Supplementary Materials

Effects of Physical Exercise on Adults with Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis

S1. PRISMA checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Line 2
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Line 16
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Line 50
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Line 107
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Line 128
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Line 234
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Line 234
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.	Line 257
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.	Line 269
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.	Line 269

Section and Topic	Item #	Checklist item	Location where item is reported
	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.	Line 269
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.	Line 326
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.	Line 294
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)).	Line 357
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Line 357
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Line 368
	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.	Line 401
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Line 378
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Line 378
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Line 326
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Line 340
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,	Line 403

Section and Topic	Item #	Checklist item	Location where item is reported
		ideally using a flow diagram.	
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Line 412
Study characteristics	17	Cite each included study and present its characteristics.	Line 428
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Line 478
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.	Line 512
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Line 520
	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.	Line 520
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Line 520
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Line 520
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Line 504
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Line 597
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Line 704
	23b	Discuss any limitations of the evidence included in the review.	Line 704

Section and Topic	Item #	Checklist item	Location where item is reported
	23c	Discuss any limitations of the review processes used.	Line 715
	23d	Discuss implications of the results for practice, policy, and future research.	Line 719
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.	Line 122
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Line 122
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Line 122
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Line 778
Competing interests	26	Declare any competing interests of review authors.	Line 791
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.	Line 782

From: Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLoS Med. 2021, 18, e1003583. https://doi.org/10.1371/journal.pmed.1003583.

This work is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

S2. Full search strategies

1. PubMed

Line	Search strategy
#1	"Attention Deficit Disorder with Hyperactivity"[Mesh]
#2	("attention deficit hyperactivity disorder"[tiab] OR "attention-deficit/hyperactivity disorder"[tiab] OR ADHD[tiab] OR "attention deficit disorder"[tiab] OR "attention deficit disorder with hyperactivity"[tiab] OR "hyperkinetic disorder"[tiab] OR "hyperkinetic syndrome"[tiab])
#3	#1 OR #2
#4	"Exercise"[Mesh] OR "Exercise Therapy"[Mesh] OR "Motor Activity"[Mesh] OR "Physical Fitness"[Mesh]
#5	(exercis*[tiab] OR "physical activit*"[tiab] OR "physical exercise*"[tiab] OR sport*[tiab] OR "aerobic exercis*"[tiab] OR "aerobic train*"[tiab] OR "endurance train*"[tiab] OR "resistance train*"[tiab] OR "strength train*"[tiab] OR cycl*[tiab] OR bicycl*[tiab] OR walk*[tiab] OR run*[tiab] OR jog*[tiab] OR yoga[tiab] OR pilates[tiab] OR "tai chi"[tiab] OR taiji[tiab] OR "martial art*"[tiab] OR HIIT[tiab] OR "high intensity interval train*"[tiab] OR "high-intensity interval train*"[tiab] OR "interval train*"[tiab] OR "coordination train*"[tiab] OR "coordinative exercis*"[tiab] OR "motor skill train*"[tiab] OR "mind-body exercis*"[tiab] OR "mind body exercis*"[tiab])
#6	#4 OR #5
#7	"Adult"[Mesh] OR "Young Adult"[Mesh] OR "Middle Aged"[Mesh] OR "Aged"[Mesh]
#8	(adult*[tiab] OR "young adult*"[tiab] OR "middle aged"[tiab] OR "middle-aged"[tiab] OR "older adult*"[tiab] OR universit*[tiab] OR college*[tiab] OR undergraduate*[tiab] OR postgraduate*[tiab] OR "graduate student*"[tiab] OR "university student*"[tiab] OR "college student*"[tiab])
#9	#7 OR #8
#10	#3 AND #6 AND #9

2. Web of Science

Line	Search strategy
#1	TS=("attention deficit hyperactivity disorder" OR "attention-deficit/hyperactivity disorder" OR ADHD OR "attention deficit disorder" OR "attention deficit disorder with hyperactivity" OR "hyperkinetic disorder" OR "hyperkinetic syndrome")
#2	TS=(exercis* OR "physical activit*" OR "physical exercise*" OR sport* OR "aerobic exercis*" OR "aerobic train*" OR "endurance train*" OR "resistance train*" OR "strength train*" OR cycl* OR bicycl* OR walk* OR run* OR jog* OR yoga OR pilates OR "tai chi" OR taiji OR "martial art*" OR HIIT OR "high intensity interval train*" OR "high-intensity interval train*" OR "interval train*" OR "coordination train*" OR "coordinative exercis*" OR "motor skill train*" OR "mind-body exercis*" OR "mind body exercis*")
#3	TS=(adult* OR "young adult*" OR "middle aged" OR "middle-aged" OR "older adult*" OR universit* OR college* OR undergraduate* OR postgraduate* OR "graduate student*" OR "university student*" OR "college student*")
#4	#1 AND #2 AND #3

3. China National Knowledge Infrastructure (CNKI)

Line	Search strategy
#1	主题 =("注意缺陷多动障碍" OR "注意缺陷/多动障碍" OR "注意力缺陷多动障碍" OR "注意缺陷障碍" OR "多动症" OR "ADHD" OR "成人 ADHD")
#2	主题 =("运动" OR "体育锻炼" OR "身体活动" OR "身体运动" OR "体育运动" OR "有氧运动" OR "有氧训练" OR "耐力训练" OR "抗阻训练" OR "力量训练" OR "高强度间歇训练" OR "高强度间歇运动" OR "HIIT" OR "间歇训练" OR "骑行" OR "自行车运动" OR "步行" OR "跑步" OR "慢跑" OR "瑜伽" OR "普拉提" OR "太极" OR "太极拳" OR "武术" OR "协调训练" OR "协调运动" OR "运动技能训练" OR "身心运动")
#3	主题 =("成人" OR "成年人" OR "青年" OR "青年人" OR "年轻成人" OR "青壮年" OR "中年" OR "中年人" OR "老年" OR "老年人" OR "大学生" OR "高校学生" OR "本科生" OR "研究生" OR "硕士研究生" OR "博士研究生")
#4	#1 AND #2 AND #3

