A call to rethink the cognitive benefits of physical exercise: An umbrella review of randomized controlled trials

Luis F. Ciria[1,2*], Rafael Román-Caballero[1,2], Miguel Vadillo[3], Darias Holgado[1,2,4], Antonio Luque-Casado[5], Pandelis Perakakis[6] & Daniel Sanabria[1,2*]

[1] Mind, Brain & Behavior Research Center, University of Granada, Spain

[2] Department of Experimental Psychology, University of Granada, Spain

[3] Department of Basic Psychology, Autonomous University of Madrid, Spain

[4] Institute of sport sciences, University of Lausanne, Quartier UNIL-Centre, Bâtiment Synathlon, Lausanne, Switzerland

[5] Center for Sport Studies, King Juan Carlos University, Spain

[6] Department of Social, Work, and Differential Psychology, Complutense University of Madrid, Spain

*Corresponding authors: lciria@ugr.es and daniel@ugr.es

Summary

Background: Many meta-analytic reviews have examined the cognitive benefits of regular physical exercise across the lifespan, with a wide range of inclusion criteria and disparate analytic approaches. In fact, most global health organizations incorporate the purported cognitive benefits as part of their policies to promote regular physical exercise. We assess whether those claims are well supported by scientific evidence.

Methods: We conducted a pre-registered (PROSPERO: CRD42020191357) umbrella review of meta-analyses limited to randomized controlled trials (RCTs) on exercise–cognition. Outcomes of each meta-analysis were reanalyzed following a multilevel model approach. Publication bias was also analyzed to correct final effect sizes.

Findings: Most of the 12 reviewed meta-analyses reported a positive small effect (d = 0.21, [0.07–0.40]) of physical exercise across all cognitive domains in all life stages. However, a critical assessment of the primary studies included in these reviews and the meta-analytic approaches used indicates that the evidence might not be reliable and therefore no strong conclusions should be drawn based on the existing evidence. We found that the literature has undergone an unusually rapid growth upon underpowered RCTs and potentially biased meta-analytic reviews.

Interpretation: The exponential accumulation of low-quality evidence on the exercise–cognition topic has led to stagnation in the field, hindering the discernment of the potential impact of exercise on cognition. We urge worldwide organizations committed to public health to reconsider their recommendations on the promotion of regular physical exercise to boost cognitive functions in the healthy population until high-quality empirical evidence confirms these exercise-induced cognitive benefits.

Funding: Spanish Ministry of Science and Innovation (PID2019-105635GB-I00 and PID2020-118583GB-I00).

Keywords: meta-analysis, physical exercise, cognition, executive functions, sport, brain

Introduction

The physiological and health benefits of regular physical exercise are seemingly indisputable according to the scientific evidence accrued over the last century[1]. In addition, there has been a steady surge of studies reporting cognitive and brain benefits of regular physical exercise in healthy individuals across the lifespan[2]. These findings are driving current public health policies aimed at fostering exercise adherence[3], in consonance with the World Health Organization (WHO), which currently recommends regular exercise as a means to maintain a healthy cognitive state[4]. One would therefore dare to say that the positive effect of chronic physical exercise at the cognitive level in the healthy population is nowadays taken for granted. The question we pose here is whether those claims, policies and recommendations are well backed by scientific evidence.

Research in context

Evidence before this study

The positive impact of physical exercise on cognition seems indisputable according to the scientific evidence accrued over the last two decades. A wide range of narrative and meta-analytic reviews report cognitive gains associated with regular physical exercise in healthy populations across the lifespan. These findings are driving current public health policies aimed at fostering exercise adherence to maintain a healthy cognitive state. Here, we re-visited the gold-standard empirical evidence (i.e., meta-analyses on randomized controlled trials–RCTs) on the effect of regular physical exercise on cognitive functions, jeopardizing the established claims, policies, and recommendations. We searched Medline, Scopus, ProQuest, and Google Scholar up to December 2020, for meta-analytic reviews limited to RCTs examining the potential benefits of regular physical exercise on cognition.

Added value of this study

This umbrella review provides, for the first time, a fine-grained outline of the exercise–cognition field over the past fifty years. After systematically extracting, reviewing, and re-estimating outcomes from 12 meta-analyses (for a total of 115 primary studies and 12,121 healthy participants) we found no empirical support for the existence of a potential cognitive benefit derived from the regular practice of physical exercise in healthy populations. Our study marks a turning point in research on the exercise–cognition topic, delineating the structural weaknesses of this literature, including the marked methodological, theoretical, and communicative issues. We also highlight a series of research practices that may improve future studies in the field.

Implications of all the available evidence

International public health organizations include cognitive enhancement among the wide range of benefits derived from regular physical exercise in healthy individuals across the lifespan. Our work challenges this claim and the associated policies, appealing instead for theoretically-driven and methodologically sound research to elucidate the impact of physical exercise on cognition. In any case, we argue that there is no need to invoke the as yet unproven cognitive benefits of physical exercise to justify the crucial role of physical exercise on our physical health and psycho-social well-being.

Summarized in many narrative and systematic reviews[5] and a considerable number of meta-analyses[6–8], the main conclusions of this literature are that: (1) the regular practice of physical exercise boosts cognitive performance in children, adolescents, and older adults, with limited evidence in young adults; (2) the impact appears selective to executive functions, although effects have also been described in other cognitive domains such as memory and attention; (3) academic performance is also enhanced with exercise interventions; (4) the magnitude of the effects tends to be modest, albeit

reliable; (5) various factors might mediate and moderate the effects (e.g., exercise intensity, duration of the intervention, exercise training mode, etc.). This latter point is the principal caveat discussed in these articles, although the existence of a "true" effect is rarely questioned. However, all that glitters may not be gold in this field of research.

Support for the absence of a positive effect of regular physical exercise on cognition in healthy individuals is also available in the scientific literature, although it is not very often highlighted. For example, in their meta-analysis Young et al.[7], found no evidence that aerobic exercise interventions had any effect in cognitively healthy older adults. The expert panel that recently conducted a systematic review on the topic[9] also ended up stating that the evidence was inconclusive to claim that physical activity (a more general term that includes exercise) improves children's cognitive or academic performance, except for the case of mathematics skills. Diamond and Ling[10] were even more persuasive, claiming in a controversial article that the existing evidence shows that aerobic and resistance exercise training (arguably two of the exercise modes advocated as the best to improve cognitive performance) are inefficient tools to enhance executive function.

The present umbrella review addresses the state of the art in this scientific topic by examining meta-analytic reviews limited to randomized controlled trials (RCTs), the current gold standard to ascertain causal links, to determine whether the claims regarding the benefits of regular physical exercise on cognition are well supported by the scientific evidence.

Methods

Pre-registration

The methods and planned analyses of this umbrella review were pre-registered at PROSPERO (ref. CRD42020191357).

Literature search

We conducted a systematic literature search following the PRISMA guidelines[11] (last search in December 2020) in Medline and Scopus using the following Boolean operators: *("exercise" OR "physical activity" OR "physical exercise" OR "chronic exercise" OR "regular exercise") AND ("cognition" OR "brain" OR "executive functions" OR "memory") AND (metaanalysis OR meta-analysis)*. Additionally, we searched on Proquest and Google Scholar to identify unpublished meta-analyses meeting the inclusion criteria. Search was limited to papers published in English. The search was carried out independently by three authors independently (DH, DS, and LFC), who revised the titles and abstracts to identify possible additional publications. Subsequently, two authors (DH and

LFC) revised the full text of these articles, and discrepancies between these authors were resolved by a third author (DS).

Inclusion and exclusion criteria

We followed the Participant-Intervention-Comparison-Outcome (PICO) process to select the meta-analyses included in this umbrella review: (1) Participants: healthy participants of all ages and both sexes. Meta-analyses with clinical populations, including obesity or mild cognitive impairment, were excluded. However, if meta-analyses included healthy participants and this effect could be extracted separately, we considered only the effect on this population; (2) Intervention: RCTs studying the effects of a regular exercise program in any cognitive outcome with a minimum duration of two weeks and either involving aerobic exercise, resistance exercise, mixed exercises, or other physical activities (such as extracurricular physical activities); (3) Comparison: an active control group (in which participants completed a different exercise program or an alternative activity) or a passive control group (participants did not complete any exercise program); (4) Outcome: meta-analyses should report at least a measure of the global cognitive functioning or any specific cognitive domain (executive functions, attention, memory, etc.). To avoid confounding factors, we excluded meta-analyses or primary studies from reviewed meta-analyses involving mind-body, yoga, or exercise programs combined with any other intervention (e.g., cognitive training). We did not include meta-analyses whose data were not available in the main paper (or in supplementary material) and were not provided by the authors upon request (see **Supplementary Materials**).

Data extraction

The following data were extracted from the included meta-analyses: (1) list of authors and year of publication; (2) list of authors and year of publication from each primary article included in the meta-analysis; (3) pooled number of participants for the experimental and control group; (4) sample age; (5) type of exercise intervention; (6) program duration; (7) cognitive outcome assessed; (8) type of control condition; (9) type and estimation method of effect size; (10) effect size, standard error, and confidence interval; (11) analysis of publication bias; (12) protocol registration; and (13) availability of data. After data extraction, the resulting spreadsheets were cross-checked for accuracy. The final spreadsheet integrating all data is available at the following link: https://osf.io/e9zqf/?view_only=9f0b4d0b911c4059a0f6f5dec802d55e.

Given the variety of experimental designs used in this literature, we decided to test the influence of the type of control (passive, control physical activity, or non-physical alternative activity), type of outcome (global cognition, executive functions, or other cognitive domains), and age range of the participants (children and adolescents, young adults, or older adults).

Statistical analysis

To avoid the influence of using non-target exercise programs or samples with health conditions, we reanalyzed the effect sizes in meta-analyses only with the target interventions and healthy populations. We used the robust variance estimation method[12] (RVE) using the *robumeta*[13] package for R to conduct multilevel models for each meta-analysis. This method allows dealing with a correlated structure of outcomes from the same primary study. We repeated this approach to assess the influence of three moderating variables (type of control activity, type of cognitive outcome, and age range) by implementing separate RVE models for each level of these variables.

For publication bias analyses, we accounted for dependence using aggregates of all the effect sizes coming from the same sample[14]. We implemented the PET-PEESE method[15], the trim-and-fill method[16], and a selection model[17] to identify evidence of publication bias, and to adjust the final effect size.

Finally, we conducted an exploratory specification-curve analysis[18] with the primary studies included in the meta-analysis by Ludyga et al.[8] that met our inclusion criteria. In the specification curve, we estimated the final effect size and its significance value for a total of 78 possible combinations of six preprocessing and analytic steps.

Results

Twelve meta-analyses[7,8,19–28] meeting the inclusion criteria were selected among the 1486 records retrieved in the search (see **Figure 1**). We extracted 602 effect sizes from 115 primary studies from the included meta-analyses (**Supplementary Materials, Table S1**). The publication timeline of the reviewed meta-analyses and their respective primary studies reflects an exponential growth of the exercise–cognition topic in the last two decades (**Figure 2**).

