Supplemental material 2 - Detailed selection, data collection process, risk of bias assessment, publication bias, and methodological quality of trials

Detailed selection and data collection process

Covidence automatically recognized and excluded duplicate records. Titles and abstracts of records yielded from the search strategy were independently screened by three authors (GMB, SYS, and AB). Two authors screened each record, resulting in GMB screening 8,379 records, AB 4,017 records, and SYS 4,362 records. Next, the three authors independently reviewed full texts of the potentially eligible studies. Each full text was reviewed by two of the three authors, resulting in GMB reviewing 115 full texts, AB 47 full texts, and SYS 67 full texts. At this point, references of eligible studies and relevant systematic reviews were searched to identify additional potentially eligible studies (30 potential records identified). Disagreements were resolved as a committee between the three authors (GMB, SYS, and AB). When an agreement was not achieved, a fourth author (CKB) resolved the disagreements.

The data collection process was performed independently by three authors (GMB, SYS, and AB) using a standardized data abstraction form developed by our team. Each eligible study had its data extracted by two of the three authors, resulting in GMB extracting data from 28 articles, AB 11 articles, and SYS 17 articles. When an article provided incomplete data, we contacted the author listed as correspondence contact. We contacted eight authors[1–8] requesting additional data.

Detailed risk of bias assessment

Three authors (GMB, SYS, and AB) independently assessed the risk of bias of included studies following the Cochrane Risk of Bias tool 2.0 for randomized trials.[9] The revised tool assessments are made in five domains; within each, signaling questions address a broad range of issues. The domains are: 1) bias arising from the randomization process; 2) bias due to deviations from intended interventions; 3) bias due to missing outcome data; 4) bias in the measurement of the outcome; and 5) bias in the selection of the reported result. The risk of bias judgments for each domain are "low risk of bias," "some concerns," or "high risk of bias" based on the answers to signaling questions. The overall risk of bias, in general, corresponds to the worst risk of bias in any of the domains. However, if a study is judged to have "some concerns" for multiple domains, it might be considered at high risk of overall bias. The three authors based their judgements on the guidance document of the revised tool. The three authors assessed one study as a pilot assessment. The other 28 studies were assessed by two reviewers (GMB reviewed 28 studies, SYS reviewed 17 studies, and AB reviewed 11 studies). Disagreements were resolved as a committee between the authors.

The same three authors evaluated the methodological quality independently using the Physiotherapy Evidence Database (PEDro) Scale.[10] Each eligible study had its quality assessed by two of the three authors, resulting in GMB assessing 28 articles, AB11 articles, and SYS 18 articles. The PEDro scale was designed to evaluate the methodological quality of RCTs, including assessment of randomization, allocation concealment, blinding, attrition, design, and statistics. Each of the ten criteria were independently graded '1' for 'yes' and '0' for 'no' or 'unclear' for a maximum score of 10. Disagreements were resolved with discussions between the authors. PEDro scores of 0-3 are considered 'poor', 4-5 'fair', 6-8 'good', and 9-10 'excellent.'

For trials evaluating complex interventions (e.g., exercise training), a total PEDro score of 8-10 is optimal. It is important to note that these classifications have not been validated.[11]

Results of publication bias, risk of bias assessment and methodological quality of trials

Publication bias analyzed by funnel plot (Figure 2) demonstrates that symmetry was pronounced in studies with large standard errors, indicating a trend for larger effect sizes. This trend was not confirmed by Egger's regression test, which revealed that the study standard error was not significantly associated with the study's cognitive outcomes effect sizes (b_{SE} = 7.37, *SE* = 3.95; *p* = 0.088). For precision effect estimation with standard error, the association between the squared standard error and effect size was not statistically significant (b_{SE}= 60.05, SE = 39.21; *p* = 0.176).

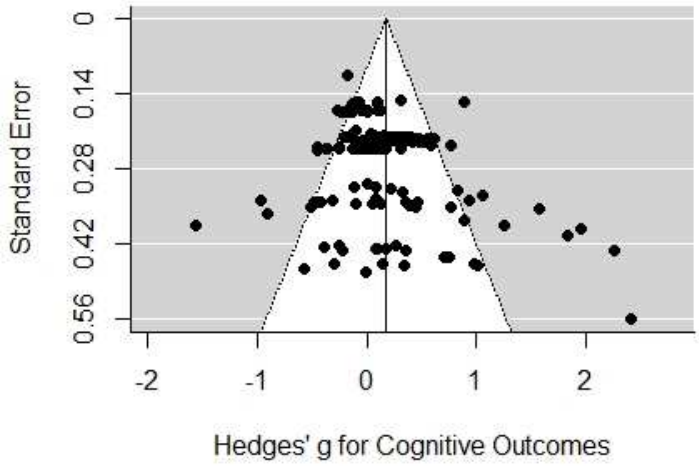

Figure 2 Funnel plot of cognitive function Hedge's g distribution of included trials.