4. Embase

Line	Search strategy
#1	'attention deficit hyperactivity disorder'/exp
#2	('attention deficit hyperactivity disorder':ti,ab,kw OR 'attention-deficit/hyperactivity disorder':ti,ab,kw OR adhd:ti,ab,kw OR 'attention deficit disorder':ti,ab,kw OR 'attention deficit disorder with hyperactivity':ti,ab,kw OR 'hyperkinetic disorder':ti,ab,kw OR 'hyperkinetic syndrome':ti,ab,kw)
#3	#1 OR #2
#4	'exercise'/exp OR 'physical activity'/exp OR 'exercise therapy'/exp OR 'aerobic exercise'/exp OR 'resistance training'/exp OR 'cycling'/exp OR 'walking'/exp OR 'running'/exp OR 'yoga'/exp OR 'tai chi'/exp
#5	(exercis*:ti,ab,kw OR 'physical activit*':ti,ab,kw OR 'physical exercise*':ti,ab,kw OR sport*:ti,ab,kw OR 'aerobic exercis*':ti,ab,kw OR 'aerobic train*':ti,ab,kw OR 'endurance train*':ti,ab,kw OR 'resistance train*':ti,ab,kw OR 'strength train*':ti,ab,kw OR cycl*:ti,ab,kw OR bicycl*:ti,ab,kw OR walk*:ti,ab,kw OR run*:ti,ab,kw OR jog*:ti,ab,kw OR yoga:ti,ab,kw OR pilates:ti,ab,kw OR 'tai chi':ti,ab,kw OR taiji:ti,ab,kw OR 'martial art*':ti,ab,kw OR hiit:ti,ab,kw OR 'high intensity interval train*':ti,ab,kw OR 'high-intensity interval train*':ti,ab,kw OR 'interval train*':ti,ab,kw OR 'coordination train*':ti,ab,kw OR 'coordinative exercis*':ti,ab,kw OR 'motor skill train*':ti,ab,kw OR 'mind-body exercis*':ti,ab,kw OR 'mind body exercis*':ti,ab,kw)
#6	#4 OR #5
#7	'adult'/exp OR 'young adult'/exp OR 'middle aged'/exp OR 'aged'/exp
#8	(adult*:ti,ab,kw OR 'young adult*':ti,ab,kw OR 'middle aged':ti,ab,kw OR 'middle-aged':ti,ab,kw OR 'older adult*':ti,ab,kw OR universit*:ti,ab,kw OR college*:ti,ab,kw OR undergraduate*:ti,ab,kw OR postgraduate*:ti,ab,kw OR 'graduate student*':ti,ab,kw OR 'university student*':ti,ab,kw OR 'college student*':ti,ab,kw)
#9	#7 OR #8
#10	#3 AND #6 AND #9

5. CENTRAL

Line	Search strategy
#1	MeSH descriptor: [Attention Deficit Disorder with Hyperactivity] explode all trees
#2	("attention deficit hyperactivity disorder" OR "attention-deficit/hyperactivity disorder" OR ADHD OR "attention deficit disorder" OR "attention deficit disorder with hyperactivity" OR "hyperkinetic disorder" OR "hyperkinetic syndrome"):ti,ab,kw
#3	#1 OR #2
#4	MeSH descriptor: [Exercise] explode all trees
#5	MeSH descriptor: [Exercise Therapy] explode all trees
#6	MeSH descriptor: [Motor Activity] explode all trees
#7	(exercis* OR "physical activit*" OR "physical exercise*" OR sport* OR "aerobic exercis*" OR "aerobic train*" OR "endurance train*" OR "resistance train*" OR "strength train*" OR cycl* OR bicycl* OR walk* OR run* OR jog* OR yoga OR pilates OR "tai chi" OR taiji OR "martial art*" OR HIIT OR "high intensity interval train*" OR "high-intensity interval train*" OR "interval train*" OR "coordination train*" OR "coordinative exercis*" OR "motor skill train*" OR "mind-body exercis*" OR "mind body exercis*"):ti,ab,kw
#8	#4 OR #5 OR #6 OR #7
#9	MeSH descriptor: [Adult] explode all trees
#10	MeSH descriptor: [Young Adult] explode all trees
#11	(adult* OR "young adult*" OR "middle aged" OR "middle-aged" OR "older adult*" OR universit* OR college* OR undergraduate* OR postgraduate* OR "graduate student*" OR "university student*" OR "college student*"):ti,ab,kw
#12	#9 OR #10 OR #11
#13	#3 AND #8 AND #12

