(1), 14–30. https://doi.org/10.1177/1073191116643164
Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging*, *21*(3), 545–557. https://doi.org/10.1037/0882-7974.21.3.545
Smith, L. H., & Vanderweele, T. J. (2019). Bounding bias due to selection. *Epidemiology (Cambridge, Mass.)*, *30*(4), 509–516. https://doi.org/10.1097/EDE.0000000000001032
Sofi, F., Valecchi, D., Bacci, D., Abbate, R., Gensini, G. F., Casini, A., & Macchi, C. (2011). Physical activity and risk of cognitive decline: A meta-analysis of prospective studies. *Journal of Internal Medicine*, *269*(1), 107–117. https://doi.org/10.1111/J.1365-2796.2010.02281.X
Sturnieks, D. L., Yak, S. L., Ratanapongleka, M., Lord, S. R., & Menant, J. C. (2018). A busy day has minimal effect on factors associated with falls in older people: An ecological randomised crossover trial. *Experimental Gerontology*, *106*, 192–197. https://doi.org/10.1016/J.EXGER.2018.03.009
Tsai, C. L., Chang, Y. C., Pan, C. Y., Wang, T. C., Ukropec, J., & Ukropcová, B. (2021). Acute effects of different exercise intensities on executive function and oculomotor performance in middle-aged and older adults: Moderate-intensity continuous exercise vs. high-intensity interval exercise. *Frontiers in Aging Neuroscience*, *13*, 743479. https://doi.org/10.3389/FNAGI.2021.743479/BIBTEX
Tsai, C. L., & Pan, C. Y. (2023). Acute and protocol-dependent effects of aerobic exercise on neurobiochemical indices and neuropsychological performance of working memory. *Mental Health and Physical Activity*, *24*, 100494. https://doi.org/10.1016/j.mhpa.2022.100494
Tsujii, T., Komatsu, K., & Sakatani, K. (2013). Acute effects of physical exercise on prefrontal cortex activity in older adults: A functional near-infrared spectroscopy study. *Advances in Experimental Medicine and Biology*, *765*, 293–298. https://doi.org/10.1007/978-1-4614-4989-8_41
Wang, C. C., Shih, C. H., Pesce, C., Song, T. F., Hung, T. M., & Chang, Y. K. (2015). Failure to identify an acute exercise effect on executive function assessed by the Wisconsin Card Sorting Test. *Journal of Sport and Health Science*, *4*(1), 64–72. https://doi.org/10.1016/j.jshs.2014.10.003
WHO guidelines on physical activity and sedentary behaviour. (2020).
Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9
Won, J., Alfini, A. J., Weiss, L. R., Callow, D. D., & Smith, J. C. (2019a). Brain activation during executive control after acute exercise in older adults. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *146*, 240–248. https://doi.org/10.1016/J.IJPSYCHO.2019.10.002
Won, J., Alfini, A. J., Weiss, L. R., Michelson, C. S., Callow, D. D., Ranadive, S. M., Gentili, R. J., & Carson Smith, J. (2019b). Semantic memory activation after acute exercise in healthy older adults. *Journal of the International Neuropsychological Society: JINS*, *25*(6), 557–568. https://doi.org/10.1017/S1355617719000171
Zimmerman, M. E. (2011). Speed–accuracy tradeoff. *Encyclopedia of Clinical Neuropsychology*, 2344–2344. https://doi.org/10.1007/978-0-387-79948-3_1247

Appendices

Appendix A. PRISMA checklist.

PRISMA full-text checklist

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

(*Continued*)

Appendix A. Continued.

PRISMA full-text checklist			
Section and topic	Item #	Checklist item	Location where item is reported
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	n/a
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	n/a
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 7
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 7
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	n/a
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	n/a
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	n/a
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	n/a
Results			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Pages 7-8, Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	n/a
Study characteristics	17	Cite each included study and present its characteristics.	Page 8, Tables 1-4
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Pages 8-9, Appendix C
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Pages 9-10, Tables 5-6
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	n/a
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	n/a
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	n/a
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	n/a

(*Continued*)

Appendix A. Continued.

