nature human behaviour

Article

https://doi.org/10.1038/s41562-026-02429-0

Individual versus group-based interventions: a systematic review and meta-analysis of physical activity, functional, psychosocial and health outcomes

Received: 2 February 2025

Accepted: 20 February 2026

Published online: 15 April 2026

Check for updates

Marlene Kritz[1,2] ✉, Hugh Riddell[2], Daryn Olsen[3], Samantha M. Harden[3], Shauna M. Burke[4,5], Nikos Ntoumanis[6,7] & Cecilie Thøgersen-Ntoumani[6,7]

Social influences, including group dynamics, social norms and peer support, are assumed to influence physical activity (PA). A previous systematic review and meta-analysis found that 'true groups' (that is, those applying group dynamics principles) had advantages over individual PA interventions. Given technological advances and virtual platforms, we updated previous findings. Systematic searches of electronic databases were completed on 19 March 2024. A meta-analysis of 71 studies (523 effect sizes) compared individual and group-based PA interventions across behavioural, functional (for example, strength and flexibility), psychosocial (for example, quality of life and loneliness) and health outcomes (for example, VO_2max). Using a three-level random effects model, group-based conditions showed a small but non-significant advantage for PA (n = 22,042; g = 0.086, 95% CI (−0.061, 0.233), P = 0.249), psychosocial outcomes (n = 18,223; g = 0.292, 95% CI (−0.171, 0.755), P = 0.214) and health outcomes (n = 31,607; g = 0.125, 95% CI (−0.023, 0.272), P = 0.096). For functional outcomes, a significant advantage for group-based interventions emerged after outlier removal (n = 14,429; g = 0.164, 95% CI (0.032, 0.297), P = 0.015). Moderation analysis indicated that different group types produced comparable effects, except that 'true groups' were associated with a larger effect size for health outcomes. Both individual and group-based approaches (in-person or online) may be effective. Future research should explore mechanisms that enhance their impact across populations and settings. This review was prospectively registered with PROSPERO (CRD42021271452).

Regular physical activity (PA) is associated with numerous physical and mental health benefits[1–3]. To achieve these benefits, the World Health Organization recommends 150–300 min of moderate-intensity or 75–150 min of vigorous-intensity aerobic activity per week, along with muscle-strengthening exercises on 2 or more days[4]. Despite such potential health gains, many adults fail to reach the recommended guidelines for PA globally[5–7]. Even among those who join a PA programme, the majority struggle with adherence and fail to sustain changes post-programme completion[8,9].

Understanding the efficacy of interventions aimed at promoting PA requires consideration of the contextual factors surrounding PA engagement. The social environment, encompassing social norms, social support, support networks and a sense of belonging, can exert a powerful influence on PA[10,11]. These social influences are central to numerous theories applied to PA, including group dynamics[12,13], social identity[14,15], self-determination[16,17] and social cognitive theories[18,19], commonly used to inform intervention design. For instance, group

A full list of affiliations appears at the end of the paper. ✉e-mail: marlene.kritz@donau-uni.ac.at

dynamics theory posits that individuals are influenced by the attitudes and behaviours of their social groups, shaping social norms and subsequently guiding behaviour. Social identity theory emphasizes how individuals derive their sense of self from group affiliations, leading them to conform to group norms and values[15]. Self-determination theory suggests that social contexts that satisfy the psychological needs for autonomy, competence and relatedness are pivotal for sustained behavioural engagement[16]. Lastly, social cognitive theory suggests that individuals model the behaviours of others, highlighting the importance of social modelling and reinforcement[18].

While social influences can profoundly shape the way PA is pursued and/or adhered to, not all individuals are attracted to group-based PA. Indeed, many choose solitary PA, either owing to personal preference or circumstances such as the COVID-19 pandemic[20]. Individualized approaches can afford individuals autonomy and flexibility, enabling them to tailor their PA programme to their specific needs, goals and schedules. However, inherently, they may lack the support, structured feedback and guidance, external accountability and social interaction that can bolster motivation and sustained PA behaviour. Participating in group activities can yield mental health benefits by fostering feelings of social connectedness, alleviating feelings of isolation and cultivating a sense of belonging[21]. Yet, reaping these benefits necessitates an environment that is conducive to these outcomes, as positive social dynamics do not always naturally emerge within group settings[15].

Burke and colleagues[22] were the first researchers to compare the effects of individual and group approaches ($k = 44$) for PA in a meta-analysis focusing on adult populations. Outcomes included adherence, social interaction, quality of life, physiological effectiveness (for example, VO_2max and body mass index) and functional effectiveness (for example, flexibility and strength). They examined four different types of PA context, two of which were individual (with and without contact/support from researchers or health professionals) and the other two were group-based ('true' groups and standard exercise classes). True groups were exercise classes where leaders intentionally apply group dynamics principles (for example, establishing group goals, fostering team identity and using group feedback), to enhance cohesion and engagement. By contrast, non-true groups involve participants exercising together without these structured group-building strategies[23,24].

Burke et al. reported a trend across different outcomes suggesting that exercising in a true group was superior to exercising in a standard exercise class, which, in turn, did not differ from exercising at home with contact (from researchers or healthcare professionals). Furthermore, the latter condition was superior to exercising at home without contact[22]. However, the number of studies that included true groups was relatively small ($k = 9$), and these studies only examined adherence-related outcomes.

An update and extension of the Burke et al.[22] meta-analysis is timely for a number of reasons. First, their review was published nearly 20 years ago, and 58 relevant new studies have since been published and included in this meta-analysis. Second, Burke et al. examined only face-to-face groups, as these were the formats generally available at the time. However, with technological developments and the introduction of online-delivered interventions[25], an updated understanding of the literature requires expanding the definition of a PA group to include online groups[26], and newer technologies such as virtual reality platforms[27]. These modalities were not widely accessible when Burke et al. conducted their review, potentially limiting the generalizability and scalability of their findings.

This shift is particularly relevant considering the post-COVID increase in remote and home-based PA formats, which offer enhanced accessibility, especially for underserved or rural populations[25]. Moreover, digital interventions may introduce different motivational dynamics (for example, altered social presence, autonomy or feedback structures) that can influence adherence and outcomes in ways that differ from traditional in-person settings[28,29]. These contextual changes highlight the need to reassess the relative effectiveness of group-based (online and in-person) versus individual interventions, and to update public health messaging to reflect this evolving delivery landscape.

Finally, since 2006, the use of device-based PA measures (for example, actigraphs) has enabled more sophisticated assessments of PA[30]. Differing from the Burke et al. meta-analysis, we excluded dyadic interventions, as their unique interpersonal dynamics differ from those in group or individual settings and could introduce conceptual heterogeneity[31,32]. We also included only studies directly comparing group-based and individual PA conditions, to specifically address the question about their relative effectiveness.

In sum, our study differed in several aspects from the Burke et al. meta-analysis[22] in terms of search strategy and eligibility criteria (and analytic methods; see the next section). Our overall aim was to provide an up-to-date quantitative synthesis of existing literature comparing the behavioural, functional, psychosocial and physical health benefits associated with individual versus group-based PA in adults. Specifically, we examined whether group-based interventions, delivered either in-person or online, differ in effectiveness from individual-based interventions.

To guide our analysis, we addressed the following sub-questions:

Do online group-based interventions perform comparably to in-person groups?

Does intervention mode (online versus in-person) moderate the relationship between group format and behavioural, functional, psychosocial and physical outcomes?

Results

Search overview

Of 16,779 records identified, 71 studies ($k = 523$ effects) were included in the meta-analysis across PA ($k = 125$), functional ($k = 157$), psychosocial ($k = 104$) and health outcomes ($k = 137$) (Fig. 1). Power curves indicated that the minimal detectable effect size at 80% power, given the number of extracted effects and assuming high heterogeneity, was $g = 0.058$ for PA outcomes, $g = 0.052$ for functional outcomes, $g = 0.062$ for psychosocial outcomes and $g = 0.050$ for health outcomes[33,34].

Study characteristics

The 71 included studies (57 experimental, 7 quasi-experimental and 7 observational) contributed a total of $k = 523$ effect sizes and were conducted across 18 countries. Most were from Europe (37%, $n = 26$), followed by the USA (23%, $n = 16$), Australia (20%, $n = 14$), Canada (11%, $n = 8$) and Asia (9.8%, $n = 7$).

Participant characteristics. Over half of the studies (55%, $n = 39$) specifically targeted participants aged 50 years and older. Only 3% ($n = 2$) of the studies exclusively targeted male participants, while 10% ($n = 7$) focused solely on female participants. Most of the studies (59%, $n = 42$) involved healthy adult populations, while 36.6% ($n = 26$) focused on adults with clinical conditions. These clinical conditions were diverse, encompassing neurological conditions ($n = 6$), cancer ($n = 4$), cardiovascular conditions ($n = 4$), diabetes ($n = 3$), musculoskeletal and joint disorders ($n = 6$), respiratory ($n = 3$) and other conditions (for example, obesity, post-operative and mixed conditions; $n = 2$).