6. SPORTDiscus

Line	Search strategy
#1	TX ("attention deficit hyperactivity disorder" OR "attention-deficit/hyperactivity disorder" OR ADHD OR "attention deficit disorder" OR "attention deficit disorder with hyperactivity" OR "hyperkinetic disorder" OR "hyperkinetic syndrome")
#2	TX (exercis* OR "physical activit*" OR "physical exercise*" OR sport* OR "aerobic exercis*" OR "aerobic train*" OR "endurance train*" OR "resistance train*" OR "strength train*" OR cycl* OR bicycl* OR walk* OR run* OR jog* OR yoga OR pilates OR "tai chi" OR taiji OR "martial art*" OR HIIT OR "high intensity interval train*" OR "high-intensity interval train*" OR "interval train*" OR "coordination train*" OR "coordinative exercis*" OR "motor skill train*" OR "mind-body exercis*" OR "mind body exercis*")
#3	TX (adult* OR "young adult*" OR "middle aged" OR "middle-aged" OR "older adult*" OR universit* OR college* OR undergraduate* OR postgraduate* OR "graduate student*" OR "university student*" OR "college student*")
#4	S1 AND S2 AND S3

S3. The inter-rater agreement for selecting studies

1. The inter-rater agreement for reading the title and abstract

Reviewer 1	Reviewer 2			Total
	Exclude	Include	Unclear	
Exclude	308	6	1	315
Include	3	37	0	40
Unclear	1	2	0	3
Total	312	45	1	358

Kappa score: 0.83 (0.75-0.92)

2. The inter-rater agreement for the full-text stage

Reviewer 1	Reviewer 2			Total
	Exclude	Include	Unclear	
Exclude	21	1	0	22
Include	1	21	0	22
Unclear	0	0	0	0
Total	22	22	0	44

Kappa score: 0.91 (0.79-1.00)

S4. Report eligibility and sample linkage

1. Full report audit and analytic placement

Report and source ID	Exposure and placement	Source evidence and analytic decision
Gapin 2015 (P01)	Acute; Narrative only	No non-exercise condition within the ADHD group; healthy participants also exercised. Within-person practice/time effects cannot be removed.
Wang 2024 (P02)	Acute; Quantitative	Drug-naive ADHD n = 24. Taipei/Taiwan setting. SSRT digitised from Figure 2; reported SEM converted to SD. Conservatively linked to Kuo; overlap not demonstrated.
Kuo 2024 (P03)	Acute; Quantitative	ADHD n = 26; healthy n = 26 not used. Taipei outpatient recruitment. Table 4 provides raw marginal SDs; adaptive stop success is not treated as a monotonic efficacy endpoint.
LaCount 2022 (P04)	Acute; Quantitative	ADHD n = 18; healthy data excluded. Table 2 cognitive measures after recovery; DASS-M and BAARS-M next morning identified explicitly.
Jones 2015 (P05)	Acute; Quantitative	Parallel randomized study, exercise n = 18 and reading n = 14. Mixed prior-diagnosis/ASRS ascertainment; Table 2 raw means and SDs.
Fritz 2016 (P06)	Acute; Quantitative	Screen-positive n = 32, only two previously diagnosed; crossover rest comparison. Earliest Post 1 retained per endpoint. Complementary accuracy/omission pairs not both counted; ARCI excluded from primary clinical/psychological synthesis.
Dinu 2023 (P07)	Acute; Narrative active-comparator	Cycling versus Hatha yoga, both active exercise. ADHD subgroup analysed separately; does not estimate exercise versus non-exercise efficacy.
Rassovsky 2019 (P08)	Acute; Narrative; sensitivity only	Published online 9 January2019 despite volume9/2018 DOI suffix. Methods 14 ADHD minus one excluded implies n = 13; Table 1 header says n = 17. ADHD column values match described changes. Analyse n = 13 and n = 17 only as sensitivity; response style excluded.

Report and source ID	Exposure and placement	Source evidence and analytic decision
Mehren 2019b (Flanker) (P09)	Acute; Quantitative	Frontiers article, ADHD n = 20. Reported SE converted to SD. Linked to Scientific Reports paper from larger project; not counted as an independent sample. Generic trial speed is separated from interference score.
Kallweit 2021 (P10)	Acute; Quantitative	ADHD n = 36, counterbalanced exercise Editing-task contrast. KL selective attention assigned to attention rather than executive function; Table 2 symptom and calmness values.
Thal 2014 (P11)	Acute; Narrative environment-comparator	Both nature and urban arms complete 20-minute walks. The contrast addresses environment, not exercise versus inactivity.
Mehren 2019a (Go/No-go) (P12)	Acute; Quantitative	Scientific Reports paper, ADHD n = 20; Methods explicitly places it within larger project including prior Mehren paper. Correct title/reference mapping; Table 2 SE converted to SD.
Huan 2024 (P13)	Acute; Quantitative	ChiCTR2300067833; matching demographic profile with Shi papers. Retain distinct n-back/PANAS outcomes in same sample cluster. SDs reconstructed approximately from 95% CI; alternative t19 conversion tested. MICT protocol describes 55% VO2max in intervention text; correct old50% entry.
Shi 2025a (P14)	Acute; Linked duplicate report	ChiCTR2300067833; n = 20 and matched demographics and Stroop table with journal report. Thesis ethics number differs; this does not establish a separate sample. Duplicate Stroop rows excluded.
Shi 2025b (P15)	Acute; Quantitative	Same registration as Huan/thesis. Retain journal Stroop data once, earliest Post 20 and incongruent rows; marginal CI-to-SD conversion documented.
Shalit 2014 (P16)	Chronic; Narrative training-congruent	Forty women randomized across EWMN n = 13, general coordination n = 14 and no-practice n = 13. Shared control; coordination outcomes are proximal to training, not symptom efficacy.
Svedell 2026 (P17)	Chronic; Narrative linked secondary report	Same START NCT05049239, n = 63 parent randomization as2025 paper; BAS MQ-E body awareness/movement quality are distinct secondary outcomes.