PRISMA full-text checklist			
Section and topic	Item #	Checklist item	Location where item is reported
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	n/a
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	n/a
Discussion			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Pages 10-14
	23b	Discuss any limitations of the evidence included in the review.	Pages 10-14
	23c	Discuss any limitations of the review processes used.	Pages 14-15
	23d	Discuss implications of the results for practice, policy, and future research.	Page 15
Other information			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 4
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 4
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	n/a
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Page 15
Competing interests	26	Declare any competing interests of review authors.	Page 15
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	n/a

Appendix B. search strategy.

MEDLINE database	
Concept	Search terms
Concept 1: physical activity	("Motor Activity"[Mesh:NoExp] OR "Exercise"[Mesh] OR "Sports"[Mesh] OR "Physical Exertion"[Mesh] OR "Motor Activit*"[tiab] OR "Physical Activit*"[tiab] OR "Locomotor Activit*"[tiab] OR "Physical Exercis*"[tiab] OR "Isometric Exercis*"[tiab] OR "Aerobic Exercis*"[tiab] OR training[tiab] OR stretching[tiab] OR "Physical Condition*"[tiab] OR "Physical fitness"[tiab] OR "Physical endurance"[tiab] OR "movement therap*"[tiab] OR running[tiab] OR jogging[tiab] OR walk*[tiab] OR bicycle[tiab] OR cycle[tiab] OR bicycling[tiab] OR cycling[tiab] OR "home training"[tiab])
Concept 2: acute	(daily[tiab] OR everyday[tiab] OR "day-to-day"[tiab] OR "day to day"[tiab] OR "within subjects"[tiab] OR acute[tiab] OR immediate[tiab] OR "within-day"[tiab] OR "within day"[tiab] OR periodic[tiab])
Concept 3: cognition	("Cognition"[Mesh:NoExp] OR "Memory"[Mesh] OR "Verbal Learning"[Mesh] OR "Executive Function"[Mesh] OR "Attention"[Mesh] OR "Problem Solving"[Mesh] OR "Learning"[Mesh:NoExp] OR "Neuropsychological Tests"[Mesh] OR "social cognition"[MeSH] OR "cognitive function*"[tiab] OR memory[tiab] OR memories[tiab] OR learning[tiab] OR learn[tiab] OR "executive function*"[tiab] OR attention[tiab] OR "problem solving"[tiab] OR neuropsycholog*[tiab] OR "social cognition"[tiab] OR "visual perception"[tiab] OR "spatial cognition"[tiab])

(*Continued*)

Appendix B. Continued.

MEDLINE database	
Concept	Search terms
Concept 4: adults aged 50 and older	("Aged"[Mesh] OR aged[tiab] OR elder*[tiab] OR eldest[tiab] OR "old age"[tiab] OR "oldest old"[tiab] OR senior[tiab] OR seniors[tiab] OR senium[tiab] OR "very old"[tiab] OR sexagenarian*[tiab] OR septuagenarian*[tiab] OR octogenarian*[tiab] OR octogenarian*[tiab] OR nonagenarian*[tiab] OR centenarian*[tiab] OR supercentenarian*[tiab] OR "older people"[tiab] OR "older subject"[tiab] OR "older subjects"[tiab] OR "older age"[tiab] OR "older adult"[tiab] OR "older adults"[tiab] OR "older man"[tiab] OR "older men"[tiab] OR "older male"[tiab] OR "older woman"[tiab] OR "older women"[tiab] OR "older female"[tiab] OR "older population"[tiab] OR "older person"[tiab] OR geriatric[tiab])
Concept 5: healthy	("Healthy Aging"[MeSH] OR "Healthy Volunteers"[MeSH] OR "Healthy Aging"[tiab] OR "Healthy Volunteers"[tiab] OR Healthy[tiab] OR unimpaired[tiab] OR non-disabled[tiab] OR non-handicapped[tiab] OR non-demented[tiab] OR "not impaired"[tiab] OR "not disabled"[tiab] OR "not demented"[tiab] OR "community dwelling"[tiab] OR "living alone"[tiab])
Combination of concepts	#1 AND #2 AND #3 AND #4 AND #5