Intervention characteristics and delivery. Intervention duration ranged from 4 to 112 weeks (most lasting 8–16 weeks). Of the studies analysed, 72% ($n = 51$) utilized aerobic activities, either alone or combined with strength, balance training and/or flexibility training, whereas 20% ($n = 14$) focused solely on strength, flexibility and/or balance training. Overall, studies were primarily conducted in community settings (66%, $n = 47$), and groups were led by health professionals (for example, physical therapists or sport scientists; 44%, $n = 31$) or exercise instructors (38%, $n = 27$).

Characteristics of group-based conditions. Face-to-face was the predominant delivery method of group conditions (79%, $n = 56$), while online delivery methods represented a smaller but significant portion

https://doi.org/10.1038/s41562-026-02429-0

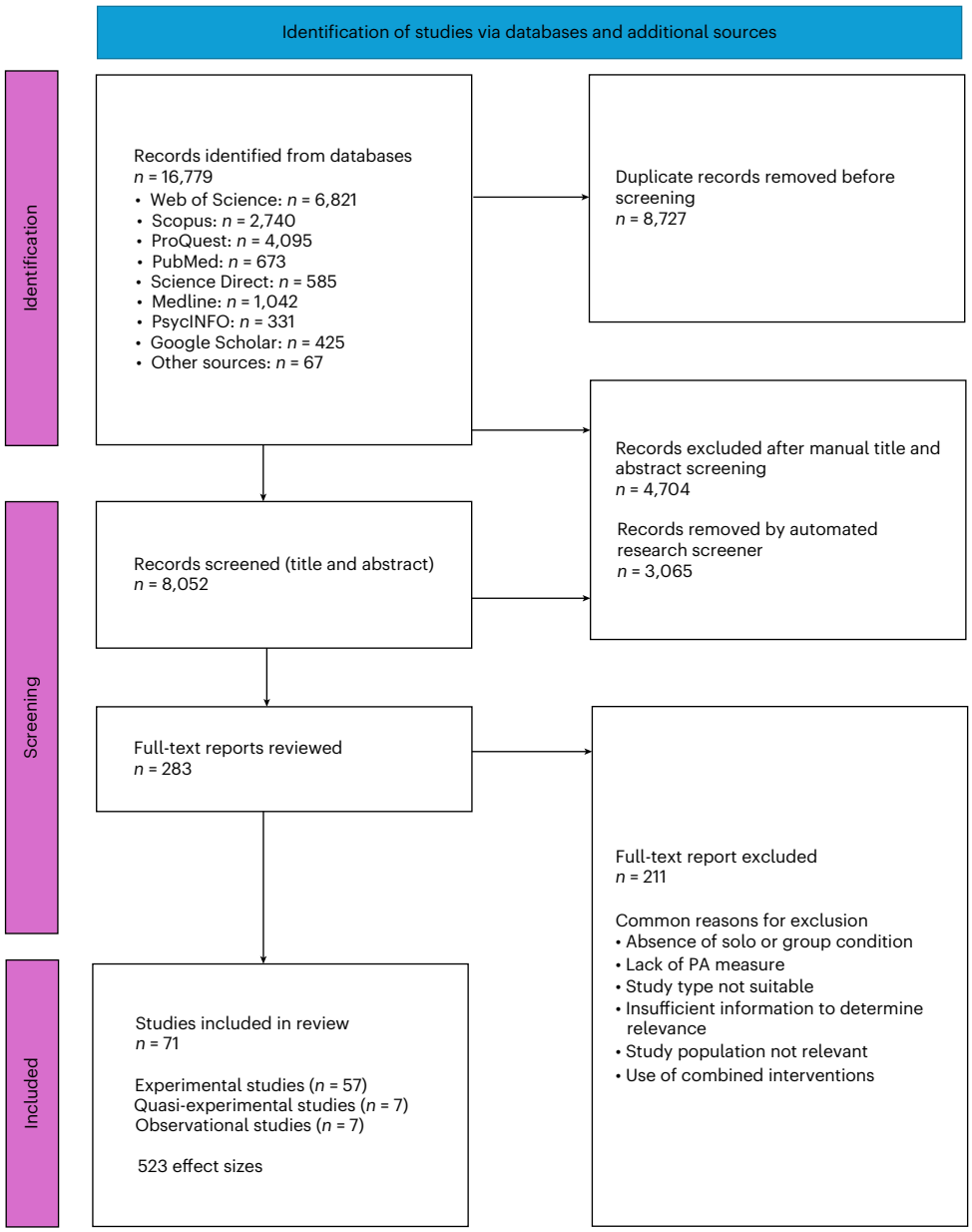


Fig. 1 | PRISMA flow diagram for study selection. Flow diagram showing identification, screening, eligibility assessment and inclusion of studies in the meta-analysis, in accordance with PRISMA guidelines.

of the studies (20%, $n = 14$). Only 14 studies (20%) used 'true groups' in which leaders actively promoted peer interaction, whereas 48 (68%) used standard classes without such facilitation. Eight studies (11%) utilized group-based designs where participants did not meet at a scheduled time but instead engaged through an interactive online community, and one study (1.4%) used a cross-sectional design examining natural group settings. Of the 65 studies describing group meetings at set times, the majority (72%, $n = 51/65$) indicated that meetings occurred 2–3 times per week. Two studies (3%) reported meetings occurring less than once per week. A minority of studies reported peer leaders ($n = 4$) or research assistants ($n = 3$) as group leaders. Group sizes in the studies that reported them (70%, $n = 50$) ranged from 3 to 30 participants per exercise group.

Characteristics of individual conditions. Most individual conditions (72%, $n = 51$) were home-based, supported by phone (35%, $n = 25$) or SMS/email (23%, $n = 16$). Contact with the facilitator varied widely, ranging from no further contact beyond the initial session ($n = 20$, 28%) to 1–3 times per week ($n = 11$, 15%). The remaining studies did not specify the number of contacts or reported contact frequencies of less than once per week. More details on study characteristics are provided in Supplementary Table 1.

Comparison of group-based and individual conditions on PA

The difference in effect sizes for PA outcomes favoured group programmes, combined (positive signed effect), but the effect was small and did not reach statistical significance ($n = 22{,}042$, $k = 125$, $g = 0.086$, SE = 0.074, $P = 0.249$, 95% CI (−0.061, 0.233); Fig. 2). There was substantial heterogeneity ($I^2 = 90.738$) between studies, which we examined through moderator analysis ('Sensitivity analyses'). Likelihood ratio tests (LRTs) indicated significant variance in effects within studies (level 2; $\sigma^2 = 0.537$, LRT = 418.378, $P < 0.001$) but not between studies (level 3; $\sigma^2 = 0.021$, LRT = 0.186, $P = 0.666$), suggesting that heterogeneity may be due, in part, to inconsistencies between the various effects reported in individual studies. Prediction intervals (95% PI (−1.399, 1.571)) suggested an almost equal likelihood of effect sizes in similar future interventions favouring either group-based or individual conditions.