Report and source ID	Exposure and placement	Source evidence and analytic decision
Oliel 2018 (P18)	Chronic; Narrative medication-comparator	Thesis dated August2018, not2019. Randomized n = 31; mixed model n = 15 exercise/n = 8 medication. Table 2 contains estimated model means/SE and transformed anxiety/depression; do not treat as raw independent post means/SD. Medication favoured for observer symptoms at post p=.026.
Svedell 2025 (P19)	Chronic; Narrative	START NCT05049239,43 exercise/20TAU;35% missing primary12-week assessment. Source reports self-rated symptom improvement; does not represent a different trial from 2026 body-awareness report.
Fritz 2022 (P20)	Chronic; Narrative	Screen-positive women n = 32, yoga/wait-list16 each;31% yoga dropout. No significant group-by-time effects for primary cognitive or secondary outcomes. Acute session affect changes are not chronic between-group effects.
Svedell 2023 (P21)	Chronic; Narrative pilot	Pilot report of the same registered START programme (NCT05049239), with 14 randomized participants (9/5). The supplied report describes a prelude to the main RCT. The extent of participant overlap with the later 63-person cohort cannot be established; not counted as independent corroboration. Percentage-change summaries contain inconsistent mean/range pairs and are not converted to raw-score SMDs.
Kouhbanani 2023 (P22)	Chronic; Narrative	60 randomized,52 analysed,25Pilates/27control;24weeks with6-month follow-up. Improvements reported in task measures; treatment-based exclusions and attrition limit causal interpretation. Attention tasks not conflated with symptom scales.
Converse 2020 (P23)	Chronic; Narrative nonrandomized	Allocation partly based on class availability; ROBINS-I instead of RoB2. Tai Chi n = 9/cardio n = 5/inactive n = 7 baseline, follow-up9/4/6. No clear primary inattention interaction, Tai Chi vs inactive beta−1.7, SE2.4, p=.5; secondary constructs include mindfulness/sleep/physical functioning.

All three Svedell reports cite NCT05049239. The 2025 and 2026 reports share the 63-person START trial; the 2023 pilot is linked to the same programme, but its participant overlap with the main cohort cannot be resolved from the available reports. They are not three established independent confirmations.

S5. Risk-of-bias profiles and supporting reasons

1. Complete RoB 2 profiles including crossover Domain S

Report	D1	D2	D3	D4	D5	S	Overall
Wang 2024	Some concerns	Low	Low	Low	Some concerns	Low	Some concerns
Huan 2024	Low	Low	Low	Low	Some concerns	Low	Some concerns
Shi 2025a	Low	Low	Low	Low	Some concerns	Low	Some concerns
Shi 2025b	Low	Low	Low	Low	Some concerns	Low	Some concerns
Kuo 2024	Low	Low	Low	Low	Some concerns	Low	Some concerns
Dinu 2023	Low	Some concerns	Some concerns	Low	Some concerns	Not applicable	Some concerns
Fritz 2016	Low	Low	Low	Low	Some concerns	Low	Some concerns
Jones 2015	Low	Low	Low	Low	Some concerns	Not applicable	Some concerns
Svedell 2023	Some concerns	Some concerns	High	Some concerns	Some concerns	Not applicable	High
Kouhbanani 2023	Some concerns	Some concerns	High	Low	Some concerns	Not applicable	High
Svedell 2025	Low	Some concerns	High	Some concerns	Some concerns	Not applicable	High
Svedell 2026	Low	Some concerns	High	Some concerns	Some concerns	Not applicable	High
Fritz & O'Connor 2022	Low	Some concerns	High	Some concerns	Some concerns	Not applicable	High
Oliel 2018	Some concerns	Some concerns	High	Some concerns	Some concerns	Not applicable	High
Shalit & Hantiu 2014	Some concerns	Some concerns	Low	Some concerns	Some concerns	Not applicable	Some concerns

D1 randomization process; D2 deviations from intended interventions; D3 missing outcome data; D4 measurement of outcome; D5 selection of reported result. S is the additional domain in the crossover version of RoB 2, addressing period and carryover effects.

2. Complete ROBINS-I domain profiles

Report	B1	B2	B3	B4	B5	B6	B7	Overall
LaCount 2022	Moderate	Moderate	Low	Low	Moderate	Low	Moderate	Moderate
Kallweit 2021	Moderate	Moderate	Low	Low	Low	Low	Moderate	Moderate
Mehren 2019b	Moderate	Moderate	Low	Low	Moderate	Low	Moderate	Moderate
Mehren 2019a	Moderate	Moderate	Low	Low	Moderate	Low	Moderate	Moderate
Rassovsky 2019	Moderate	Moderate	Low	Low	Moderate	Low	Moderate	Moderate
Gapin 2015	Serious	Moderate	Low	Low	Low	Low	Moderate	Serious
Thal 2014	Moderate	Moderate	Low	Moderate	Low	Low	Moderate	Moderate
Converse 2020	Serious	Serious	Low	Serious	Moderate	Serious	Moderate	Serious

B1 confounding; B2 selection of participants; B3 classification of interventions; B4 deviations from intended interventions; B5 missing data; B6 measurement of outcomes; B7 selection of reported result.