Figures 3 and 4 depict the risk of bias by domain for studies with randomization at the individual- and cluster-level, respectively. Figure 5 presents a summary of risk of bias of studies included in the qualitative analysis

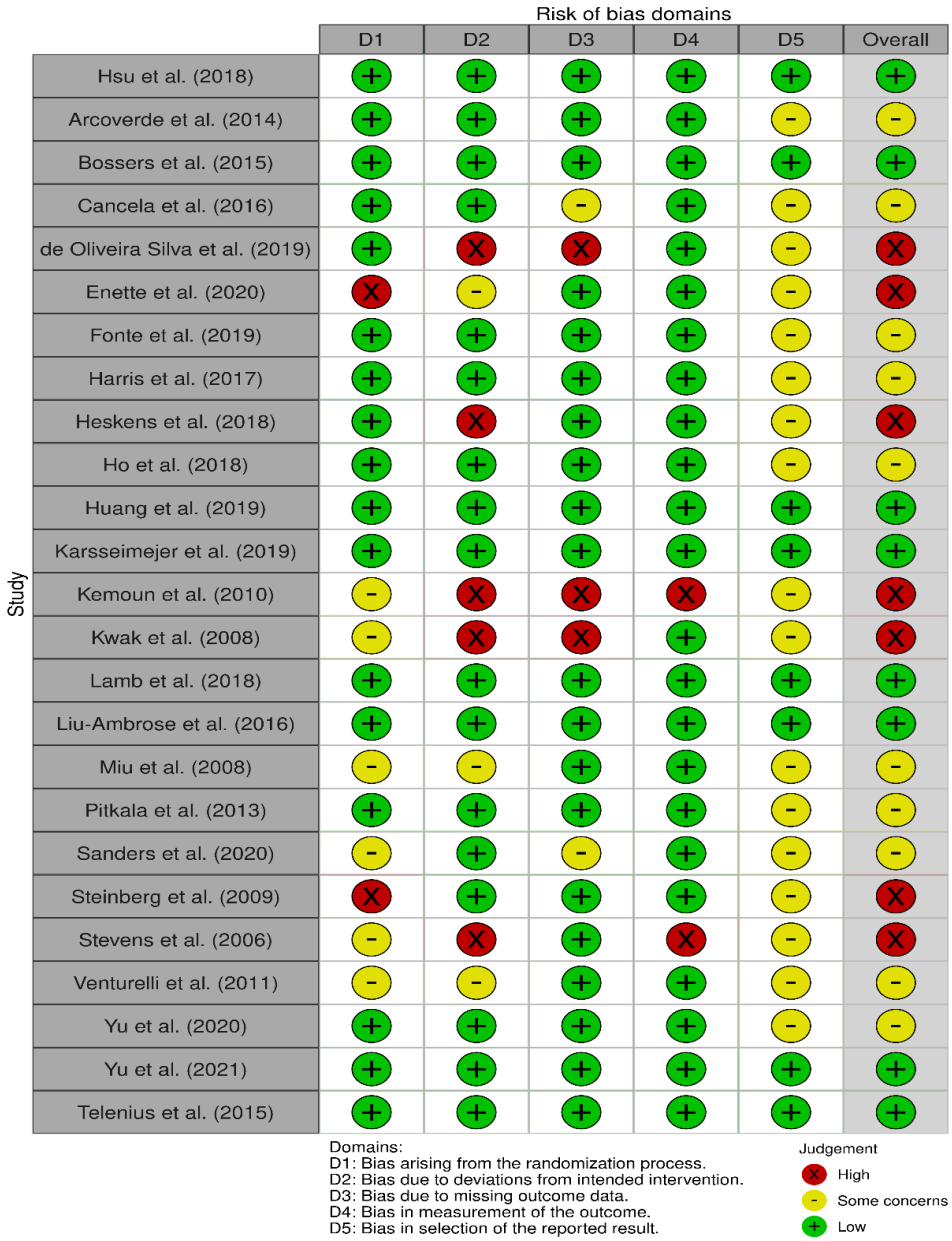

Figure 3 Risk of bias by domains of trials with randomization at the individual-level.