Embase database	
Concept	Search terms
Concept 1: physical activity	('motor activity'/exp OR 'exercise'/exp OR 'sport'/exp OR 'physical activity'/mj OR 'training'/exp OR 'stretching'/exp OR 'endurance'/exp OR 'movement therapy'/exp OR 'running'/exp OR 'jogging'/exp OR 'walking'/mj OR 'cycling'/exp OR 'swimming'/exp OR 'motor activit*':ti,ab,kw OR 'locomotor activit*':ti,ab,kw OR exercise:ti,ab,kw OR 'physical exercis*':ti,ab,kw OR 'isometric exercis*':ti,ab,kw OR sports:ti,ab,kw OR 'physical exertion':ti,ab,kw OR 'physical activit*':ti,ab,kw OR 'physical fitness':ti,ab,kw OR training:ti,ab,kw OR stretching:ti,ab,kw OR 'physical endurance':ti,ab,kw OR 'aerobic exercis*':ti,ab,kw OR 'physical condition*':ti,ab,kw OR 'movement therap*':ti,ab,kw OR running:ti,ab,kw OR jogging:ti,ab,kw OR walk*:ti,ab,kw OR bicycle:ti,ab,kw OR cycle:ti,ab,kw OR bicycling:ti,ab,kw OR cycling:ti,ab,kw OR 'home training':ti,ab,kw)
Concept 2: acute	(daily OR everyday OR 'day-to-day' OR 'day to day' OR 'within subjects' OR acute OR immediate OR 'within-day' OR 'within day' OR periodic)
Concept 3: cognition	('social cognition' OR cognition OR attention OR 'cognitive flexibility' OR 'executive function*' OR learning OR memory OR 'mental performance' OR thinking OR 'problem solving' OR 'neuropsychological test' OR 'cognitive function*' OR memories OR learn OR executive OR neuropsycholog* OR 'visual perception' OR 'spatial cognition')
Concept 4: adults aged 50 and older	('aged'/mj OR 'very elderly'/exp OR aged:ti,ab,kw OR 'very elderly':ti,ab,kw OR elder*:ti,ab,kw OR eldest:ti,ab,kw OR 'old age':ti,ab,kw OR 'oldest old':ti,ab,kw OR senior:ti,ab,kw OR seniors:ti,ab,kw OR senium:ti,ab,kw OR 'very old':ti,ab,kw OR sexagenarian*:ti,ab,kw OR septuagenarian*:ti,ab,kw OR octogenarian*:ti,ab,kw OR octogenarian*:ti,ab,kw OR nonagenarian*:ti,ab,kw OR centenarian*:ti,ab,kw OR supercentenarian*:ti,ab,kw OR 'older people':ti,ab,kw OR 'older subject':ti,ab,kw OR 'older subjects':ti,ab,kw OR 'older age':ti,ab,kw OR 'older adult':ti,ab,kw OR 'older adults':ti,ab,kw OR 'older man':ti,ab,kw OR 'older men':ti,ab,kw OR 'older male':ti,ab,kw OR 'older woman':ti,ab,kw OR 'older women':ti,ab,kw OR 'older female':ti,ab,kw OR 'older population':ti,ab,kw OR 'older person':ti,ab,kw OR geriatric:ti,ab,kw)
Concept 5: healthy	('healthy aging'/exp OR 'normal human'/exp OR 'healthy aging':ti,ab,kw OR 'normal human':ti,ab,kw OR 'Healthy Volunteers':ti,ab,kw OR Healthy:ti,ab,kw OR unimpaired:ti,ab,kw OR 'non-disabled':ti,ab,kw OR 'non-handicapped':ti,ab,kw OR 'non-demented':ti,ab,kw OR 'not disabled':ti,ab,kw OR 'not demented':ti,ab,kw OR 'not impaired':ti,ab,kw OR 'not handicapped':ti,ab,kw OR 'community dwelling':ti,ab,kw OR 'living alone':ti,ab,kw)
Combination of concepts	#1 AND #2 NEAR/2 #3 AND #4 AND #5

Note: concepts 2 and 3 do not contain field codes, since they are not allowed when using the NEAR operator.