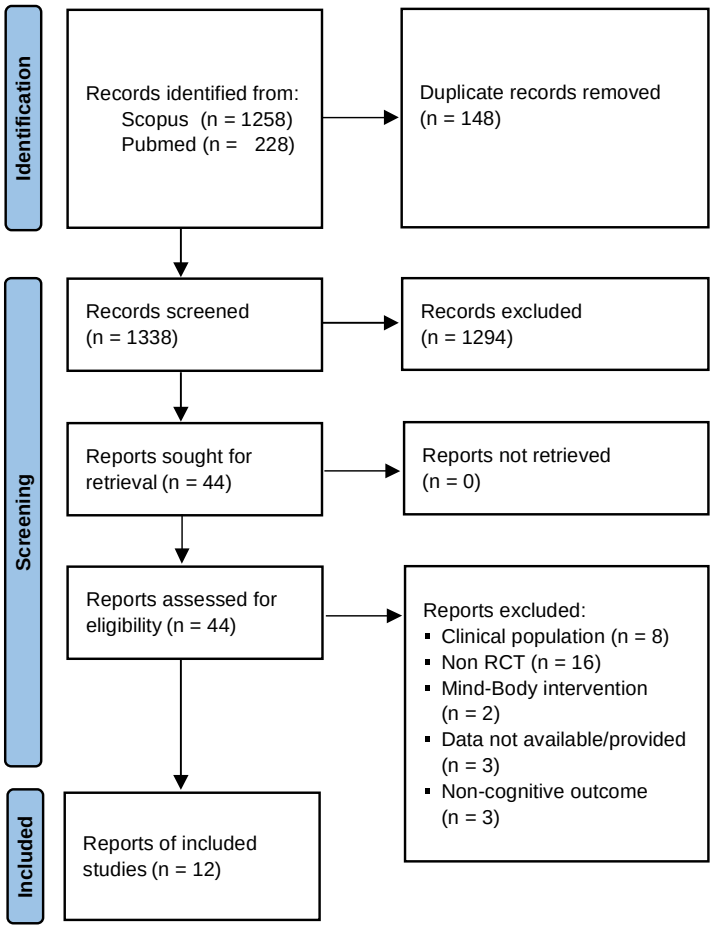


Figure 1. PRISMA flow chart for study inclusion. The initial search retrieved 1486 records (Identification), among which 148 were removed as duplicates. After the screening of title and abstract, 44 records were chosen for full-article reading (Screening). Finally, twelve meta-analyses met the inclusion criteria of the umbrella review.

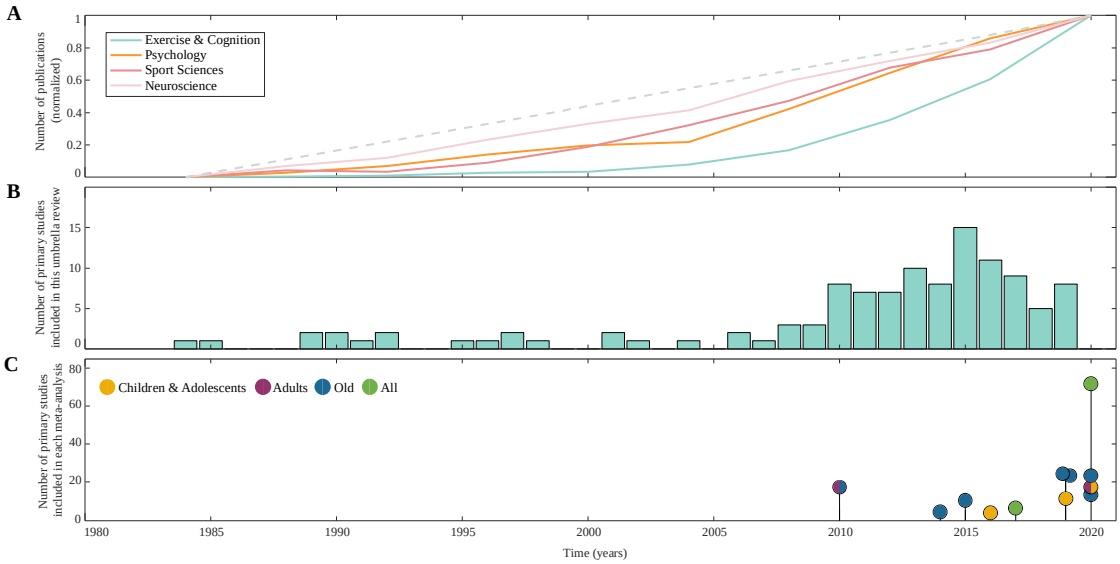


Figure 2. Evolution of the scientific literature. (A) Publication growth of Scopus-indexed articles in the areas of psychology (orange), sport sciences (red), neuroscience (pink), and exercise–cognition (turquoise) [using the search equation *(sport OR exercise OR "physical activity") AND (cognition OR "executive functions" OR "executive control")* in December 2021]. The number of publications per year was normalized to express the values of all the categories in a common scale from 0 to 1. Whereas the growth of the general categories closely followed a linear trend (gray dashed line), publications on the exercise–cognition topic depict an exponential proliferation highlighting the great interest generated by the topic in the last two decades. (B) Primary studies (RCTs) included in the present umbrella review. The 115 RCTs included show a similar exponential growth over the last 25 years, with a peak in 2015. (C) Timeline of meta-analyses included in this umbrella review along with the number of primary studies used by each meta-analysis to estimate the effect of regular exercise on cognition. Dot color depicts the target age range of each meta-analysis.

Almost all of the 12 meta-analyses found a significant positive effect of regular exercise on cognition with healthy participants (average Cohen's d = 0.21), concluding that exercise may improve cognitive skills (10[8,19–25,27,28] out of 12). When we re-estimated the effect size in all these meta-analyses, the mean final effect remained small (d = 0.21; range 0.07–0.40; **Figure 3A**) and, in general, with small-to-medium heterogeneity (mean I^2 = 40.87%; range 0–80.84%; **Figure 3B**). Part of the dispersion in the final effect and heterogeneity could be due to the inclusion criteria adopted in each meta-analysis (**Table S1**), differing in the type of physical exercise, age range of the participants, or cognitive outcome. Other sources of variability could come from the way the individual effect sizes were estimated or the strategy adopted in the meta-analyses to deal with the dependence generated by the inclusion of several outcomes from the same sample of participants. Finally, the divergence in the inclusion of primary studies (see **Variations in Study Sampling**) could affect the results of the meta-analyses[i]. On the other hand, the number of primary studies per meta-analysis (mean of 19 studies, range 3–71; **Figure 3C**) increased as a function of the year of publication (Kendall's τ = 0.45, p = 0.05), in parallel to the exponential growth of primary articles. Regarding the number of participants per effect, it did not increase over the years (Kendall's τ = 0.28, p = 0.23) and, more important, the number of participants were below the sample size required to achieve an acceptable statistical power for a d = 0.2 (i.e., at least 394 participants per group, 788 in total[ii]; mean of 111 participants, range 56–375; **Figure 3D**). At the primary study level, only one study reached acceptable statistical power (i.e., > 788 participants; **Figure 3E**).

[i] The full assessment of the quality evaluation of each meta-analytic review included is available at the following link: https://osf.io/e9zqf/?view_only=9f0b4d0b911c4059a0f6f5dec802d55e

[ii] We estimated that 788 participants are required to achieve a power of 0.80 for a small effect size such as d = 0.2 in a two-tailed two-sample t-test and an α of 0.05. This analysis corresponds with the classic between-group comparison in the pre-posttest change of the cognitive measure. However, another common analysis is to conduct a two-way mixed ANOVA with a between-group factor (i.e., treatment vs. control group) and moment as a within-participant factor (i.e., pretest vs. posttest). There, the key effect is the group-by-moment interaction, in which at least 790 participants are necessary to achieve an adequate power for a Cohen's f = 0.1 (or, equivalently, an η^2 = 0.01).

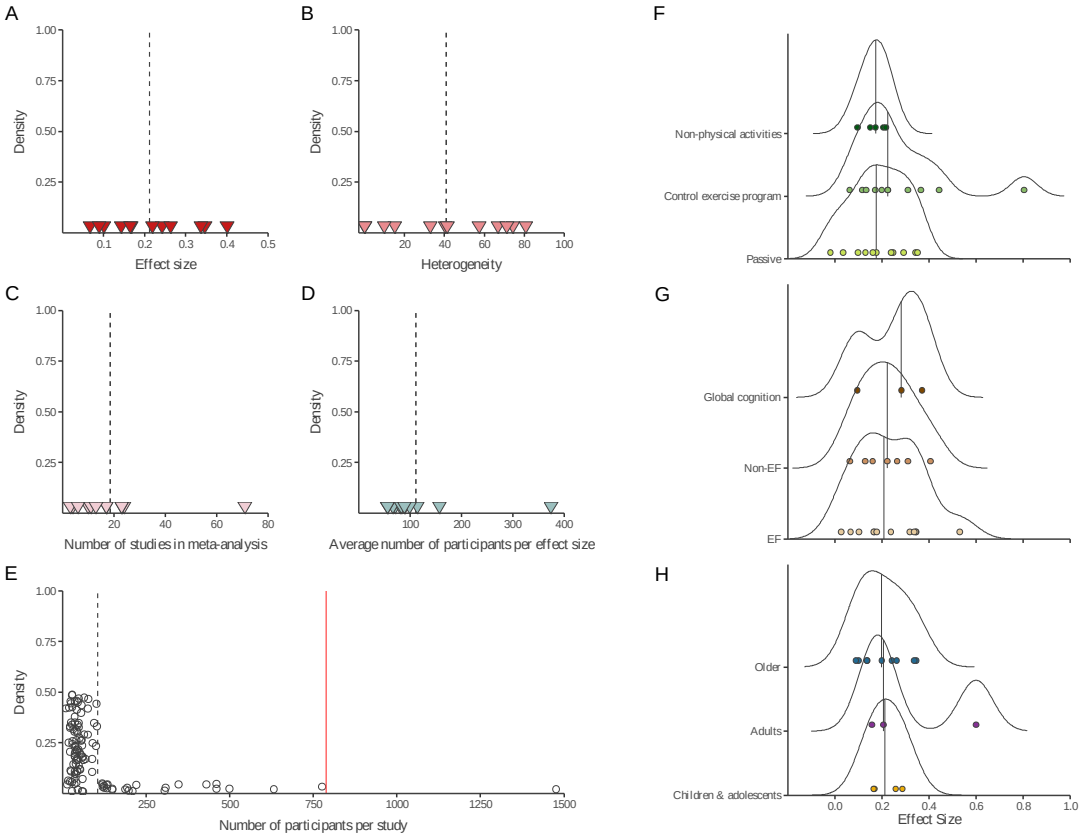


Figure 3. Reanalysis of the meta-analyses included in the umbrella review and influential variables. Distribution of meta-analytic effect sizes (A) and heterogeneity, expressed in I^2 (B), estimated from the 12 meta-analyses. Distribution of the number of studies (C), and the average number of participants per effect size in each meta-analysis (D) and per primary studies (E). Dashed lines indicate the average value for all the meta-analyses. The red line indicates the required sample size for an acceptable power[2]. The averages showed a small mean effect size (d = 0.21) and with substantial mean heterogeneity (I^2 = 40.87%). All but one the primary studies used underpowered designs to assess a small effect size individually. Note that we excluded primary studies (from reviewed meta-analyses) involving mind-body, yoga (or similar), exercise programs combined with any other intervention (e.g., cognitive training), or samples of participants with medical conditions to avoid confounding factors. Therefore, the effect sizes represented might differ slightly from those originally reported by each meta-analysis. (F) Estimated final effects of the included meta-analyses as a function of the type of control program, (G) the type of cognitive outcome, (H) and the age range of the included samples. Vertical solid lines indicate the median of each category and the circles at the bottom the effect of each meta-analysis. The substantial overlap between the distributions indicates a reduced impact of these factors in the final outcome of the meta-analyses.