3. Source reasons for overall profiles

Report	Framework	Overall	Source rationale
Wang 2024	RoB 2 crossover	Some concerns	Random order and ≥1-week separation reported; sequence-generation/allocation details limited. Objective SST/TMS outcomes.
Huan 2024	RoB 2 crossover	Some concerns	Randomized crossover design, 7-day washout, complete participation, and objective working-memory outcomes.
Shi 2025a	RoB 2 crossover	Some concerns	Randomized crossover design, ≥7-day washout, complete outcome data, and objective Stroop/fNIRS outcomes.
Shi 2025b	RoB 2 crossover	Some concerns	Randomized crossover design, ≥7-day washout, complete outcome data, and objective Stroop/HRV outcomes.
Kuo 2024	RoB 2 crossover	Some concerns	Within-subject sessions were counterbalanced, the order of study parts was randomized using a web-based random-number generator, all participants completed the sessions, and cognitive outcomes were objective. No clearly accessible prespecified analysis plan for all reported outcomes was identified in the article.

Report	Framework	Overall	Source rationale
Dinu 2023	RoB 2 parallel-group	Some concerns	Web-based randomization was stratified by group with 1:1 allocation to cycling or yoga. The trial was unblinded, minor task data were missing and handled by last observation carried forward, and protocol changes were reported. Objective cognitive outcomes reduce measurement concerns.
LaCount 2022	ROBINS-I / non-randomized crossover	Moderate	The two sessions were counterbalanced but random sequence generation was not clearly reported. Three participants were excluded because of adverse responses during HIIT, including two ADHD participants. Objective CPT outcomes and standardized symptom measures were used.
Kallweit 2021	ROBINS-I / non-randomized within-subject	Moderate	Standard and experimental sessions were counterbalanced and separated by one week, but randomization was not clearly reported. The exercise manipulation was verified physiologically; outcomes were objective task performance and task-specific symptom ratings.
Mehren 2019b	ROBINS-I / non-randomized crossover	Moderate	Exercise and active-control conditions were counterbalanced but random sequence generation was not clearly reported. Three ADHD participants were excluded for technical problems/head motion. Behavioral outcomes were objective and prespecified primary outcomes were described.
Mehren 2019a	ROBINS-I / crossover	Moderate	Exercise and movie sessions were counterbalanced, but random sequence generation was not reported. Three ADHD participants were excluded from analysis (task misconception/head motion). Objective flanker outcomes were used. The revised analysis links the Mehren reports at participant-sample level.
Rassovsky 2019	ROBINS-I / within-subject	Moderate	Sitting and treadmill conditions were counterbalanced. One ADHD age outlier was excluded. The paper contains an internal inconsistency in group sample-size labels (text vs Table 1), which warrants caution.
Fritz 2016	RoB 2 crossover	Some concerns	Treatment order was block randomized in a within-participant crossover. Outcomes were repeatedly assessed with objective cognitive tests and standardized mood measures. The main population was screening-positive for ADHD rather than clinician-confirmed ADHD.
Gapin 2015	ROBINS-I / pre-post	Serious	The ADHD group received a single pre-post exercise exposure without a no-exercise control condition, making practice/time effects inseparable from exercise effects. Small n=10 ADHD sample.

Report	Framework	Overall	Source rationale
Thal 2014	ROBINS-I / parallel field trial	Moderate	Participants were assigned by alternating nature/urban designation and day rather than concealed random allocation. Both arms walked for 20 minutes, so the contrast primarily tests environment rather than exercise versus no exercise.
Jones 2015	RoB 2 parallel-group	Some concerns	Participants were randomly assigned to exercise or reading control. Outcomes were assessed before and immediately after the assigned condition. Diagnostic eligibility allowed either prior ADHD diagnosis or ASRS screening positivity, which affects population definition rather than internal validity.
Svedell 2023	RoB 2	High	Randomized pilot trial but detailed sequence/concealment is limited. Three of nine intervention participants dropped before treatment and another low-adherence participant was excluded from analyses; analyses were largely per-protocol. Self-report outcomes could not be blinded, although evaluator blinding was attempted.
Kouhbanani 2023	RoB 2	High	Randomization is stated but allocation concealment details are not reported. After randomization, five Pilates participants were removed for nonattendance and three controls for stimulant medication use; analyses used 25/27 completers. Objective cognitive outcomes were used, but post-randomization exclusions create substantial bias concerns.
Svedell 2025	RoB 2	High	Computer-generated block randomization was prepared by an independent statistician. Participants and intervention clinicians were not blinded. Only 41/63 had complete primary outcome data at week 12; modified ITT used LOCF from week 6. Large attrition and imputation materially affect confidence.
Svedell 2026	RoB 2	High	Same randomized START trial. Complete-case analysis included 39/63 for body awareness and 38 for movement quality; one movement observation was excluded as an error outlier. Participants and physiotherapist assessors were not blinded, and one-tailed testing was used.
Fritz & O'Connor 2022	RoB 2	High	Randomization used Randomizer.org; differential attrition was substantial; 6-week ITT used LOCF.
Converse et al. 2020	ROBINS-I	Serious	Class-availability allocation and unequal expectancy/contact create confounding/selection concerns; low attendance, missing followup and unblinded self-report. See full ROBINS-I domains.

Report	Framework	Overall	Source rationale
Oliel 2018	RoB 2	High	Random assignment reported but concealment details limited; substantial/differential attrition; mixed models handled missingness.
Shalit & Hantiu 2014	RoB 2	Some concerns	Random division reported but sequence/concealment not described; no attrition reported; assessor blinding unclear.