Appendix B. Continued.

Cochrane CENTRAL database	
Concept	Search terms
Concept 1: physical activity	([mh ^"Motor Activity"] OR [mh Exercise] OR [mh Sports] OR [mh "Physical Exertion"] OR "motor activit*":ti,ab,kw OR "locomotor activit*":ti,ab,kw OR exercise:ti,ab,kw OR "physical exercis*":ti,ab,kw OR "isometric exercis*":ti,ab,kw OR sports:ti,ab,kw OR "physical exertion":ti,ab,kw OR "physical activit*":ti,ab,kw OR "physical fitness":ti,ab,kw OR training:ti,ab,kw OR stretching:ti,ab,kw OR "physical endurance":ti,ab,kw OR "aerobic exercis*":ti,ab,kw OR "physical condition*":ti,ab,kw OR "movement therap*":ti,ab,kw OR running:ti,ab,kw OR jogging:ti,ab,kw OR walk*:ti,ab,kw OR bicycle:ti,ab,kw OR cycle:ti,ab,kw OR bicycling:ti,ab,kw OR cycling:ti,ab,kw OR "home training":ti,ab,kw)
Concept 2: acute	(daily:ti,ab,kw OR everyday:ti,ab,kw OR "day-to-day":ti,ab,kw OR "day to day":ti,ab,kw OR "within subjects":ti,ab,kw OR acute:ti,ab,kw OR immediate:ti,ab,kw OR "within-day":ti,ab,kw OR "within day":ti,ab,kw OR periodic:ti,ab,kw)
Concept 3: cognition	([mh ^"Cognition"] OR [mh "Memory"] OR [mh "Verbal Learning"] OR [mh "Executive Function"] OR [mh "Attention"] OR [mh "Problem Solving"] OR [mh ^"Learning"] OR [mh "Neuropsychological Tests"] OR [mh "social cognition"] OR cognition:ti,ab,kw OR attention:ti,ab,kw OR "cognitive flexibility":ti,ab,kw OR "executive function*":ti,ab,kw OR learning:ti,ab,kw OR memory:ti,ab,kw OR "mental performance":ti,ab,kw OR thinking:ti,ab,kw OR "problem solving":ti,ab,kw OR "neuropsychological test":ti,ab,kw OR "cognitive function*":ti,ab,kw OR memories:ti,ab,kw OR learn:ti,ab,kw OR executive:ti,ab,kw OR neuropsycholog*:ti,ab,kw OR "visual perception":ti,ab,kw OR "spatial cognition":ti,ab,kw OR "social cognition":ti,ab,kw)
Concept 4: adults aged 50 and older	([mh "Aged"] OR aged:ti,ab,kw OR "very elderly":ti,ab,kw OR elder*:ti,ab,kw OR eldest:ti,ab,kw OR "old age":ti,ab,kw OR "oldest old":ti,ab,kw OR senior:ti,ab,kw OR seniors:ti,ab,kw OR senium:ti,ab,kw OR "very old":ti,ab,kw OR sexagenarian*:ti,ab,kw OR septuagenarian*:ti,ab,kw OR octogenarian*:ti,ab,kw OR octogenarian*:ti,ab,kw OR nonagenarian*:ti,ab,kw OR centenarian*:ti,ab,kw OR supercentenarian*:ti,ab,kw OR "older people":ti,ab,kw OR "older subject":ti,ab,kw OR "older subjects":ti,ab,kw OR "older age":ti,ab,kw OR "older adult":ti,ab,kw OR "older adults":ti,ab,kw OR "older man":ti,ab,kw OR "older men":ti,ab,kw OR "older male":ti,ab,kw OR "older woman":ti,ab,kw OR "older women":ti,ab,kw OR "older female":ti,ab,kw OR "older population":ti,ab,kw OR "older person":ti,ab,kw OR geriatric:ti,ab,kw)
Concept 5: healthy	([mh "Healthy Aging"] OR [mh "Healthy Volunteers"] OR "healthy aging":ti,ab,kw OR "normal human":ti,ab,kw OR "Healthy Volunteers":ti,ab,kw OR Healthy:ti,ab,kw OR unimpaired:ti,ab,kw OR "non-disabled":ti,ab,kw OR "non-handicapped":ti,ab,kw OR "non-demented":ti,ab,kw OR "not disabled":ti,ab,kw OR "not demented":ti,ab,kw OR "not impaired":ti,ab,kw OR "not handicapped":ti,ab,kw OR "community dwelling":ti,ab,kw OR "living alone":ti,ab,kw)
Combination of concepts	#1 AND #2 AND #3 AND #4 AND #5