Variations in study sampling

The network visualization (**Figure 4A**) of the reviewed meta-analyses reveals that the variability across meta-analyses might be in part due to the divergences in the inclusion of primary studies. Altogether, the 12 meta-analyses comprise a total of 115 different primary studies, of which 52 (i.e., 45.2%) were only included in one meta-analysis, and 32 (27.8%) in two of them. This high variability is noticeable even between meta-analyses addressing the same age range and cognitive outcome. For instance, the three meta-analyses focused on aging and general cognitive domains[23,24,26] share only one primary study[29] of the 23, 34, and 25 primary studies used, respectively. Although these three meta-analyses address the same topic with a similar approach (meta-analysis of RCTs) in a comparable time frame (between 2019 and 2020), their results are based on radically different bodies

of evidence. To further test this disconnection between meta-analytic reviews, we assessed the number of primary studies that were available at the time the meta-analyses carried out their last search and that met the inclusion criteria for each review (**Figure 4B**). Most of the meta-analyses included only half of the available studies meeting their criteria (mean of 51.3% of studies). Only two meta-analyses included most of the available studies (Haverkamp et al.[28]: 81.3%; Young et al.[7]: 83.3%), whereas in several cases the analyzed studies represented a third or less of all the potential targets (Jackson et al.[21]: 33.3%; Rathore & Lom[22]: 12.5%; Sanders et al.[24]: 30.3%; Scherder et al.[20]: 36.4%). This lack of overlap in the meta-analytic literature on the topic suggests that the conclusions drawn from these quantitative reviews cannot be taken as the empirical evidence accumulated over years, but as selective slices of it.

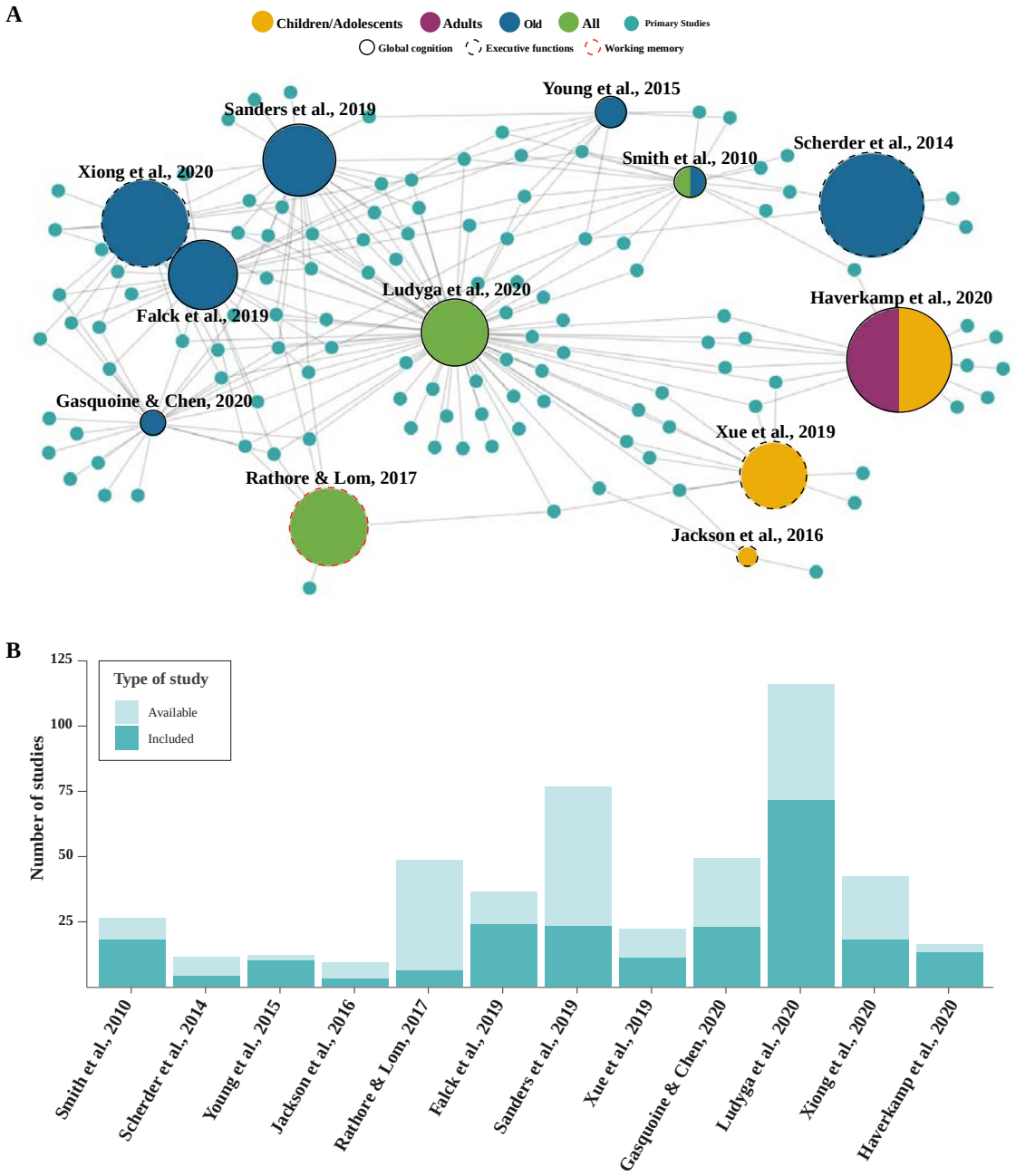


Figure 4. Network interaction among the meta-analyses included in the umbrella review. (A) Network of meta-analyses targeting RCTs on the relationship between regular physical exercise and cognitive functions in healthy populations included in this review. Meta-analyses are represented as nodes, with color indicating age range and edges depicting the cognitive domain addressed. The size of the nodes (for meta-analyses only) is proportional to the reported effect size and the color indicates the age range addressed. Note that most meta-analyses report the effect of exercise on general cognitive domains, but some only report the effect of exercise on executive functions (Jackson et al.[21]; Scherder et al.[20]; Xiong et al.[27]; Xue et al.[25]) or working memory (Rathore & Lom[22]). The primary studies used by each meta-analysis to estimate their reported effect size are represented by turquoise nodes (the size of these nodes is fixed). The inclusion of a primary study in a meta-analysis is represented by a gray line between the meta-analysis node and the primary study node. Spatially close meta-analyses share a greater number of primary studies, while more distant meta-analyses share few or none. It is important to highlight that this network visualization is basically descriptive and does not depict a complete picture of the field. It is intended to help clarify the state-of-the-art of the field rather than refute any particular hypothesis. (B) Number of primary studies included in the reviewed meta-analyses (darker turquoise nodes) compared to the studies that met the inclusion

criteria and were available by the moment when the meta-analysis search was carried out (light turquoise nodes). Meta-analyses are sorted by year of publication.

Influential variables and publication bias

The final effect did not substantially vary with regard to the type of comparison group (control physical exercise: median d = 0.28; non-physical activities: d = 0.13; passive control: d = 0.20; **Figure 3F**), the type of cognitive outcomes (executive functions: d = 0.24; non-executive functions: d = 0.24; global cognition: d = 0.25; **Figure 3G**) or between age cohorts (children and adolescents: d = 0.22; adults: d = 0.32; older: d = 0.21; **Figure 3H**). The latter result was obtained even when the duration of the interventions for older participants were on average longer (mean of 23 weeks, range 4–104) than those with children and adolescents (mean of 20 weeks, 6–44), and especially with young adults (mean of 10 weeks, 4–24).

On the other hand, only 60% of the meta-analyses reported publication-bias analyses (7[8,20,23–25,27,28] out of 12), most of them using funnel plot-based methods. Among those that assessed it, almost all found evidence of publication bias (6[8,23–25,27,28] out of 7), but many of them did not adjust the final effect (3[24,25,27] out of 7). We subsequently assessed the presence of publication bias in meta-analyses with ten or more primary studies (8[8,19,23–28] of the 12 meta-analyses); **Figure 5**). Both meta-regression-based methods for testing funnel-plot asymmetry (i.e., PET and PEESE) and the trim-and-fill method suggested publication bias in four and five meta-analyses, respectively, while the Vevea and Hedges' selection model did not detect any publication bias. In general, the final effect was reduced after bias correction (**Figure 5**), with the smallest values for the PET method (mean of -0.002), followed by PEESE (mean of 0.10), and the trim-and-fill method (mean of 0.15). Therefore, besides the fast growth of this literature, the publication process seems to have favored the reporting of positive and significant results over null results, especially in small studies. These findings suggest that the true effect of exercise over cognition is probably smaller than originally reported in the meta-analyses.

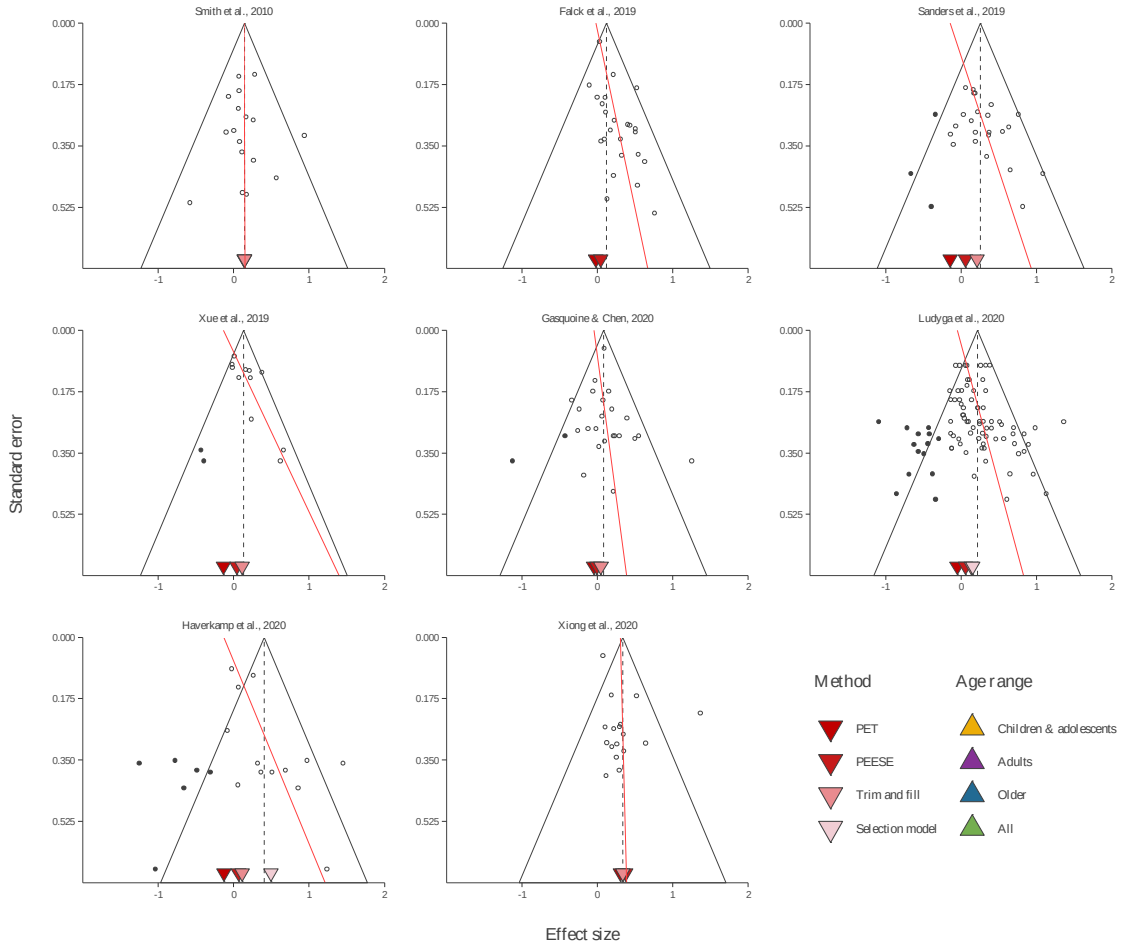


Figure 5. Assessment of publication bias across the meta-analyses included in the umbrella review. Funnel plot of eight of the 12 meta-analyses. The aggregated outcomes of the primary studies are depicted with empty circles, whereas black circles represent missing studies identified with the trim-and-fill method. The red line represents the effect predicted by the PET meta-regressive model (using the standard error of the effect as moderator), where a non-zero slope may indicate publication bias. The adjusted final effects with the multiple methods are depicted with triangles at the bottom of each funnel plot. In general, there was evidence of publication bias suggesting that in most cases the benefits of physical activity over cognition was potentially overestimated. Note that trim-and-fill could not be applied in one meta-analysis (Falck et al.[23]) due to a problem of convergence, and the Vevea and Hedges' selection model only could be applied to two meta-analyses (Haverkamp et al.[28]; Ludyga et al.[8]) due to problems of convergence and the reduced number of significant individual effect sizes (i.e., less than three).