https://doi.org/10.1038/s41562-026-02429-0

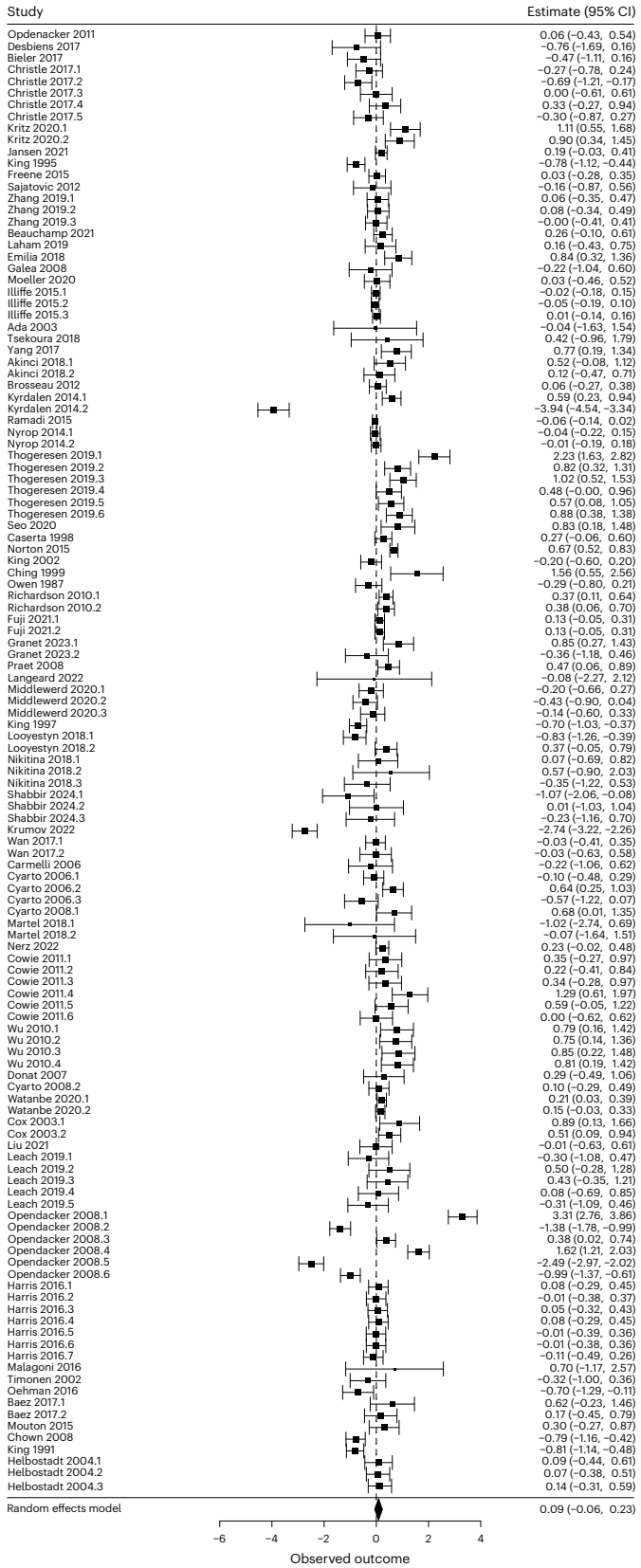


Fig. 2 | Forest plot of PA for combined effect sizes derived from proportions and standardized mean differences. Forest plot of effect sizes (Cohen's *d*) derived from two-sided meta-regression analyses comparing PA outcomes between group-based and individual interventions. Squares represent study-specific effect size estimates, with square size proportional to study weight. Horizontal lines indicate 95% confidence intervals. The diamond represents the pooled effect size estimate with corresponding 95% confidence interval.

Table 1 | Effects of moderators on PA outcomes

	g	SE	*P* value	CI lower	CI upper
Moderator: group type					
Intercept (standard face-to-face group)	0.059	0.098	0.549	−0.135	0.252
Not specified	0.071	0.564	0.900	−1.044	1.187
Online/tele group	0.060	0.180	0.741	−0.297	0.417
True face-to-face group	0.063	0.236	0.789	−0.404	0.531
Moderator: group leader					
Intercept (not specified)	−0.077	0.104	0.460	−0.284	0.129
No group leader/other	0.156	0.194	0.424	−0.229	0.540
Researcher	0.210	0.290	0.470	−0.364	0.784
Trained group exercise leader/instructor	0.478	0.178	0.008	0.126	0.829
Moderator: health status					
Intercept (healthy)	0.165	0.089	0.065	−0.010	0.340
Intervention targeted at specific health condition	−0.255	0.160	0.113	−0.571	0.061
Moderator: age					
Intercept (continuous)	−0.072	0.442	0.872	−0.952	0.809
Age	0.002	0.003	0.780	−0.011	0.015
Moderator: gender					
Intercept (continuous)	0.099	0.093	0.289	−0.086	0.284
Female (%)	−0.004	0.007	0.517	−0.017	0.009
Moderator: number of sessions					
Intercept (3+ times per week)	0.038	0.125	0.765	−0.210	0.285
1–2 times per week	0.094	0.175	0.592	−0.252	0.440
1× per week or less	−0.089	0.309	0.774	−0.700	0.522
2–3 times per week	0.247	0.618	0.691	−0.978	1.471
Not specified/other	0.080	0.234	0.732	−0.382	0.543
Moderator: intervention duration					
Intercept (continuous)	0.149	0.120	0.218	−0.089	0.386
Number of weeks	−0.004	0.005	0.482	−0.014	0.007
Moderator: study quality					
Intercept (continuous)	−0.210	0.557	0.707	−1.311	0.892
Study quality	0.378	0.703	0.592	−1.013	1.769

Two-sided meta-regression results reporting effect sizes (Cohen's *d*) for moderators comparing group-based and individual interventions; exact *P* values and 95% confidence intervals are shown. Text in brackets describes the reference category for each intercept; SE = standard error; CI = 95% confidence interval.

We also examined differences in PA outcomes between group-based and individual conditions in the subset of effect sizes derived from experimental and quasi-experimental studies only (and adjusting for pre-intervention levels of PA). Effect sizes were larger but remained non-significant ($n = 16{,}886$, $k = 55$, $g = 0.141$, SE = 0.134, $P = 0.294$, 95% CI (−0.126, 0.410)).

Sensitivity analyses

Controlling for significant moderators. We examined the effect of the following moderators on overall effect sizes: group type, leader type, sample clinical status, age, gender, number of sessions per week, method of intervention delivery, intervention duration and study quality (Table 1). Group leader was a significant moderator, with studies involving trained group exercise leaders/instructors showing larger

https://doi.org/10.1038/s41562-026-02429-0

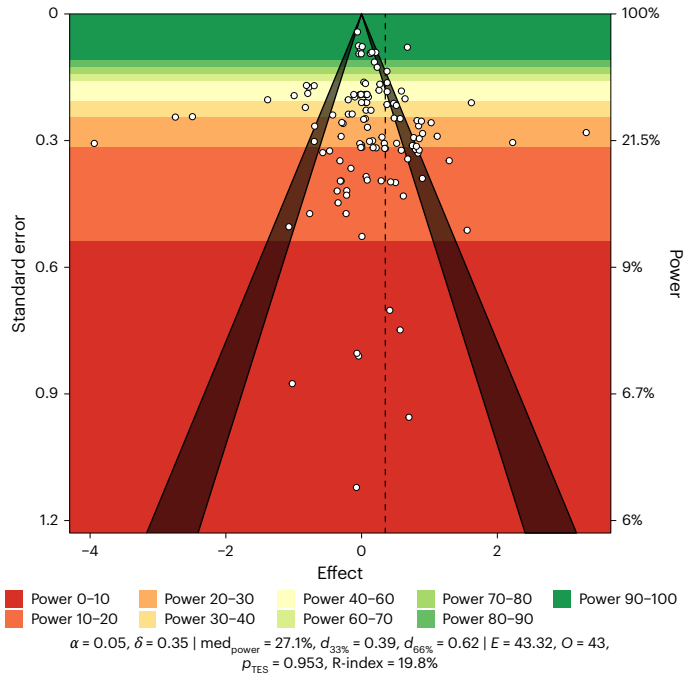


Fig. 3 | Sunset-enhanced funnel plot of included studies for PA outcome. Sunset-enhanced funnel plot showing study-level effect sizes (Cohen's *d*) and precision for PA outcomes. Shaded regions indicate statistical significance based on two-sided tests. Meta-regression and associated inferential statistics are reported with exact *P* values and 95% confidence intervals in Tables 1–4.

effects in favour of group conditions than those led by health professionals. However, including this moderator did not significantly improve model fit (omnibus test; LRT = 7.396, $P = 0.060$), suggesting that leader type was not sufficient to account for differences in PA outcomes across conditions. The variance explained by this moderator, partitioned across levels of the model, was (level 2) pseudo-$R^2 = 0.022$ and (level 3) pseudo-$R^2 = 0.346$.

Outliers. We identified six outlying effects based on standardized residuals (±3 SD from the model estimate) or Cook's distances larger than three times the mean. These outliers contributed to heavy-tailed (that is, non-normal) effect size distributions, which can influence meta-analytic results (Liu et al., 2023). Aligning with our protocol, we conducted a sensitivity analysis to examine the impact of these outlying studies. After removing these cases, the pooled effect size slightly favoured group-based over individual conditions but remained small and non-significant ($n = 21{,}122$, $k = 119$, $g = 0.094$, SE = 0.059, $P = 0.113$, 95% CI (−0.023, 0.211)). We provide diagnostic plots (Supplementary Figs. 4–11) and a side-by-side comparison of models (Supplementary Tables 2–5) with and without outliers in the supplementary file.