Appendix C. risk of bias assessment with component scores and total ratings.

Study	Selection bias[a]	Designb	Confounders[c]	Blindingd	Data collection methodse	Withdrawals and drop-outsf	Global rating
Abe et al. (2018)	W	S	n/a	M	W	S	W
Barella et al. (2010)	W	S	S	M	W	S	W
Callow et al. (2023)	Not reported	M	n/a	M	W	S	M
Chang et al. (2014a)	Not reported	M	S	S	W	W	W
Chang et al. (2012a)	Not reported	M	n/a	M	S	W	M
Chang et al. (2014b)	Not reported	M	n/a	M	W	W	W
Chen et al. (2018)	W	M	n/a	S	W	W	W
Chu et al. (2015)	Not reported	M	S	S	W	W	W
Córdova et al. (2009)	W	S	S	M	M	W	W
Emery et al. (2001)	Not reported	S	S	M	S	S	S
Faulkner et al. (2017)	W	S	S	M	W	W	W
Hatta et al. (2013)	Not reported	M	n/a	M	S	W	M
Hogan et al. (2013)	W	S	S	M	W	S	W
Hsieh et al. (2016)	W	M	n/a	M	W	S	W
Hsieh et al. (2016)	W	M	n/a	M	W	W	W
Hyodo et al. (2012)	Not reported	M	n/a	M	W	W	W
Hyodo et al. (2021)	W	M	n/a	M	W	S	W
Ji et al. (2019)	W	M	n/a	M	W	W	W
Johnson et al. (2016)	W	S	S	M	W	W	W
Kamijo et al. (2009)	Not reported	M	S	M	W	S	M
Kimura and Hozumi (2012)	Not reported	S	S	M	W	S	M
Lebeau et al. (2022)g	M	M	S	M	M	S	S
Luo et al. (2023)	W	M	n/a	M	S	S	M
Martelli et al. (2021)	Not reported	S	S	M	S	W	M
McSween et al. (2021)	W	S	S	S	W	S	W
Naderi et al. (2019)	W	S	S	M	W	W	W
Netz et al. (2009)	W	M	W	M	W	W	W
Nielson et al. (2014)	W	M	S	M	S	S	M
Nouchi et al. (2020)	W	S	S	M	M	S	M
O'Brien et al. (2017)	W	M	S	M	M	W	W
Olivo et al. (2021)	W	S	S	S	W	S	W
Peiffer et al. (2015)	Not reported	M	n/a	M	M	S	S
Pellegrini-Laplagne et al. (2022)	Not reported	S	n/a	M	W	S	M
Phillips et al. (2016)	Not reported	M	n/a	M	M	S	S
Schramke and Bauer (1997)	W	S	S	M	S	W	W
Sturnieks et al. (2018)	W	S	n/a	M	S	S	M
Tsai et al. (2021)	W	S	n/a	M	W	W	W
Tsai and Pan (2023)	W	S	n/a	M	W	W	W
Tsujii et al. (2013)	Not reported	M	n/a	M	W	W	W
Wang et al. (2015)	W	S	S	M	S	S	M
Won et al. (2019a)	W	M	n/a	M	W	S	W
Won et al. (2019b)	W	M	n/a	M	W	S	W