Specification-curve analysis

As we described in the previous sections, the reviewed meta-analyses differed in many aspects. One of them was their inclusion criteria (**Figure 4**). In relation to the statistical process itself, the meta-analyses differed in the decisions of multiple preprocessing and analytic steps, such as the way they estimated the effect size and their variance, the method for assessing publication bias, or the approach used to deal with the within-effects dependence.

To examine the impact of all these decisions on the meta-analytic outcome, we conducted an exploratory (not pre-registered) specification-curve analysis with the studies in Ludyga et al.[8] as the

largest and one of the most recent meta-analyses. The analysis showed that the final effect could vary greatly depending on preprocessing and analytic decisions[i] (from d = -0.12 to d = 0.20; **Figure 6**). Some common specifications in the reviewed meta-analyses (such as using aggregates to deal with within-effects dependence, or not correcting for publication bias) led to larger effects, more likely to be significant. In contrast, some decisions rarely adopted in the meta-analyses reduced the final effect substantially: the use of multilevel models[8,23], and correcting for publication bias with the PET-PEESE method (the adjusted effect and its significance were only reported in Ludyga et al.[8]). Therefore, most of the meta-analyses opted for specifications in some steps that tend to find more positive and significant effects, whereas meta-analyses with more conservative decisions were underrepresented in the literature.

[i] The reviewed meta-analyses made decisions at three levels of preprocessing: (a) the standardized effect size was based on pre-posttest change or only on the posttest difference between groups; (b) the use of Cohen's *d* or Hedges' *g*; and (c) the estimation of the variance of the effect size followed the classic formulation or the Morris' proposal[30]. For the specification curve, we selected these three preprocessing steps along with two additional analytic decision levels: (d) how to deal with within-effects dependence (none strategy or assuming within-effects independence, a RVE multilevel model, and fitting an univariate model with aggregates); and (e) the strategies to assess and correct the final outcome for publication bias (PET and PEESE methods, trim-and-fill approach, selection model, and no correction). Following the recommendations of Pustejovsky and Rodgers[31], we used an alternative formula for the variance/standard error (W and its square root, respectively) when they were included as regressor in PET-PEESE methods to prevent the artifactual dependence between the effect size and its precision estimate. The analysis led to 78 different combinations of specifications. Note that the trim-and-fill method and selection model cannot be conducted in multilevel model, and the Morris' variance was only applied to pre-posttest estimates of the effect size (reducing the number of possible combinations from to 120 to 78).

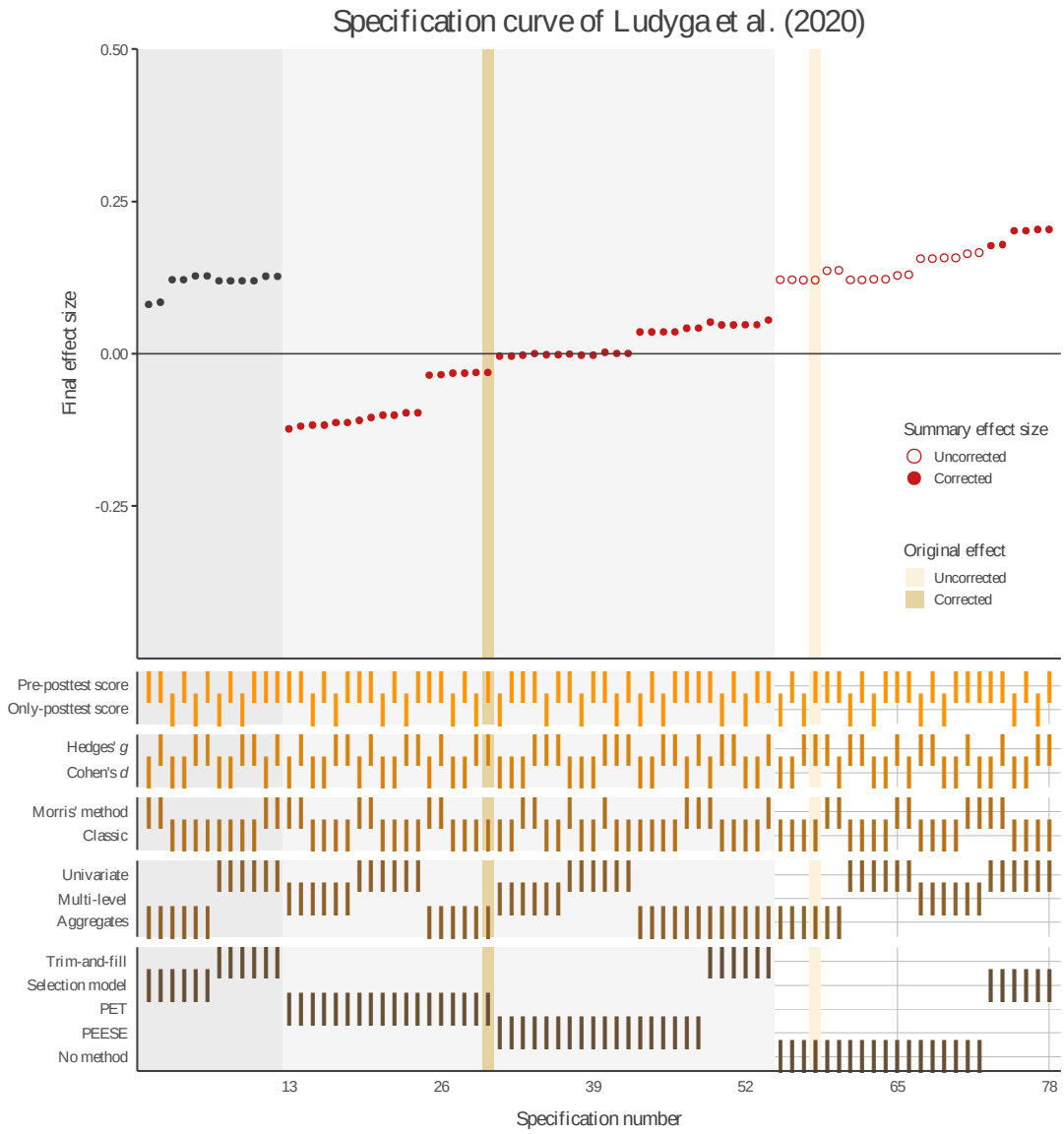


Figure 6. Specification curve of Ludyga et al.'s meta-analysis. Summary effect size of the target studies (and its 95% confidence interval) varied across the multiple combinations of preprocessing and analytic decisions. Gray effects denote those where there was no evidence of publication bias. Red empty effects represent the outcome of models without publication-bias adjustment, whereas red and black filled effects show the corrected summary effect. The light gray rectangular shade distinguishes non-significant results from significant ones. The dark and light golden vertical shades highlight the effects with the original specifications in Ludyga et al. with and without publication bias correction, respectively. It is apparent the high variability of results and the disparity of conclusions that can be extracted from them. The present specification-curve analysis highlights the great impact of preprocessing and analytic decisions on the final outcome. Note that our result could differ from the effect size in the original meta-analysis due to our inclusion criteria.

Discussion

In this umbrella review, we questioned the settled claim that regular physical exercise leads to cognitive gains across the lifespan. After reanalyzing 12 meta-analyses of RCTs, including a total of 115 primary studies and 12121 healthy participants, we found inconclusive evidence supporting the

existence of a potential cognitive benefit derived from the regular practice of physical exercise in healthy populations. Consequently, the current recommendations of worldwide organizations such as the WHO, framed in the policies of promoting regular physical exercise for improving or preserving cognitive functions in healthy individuals, are not as strongly supported by empirical evidence as previously thought.

This review provides, for the first time, a fine-grained outline of the exponential growth of the exercise–cognition field over the past fifty years. While this rapid growth has provided insight into the potential benefits, risks and pitfalls of implementing exercise-based interventions in the general population, it has not been accompanied by a strengthening of scientific practices resulting in a vast body of low-quality evidence. This evidence, often from underpowered RCTs and potentially biased meta-analyses, shows signs of unusual clustered-like growth (i.e., meta-analyses on the topic do not share primary sources of evidence), with a marked lack of consideration for critical voices[10] and the neglect of mixed or contradictory findings[32]. In line with recent accounts[33], we believe this exponential accumulation of low-quality evidence has led to stagnation rather than advance in the field hindering the discernment of the real existing effect. This umbrella review marks a turning point in research on the exercise–cognition topic, delineating the structural weaknesses of this literature, including the marked methodological, theoretical, and communicative issues. Below, we briefly develop a series of opportunities for improvement that may guide future studies in the field.

Methodological issues

The exercise–cognition field has been flooded with individual experiments addressing this relationship on designs with low statistical power that yield estimates with low precision and stability. While conducting intervention studies with remarkably large sample sizes (see *Footnote 2*) is not within the reach of almost any individual laboratory, simplification of experimental designs to increase statistical power, standardization of experimental designs to facilitate comparison of results and replication, proper active control groups, pre-registration, or multi-laboratory initiatives can definitely enhance the field.

In addition, it is important to cautiously reconsider the role of meta-analyses and the extent to which their results shed light on this particular topic[34,35]. Even though meta-analytic approaches have the potential to minimize some of the shortcomings of individual intervention studies[36], their results largely depend on the quality of the included reports as well as on the methodological decisions followed to estimate a particular effect. Thus, meta-analytic results do not necessarily represent the true effect of a particular phenomenon. As we show here, the particular conclusions from the different meta-analyses cannot be taken as the empirical evidence accumulated over years, but as selective slices of it.

For conclusions of future meta-analyses to be translated into social recommendations and policies, the reviews need to be as comprehensive as possible. This means that they should incorporate not only the knowledge from the primary literature, but also from previous reviews. Bearing in mind the marked publication bias in the exercise–cognition field, forthcoming meta-analyses should include gray literature contributing, predictably, with less optimistic results. Finally, at the preprocessing and analytic levels, there are many choices that are arguably preferred. For example, the use of an effect size estimate based on the pre-posttest changes to control for preexisting differences at baseline. We also encourage the use of multilevel models (e.g., RVE approach) accounting for the correlated structure of effects without averaging information. Regarding the assessment of publication bias, none of the available methods stands as superior in terms of performance. It depends on factors such as the number of studies, heterogeneity, or the degree of publication bias present in the literature[27]. A reasonable strategy is to use several of them in combination and interpret their results regarding the conditions of the meta-analysis. The exercise–cognition literature is currently big enough for the implementation of this type of analysis and to adjust in consequence the final effect.