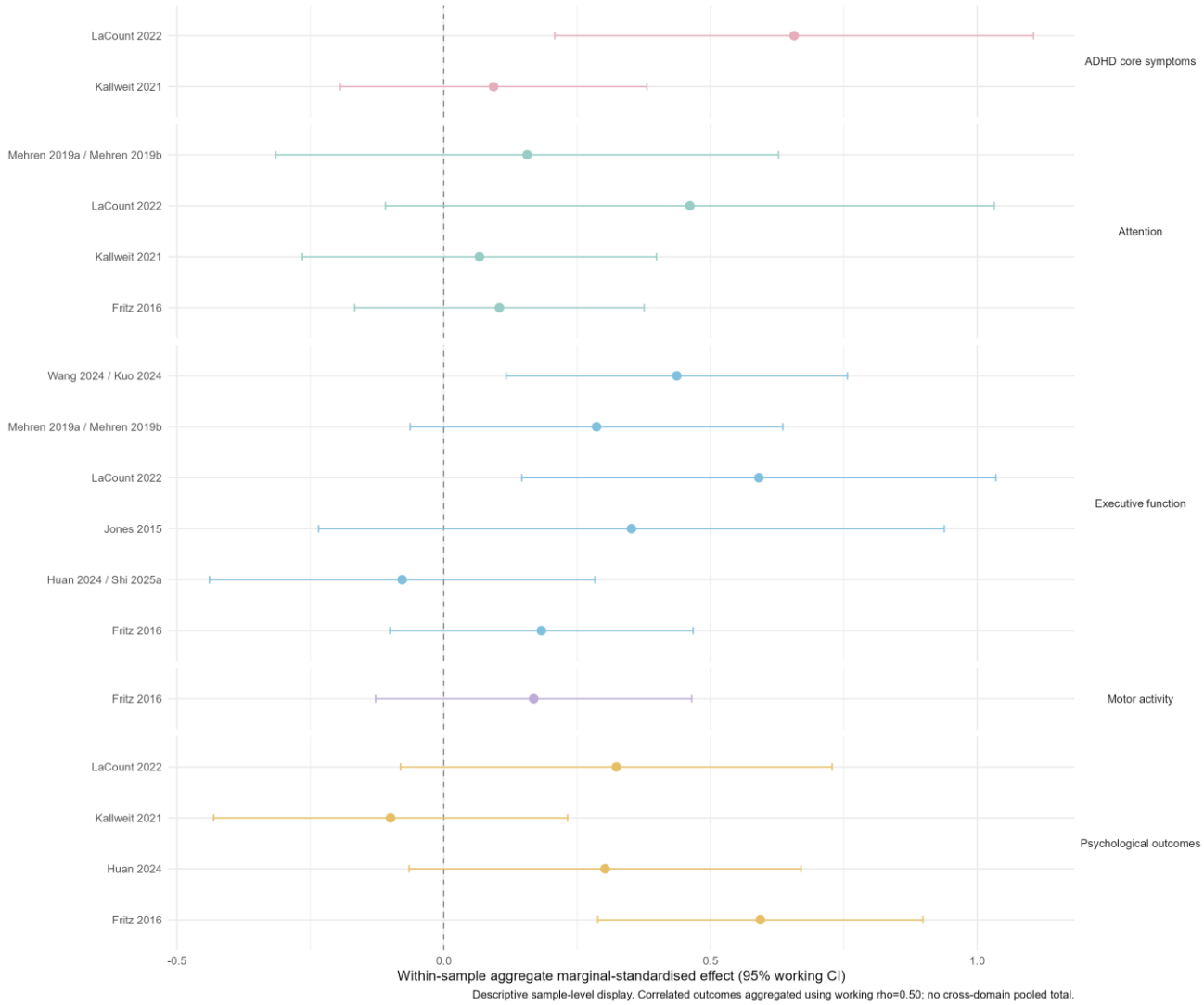
S6. Primary acute domain results

Outcome	k / effects	g (95% CI)	df / p	Approx. 95% PI	Interpretation
ADHD core symptoms	2 / 4	Not pooled	Not applicable	Not estimated	Narrative synthesis; sample-level estimates in S7
Executive function	6 / 40	0.26 (-0.01, 0.53)	4.34 / 0.056	-0.84, 1.36	CR2 primary
Attention	4 / 12	0.13 (-0.06, 0.31)	2.19 / 0.099	-0.07, 0.33	Exploratory, effective df below 4
Psychological outcomes	4 / 15	0.40 (-0.13, 0.93)	2.44 / 0.088	-1.33, 2.13	Exploratory, effective df below 4
Motor activity	1 / 4	Not pooled	Not applicable	Not estimated	Narrative synthesis; sample-level estimates in S7

Core symptoms and motor activity are synthesised narratively; no pooled mean or between-study prediction interval is reported for these outcomes. Variance components (between-sample / within-sample τ^2, rounded to three decimals): executive function 0.000 / 0.148; attention 0.000 / 0.000; psychological outcomes 0.000 / 0.141. No variance component or prediction interval is estimated for narrative rows.

S7. Sample-level estimates

1. Descriptive sample-level estimates by acute outcome domain



Positive effects favour exercise. The working intervals describe within-sample aggregates, not additional independent studies or a cross-domain pooled treatment effect.

Sample-level composites for working memory and response-time variability

Working memory. Huan 2024 (g = 0.06, 95% working CI −0.32 to 0.44) and Jones 2015 (g = −0.33, 95% working CI −1.03 to 0.37).