[a]Based on the selection method and percentage of individuals that agreed to participate.
[b]Based on the strength of the design type. [c]Based on differences between groups and number of confounders controlled for.
[d]Based on the blinding of outcome assessors and participants.
[e]Based on validity and reliability of used methods.
[f]Based on reporting of withdrawals and drop-outs.
[g]Study consisted of 2 designs: RCT and one group pretest-posttest design. Only the latter design was included in this review and scored for the risk of bias assessment.
Abbreviations: S = strong, M = moderate, W = weak.

Appendix D. clarification to Table 5

	Condition, score, outcome	Definition
Stroop test		
a	Interference score	RT or accuracy score calculated by subtracting the RT or accuracy score of neutral trials from RT or accuracy score of incongruent trials.
b	Color condition	Condition in which participants must indicate the color of a string of four colored circles.
c	Interference condition	Condition in which color names are printed in an ink color that does not match the color name. Participants must indicate the color.
d	Inhibition condition	Same condition as the interference condition, but the color of the word for each trial is the same as the color name on the previous trial.
e	Congruent condition	Condition in which color names are printed in the same ink color as the color name. Participants must indicate the color.
f	Word condition	Condition in which the color names are printed in black ink. Participants must indicate the color.
g	Color condition	Condition in which participants must indicate the color of a colored rectangle.
h	Neutral condition	Condition in which words unrelated to the colors are printed in colored ink. Participants must indicate the color.
i	Incongruent condition	Condition in which color names are printed in an ink color that does not match the color name. Participants must indicate the color.
j	Interference condition	Condition in which color names are printed in an ink color that does or does not match the color name. Participants must indicate the color.
k	Naming condition	Condition in which participants must indicate the color of a string of three colored x's.
l	Executive condition	Condition in which color names are printed in a color that does or does not match the color name. Participants must indicate the color of the ink when no rectangle is around the color word or read the color name when a rectangle is around the word.
m	Color condition	Condition in which participants must indicate the color of a string of colored x's.
n	Word interference condition	Condition in which color names are printed in an ink color that does not match the color name. Participants must indicate the color.
o	Inhibition condition	Same condition as the interference condition, but the color of the word for each trial is the same as the color name on the previous trial.
p	Stroop condition	Condition in which color names are printed in an ink color that does not match the color name. Participants must indicate the color.
q	Reverse Stroop condition	Condition in which color names are printed in a color that does not match the color name. Participants must indicate the color name.
r	Naming condition	Condition in which the color names are printed in the same ink color as the word. Participants must indicate the color.
s	Inhibition condition	Condition in which color names are printed in an ink color that does not match the color name. Participants must indicate the color.
t	Inhibition score	Inhibition block's score minus naming blocks' score.
u	Switching/flexibility condition	Condition in which color names are printed in a color. Participants must indicate the color of the ink when the usual fixation cross precedes the stimulus or read the color name when a square replaces the fixation cross.
v	Switching/flexibility score	Switching block's score minus inhibition blocks' score.
Trail Making test		
w	Total score	Score calculated as the difference between TMT-A and TMT-B.
Eriksen Flanker Test		
x	Congruent condition	Condition in which a row of five arrows all point in the same direction as the middle arrow (arrow of reference). Participants must indicate in which direction the middle arrow is pointing.
y	Incongruent condition	Condition in which a row of five arrows can point in the same or opposite direction as the middle arrow (arrow of reference). Participants must indicate in which direction the middle arrow is pointing.

(*Continued*)

Appendix D. Continued.