Theoretical issues

It is surprising, given the relatively large number of published experiments and reviews on this topic, the existing gap in the literature on the mechanisms involved in exercise-induced cognitive improvements. Different theoretical models have been proposed to explain these potential cognitive benefits. The cardiovascular hypothesis[38], cognitive training[39], or neuroselection[40] are some of these attempts. Although there is data supporting each of these hypotheses, to date, none of them has been able to fully account for all of the existing evidence. This absence of consensus on a theoretical framework has been further accentuated by the myriad of experimental approaches that hamper the reconciliation of different findings. Certainly, it is necessary to shift away from the metaphor of the brain as a muscle and develop comprehensive theoretical models on the cognitive and neural mechanisms of these potential exercise-induced cognitive improvements.

Communicative issues

The pressure for publishing is an endemic issue shared by researchers and mass media. In both cases, there is an urgency for publishing novel and eye-catching findings to attract public attention, which often leads to oversimplification, misrepresentation, or overdramatization of scientific results without the nuances and limitations essential for proper interpretation. The exercise–cognition topic is not immune to this. Transparent practices throughout the research process (e.g., reporting bias-corrected effects) and accurate dissemination of scientific findings through the media would definitely improve the situation[41]. Nevertheless, this is a necessary but not sufficient step. Without the willingness of researchers to transparently report their results and databases, the collaborative efforts of editors to

publish meaningful results (regardless of whether they are positive or not), and the commitment of the media to move away from hyped headlines and clickbait, it is a futile endeavor.

Final remark

In light of the present findings, we urge worldwide organizations committed to public health (e.g. WHO, NIH) to reconsider their recommendations on the promotion of regular physical exercise to boost cognitive functions in the healthy population. The positive impact of physical exercise on human well-being, especially regarding physical health, seems indisputable. These physical benefits are in themselves sufficient to justify its regular application in our daily lives[42]. There is no need to appeal to the as yet unproven cognitive benefits of physical exercise, especially when the current evidence suggests that, if the effect exists, it is notably small to rethink its practical relevance[i]. Further, engaging in physical exercise brings not only physical but also social benefits, as we connect with others by forging social bonds, participating in collective activities that give us a sense of belonging and build new sources of social support. Above all, we strongly believe in the pleasure of doing something for its own sake. The value of exercising may lie simply in its enjoyable nature.

Contributors

LFC, MV, DH, ALC, PP and DS were involved in the original conceptualization. LFC, RRC, MV, DH, ALC, PP and DS were responsible for developing the study methodology. LFC, DH and DS did the literature search. LFC, RRC, MV, DH, ALC, PP and DS were responsible for data curation. RRC and MA did the formal statistical analysis. LFC, RRC, DH and DS wrote the original draft. MV, DH, ALC and PP edited and reviewed the manuscript. MA and DS acquired the funding. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

All authors declare no competing interests.

Data sharing

The hypotheses and analysis plan were pre-registered in the OSF repository (10.17605/OSF.IO/HFRPC). All deviations from the pre-registered procedures and analysis plans are transparently identified in the manuscript. Data and codes used for the analyses presented here are available at the OSF repository (https://osf.io/e9zqf/?view_only=9f0b4d0b911c4059a0f6f5dec802d55e).

Acknowledgments

This research was supported by two research grants from Spanish Ministerio de Economía y Competitividad to D.S and M.V., respectively.

[i] A Cohen's *d* of 0.125, such as the adjusted effect found in Ludyga et al.[8], represents an explanation less than 1% of the variability between groups, or also corresponds with a gain of 2 IQ scores.

References

1 Sallis JF, Cerin E, Conway TL, *et al.* Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. *The Lancet* 2016; **387**: 2207–17.
2 Guiney H, Machado L. Benefits of regular aerobic exercise for executive functioning in healthy populations. *Psychon Bull Rev* 2013; **20**: 73–86.
3 Ding D, Mutrie N, Bauman A, Pratt M, Hallal PR, Powell KE. Physical activity guidelines 2020: comprehensive and inclusive recommendations to activate populations. *The Lancet* 2020; **396**: 1780–2.
4 Bull FC, Al-Ansari SS, Biddle S, *et al.* World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br J Sports Med* 2020; **54**: 1451–62.
5 Tomporowski PD, Pesce C. Exercise, sports, and performance arts benefit cognition via a common process. *Psychol Bull* 2019; **145**: 929.
6 Angevaren M, Aufdemkampe G, Verhaar HJ, Aleman A, Vanhees L. Physical activity and enhanced fitness to improve cognitive function in older people without known cognitive impairment. *Cochrane Database Syst Rev* 2008; **3**. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD005381.pub3/full (accessed June 29, 2017).
7 Young J, Angevaren M, Rusted J, Tabet N. Aerobic exercise to improve cognitive function in older people without known cognitive impairment. *Cochrane Database Syst Rev* 2015.
8 Ludyga S, Gerber M, Pühse U, Looser VN, Kamijo K. Systematic review and meta-analysis investigating moderators of long-term effects of exercise on cognition in healthy individuals. *Nat Hum Behav* 2020; **4**: 603–12.
9 Singh AS, Saliasi E, Van Den Berg V, *et al.* Effects of physical activity interventions on cognitive and academic performance in children and adolescents: a novel combination of a systematic review and recommendations from an expert panel. *Br J Sports Med* 2019; **53**: 640–7.
10 Diamond A, Ling DS. Aerobic-Exercise and resistance-training interventions have been among the least effective ways to improve executive functions of any method tried thus far. 2019.
11 Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Med* 2009; **6**: e1000097.
12 Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. *Res Synth Methods* 2010; **1**: 39–65.
13 Fisher Z, Tipton E, Zhipeng H, Fisher MZ. Package 'robumeta'. 2017.
14 Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to meta-analysis. John Wiley & Sons, 2021.
15 Stanley TD, Doucouliagos H. Meta-regression approximations to reduce publication selection bias. *Res Synth Methods* 2014; **5**: 60–78.
16 Duval S, Tweedie R. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000; **56**: 455–63.
17 Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995; **60**: 419–35.
18 Simonsohn U, Simmons JP, Nelson LD. Specification curve analysis. *Nat Hum Behav* 2020; **4**: 1208–14.
19 Smith PJ, Blumenthal JA, Hoffman BM, *et al.* Aerobic Exercise and Neurocognitive Performance: A Meta-Analytic Review of Randomized Controlled Trials: *Psychosom Med* 2010; **72**: 239–52.
20 Scherder E, Scherder R, Verburgh L, *et al.* Executive functions of sedentary elderly may benefit from walking: a systematic review and meta-analysis. *Am J Geriatr Psychiatry* 2014; **22**: 782–91.
21 Jackson WM, Davis N, Sands SA, Whittington RA, Sun LS. Physical activity and cognitive development: a meta-analysis. *J Neurosurg Anesthesiol* 2016; **28**: 373–80.
22 Rathore A, Lom B. The effects of chronic and acute physical activity on working memory performance in healthy participants: a systematic review with meta-analysis of randomized controlled trials. *Syst Rev* 2017; **6**: 1–16.
23 Falck RS, Davis JC, Best JR, Crockett RA, Liu-Ambrose T. Impact of exercise training on physical and cognitive function among older adults: a systematic review and meta-analysis. *Neurobiol Aging* 2019; **79**: 119–30.
24 Sanders LM, Hortobagyi T, la Bastide-van Gemert S, van der Zee EA, van Heuvelen MJ. Dose-response relationship between exercise and cognitive function in older adults with and without cognitive impairment: a systematic review and meta-analysis. *PloS One* 2019; **14**: e0210036.
25 Xue Y, Yang Y, Huang T. Effects of chronic exercise interventions on executive function among children and adolescents: a systematic review with meta-analysis. *Br J Sports Med* 2019; **53**: 1397–404.
26 Gasquoine PG, Chen P-Y. Effect of physical exercise on popular measures of executive function in older, nonclinical, participants of randomized controlled trials: A meta-analytic review. *Appl Neuropsychol Adult* 2020; **29**: 1–9.
27 Xiong J, Ye M, Wang L, Zheng G. Effects of physical exercise on executive function in cognitively healthy older adults: A systematic review and meta-analysis of randomized controlled trials. *Int J Nurs Stud* 2021; **114**: 103810.
28 Haverkamp BF, Wiersma R, Vertessen K, van Ewijk H, Oosterlaan J, Hartman E. Effects of physical activity interventions on cognitive outcomes and academic performance in adolescents and young adults: A meta-analysis. *J Sports Sci* 2020; **38**: 2637–60.
29 Ansai JH, Rebelatto JR. Effect of two physical exercise protocols on cognition and depressive symptoms in oldest-old people: A randomized controlled trial. *Geriatr Gerontol Int* 2015; **15**: 1127–34.
30 Morris SB. Estimating effect sizes from pretest-posttest-control group designs. *Organ Res Methods* 2008; **11**: 364–86.
31 Pustejovsky JE, Rodgers MA. Testing for funnel plot asymmetry of standardized mean differences. *Res Synth Methods* 2019; **10**: 57–71.
32 Coen RF, Lawlor BA, Kenny R. Failure to demonstrate that memory improvement is due either to aerobic exercise or

increased hippocampal volume. *Proc Natl Acad Sci* 2011; **108**: E89–E89.
33 Chu JS, Evans JA. Slowed canonical progress in large fields of science. *Proc Natl Acad Sci* 2021; **118**.
34 Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *The Lancet* 1998; **351**: 123–7.
35 Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *The Lancet* 1998; **351**: 47–52.
36 Hunter JE, Schmidt FL. Methods of meta-analysis: Correcting error and bias in research findings. Sage, 2004.
37 Carter EC, Schönbrodt FD, Gervais WM, Hilgard J. Correcting for bias in psychology: A comparison of meta-analytic methods. *Adv Methods Pract Psychol Sci* 2019; **2**: 115–44.
38 Tarumi T, Zhang R. The role of exercise-induced cardiovascular adaptation in brain health. *Exerc Sport Sci Rev* 2015; **43**: 181–9.
39 Mann DT, Williams AM, Ward P, Janelle CM. Perceptual-cognitive expertise in sport: A meta-analysis. *J Sport Exerc Psychol* 2007; **29**: 457–78.
40 Belsky DW, Caspi A, Israel S, Blumenthal JA, Poulton R, Moffitt TE. Cardiorespiratory fitness and cognitive function in midlife: Neuroprotection or neuroselection?: Fitness and Cognitive Function. *Ann Neurol* 2015; **77**: 607–17.
41 Román-Caballero R, Sanabria D, Ciria LF. Let's go beyond "the effect of" physical exercise and musical training on cognition: much more than a significant effect. *PsyArXiv* 2021.
42 Ekkekakis P. Why Is Exercise Underutilized in Clinical Practice Despite Evidence It Is Effective? Lessons in Pragmatism From the Inclusion of Exercise in Guidelines for the Treatment of Depression in the British National Health Service. *Kinesiol Rev* 2020; **10**: 29–50.

SUPPLEMENTARY MATERIALS

Study inclusion

Despite meeting the inclusion criteria, three meta-analyses were not included in the final analyses because they did not report outcome effect sizes relative to the primary studies in the original manuscript or supplementary material. The corresponding authors of these articles did not reply (Barha et al., 2017), did not provide (Chen et al., 2020) or replied that data had been misplaced and was no longer available (Colcombe & Kramer, 2003) upon email request.