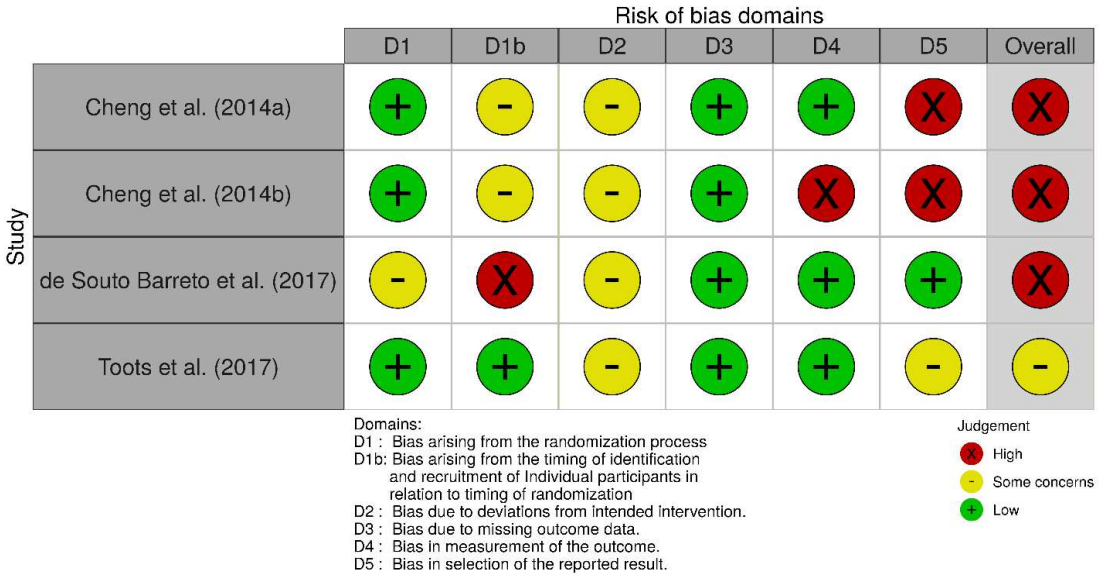

Figure 4 Risk of bias by domains of trials with randomization at the cluster-level.

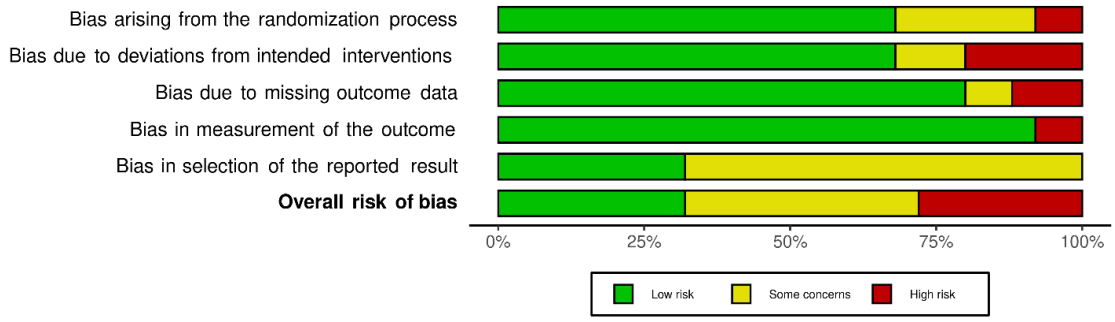

Figure 5 Summary of risk of bias of studies included in the qualitative analysis (systematic review).

Table 3 depicts the detailed methodological quality of included trials by study and

Physiotherapy Evidence Database Scale criteria.

Table 3 Detailed methodological quality of included trials (Physiotherapy Evidence Database Scale).

Study	Eligibility criteria	Random allocation	Concealed allocation	Similar at baseline	Participants blinded	Therapists blinded	Assessors blinded	<15% dropouts	ITT analysis	Between-group comparisons	Point measures and variability data	Total	Classification
Hsu et al. (2018)	1	1	1	1	0	0	1	0	0	1	1	6	Good
Arcoverde et al. (2014)	1	1	1	1	0	0	0	1	1	1	1	7	Good
Bossers et al. (2015)	1	1	1	1	0	0	1	1	1	1	1	8	Optimal
Cancela et al. (2016)	1	1	1	1	0	0	1	0	1	1	1	7	Good
Cheng et al. (2014a)	1	1	0	1	0	0	0	1	1	1	1	6	Good
Cheng et al. (2014b)	1	1	0	1	0	0	0	1	1	1	1	6	Good
de Oliveira Silva et al. (2019)	1	1	1	1	0	0	0	0	0	1	1	5	Fair
de Souto Barreto et al. (2017)	1	1	1	0	0	0	0	1	1	1	1	6	Good
Enette et al. (2020)	1	1	0	1	0	0	1	1	0	1	1	6	Good
Fonte et al. (2019)	1	1	1	1	0	0	1	1	1	1	1	8	Optimal
Harris et al. (2017)	1	1	1	1	0	0	1	1	1	1	1	8	Optimal
Henskens et al. (2018)	1	1	1	1	0	0	1	0	1	1	1	7	Good
Ho et al. (2020)	1	1	1	1	0	0	1	1	1	1	1	8	Optimal
Huang et al. (2019)	1	1	1	1	0	0	1	1	0	1	1	7	Good
Karssemeijer et al. (2019)	1	1	0	1	0	0	1	1	1	1	1	7	Good
Kemoum et al. (2010)	1	1	0	1	0	0	0	0	0	1	1	4	Fair
Kwak et al. (2008)	1	1	0	1	0	0	0	1	0	1	1	5	Fair
Lamb et al. (2018)	1	1	1	1	0	0	1	0	1	1	1	7	Good