Publication bias. A sunset-enhanced funnel plot was used to assess publication bias (Fig. 3). The plot appeared largely symmetrical, with effect sizes from studies with larger standard errors showing only slightly more dispersion than those from studies with more precise estimates of effect size. Egger et al.[35] test of funnel plot asymmetry was conducted and indicated a small risk of publication bias (Intercept B0 = −0.035, $P = 0.745$, 95% CI (−0.177, 0.247)). However, only few studies had >80% power to detect small to moderate effect sizes ($g = 0.35$). In addition, a *z*-curve analysis was used to estimate the percentage of significant results expected if there was no selection bias (expected discovery rate (EDR) = 48.4%), as well as the estimated percentage of significant studies that would replicate if repeated perfectly (expected replication rate (ERR) = 68.8%)[36]. This analysis suggests that studies were approximately as likely to produce significant results as non-significant results, with a moderate level of replicability.

Testing equivalence between group and individual PA conditions. The absence of a statistically significant difference between group-based and individual conditions for PA outcomes does not imply support for the null hypothesis. Hence, we used the two one-sided test (TOST) procedure for equivalence testing as a follow-up procedure to formally test for the absence of a meaningful difference between condition types[37]. We used the model that combined mean differences

Table 2 | Effects of moderators on functional outcomes

	g	SE	*P* value	CI lower	CI upper
Moderator: group type					
Intercept (standard face-to-face)	0.195	0.097	0.046	0.003	0.386
Not specified	−4.381	0.612	<0.001	−5.590	−3.173
Online/tele group	−0.174	0.285	0.541	−0.737	0.388
True face-to-face group	0.103	0.288	0.722	−0.466	0.671
Moderator: group leader					
Intercept (health professional)	0.182	0.100	0.070	−0.015	0.380
No group leader/other	−3.265	0.562	<0.0001	−4.374	−2.155
Researcher	−0.056	0.478	0.907	−1.001	0.889
Trained group exercise leader/instructor	0.070	0.275	0.801	−0.474	0.613
Moderator: health status					
Intercept (healthy)	−0.196	0.208	0.347	−0.606	0.214
Intervention targeted at specific health condition	0.566	0.305	0.065	−0.035	1.168
Moderator: age					
Intercept	0.010	1.285	0.981	−2.577	2.515
Age (continuous)	−0.001	0.019	0.997	−0.037	0.037
Moderator: gender					
Intercept	−0.273	0.411	0.508	−1.086	0.540
Female (%) (continuous)	0.581	0.599	0.333	−0.603	1.765
Moderator: number of sessions					
Intercept (3+ times per week)	0.081	0.1625	0.619	−0.240	0.402
1–2 times per week	0.1405	0.2014	0.486	−0.257	0.538
1× per week or less	0.6316	0.4431	0.156	−0.244	1.507
2–3 times per week	−0.0951	0.4235	0.823	−0.932	0.742
Not specified/other	−3.1631	0.5738	<0.0001	−4.297	−2.030
Moderator: intervention duration					
Intercept	0.184	0.133	0.170	−0.079	0.447
Weeks (continuous)	0.000	0.006	0.977	−0.012	0.012
Moderator: study quality					
Intercept (continuous)	0.068	0.145	0.641	−0.218	0.354
Study quality	−0.737	1.824	0.687	−4.340	2.866

Two-sided meta-regression results reporting effect sizes (Cohen's *d*) for moderators comparing group-based and individual interventions; exact *P* values and 95% confidence intervals are shown. Text in brackets describes the reference category for each intercept; SE = standard error; CI = 95% confidence interval.

https://doi.org/10.1038/s41562-026-02429-0

Table 3 | Effects of moderators on psychosocial outcomes

	g	SE	*P* value	CI lower	CI upper
Moderator: group type					
Intercept (standard face-to-face)	0.077	0.103	0.458	−0.128	0.282
Not specified	5.677	0.384	<0.001	4.916	6.439
Online/tele group	0.081	0.235	0.733	−0.386	0.547
True face-to-face group	−0.221	0.279	0.430	−0.774	0.332
Moderator: group leader					
Intercept (health professional)	0.067	0.292	0.818	−0.512	0.647
No group leader/other	1.737	0.640	0.008	0.468	3.006
Researcher	0.164	0.826	0.843	−1.474	1.803
Trained group exercise leader/instructor	−0.112	0.513	0.827	−1.131	0.906
Moderator: health status					
Intercept (healthy)	0.369	0.299	0.219	−0.223	0.961
Intervention targeted at specific health condition	−0.207	0.488	0.672	−1.176	0.761
Moderator: age					
Intercept	−0.087	1.375	0.950	−2.822	2.648
Age (continuous)	0.006	0.020	0.771	−0.034	0.046
Moderator: gender					
Intercept	0.291	0.247	0.241	−0.199	0.781
Female (%) (continuous)	0.004	0.008	0.578	−0.011	0.020
Moderator: number of sessions					
Intercept (3+ times per week)	0.041	0.318	0.898	−0.590	0.672
1–2 times per week	0.092	0.422	0.828	−0.744	0.929
1× per week or less	−0.517	0.766	0.501	−2.036	1.002
Not specified/other	2.489	0.677	<0.001	1.147	3.832
Moderator: intervention duration					
Intercept	0.276	0.163	0.094	−0.048	0.599
Weeks (continuous)	−0.013	0.008	0.133	−0.029	0.004
Moderator: study quality					
Intercept (continuous)	−1.291	2.587	0.619	−6.423	3.840
Study quality	1.950	3.173	0.540	−4.344	8.243

Two-sided meta-regression results reporting effect sizes (Cohen's *d*) for moderators comparing group-based and individual interventions; exact *P* values and 95% confidence intervals are shown. Text in brackets describes the reference category for each intercept; SE=standard error; CI=95% confidence interval.

and proportional effects and removed outlying effect sizes for this analysis, as outlying studies caused a heavy-tailed distribution, which violated the assumption of normality for the test. The equivalence test was significant ($z = -1.667$, $P = 0.048$), indicating that group-based and individual conditions produce statistically equivalent changes in PA.

Comparison of group-based and individual conditions on functional, psychosocial and health outcomes

Pooled effect sizes for differences in functional outcomes were moderate and favoured group-based conditions, but were not statistically significant ($n = 25{,}773$, $k = 157$, $g = 0.066$, SE = 0.142, $P = 0.065$, 95% CI (−0.215, 0.346)). There were small but non-significant differences favouring group programmes for psychosocial outcomes ($n = 18{,}223$, $k = 104$, $g = 0.292$, SE = 0.233, $P = 0.214$, 95% CI (−0.171, 0.755)) and health-related outcomes ($n = 31{,}607$, $k = 137$, $g = 0.125$, SE = 0.075, $P = 0.096$, 95% CI (−0.023, 0.272)); see the Open Science Framework (OSF) page for forest plots[38]. There was moderate to substantial heterogeneity between studies for functional and health outcomes ($I^2_{func} = 74.480$; $I^2_{health} = 43.384$), and moderate to low heterogeneity for psychosocial outcomes ($I^2_{psych} = 17.268$). Prediction intervals and variance partitioned across levels of the model are presented in Supplementary Table 6. Likelihood ratio tests indicated significant variance in effects within studies (level 2) and between studies (level 3), suggesting that the partitioning of variance across levels is meaningful.

A re-analysis of the data in which we examined only experimental and quasi-experimental studies and adjusted for baseline differences between groups did not reveal any substantive changes to our results (that is, effect size direction and statistical significance remained the same), with the exception of functional outcomes, for which a small, significant effect favouring group-based interventions emerged ($n = 19{,}707$, $k = 140$, $g = 0.175$, SE = 0.086, $P = 0.044$, 95% CI (0.005, 0.345)). Detailed results of these analyses are available in the project's OSF repository[38].

Sensitivity analyses

Controlling for significant moderators. We examined the same moderators for our secondary outcomes as for the primary PA outcome (Tables 2–4). Group type, group leader and session frequency significantly moderated functional and psychosocial outcomes, whereas only group type significantly influenced health outcomes. For functional and psychosocial outcomes, studies that did not specify group type, group leader or session frequency tended to show stronger effect sizes compared with the respective reference categories (standard face-to-face groups, health professional-led interventions or ≥3 sessions per week). For health outcomes, studies using true face-to-face groups produced stronger effects favouring group over individual formats.

To assess the influence of moderating factors on our results for each outcome, we constructed models that controlled for the combined effects of all significant moderators and compared them with models that did not contain moderator variables, using likelihood ratio tests. The omnibus test of the moderator model indicated a significant improvement over the baseline model for functional outcomes (LRT = 42.253, $P < 0.001$); the effect size adjusted for moderating variables was $g = 0.039$ (SE = 0.170, $P = 0.819$, 95% CI (−0.298, 0.376)). The omnibus test of moderators indicated a significant improvement over the no-moderator model for psychosocial outcomes (LRT = 65.862, $P = <0.001$); the adjusted effect size was $g = 0.150$ (SE = 0.161, $P = 0.355$, 95% CI (−0.170, 0.470)). The omnibus test of moderators did not indicate a substantial improvement over the no-moderator model for health-based outcomes (LRT = 5.837, $P = 0.120$). The proportion of variance explained by moderators for each secondary outcome at each level of the respective models (pseudo-R^2) is provided in Supplementary Table 6.