Methodological quality evaluation

The quality of the meta-analyses included in this review was assessed using the validated scale Assessing the Methodology Quality of Systematic Review 2 (AMSTAR 2) checklist. The AMSTAR 2 scale is developed to assess the quality of systematic reviews and meta-analyses that includes both randomized and non-randomized controlled trials. The scale is composed of 16 items, assessing questions regarding the use of the PICO criteria, registration of the review protocol, assessment of the risk of bias in the literature, adequate analysis of the data, analysis publication bias, etc. Each item can be scored as “yes” and “no”, if they fulfill several criteria. Although a final score might be calculated, we followed the authors' recommendation of assessing the quality of the reviews based on critical domains rather than by a final score (Shea et al., 2017). Nonetheless, we provide the percentage of articles that fulfill each criteria. Additionally, in our pre-registration form, we stated that we would also assess the quality of the included articles with another scale (Grading of Recommendations Assessment, Development and Evaluation (Guyatt et al., 2008), but given that many of the items overlap with the AMSTAR 2 scale, we decided to omit it.

The results of the critical domains assessed in the AMSTAR 2 scale (https://osf.io/e9zqf/?view_only=9f0b4d0b911c4059a0f6f5dec802d55e) revealed that in 9 out of 12 meta-analyses (Falck et al., 2019; Jackson et al., 2016; Ludyga et al., 2020; Rathore & Lom, 2017; Sanders et al., 2019; Scherder et al., 2014; Smith et al., 2010; Xue et al., 2019; Young et al., 2015) the authors explained their selection of the study designs for inclusion in the review. Only in one of the meta-analyses (Gasquoine & Chen, 2020) the authors did not select the studies in duplicate, and in five (Gasquoine & Chen, 2020; Jackson et al., 2016; Sanders et al., 2019; Scherder et al., 2014; Xue et al., 2019) the authors did not extract the data in duplicate. In six of the meta-analyses (Gasquoine & Chen, 2020; Haverkamp et al., 2020; Sanders et al., 2019; Scherder et al., 2014; Smith et al., 2010), the authors did not assess the potential impact of risk of bias in individual studies on the results of the meta-analysis, and six of the meta-analyses (Haverkamp et al., 2020; Ludyga et al., 2020; Sanders et

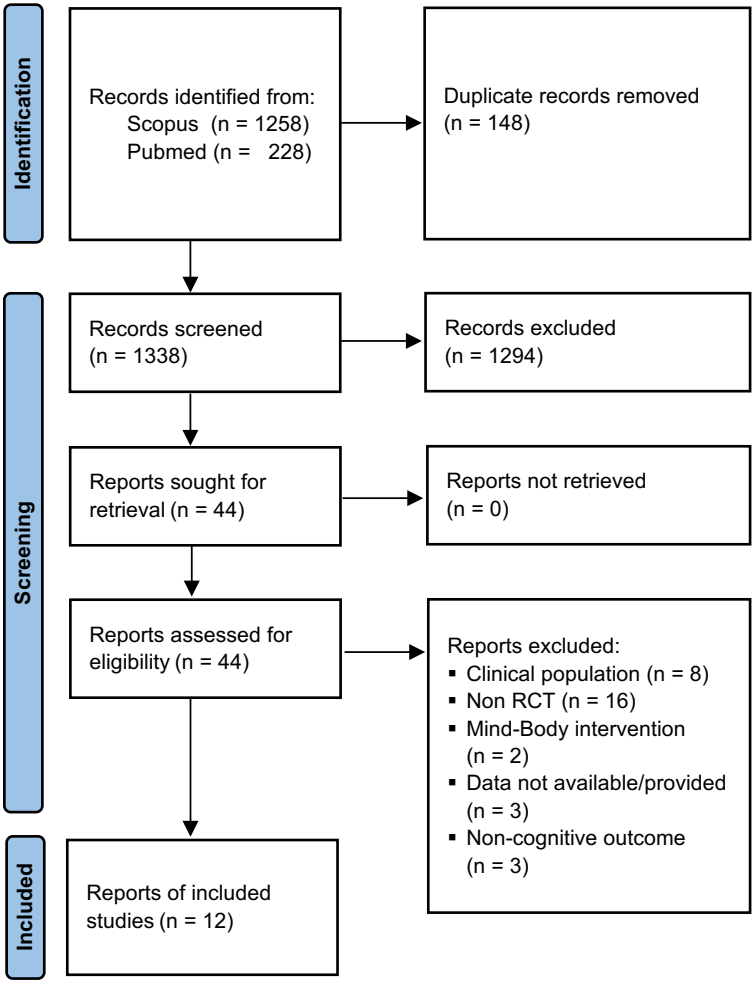
al., 2019; Scherder et al., 2014; Smith et al., 2010; Xiong et al., 2020) did not either account for the risk of bias when interpreting the results. In one meta-analysis (Scherder et al., 2014), the authors did not provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review. Finally, only in one meta-analysis (Scherder et al., 2014), the authors did not report any potential sources of conflict of interest.