	Condition, score, outcome	Definition
z	Interference RT	Score calculated as ([incongruent RT − congruent RT)/congruent RT] * 100.
aa	Incongruent condition	Condition in which a row of five arrows point in the opposite direction as the middle arrow (arrow of reference). Participants must indicate in which direction the middle arrow is pointing.
n-back test		
bb	Performance score	Score calculated as the sum of correct hits and correct rejections, divided by the total number of stimuli.
Wisconsin Card Sorting Test		
cc	Perseverative errors (of Nelson)	Error made when a participant keeps on sorting the response cards using a previous matching rule after the matching rule has been changed.
dd	Perseverative responses	Preservative error of Nelson that is correct because the response card matches the stimulus card in multiple dimensions. This leads to an ambiguous answer, as the test administrator cannot determine which dimension the participant uses to sort the card. A preservative response is therefore defined as an ambiguous response that follows and is followed by an unambiguous perseverative error, and matches the perseverated-to principle.
ee	Non-perseverative errors	Random errors.
Digit Span Test		
ff	Forward digit span	Test version in which the participants must repeat the numbers in the same order as previously presented.
gg	Digit span score	Score determined as the last block where the participant correctly repeated at least one of the two strings of numbers.
hh	Digit span product score	Score calculated as the number of correct strings reported.
Mnemonic Similarity Task		
ii	Object recognition memory	Score calculated as difference between correctly recognized images in the retrieval phase as the same image as in the encoding phase and wrongly identified new images in the retrieval phase as images that were shown in the encoding phase.
jj	Mnemonic discrimination	Score calculated as the difference between correctly recognized images as 'similar' in the retrieval phase (similar image to an image that was shown in the encoding phase) and wrongly identified new images in the retrieval phase as images that were similar to what was shown in the encoding phase.
Tower of London Test		
kk	Total move score	Score calculated as the sum of differences between the number of actual ball moves and the minimum number of moves for each problem.
ll	Total correct score	Score calculated as the number of problems solved where the criteria of the minimum number of moves were reached.
mm	Rule violation score	Score calculated as a combined score of two types of rule violations: (1) placing or trying to place more balls on a peg than it can physically support and (2) removing two balls from the peg at the same time.
nn	Time violation score	This score was given when a problem was not finished in under one minute.
oo	Total initial time	Score calculated as the time between the presentation of the goal configuration by the examiner and the participant lifting the first ball of a post.
pp	Total execution time	Score calculated as the time between the first ball being lifted and the successful completion of a given problem.
qq	Total planning-solving time	Score calculated as the sum of the total initial time and the total execution time.
Task Switching Paradigm		
rr	Homogeneous condition	Testing condition in which the participant must follow one rule at a time.
ss	Global switching	This effect was calculated by using the performance of the homogeneous and heterogeneous conditions.
tt	Heterogeneous condition	Testing condition in which the participant must switch between two different rules.

(*Continued*)

Appendix D. Continued.

	Condition, score, outcome	Definition
uu	Local switching	This effect was calculated by using the performance of the non-switch and switch trials within the heterogeneous condition.
Go/no-go Sustained Attention to Response Test		
vv	Go trials	Trial in which a participant must make a response when presented with a certain stimulus.
ww	Rate of commission errors	Score calculated as the percentage of no-go symbols (no response required) responded to.
xx	Rate of omission errors	Score calculated as the percentage of go symbols not responded to.
Task-switching Reaction Time Test		
yy	Switch cost	Score calculated as the difference between the RTs under the two conditions.
More-odd		
zz	Switch cost	The difference between the average RTs of the switch trials in the shifting block and the average RTs of the non-switch trials in the control blocks.
d2		
aaa	Working speed	Score calculated as the overall number of marked letters.
bbb	Attention span	Score calculated as the number of accurate answers minus confusion errors.
Saccadic Paradigm		
ccc	Prosaccade task	Task during which the participant must move their eyes to the location of a stimulus.
ddd	Antisaccade task	Task during which the participant must move their eyes to a location opposite of a stimulus.

Abbreviations: RT = reaction time.