Liu-Ambrose et al. (2016)	1	1	1	1	0	0	1	1	1	1	1	8	Optimal
Miu et al. (2008)	1	1	0	1	0	0	1	1	1	1	1	7	Good
Pitkälä et al. (2013)	1	1	1	1	0	0	1	1	1	1	1	8	Optimal
Sanders et al. (2020)	1	1	0	1	0	0	1	0	1	1	1	6	Good
Steinberg et al. (2009)	1	1	0	0	0	0	1	1	1	1	0	5	Fair
Stevens et al. (2006)	1	1	0	0	0	0	0	0	0	1	0	2	Poor
Toots et al. (2017)	1	1	1	1	0	0	1	1	1	1	1	8	Optimal
Venturelli et al. (2011)	1	1	0	1	0	0	1	1	0	1	1	6	Good
Yu et al. (2020)	1	1	1	1	0	0	1	1	0	1	1	7	Good
Yu et al. (2021)	1	1	1	1	0	0	1	1	0	1	1	7	Good
Telenius et al. (2015)	1	1	1	1	0	0	1	1	1	1	1	8	Optimal

References

1. Telenius EW, Engedal K, Bergland A. Effect of a high-intensity exercise program on physical function and mental health in nursing home residents with dementia: An assessor blinded randomized controlled trial. *PLoS One*. 2015;10(5):1-18. doi:10.1371/journal.pone.0126102
2. Yu F, Vock DM, Zhang L, et al. Cognitive effects of aerobic exercise in Alzheimer's disease: A pilot randomized controlled trial. *J Alzheimer's Dis*. 2021;80(1):233-244. doi:10.3233/JAD-201100
3. Cheng ST, Chow PK, Song YQ, et al. Mental and physical activities delay cognitive decline in older persons with dementia. *Am J Geriatr Psychiatry*. 2014;22(1):63-74. doi:10.1016/j.jagp.2013.01.060
4. Steinberg M, Sheppard Leoutsakos JM, Podewills LJ, Lyketsos CG. Evaluation of a home-based exercise program in the treatment of Alzheimer's disease: The Maximizing Independence in Dementia (MIND) study. *Int J Geriatr Psychiatry*. 2009;24(7):680-685. doi:10.1002/gps.2175
5. PitkäÏa KH, Pöysti MM, Laakkonen ML, et al. Effects of the Finnish Alzheimer Disease Exercise Trial (FINALEX): A randomized controlled trial. *JAMA Intern Med*. 2013;173(10):894-901. doi:10.1001/jamainternmed.2013.359
6. Harris JB, Johnson CS. The impact of physical versus social activity on the physical and cognitive functioning of seniors with dementia. *Act Adapt Aging*. 2017;41(2):161-174. doi:10.1080/01924788.2017.1306383
7. Bossers WJR, Van Der Woude LHV, Boersma F, Hortobágyi T, Scherder EJA, Van Heuvelen MJG. A 9-Week Aerobic and Strength Training Program Improves Cognitive and Motor Function in Patients with Dementia: A Randomized, Controlled Trial. *Am J Geriatr Psychiatry*. 2015;23(11):1106-1116. doi:10.1016/j.jagp.2014.12.191
8. Fonte C, Smania N, Pedrinolla A, et al. Comparison between physical and cognitive treatment in patients with MCI and Alzheimer's disease. *Aging (Albany NY)*. 2019;11(10):3138-3155. doi:10.18632/aging.101970
9. Sterne JAC, Savović J, Page MJ, et al. RoB 2: A revised tool for assessing risk of bias in randomised trials. *Br Med J*. 2019;366:1-8. doi:10.1136/bmj.l4898
10. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther*. 2003;83(8):713-721. doi:10.1093/ptj/83.8.713
11. Cashin AG, McAuley JH. Clinimetrics: Physiotherapy Evidence Database (PEDro) Scale. *J Physiother*. 2019;66(1):59-59. doi:10.1016/j.jphys.2019.08.005