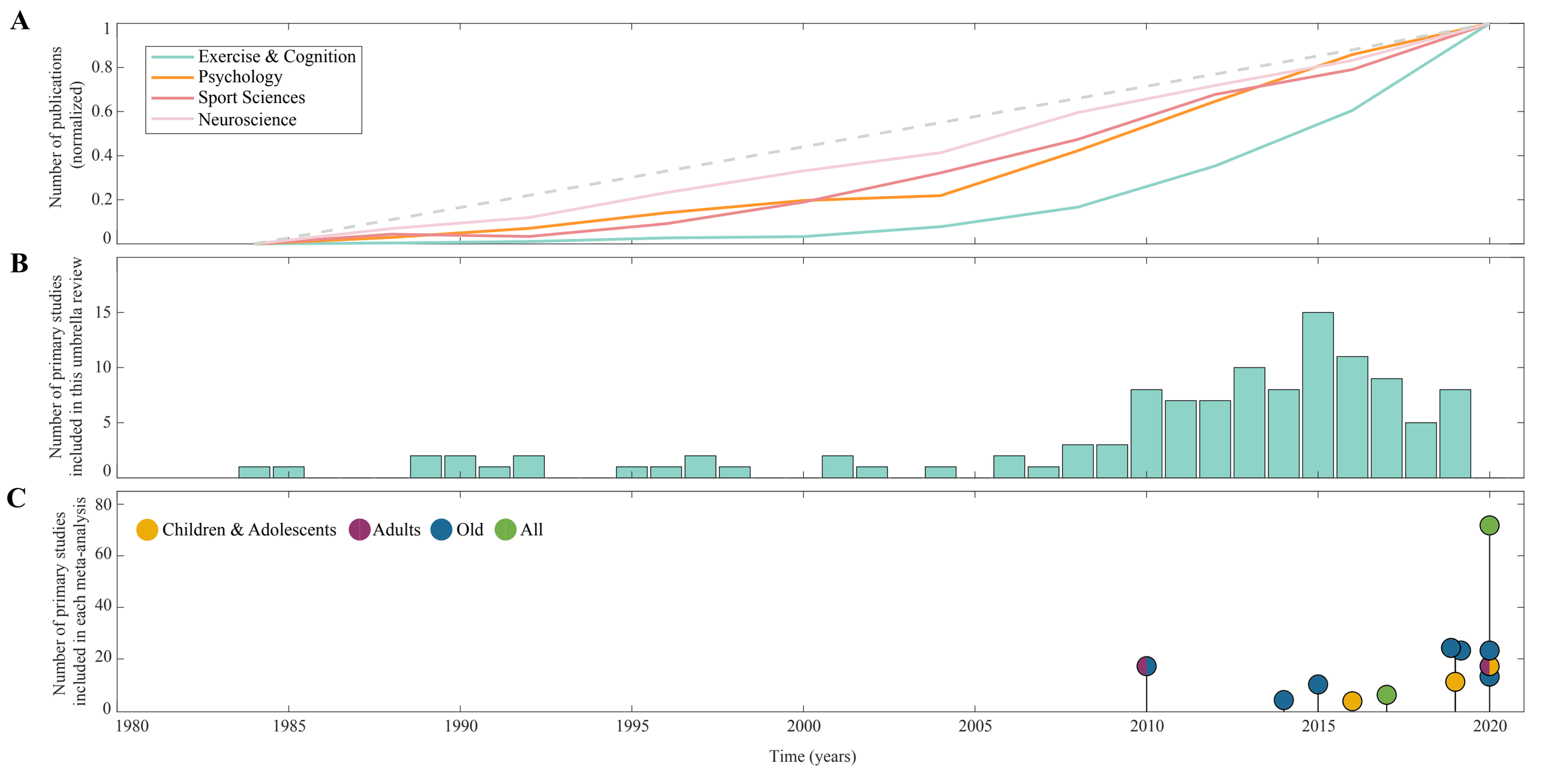
References supplementary material

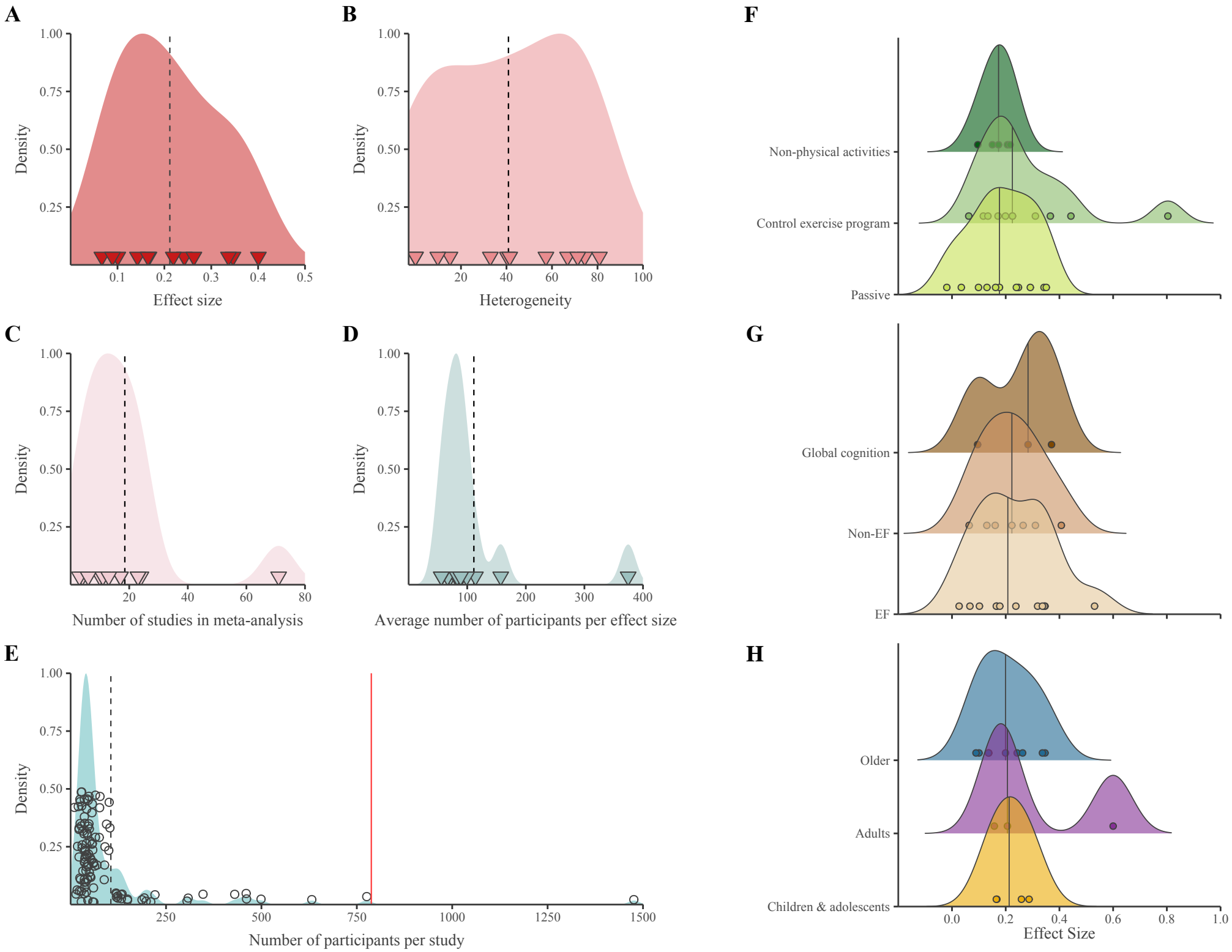
- Chen, F. T., Etnier, J. L., Chan, K. H., Chiu, P. K., Hung, T. M., & Chang, Y. K. Effects of exercise training interventions on executive function in older adults: a systematic review and meta-analysis. *Sports Medicine* 2020; **50**: 1451-1467.
- Barha, C. K., Davis, J. C., Falck, R. S., Nagamatsu, L. S., & Liu-Ambrose, T. Sex differences in exercise efficacy to improve cognition: a systematic review and meta-analysis of randomized controlled trials in older humans. *Frontiers in neuroendocrinology* 2017; **46**: 71-85.
- Colcombe, S., & Kramer, A. F. Fitness effects on the cognitive function of older adults: a meta-analytic study. *Psychological science* 2003; **14**: 125-130.
- Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., ... & Henry, D. A. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *Bmj* 2017; **358**.
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj* 2008; **336**: 924-926.
- Falck RS, Davis JC, Best JR, Crockett RA, Liu-Ambrose T. Impact of exercise training on physical and cognitive function among older adults: a systematic review and meta-analysis. *Neurobiol Aging* 2019; **79**: 119–30.
- Ludyga S, Gerber M, Pühse U, Looser VN, Kamijo K. Systematic review and meta-analysis investigating moderators of long-term effects of exercise on cognition in healthy individuals. *Nat Hum Behav* 2020; **4**: 603–12.
- Rathore A, Lom B. The effects of chronic and acute physical activity on working memory performance in healthy participants: a systematic review with meta-analysis of randomized controlled trials. *Syst Rev* 2017; **6**: 1–16.
- Sanders LM, Hortobagyi T, la Bastide-van Gemert S, van der Zee EA, van Heuvelen MJ. Dose-response relationship between exercise and cognitive function in older adults with and without cognitive impairment: a systematic review and meta-analysis. *PloS One* 2019; **14**: e0210036.
- Scherder E, Scherder R, Verburgh L, et al. Executive functions of sedentary elderly may benefit from walking: a systematic review and meta-analysis. *Am J Geriatr Psychiatry* 2014; **22**: 782–91.
- Smith PJ, Blumenthal JA, Hoffman BM, et al. Aerobic Exercise and Neurocognitive Performance: A Meta-Analytic Review of Randomized Controlled Trials: *Psychosom Med* 2010; **72**: 239–52.
- Xue Y, Yang Y, Huang T. Effects of chronic exercise interventions on executive function among children and adolescents: a systematic review with meta-analysis. *Br J Sports Med* 2019; **53**: 1397–404.
- Young J, Angevaren M, Rusted J, Tabet N. Aerobic exercise to improve cognitive function in older people without known cognitive impairment. *Cochrane Database Syst Rev* 2015.
- Gasquoine PG, Chen P-Y. Effect of physical exercise on popular measures of executive function in older, nonclinical, participants of randomized controlled trials: A meta-analytic review. *Appl Neuropsychol Adult* 2020; **29**: 1–9.
- Jackson WM, Davis N, Sands SA, Whittington RA, Sun LS. Physical activity and cognitive development: a meta-analysis. *J Neurosurg Anesthesiol* 2016; **28**: 373–80.
- Haverkamp BF, Wiersma R, Vertessen K, van Ewijk H, Oosterlaan J, Hartman E. Effects of physical activity interventions on cognitive outcomes and academic performance in adolescents and young adults: A meta-analysis. *J Sports Sci* 2020; **38**: 2637–60.
- Xiong J, Ye M, Wang L, Zheng G. Effects of physical exercise on executive function in cognitively healthy older adults: A systematic review and meta-analysis of randomized controlled trials. *Int J Nurs Stud* 2021; **114**: 103810.

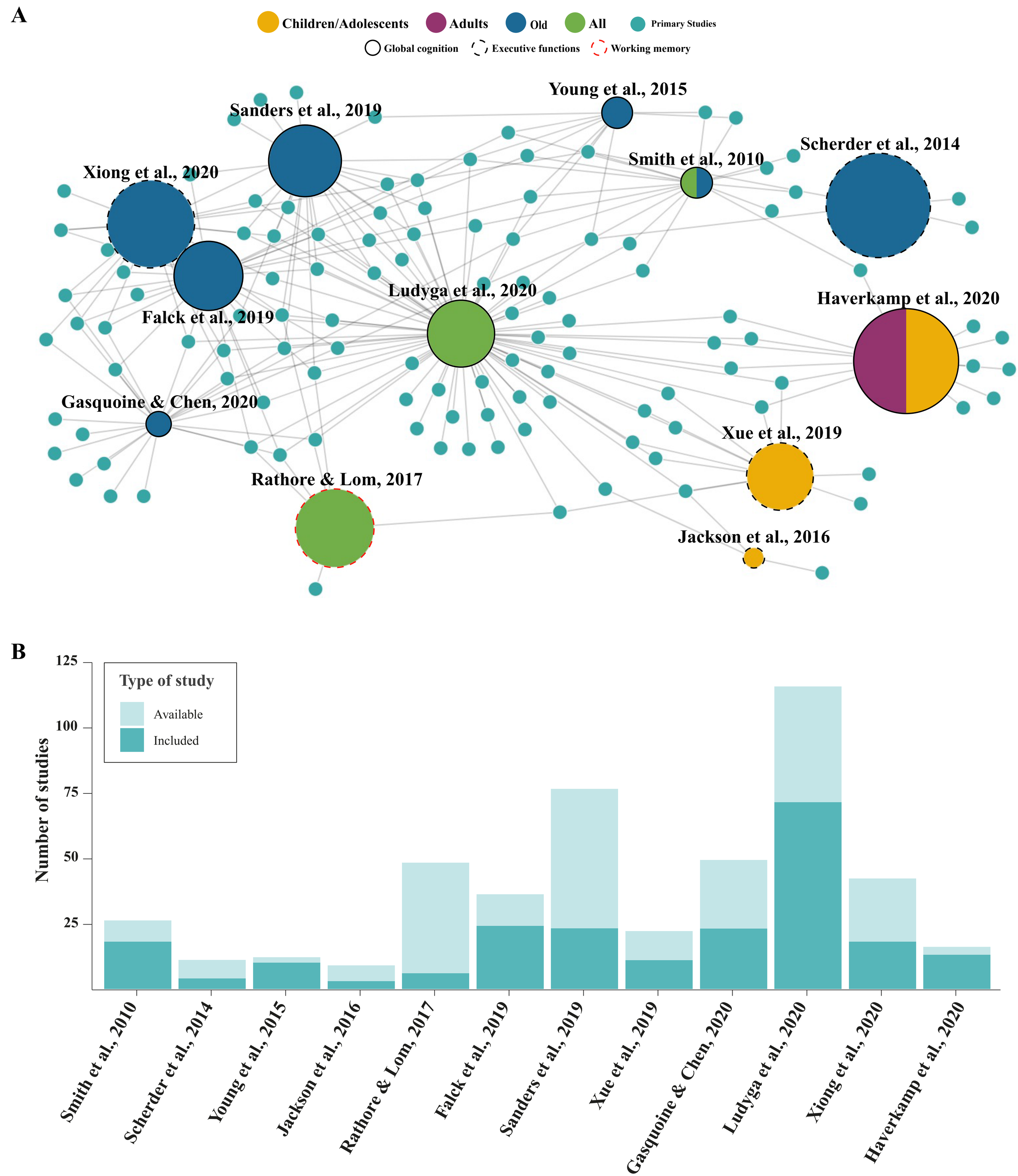
Meta-Analysis	Type of Exercise	Age Range	Cognitive Domains	Other Inclusion Criteria	Studies Originally Included	Studies Included in the Umbrella Review	Effect Sizes Included in the Umbrella Review	Original Final Effect	Estimated Final Effect	Heterogeneity (I^2)	Average Participants per Effect Size	Average Program Duration (Weeks)	Dependence Dealing
Smith et al. (2010)[19]	Aerobic	Adults & Older (≥ 18 years)	Multiple	• Only cognitively healthy • Program duration > 1 month	29	17	38	0.11	0.14 [0.02, 0.26]	15.16	71 (range 10–187)	18 (range 6–42)	Aggregates
Scherder et al. (2014)[20]	Walking	Older (> 55 years)	Executive functions	• With and without cognitive impairment	5	4	4	0.36	0.40 [0.11, 0.58]	0	81 (range 43–124)	28 (range 12–48)	Aggregates
Young et al. (2015)[7]	Aerobic	Older (≥ 55 years)	Multiple	• Only cognitively healthy • Measure of cardiorespiratory fitness	12	10	49	0.10	0.10 [-0.19, 0.39]	57.34	66 (range 10–124)	16 (range 8–26)	None / unknown
Jackson et al. (2016)[21]	Multiple	Children & Adolescents (7–12 years)	Executive functions	• Program duration ≥ 1 month	8	3	3	0.20	0.17 [0.07, 0.26]	0	90 (range 23–221)	27 (range 8–36)	Aggregates
Rathore & Lom (2017)[22]	Multiple	All	Working memory	• Only cognitively and physically healthy	8	6	10	0.27	0.07 [-0.27, 0.41]	32.91	56 (range 36–92)	15 (range 4–24)	Aggregates
Falck et al. (2019)[23]	Multiple	Older (≥ 60 years)	Multiple	• Only cognitively and physically healthy • Measure of cardiorespiratory fitness	58	24	212	0.24	0.24 [0.15, 0.34]	66.83	114 (range 14–1476)	26 (range 8–104)	Multilevel
Sanders et al. (2019)[24]	Multiple	Older (≥ 18 years)	Multiple	• With and without cognitive impairment • Program duration ≥ 1 month • The physical intervention included a non-physical component	36	23	44	0.25	0.26 [0.15, 0.37]	9.81	56 (range 16–120)	25 (range 4–52)	Multilevel
Xue et al. (2019)[25]	Multiple	Children & Adolescents (6–17 years)	Executive Functions	• Only cognitively and physically healthy • Program duration ≥ 6 weeks	19	11	22	0.23	0.17 [0.05, 0.29]	74.39	375 (range 36–776)	27 (range 6–44)	Measure selection
Gasquoine & Chen (2020)[26]	Multiple	Older	Multiple	MOST USED NEUROPSYCHOLOGICAL TESTS	26	23	45	-0.09	0.09 [-0.02, 0.20]	40.40	99 (range 18–1476)	26 (range 12–104)	Aggregates
Ludyga et al. (2020)[8]	Multiple	All	Multiple	• Only cognitively and physically healthy • Program duration ≥ 1 month	80	71	100	0.23	0.22 [0.15, 0.29]	41.43	83 (range 15–632)	21 (range 4–52)	Aggregates
Xiong et al. (2020)[27]	Multiple	Older (≥ 60 years)	Executive functions	• Only cognitively healthy • Program duration ≥ 1 month, ≥ 3 days/week, ≥ 20 min/session	21	17	48	0.30	0.34 [0.16, 0.51]	80.84	157 (range 24–1476)	21 (range 4–48)	Measure selection
Haverkamp et al. (2020)[28]	Multiple	Adolescents & Adults (12–30 years	Multiple	• Only cognitively healthy • Controlled design, with or without random allocation	27	13	27	0.36	0.40 [0.11, 0.69]	71.29	89 (range 20–632)	11 (range 4–28)	Aggregates