Outliers. We identified outlying effect sizes for all secondary outcomes, which contributed to heavy-tailed (that is, non-normal) effect size distributions. Aligning with our protocol, we conducted sensitivity analyses to examine the impact of these effects. Regarding functional outcomes, six cases were identified as potentially influential using the same criteria as for PA outcomes. Removing these cases resulted in a pooled effect size, which was small and statistically significant, in favour of group-based conditions ($n = 14{,}429$, $k = 151$, $g = 0.164$, SE = 0.067, $P = 0.015$, 95% CI (0.032, 0.297)). Regarding psychosocial outcomes, 12 cases were identified as potentially influential. Removing these cases resulted in a small, non-significant effect size favouring group-based conditions ($n = 14{,}429$, $k = 92$, $g = 0.221$, SE = 0.142, $P = 0.124$, 95% CI (−0.062, 0.503)). Finally, regarding health-based outcomes, 12 cases were identified as outliers. After removing these

https://doi.org/10.1038/s41562-026-02429-0

Table 4 | Effects of moderators on health-related outcomes

	g	SE	P value	CI lower	CI upper
Moderator: group type					
Intercept (standard face-to-face)	0.071	0.082	0.388	−0.091	0.234
Online/tele group	−0.095	0.213	0.656	−0.516	0.326
True face-to-face group	0.457	0.203	0.026	0.055	0.859
Moderator: group leader					
Intercept (health professional)	0.067	0.097	0.490	−0.125	0.260
No group leader	0.242	0.306	0.430	−0.362	0.846
Researcher	0.144	0.594	0.809	−1.031	1.318
Trained group exercise leader/instructor	0.136	0.175	0.439	−0.210	0.482
Moderator: health status					
Intercept (healthy)	0.198	0.100	0.051	−0.001	0.396
Intervention targeted at specific health condition	−0.159	0.148	0.286	−0.451	0.134
Moderator: age					
Intercept	0.058	0.472	0.903	−0.883	0.999
Age (continuous)	0.002	0.007	0.800	−0.013	0.016
Moderator: gender					
Intercept	0.130	0.178	0.468	−0.223	0.482
Female (%) (continuous)	0.090	0.275	0.744	−0.455	0.635
Moderator: number of sessions					
Intercept (3+ times per week)	0.156	0.107	0.148	−0.056	0.368
1–2 times per week	−0.078	0.165	0.636	−0.403	0.247
2–3 times per week	−0.456	0.498	0.362	−1.441	0.530
Not specified/other	0.153	0.309	0.621	−0.458	0.765
Moderator: intervention duration					
Intercept	−0.021	0.117	0.860	−0.253	0.211
Weeks (continuous)	0.008	0.005	0.121	−0.002	0.018
Moderator: study quality					
Intercept (continuous)	1.285	0.681	0.061	−0.062	2.631
Study quality	−1.440	0.840	0.089	−3.101	0.222

Two-sided meta-regression results reporting effect sizes (Cohen's *d*) for moderators comparing group-based and individual interventions; exact *P* values and 95% confidence intervals are shown. Text in brackets describes the reference category for each intercept; SE=standard error; CI=95% confidence interval.

cases, the pooled effect size was small but statistically non-significant, in favour of group-based conditions (n = 30,203, k = 125, g = 0.104, SE = 0.054, P = 0.055, 95% CI (−0.002, 0.211)). Side-by-side comparisons of models with and without outliers, as well as diagnostic plots, are provided in Supplementary Tables 2–5.

Publication bias. Visual inspection of the funnel plots for all secondary outcomes revealed some asymmetry, indicating potential publication bias (see Supplementary Figs. 1–3 for functional, psychosocial and health outcomes, respectively). The test of funnel plot asymmetry[35] indicated a significant risk of publication bias for functional (Intercept B0 = −0.635, P = 0.010, 95% CI (−1.113, −0.157)) outcomes, but not for health-related (Intercept B0 = 0.115, P = 0.351, 95% CI (−0.128, 0.357)) or psychosocial (Intercept B0 = −0.228, P = 0.570, 95% CI (−0.568, 1.026)) outcomes. For all three secondary outcomes, few studies had >80% power to detect small to moderate effect sizes (g = 0.35). For all outcomes, z-curve analyses were used to estimate the ERR (functional = 76.0%; psychosocial = 75.2; health = 52.7%) and EDR (functional = 25.4%; psychosocial = 26.3%; health = 12.9%). These values indicated that studies may be likely to produce significant results but had potentially low replicability.

Testing equivalence between group and individual PA studies for psychosocial and health outcomes. When using the TOST procedure to test equivalence for secondary outcomes, we applied models with outliers removed to address violations of normality in the effect size distribution[37]. The equivalence test was non-significant for both psychosocial outcomes (z = 0.215, P = 0.585) and health-related outcomes (z = −1.596, P = 0.055), indicating insufficient precision in the data to either accept or reject the null hypothesis that group-based and individual conditions produce statistically similar differences in these outcomes. We did not conduct an equivalence test for functional outcomes because a significant difference between conditions emerged (after the removal of outliers), indicating that there was sufficient evidence to conclude a difference between group-based and individual conditions.

Methodological quality and certainty of evidence

Quality ratings of included studies ranged from 0.36 to 0.95 with an overall mean of $SQ_{Overall}$ = 0.78 (SD = 0.11). This moderate to high level of study quality suggests that, on aggregate, studies included in this meta-analysis were well conducted, appropriately analysed and clearly reported. When examining study quality for each individual outcome, average study quality was moderate to high for studies reporting PA outcomes (SQ_{PA} = 0.782, SD = 0.11), and high for studies reporting functional outcomes (SQ_{Func} = 0.82, 0.08), psychosocial outcomes (SQ_{Psyc} = 0.82, 0.07) and health outcomes (SQ_{Health} = 0.81, SD = 0.09). This did not change substantively when considering only experimental/quasi-experimental studies for any of the outcomes (SQ_{PA} = 0.78, SD = 0.11; RoB_{Func} = 0.82, SD = 0.08; RoB_{Psyc} = 0.81, SD = 0.07; RoB_{Health} = 0.82, SD = 0.07).

We assessed the certainty of evidence for randomized controlled trials (RCTs) and observational studies separately, following GRADE guidelines (Supplementary Table 1). For RCTs, the certainty was moderate for PA, functional and health outcomes, but low for psychosocial outcomes. Imprecision and risk of bias were common reasons for downgrading, with small sample sizes contributing to wide confidence intervals, making it difficult to determine the true effect of interventions. RCTs were downgraded owing to insufficient reporting of blinding and randomization, while non-experimental studies showed bias and inconsistency from lack of control for confounding factors and from using varied outcome measures. For PA, RCTs were rated moderate in terms of certainty of evidence, downgraded for inconsistency owing to varying activity measures. Non-experimental studies were rated low for often relying on adherence as a proxy for activity behaviour and for using self-reported measures. Functional and health outcomes were rated moderate in terms of certainty of evidence for both study designs, downgraded owing to risk of bias and imprecision. Psychosocial outcomes were rated low, driven by heterogeneous self-reported measures prone to bias and indirectness. Specifically, many studies measured general well-being instead of specific psychosocial or social outcomes such as mental health or loneliness.

Discussion

The main aim of the present study was to conduct an up-to-date meta-analysis comparing group versus individual approaches for promoting PA, functional, health and psychosocial outcomes in adults. This review extends Burke et al.[22] by incorporating studies with new technologies, delivery formats (for example, online and virtual) and more diverse outcome measures. Across 71 studies, we found statistically equivalent outcomes for PA and health, regardless of delivery

format. Group-based formats showed a small, significant advantage for functional outcomes after outlier removal. Moderation analyses indicated slightly larger health effects for 'true groups'.

Our findings—that group-based and individual approaches did not differ significantly, except for functional outcomes (where group approaches showed superiority after removing outliers)—align with those of Burke et al.[22]. The latter reported no differences between 'standard' exercise classes (in which group dynamics strategies were not used to enhance group cohesion) and home-based programmes with contact from researchers or health professionals.