Response-time variability. LaCount 2022 (g = 0.65, 95% working CI 0.14 to 1.16) and Mehren 2019a / Mehren 2019b (g = 0.50, 95% working CI 0.11 to 0.89).

These descriptive composites use a working within-sample sampling-error correlation of 0.50. S8 reports the primary subdomain results and S9 reports exploratory two-sample pooling.

S8. Primary executive-function subdomain results

Outcome	k / effects	g (95% CI)	df / p	Approx. 95% PI	Interpretation
Inhibitory control	6 / 19	0.19 (-0.21, 0.60)	4.67 / 0.272	-0.90, 1.29	CR2 primary
Working memory	2 / 9	Not pooled	Not applicable	Not estimated	Not pooled in primary analysis
Processing speed	4 / 7	0.39 (0.03, 0.75)	2.65 / 0.043	-0.56, 1.34	Exploratory, effective df below 4
Response-time variability	2 / 5	Not pooled	Not applicable	Not estimated	Not pooled in primary analysis

Working memory and response-time variability are not pooled in the primary analysis; sample-level composites are in S7 and exploratory two-sample summaries are in S9. Variance components (between-sample / within-sample τ^2, rounded to three decimals): inhibitory control 0.076 / 0.056; processing speed 0.000 / 0.038. No primary variance component or prediction interval is estimated for the two-sample narrative rows.

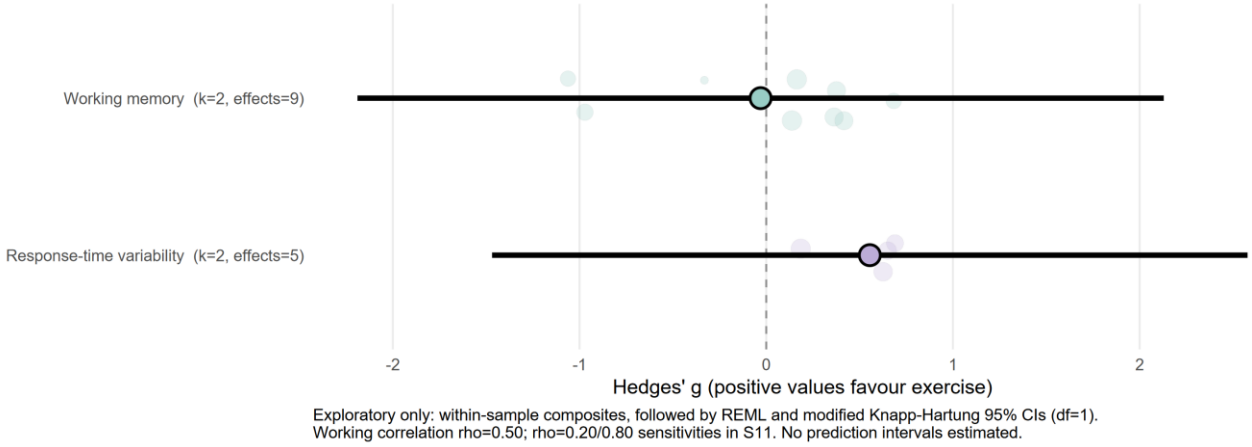
S9. Exploratory two-sample executive-function pooling

1. Correlation sensitivity of exploratory pooling

Outcome	k / effects	rho	g	95% CI	df	p
Working memory	2 / 9	0.2	0.003	-1.703, 1.709	1	0.985
Response-time variability	2 / 5	0.2	0.540	-1.199, 2.279	1	0.158
Working memory	2 / 9	0.5	-0.030	-2.189, 2.129	1	0.888
Response-time variability	2 / 5	0.5	0.555	-1.468, 2.578	1	0.178
Working memory	2 / 9	0.8	-0.059	-2.546, 2.429	1	0.815
Response-time variability	2 / 5	0.8	0.566	-1.650, 2.782	1	0.190

At rho = 0.50, working memory was g = -0.03 (95% CI -2.19 to 2.13, p = 0.888), and response-time variability was g = 0.55 (95% CI -1.47 to 2.58, p = 0.178). All correlation-sensitivity intervals included zero. These extremely imprecise results do not demonstrate benefit, equivalence or a reliable ranking of executive subdomains.

2. Exploratory two sample executive function estimates



Exploratory two-sample orchard display. Pastel bubbles are correlated selected effects; the summary centre is the REML mean of the two sample composites and the thick line is its modified Knapp-Hartung 95% confidence interval (df = 1). No prediction interval is shown. This display is supplementary, not a main-result figure.