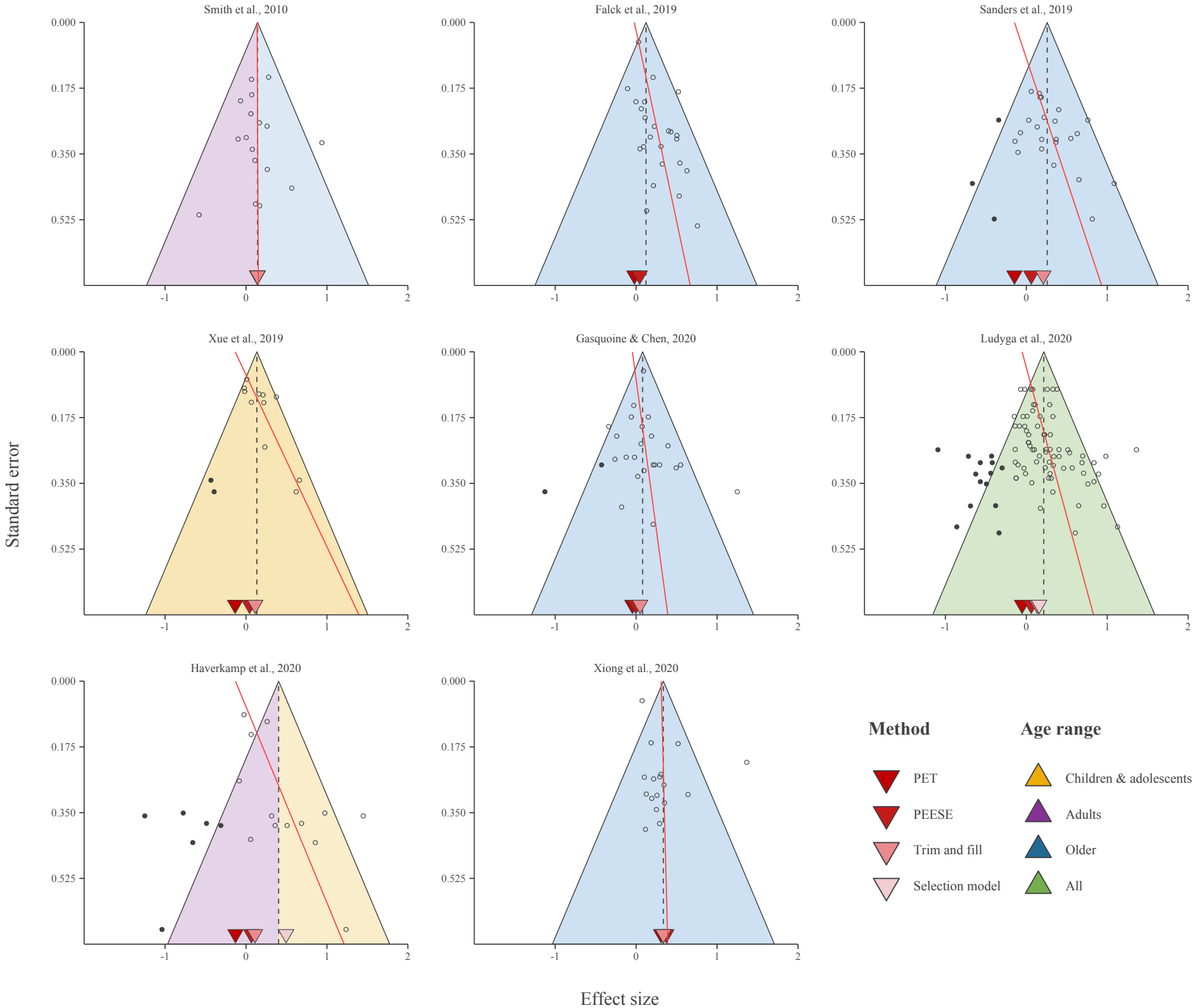
Table S1. Characteristics of the meta-analyses included in the umbrella review.

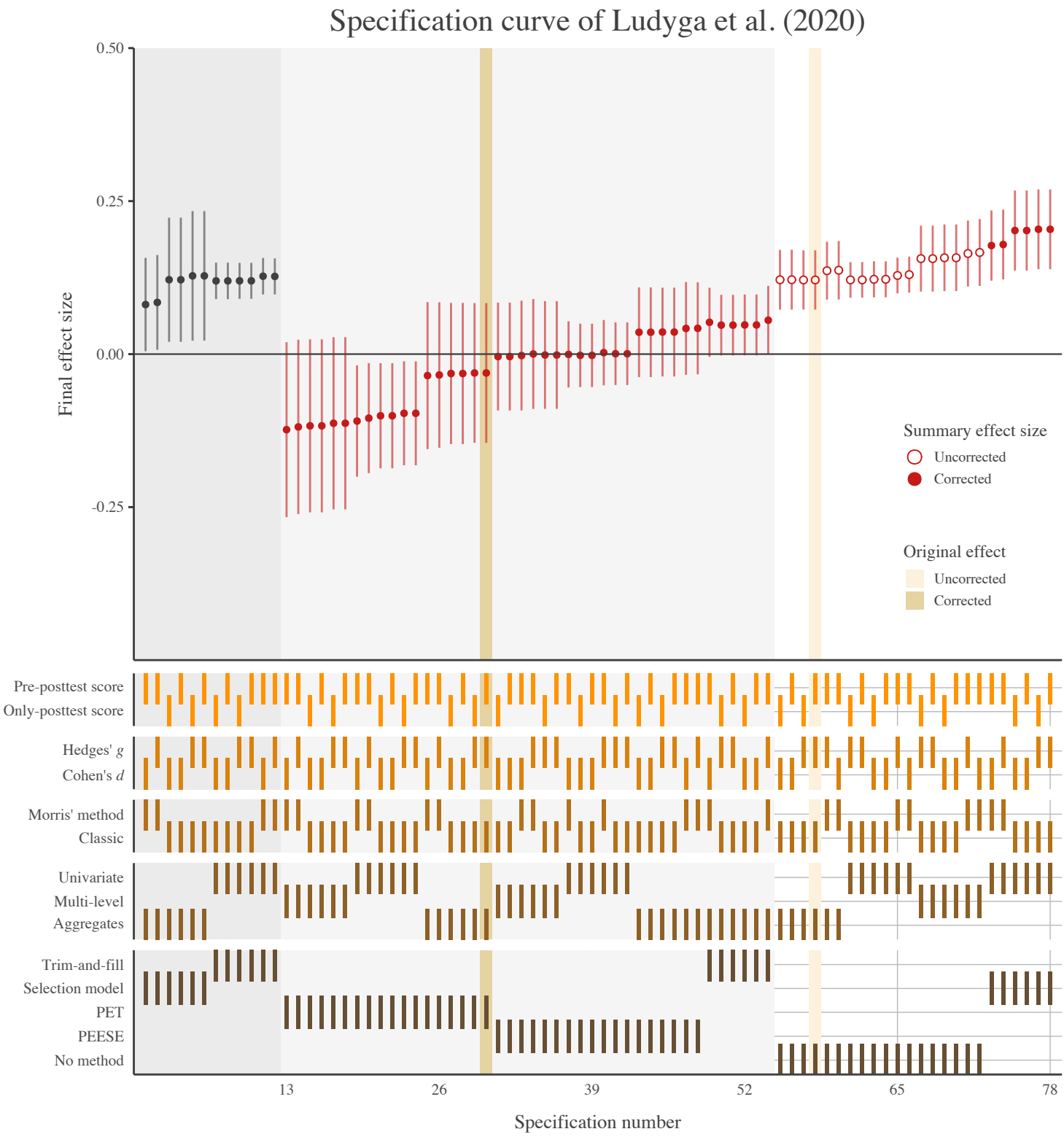

Identification
Records identified from:
Scopus (n = 1258)
Pubmed (n = 228)
Duplicate records removed
(n = 148)
Screening
Records screened
(n = 1338)
Records excluded
(n = 1294)
Reports sought for
retrieval (n = 44)
Reports not retrieved
(n = 0)
Reports assessed for
eligibility (n = 44)
Reports excluded:
Clinical population (n = 8)
Non RCT (n = 16)
Mind-Body intervention
(n = 2)
Data not available/provided
(n = 3)
Non-cognitive outcome
(n = 3)
Included
Reports of included
studies (n = 12)


A
Number of publications (normalized)
Exercise & Cognition
Psychology
Sport Sciences
Neuroscience
B
Number of primary studies included in this umbrella review
C
Number of primary studies included in each meta-analysis
Children & Adolescents
Adults
Old
All
Time (years)


A
Density
1.00
0.75
0.50
0.25
0.1
0.2
0.3
0.4
0.5
Effect size
B
Density
1.00
0.75
0.50
0.25
20
40
60
80
100
Heterogeneity
C
Density
1.00
0.75
0.50
0.25
20
40
60
80
Number of studies in meta-analysis
D
Density
1.00
0.75
0.50
0.25
100
200
300
400
Average number of participants per effect size
E
Density
1.00
0.75
0.50
0.25
250
500
750
1000
1250
1500
Number of participants per study
F
Non-physical activities
Control exercise program
Passive
G
Global cognition
Non-EF
EF
H
Older
Adults
Children & adolescents
0.0
0.2
0.4
0.6
0.8
1.0
Effect Size


A
Children/Adolescents
Adults
Old
All
Primary Studies
Global cognition
Executive functions
Working memory
Young et al., 2015
Sanders et al., 2019
Scherder et al., 2014
Smith et al., 2010
Xiong et al., 2020
Falck et al., 2019
Ludyga et al., 2020
Haverkamp et al., 2020
Gasquoine & Chen, 2020
Xue et al., 2019
Rathore & Lom, 2017
Jackson et al., 2016
B
Type of study
Available
Included
Number of studies
125
100
75
50
25
Smith et al., 2010
Scherder et al., 2014
Young et al., 2015
Jackson et al., 2016
Rathore & Lom, 2017
Falck et al., 2019
Sanders et al., 2019
Xue et al., 2019
Gasquoine & Chen, 2020
Ludyga et al., 2020
Xiong et al., 2020
Haverkamp et al., 2020


Smith et al., 2010
Falck et al., 2019
Sanders et al., 2019
Xue et al., 2019
Gasquoine & Chen, 2020
Ludyga et al., 2020
Haverkamp et al., 2020
Xiong et al., 2020
0.000
0.175
0.350
0.525
-1
0
1
2
Standard error
Effect size
Method
PET
PEESE
Trim and fill
Selection model
Age range
Children & adolescents
Adults
Older
All


Specification curve of Ludyga et al. (2020)
Final effect size
0.50
0.25
0.00
-0.25
Summary effect size
Uncorrected
Corrected
Original effect
Uncorrected
Corrected
Pre-posttest score
Only-posttest score
Hedges' g
Cohen's d
Morris' method
Classic
Univariate
Multi-level
Aggregates
Trim-and-fill
Selection model
PET
PEESE
No method
13
26
39
52
65
78
Specification number