Further, our moderation analysis on group type showed that different types of group resulted in broadly similar outcomes. The only exception was 'true groups' (that is, those using group dynamics principles). Consistent with Burke et al.'s finding that these groups outperform standard exercise groups, our moderator analysis showed that effect sizes for health outcomes (but not other outcomes) were slightly larger for true groups compared with standard face-to-face groups. However, the inclusion of group type as a moderator did not substantially influence the overall effect size, suggesting that any variation owing to different group types is not sufficient to account for differences in health-related outcomes between group-based and individual conditions. It should be noted, nevertheless, that there was a limited number of true group conditions ($k = 14$) and in these studies it was not possible to determine the number of group-based strategies used. This is unsurprising since previous systematic reviews of general PA promotion programmes[39] and PA programmes for cancer survivors[40] have shown that many 'group based' PA promotion programmes do not fully use a theoretical approach or provide replicable strategies that may enhance group influence or perceptions. Given the limited number of true group conditions, we were unable to examine differences between online and face-to-face true groups, although this would be an interesting avenue for future research.

The results of the TOST showed statistical equivalence between individual and group conditions for PA outcomes, but not for health or psychosocial outcomes. This implies that we can be reasonably confident that group-based and individual interventions have similar effects in terms of PA outcomes. However, it was not possible to conclude whether group or individual interventions are superior for health or psychosocial outcomes. This could be due to heterogeneity in how these outcomes were assessed, poorly described groups or different types of intervention. This finding underscores the need for consistent definitions and standardized measurement to improve comparability across studies.

The equivalence in PA outcomes has important practical implications in terms of intervention flexibility and scalability. Group interventions can foster social interaction and accountability, potentially have a greater reach and are scalable[39]. However, individual formats can offer greater accessibility, autonomy and personalization. Our findings suggest that both formats can be viable pathways to PA behaviour change. Offering choice-based or hybrid models, where individuals select the format that best fits their preferences, needs and life circumstances, may help expand the reach and inclusivity of PA programmes. For example, individuals with limited time, social anxiety or a preference for exercising alone could benefit from self-guided or online formats, particularly in rural or resource-limited settings.

From a theoretical perspective, these findings can be interpreted through the lens of self-determination theory, which highlights the importance of satisfying the psychological needs for autonomy, competence and relatedness for the facilitation of behaviour change maintenance[16,17]. The absence of clear differences in PA between group and individual formats may suggest that both formats are capable of supporting these needs, albeit through different mechanisms. Individual formats may promote autonomy and competence through flexible, self-paced engagement, while group settings may foster relatedness and competence via social support and structured guidance. It is possible that when these needs are sufficiently met, the delivery format becomes less critical to ensuing outcomes.

For functional outcomes (for example, balance, coordination and muscle strength), we did not find a substantive difference in our primary analysis; however, sensitivity analyses indicated that group-based approaches became significantly more effective after removing outliers, as well as when restricting the analysis to experimental/quasi-experimental studies (that is, excluding correlational studies). While results of sensitivity analyses should be interpreted with caution, they can be useful for hypothesis generation[41]. We suggest that in some cases (for example, when groups are formalized such as in experimental studies), group structure and support may better promote functional gains compared with individual contexts, as individuals in a group can benefit from observing others and receiving feedback.

Our findings can also be explained via the lens of social cognitive theory[18,19], which emphasizes the role of observational learning, group modelling and vicarious reinforcement in enhancing self-efficacy and behaviour change. Within group-based exercise settings, individuals may feel more capable and motivated after observing peers succeed at a task, receiving real-time feedback from an instructor or perceiving shared progress within the group[18]. Previous research supports this, showing that the presence of others in an exercise class can facilitate skill mastery and confidence[42]. These mechanisms may be particularly salient for functional tasks requiring coordination or technique, which benefit from visual learning and real-time adjustment compared with more general PA outcomes, where intensity and consistency may matter more than group dynamics. Such functional advantages suggest that group-based approaches could be valuable for rehabilitation programmes targeting older adults and clinical populations, where preserving functional ability is critical[43].

Another important new finding of our review was that online and standard face-to-face groups did not differ in their effectiveness. This finding suggests that the critical elements of group conditions, such as social support, accountability and shared goals, may be effectively facilitated in virtual settings. From the perspective of self-determination theory, relatedness may be fostered through virtual communities, live sessions or peer forums embedded in these platforms where users can share progress, exchange encouragement and engage in group challenges. Aligning with suggestions by other researchers[44], it is possible that the quality of interaction may be more important than the mode of delivery in driving successful outcomes. Online formats could potentially support also autonomy via flexible scheduling and self-paced engagement. Competence could be supported through structured guidance and tailored feedback, for example, through virtual coaching platforms that provide personalized goals and progress tracking.

The similar effectiveness of online and face-to-face groups across psychosocial, functional and PA-related outcomes supports the potential of online interventions to reach populations who face barriers to in-person participation, such as mobility limitations, transportation challenges or geographic isolation. Online formats address some logistical demands (for example, transport and staffing), expand scalability and extend reach to underserved populations, thereby supporting both cost-effectiveness and equity. Building on these findings, policy initiatives could integrate virtual PA programmes into public health strategies, for example, by offering virtual and group-based rehabilitation programmes for older adults or embedding digital PA support into chronic disease prevention and management initiatives. Workplace wellness policies could also leverage virtual, group-based PA programmes to enhance reach and scalability, particularly for national PA promotion efforts in rural or resource-limited settings.

Strengths, limitations and future research directions

Our meta-analysis had several strengths. Specifically, we used a three-level meta-analysis framework that enabled us to account for dependencies among effect sizes within studies, thereby providing a

more nuanced analysis that captured variability at different levels, such as within studies, between studies and between subgroups. This added layer of precision allows for a richer understanding of the data and helped avoid biases that may arise from treating all effect sizes as independent. The integration of the TOST equivalence test further strengthens the analysis by allowing us not only to detect significant differences but also to assess practical equivalence between conditions. This is particularly valuable for understanding whether interventions are similarly effective, supporting more definitive and actionable conclusions that can be translated into informed, evidence-based decisions.

Study quality was assessed using the QualSyst tool and overall literature quality using GRADE. Average quality was relatively high, indicating that studies were generally well conducted, analysed and reported. However, GRADE ratings showed moderate certainty for all outcomes, except for psychosocial outcomes, which was low.

Although most studies were experimental, a small number were observational or cross-sectional, introducing potential limitations such as measurement imprecision and confounding. In line with our protocol, we aggregated both experimental and non-experimental studies[45]. Secondary analyses restricted to experimental and quasi-experimental studies yielded results consistent with the main analysis, suggesting that the inclusion of observational studies had minimal impact on overall findings. However, methodological concerns across the literature, including reliance on cross-sectional designs in some outcome domains, risk of bias, high heterogeneity, limited control for confounders, self-reported outcomes, use of indirect proxies (for example, class attendance as a proxy for PA) and small sample sizes, indicate scope for methodological improvement in future research.

Psychosocial outcomes in our review were particularly affected by imprecision, risks of bias (for example, lack of blinding) and heavy reliance on self-report measures, raising concerns about the robustness of the associated findings. While self-report remains a necessary method for assessing many psychological and social constructs, future research may benefit from prioritizing more specific and well-validated measures (for example, affect and loneliness) over broad or diffuse constructs (for example, general psychological well-being). In addition, meta-analytical comparisons of group-based versus individual interventions targeting clearly defined psychological and social outcomes could offer greater granularity regarding intervention effects and advance construct-specific conclusions.

We observed imprecision in both primary and secondary outcome effect size estimates, indicating uncertainty around the true size of differences between group and individual intervention programmes. This issue is potentially attributable to the substantial heterogeneity observed in all effect sizes, which underscores variability in study designs, populations and intervention implementation across studies. Differences in participant demographics, settings (for example, community versus clinical) and intervention characteristics (for example, frequency and duration) likely contributed to inconsistent results.

Although our sensitivity analyses indicated that the overall pattern of findings was largely robust to between-study differences, heterogeneity remains an important limitation. Inconsistent or incomplete reporting of intervention settings (for example, community, workplace and clinical) and intervention characteristics (for example, group size) prevented us from examining whether these factors moderated intervention effects. To improve interpretability and better isolate the conditions under which group or individual formats are most effective, future research should adopt clearer reporting standards for intervention context and consider standardizing key intervention characteristics or incorporating planned subgroup/moderator analyses. Such steps would enhance comparability across studies and improve the precision of effect size estimates.

Finally, our obtained effect sizes were not disattenuated for measurement error. We suggest that this is likely to have a minimal impact on results, since measurement error should affect both individual PA and group-based PA equally; however, we acknowledge that correcting for this would have provided a purer assessment of effects. In addition, we excluded grey/unpublished literature, including dissertations and conference proceedings, as these typically provide insufficient methodological detail or extractable data. While this decision aligns with our methodological rigour, it may bias results by overrepresenting published studies, which are more likely to report positive effects.