S10. Sensitivity analyses and exploratory moderators

1. Executive function sensitivity analyses

Sensitivity	k	g	95% CI	df	p
paired_r=0.3; sampling_rho=0.2	6	0.260	-0.017, 0.538	4.48	0.060
paired_r=0.3; sampling_rho=0.5	6	0.265	0.005, 0.525	4.49	0.047
paired_r=0.3; sampling_rho=0.8	6	0.272	0.028, 0.516	4.59	0.036
paired_r=0.5; sampling_rho=0.2	6	0.259	-0.021, 0.540	4.45	0.063
paired_r=0.5; sampling_rho=0.5	6	0.258	-0.011, 0.528	4.34	0.056
paired_r=0.5; sampling_rho=0.8	6	0.265	0.013, 0.518	4.45	0.043
paired_r=0.7; sampling_rho=0.2	6	0.258	-0.024, 0.539	4.42	0.065
paired_r=0.7; sampling_rho=0.5	6	0.254	-0.020, 0.528	4.26	0.062
paired_r=0.7; sampling_rho=0.8	6	0.256	-0.005, 0.517	4.25	0.053
Clinically diagnosed only	3	0.399	0.010, 0.787	1.81	0.048
Randomized only	4	0.196	-0.313, 0.705	2.64	0.288
Exclude graph-digitised data	6	0.233	-0.029, 0.496	4.26	0.070
Same-day only	6	0.258	-0.011, 0.528	4.34	0.056
Wang and Kuo separate clusters	7	0.270	0.029, 0.510	5.19	0.034
Add Rassovsky ADHD n13	7	0.227	-0.033, 0.488	4.73	0.074
Add Rassovsky table n17	7	0.224	-0.038, 0.486	4.76	0.079
Exclude all baseline-change design reports	4	0.196	-0.203, 0.595	2.72	0.205
CI-to-SD uses t19 rather than normal	6	0.259	-0.012, 0.529	4.33	0.057

2. Leave one sample out analyses

Omitted participant sample	Remaining k	g	95% CI	p
Taipei_possible_overlap	5	0.210	-0.119, 0.539	0.142
ChiCTR2300067833	5	0.341	0.138, 0.544	0.011
Fritz 2016	5	0.277	-0.093, 0.648	0.102
Jones 2015	5	0.253	-0.053, 0.558	0.082

Omitted participant sample	Remaining k	g	95% CI	p
LaCount 2022	5	0.221	-0.082, 0.524	0.109
Mehren_project	5	0.253	-0.148, 0.654	0.147

3. Exploratory executive function moderators

Moderator term	Coefficient	CR2 SE	Effective df	p
intrcpt	0.403	0.081	1.77	0.049
factor(diagnosis)screen_positive	-0.353	0.155	2.32	0.132
intrcpt	0.378	0.135	1.00	0.218
factor(randomized)TRUE	-0.176	0.195	2.05	0.460

S11. Certainty of evidence

1. GRADE evidence profiles

Outcome	k	Risk of bias	Inconsistency	Indirectness	Imprecision	Reporting bias	Certainty
Acute core symptoms	2	Serious	Not assessable	Not additionally downgraded for diagnostic ascertainment; short-term symptom context limits inference	Very serious	Unresolved, too few samples	Very low
Acute executive function	6	Serious	Serious	Serious: screening-based populations and/or task-to-everyday-function extrapolation	Serious: few effective independent units and clinically different effects compatible with CI	Unresolved, too few samples	Very low
Acute attention	4	Serious	Not additionally downgraded	Serious: screening-based populations and/or task-to-everyday-function extrapolation	Serious: few effective independent units and clinically different effects compatible with CI	Unresolved, too few samples	Very low
Acute psychological	4	Serious	Not additionally downgraded	Serious: screening-based populations and/or task-to-everyday-function extrapolation	Serious: few effective independent units and clinically different effects compatible with CI	Unresolved, too few samples	Very low

Outcome	k	Risk of bias	Inconsistency	Indirectness	Imprecision	Reporting bias	Certainty
Acute motor activity	1	Serious	Not assessable	Serious: screening-based populations and/or task-to-everyday-function extrapolation	Very serious	Unresolved, too few samples	Very low
Chronic ADHD symptoms	—	Very serious: attrition, nonrandom allocation and/or unblinded ratings	Important clinical/design differences; no common pooled estimate	Outcome/comparator-specific applicability limitations	Serious: small, selectively analysed samples	Unresolved, too few trials	Very low
Chronic cognition/attention	—	Very serious: attrition, nonrandom allocation and/or unblinded ratings	Important clinical/design differences; no common pooled estimate	Outcome/comparator-specific applicability limitations	Serious: small, selectively analysed samples	Unresolved, too few trials	Very low
Chronic psychological outcomes	—	Very serious: attrition, nonrandom allocation and/or unblinded ratings	Important clinical/design differences; no common pooled estimate	Outcome/comparator-specific applicability limitations	Serious: small, selectively analysed samples	Unresolved, too few trials	Very low
Chronic physical function	—	Very serious: attrition, nonrandom allocation and/or unblinded ratings	Important clinical/design differences; no common pooled estimate	Outcome/comparator-specific applicability limitations	Serious: small, selectively analysed samples	Unresolved, too few trials	Very low

Outcome	k	Risk of bias	Inconsistency	Indirectness	Imprecision	Reporting bias	Certainty
Chronic quality of life	—	Very serious: attrition, nonrandom allocation and/or unblinded ratings	Important clinical/design differences; no common pooled estimate	Outcome/comparator-specific applicability limitations	Serious: small, selectively analysed samples	Unresolved, too few trials	Very low
Chronic global severity	—	Very serious: attrition, nonrandom allocation and/or unblinded ratings	Important clinical/design differences; no common pooled estimate	Outcome/comparator-specific applicability limitations	Serious: small, selectively analysed samples	Unresolved, too few trials	Very low

Outcome-level GRADE judgements apply to the primary evidence synthesis. k counts linked sample clusters for the acute quantitative domains; the chronic evidence is a narrative synthesis of linked reports and heterogeneous comparisons. A dash in the k column denotes narrative synthesis without a pooled independent-sample count.

Outcome judgements consider the body of evidence rather than applying a single downgrade solely because a prediction interval crosses zero. Some domains have multiple independent reasons for very low certainty; overlapping reasons are not counted twice. Chronic certainty applies to confidence in the available narrative evidence, not a newly calculated pooled effect.