Future research could explore potential moderators to better understand which individuals respond best to group versus individual PA conditions. Factors such as personal preference, personality traits, physical and psychological capability, and social or physical opportunity may influence individual responses[46,47]. Interaction effects between moderators (for example, group type × leader role or duration × setting) may offer additional explanatory insight, but limited reporting in the primary studies precluded reliable testing. Larger datasets or access to the raw data from all primary studies may enable future analyses to systematically explore these interactions.

It may also be valuable to determine whether specific group formats (for example, synchronous versus asynchronous, instructor-led versus peer-supported) better align with individual characteristics or contextual demands. Hybrid models that combine group and individual elements (for example, periodic group sessions alongside personal exercise plans) could be tested for synergistic benefits[48]. Where integration is not feasible, formats should be tailored to individual preferences, intended outcomes (for example, functional, psychological and social) and contextual constraints (for example, personnel time, facility space, available social support and scalability)[49].

Methods

Protocol and registration

This meta-analysis was pre-registered with PROSPERO (CRD42021271452) and adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[50]. We followed the protocol[51] for objectives, eligibility and primary analyses, with a small number of refinements: the search period was extended, screening was restricted to English-language studies, GRADE ratings were added to supplement QualSyst, and equivalence testing (TOST) was incorporated. In addition, psychological and social outcomes were merged and reported as 'psychosocial', and physical health outcomes were separated into functional and health domains. The originally pre-registered outcome structure (separate psychological and social outcomes, and a single physical outcome domain) was not analysed, as substantial overlap in outcome measures and inconsistent reporting across primary studies rendered these distinctions methodologically uninformative and analytically unfeasible. Full details are provided in Supplementary Table 7.

Eligibility criteria

To be eligible for this meta-analysis, the authors of the studies had to:

1) Report any type of PA or exercise condition or intervention, performed in at least one individual (home-based, online or face-to-face) and at least one group setting. Group settings could include face-to-face or online formats and needed to include at least three group members. Dyadic interventions were excluded as they have been meta-analysed recently[31], and also for the reasons explained earlier. Studies that described multi-component programmes (for example, combining PA with nutrition counselling) were excluded unless they provided separate effect sizes for the PA component(s).
2) Report at least one measure of PA behaviour/adherence (primary outcome of interest), which could be based on self-reports or device-based measurements.
3) Include adults aged 18 and older, regardless of their health status. Studies that reported data for mixed age groups but included a clear breakdown for adults were also eligible.

We included quantitative studies with experimental, quasi-experimental and correlational designs (longitudinal or cross-sectional). Both intervention and observational studies were eligible if they reported sufficient information to compute effect sizes. Studies were excluded if they focused on competitive sports, if they lacked sufficient data to compute an effect size (and the authors could not provide this data upon request), or if they were unavailable in English. Conference proceedings and theses/dissertations were excluded owing to limited methodological detail and incomplete data reporting. Only one relevant abstract was identified, which lacked usable data.

Information sources

Following best practices for systematic reviews and meta-analyses, we conducted a search of multiple databases via platforms such as Web of Science and ProQuest, ensuring broad coverage and methodological rigour[52]. No date restrictions were applied, and all searches were conducted in English. We included studies from inception given that we used somewhat different criteria from the 2006 meta-analysis[22]. In addition, manual searches of reference lists of included studies were conducted, and forward and backward citation tracking was used. Google Scholar was used to identify any additional relevant literature. The final search was completed on 19 March 2024.

Search strategy

The search strategy was designed by the first author in collaboration with an experienced research librarian and reviewed by the research team. Literature searches were conducted in Web of Science and ProQuest from database inception to 19 March 2024. Search terms combined keywords and subject headings related to PA (for example, PA and exercise), intervention format (for example, group, collective, group-based, individual, solo, home-based and online) and outcomes (for example, adherence, fitness, functional, health, psychological and social). Reference lists of included studies were manually searched, and forward and backward citation tracking was conducted using Google Scholar. Full search strings for all databases are provided in the supplementary file (Part 1: Search strategy)

Study selection

All identified records were imported into Endnote v.21 to remove duplicates. The remaining studies (n = 8,052) were screened in 2 stages.

Title and abstract screening. Two independent reviewers (M.K. and D.O.) screened the titles and abstracts of 62% of the retrieved studies (4,987/8,052) based on eligibility criteria. The Research Screener software[53,54], a validated machine learning tool[53] trained on eligible studies, was used to assist with the initial screening process by automating the exclusion of irrelevant studies. Using this tool, in the second step, the reviewers screened 51% of the imported studies (3,152/6,217), manually reviewing those flagged as potentially eligible. This percentage far exceeded the 5%–35% recommended by the authors of the tool[53], as well as generic recommendations of a minimum of 20% for automated screening tools[54]. All AI-flagged potentially eligible abstracts were manually reviewed, and final inclusion decisions were verified by the two independent human reviewers (M.K. and D.O.). The tool excluded 3,065 studies as irrelevant. In an updated search, an additional 1,835 studies were manually screened by the reviewers using Endnote. Any conflicts between reviewers (n = 212) were resolved through discussion or with input from a senior researcher (C.T.-N.).

Full-text screening. Full texts of the studies deemed eligible after the title and abstract screening (n = 283) were retrieved and assessed for eligibility by at least 2 independent reviewers (M.K. and D.O.). Disagreements were resolved either through discussion (n = 55) or consultation with a third reviewer (H.R., C.T.-N. or N.N.; n = 16). Agreement was reached in 213 cases (75%), and interrater reliability was moderate (Cohen's κ = 0.47). Reasons for exclusion at this stage are documented in Fig. 1.

Data extraction process. A standardized Excel form was used for data extraction, which is available on the OSF[38]. Three independent reviewers (M.K., H.R. and D.O.) extracted study and participant details (for example, author, design, sample size, age, gender and health status), PA intervention characteristics (for example, type, duration, intensity and frequency), setting (for example, community and healthcare), delivery method (in-person versus online), facilitator contact, group and individual format characteristics, and reported outcomes. Extending the work of Burke et al.[22], groups were categorized into true face-to-face (n = 11), standard face-to-face (n = 45), true online (n = 3), standard online (n = 3), online community with interactive components (n = 8) and other (n = 1). All studies were screened by at least two reviewers, with disagreements resolved by consensus or consultation with a senior researcher (C.T.-N. or N.N.). We contacted the authors of five studies with missing or unclear data, sending two email reminders spaced 2 weeks apart. Two authors did not respond, and their studies were excluded owing to insufficient information.

Outcomes and grouping. Primary outcomes for our review included self-reported (for example, frequency, duration and type of PA) or device-based measures of PA (for example, step counts and accelerometer data). Secondary outcomes were grouped into three distinct categories: (1) psychosocial outcomes, including measures of psychological well-being (for example, life satisfaction and mental health) and social interaction (for example, social support and group cohesion); (2) health outcomes, referring to physical indicators such as VO_2max, body mass index and blood pressure; and (3) functional outcomes, such as strength, balance and mobility. Details of these groupings and the specific variables included in each category are provided on the project's OSF page[38].

Moderators. We pre-specified and a priori registered the following moderators for later meta-regression: group type, leader type, clinical status, age, gender, session frequency, delivery format, duration and study quality.

Quality and risk of bias of individual studies

Methodological and reporting quality. The quality of the included studies was assessed by three independent reviewers (M.K., H.R. and C.T.-N.) using the 14-item Standard Quality Assessment Criteria for Primary Research Papers (QualSyst[55]). QualSyst provides a standardized checklist and manual specifically for evaluating the methodological rigour and reporting quality of a wide range of study designs based on items tapping study size, outcome appropriateness, control of confounding factors and clarity in reporting of results. Scores for this tool range between zero and one with a higher score indicating higher study quality.

Certainty of evidence and risk of bias. The risk of bias of the whole body of literature was independently assessed by two reviewers (M.K. and H.R.) using the GRADE tool[56], which evaluates factors such as study limitations, consistency of results, imprecision, indirectness and publication bias. The final grading was reviewed by the other team members. Certainty of evidence was evaluated separately for experimental and observational studies (Supplementary Table 1), and independently for PA, functional, psychosocial and health outcomes. The evidence was downgraded, where needed, based on factors such as serious study limitations, inconsistency, indirectness, imprecision and a high risk of publication bias.

Data analyses

Calculation of effect sizes. We calculated standardized mean differences to have a comparable index of effect size for outcomes (for example, PA) measured using different metrics, scales or tools. We used Hedges' g as the effect size metric to account for the relative size of each sample[57]. Effect sizes were calculated from means, standard

deviations and sample sizes of experimental or intervention groups at post-intervention, using existing formulas[58]. The values of 0.2, 0.5 and 0.8 or greater were deemed as small, medium and large, respectively[59]. In addition, many of the included studies reported proportional metrics of PA adherence and some non-experimental studies reported correlational effect sizes. We calculated effect sizes for these proportional measures as odds ratios, which were converted to *g* (ref. 58), an approach that was also used for correlational effect sizes. We combined converted effects with effect sizes derived from mean differences. Our primary analysis used all available effect sizes (for example, from observational/cross-sectional and intervention studies) and combined effect sizes derived from mean differences and proportional data. Data analysis scripts, extracted datasets and additional methodological materials are available via the OSF[38].

Statistical synthesis of effect sizes. The inclusion of multiple effect sizes from a single study, as was the case in our review, violates the assumption of independence in effect sizes in traditional meta-analyses. Consequently, we used a 3-level random effects model to decompose sampling variance of the observed effect sizes (level 1), and variance within studies (level 2) and between studies (level 3), thus accounting for interdependences in effect sizes from the same study[60]. We conducted all statistical analyses metafor, metaSEM and metaviz packages[61] in R (version 4.3.1)[62], following guidelines[63]. In addition, because non-significant effects obtained from the random effects model cannot be used to infer evidence for the null hypothesis (that is, non-significant effects do not necessarily infer an absence of a difference), we used a TOST procedure for equivalence testing[37] to formally test for the absence of a meaningful difference between individual and group conditions, when the main effect sizes were non-significant. To provide a brief primer on the interpretation of this procedure, a significant TOST test indicates that the two conditions produce statistically equivalent effect sizes, and we can thus accept the null hypothesis that there is no difference between the two. A non-significant TOST test indicates that there is insufficient precision to conclude either for or against the null hypothesis. We specified equivalence bounds (the upper and lower limits of effect sizes that we would consider meaningful) at $d = \pm 0.19$, corresponding to the 25th percentile of effect sizes 6 months post PA intervention, based on a relevant meta-analysis[64]. We used the metaviz[65] that produces 'sunset' funnel plots to show information on the statistical power of each individual study included in the synthesis.

Sensitivity analyses. In addition to our primary analyses, we conducted several sensitivity analyses for primary and secondary outcomes to assess the robustness of findings and examine potential sources of heterogeneity[63]. First, we assessed the influence of outlying effect sizes on the model. Outliers were defined as effects with standardized residuals exceeding three standard deviations from the model estimate or Cook's distances larger than three times the mean. We conducted these analyses to assess the impact of statistically extreme effect sizes that may disproportionately influence pooled estimates or produced non-normal effect size distributions. Second, we examined effects of potential moderators by including the following variables as covariates in a meta-regression: group type (the style of group setting used, such as online or standard face-to-face groups), group leader (the individual leading the group), health status (whether participants were healthy or had a specific medical condition), age (average age of the sample), gender (the proportion of female identifying participants in the sample), sessions per week (the frequency that groups or individuals exercised at), intervention duration (the number of weeks that the intervention ran for) and study quality (the study's methodological and reporting quality rating). This moderator analysis allowed us to compare effect sizes across different moderator categories/quantities.

We made an a priori decision to include both experimental and observational studies to ensure comprehensive coverage of the literature and to reflect the full range of evidence on PA interventions across real-world contexts. While experimental designs typically provide greater evidentiary strength, observational studies can enhance ecological validity[45]. To assess whether their inclusion biased results, we conducted a secondary analysis restricted to experimental and quasi-experimental studies, adjusted for pre-intervention PA levels. Results were consistent with the main analysis, indicating minimal impact on overall findings.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset generated and analysed during this meta-analysis and additional materials are publicly available via the Open Science Framework (OSF) at https://doi.org/10.17605/OSF.IO/XT2G4 (ref. 38).

Code availability

All custom code used for data processing and statistical analyses in this meta-analysis is publicly available via the Open Science Framework (OSF) at https://doi.org/10.17605/OSF.IO/XT2G4 (ref. 38).

References

1. Kariuki, J. K. et al. Impact of lifestyle-based interventions on absolute cardiovascular disease risk: a systematic review and meta-analysis. *JBI Evid. Synth.* **22**, 4–65 (2024).
2. Noetel, M. et al. Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials. *Br. Med. J.* **384**, e075847 (2024).
3. Posadzki, P. et al. Exercise/physical activity and health outcomes: an overview of Cochrane systematic reviews. *BMC Public Health* **20**, 1724 (2020).
4. Bull, F. C. et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br. J. Sports Med.* **54**, 1451–1462 (2020).
5. Garcia-Hermoso, A. et al. Adherence to aerobic and muscle-strengthening activities guidelines: a systematic review and meta-analysis of 3.3 million participants across 32 countries. *Br. J. Sports Med.* **57**, 225–229 (2023).
6. Nikitara, K. et al. Prevalence and correlates of physical inactivity in adults across 28 European countries. *Eur. J. Public Health* **31**, 840–845 (2021).
7. Strain, T. et al. National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5·7 million participants. *Lancet Glob. Health* **12**, e1232–e1243 (2024).
8. Murray, J. M. et al. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: a systematic review and meta-analysis. *Soc. Sci. Med.* **192**, 125–133 (2017).
9. Olson, J. L. et al. Challenges to engaging older adults in a group-based walking intervention: lessons from the residents in action trial. *J. Aging Phys. Act.* **30**, 788–798 (2022).
10. Pang, B. et al. Barriers and facilitators to participation in vigorous lifestyle physical activity in adults aged 55-75 years: a scoping review. *J. Aging Phys. Act.* **32**, 446–459 (2024).
11. Prochnow, T. & Patterson, M. S. Assessing social network influences on adult physical activity using social network analysis: a systematic review. *Am. J. Health Promot.* **36**, 537–558 (2022).
12. Carron, A. V. & Brawley, L. R. in *Advances in Sport Psychology* (ed. Horn, T. S.) 213–237 (Human Kinetics, 2008).
13. Eys, M., Bruner, M. W. & Martin, L. J. The dynamic group environment in sport and exercise. *Psychol. Sport Exerc.* **42**, 40–47 (2019).

14. Abrams, D. & Hogg, M. A. *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes* (Routledge, 2006).
15. Stevens, M. et al. A social identity approach to understanding and promoting physical activity. *Sports Med.* **47**, 1911–1918 (2017).
16. Ryan, R. M. & Deci, E. L. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness* (The Guilford Press, 2017).
17. Ntoumanis, N. et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. *Health Psychol. Rev.* **15**, 214–244 (2021).
18. Bandura, A. Social cognitive theory of self-regulation. *Organ. Behav. Hum. Decis. Process.* **50**, 248–287 (1991).
19. Beauchamp, M. R., Crawford, K. L. & Jackson, B. Social cognitive theory and physical activity: mechanisms of behavior change, critique, and legacy. *Psychol. Sport Exerc.* **42**, 110–117 (2019).
20. Cheval, B. et al. Relationships between changes in self-reported physical activity, sedentary behaviour and health during the coronavirus (COVID-19) pandemic in France and Switzerland. *J. Sports Sci.* **39**, 699–704 (2021).
21. Sivaramakrishnan, H. et al. Psychosocial outcomes of sport participation for middle-aged and older adults: a systematic review and meta-analysis. *Int. Rev. Sport Exerc. Psychol.* **17**, 398–419 (2024).
22. Burke, S. M., Carron, A. V., Eys, M. A., Ntoumanis, N. & Estabrooks, P. A. Group versus individual approach? A meta-analysis of the effectiveness of interventions to promote physical activity. *Sport Exerc. Psychol. Rev.* **2**, 13–29 (2006).
23. Carron, A. V. & Hausenblas, H. A. *Group Dynamics in Sport* (Fitness Information Technology, 1998).
24. Carron, A. V. & Spink, K. S. Team building in an exercise setting. *Sport Psychol.* **7**, 8–18 (1993).
25. Davies, C. A., Spence, J. C., Vandelanotte, C., Caperchione, C. M. & Mummery, W. K. Meta-analysis of internet-delivered interventions to increase physical activity levels. *Int. J. Behav. Nutr. Phys. Act.* **9**, 52 (2012).
26. Lim, S. E. R. et al. Volunteer-led online group exercise for community-dwelling older people: a feasibility and acceptability study. *BMC Geriatr.* **23**, 461 (2023).
27. Mocco, A., Valmaggia, L., Bernardi, L., Alfieri, M. & Tarricone, I. Enhancing physical activity with immersive virtual reality: a systematic review. *Cyberpsychol. Behav. Soc. Netw.* **27**, 303–317 (2024).
28. Hajder, Đ. et al. A systematic review and meta-analysis of virtual and traditional physical activity programs: effects on physical, health, and cognitive outcomes. *Healthcare* **13**, 711 (2025).
29. Southcott, E. & Jooste, J. Unveiling the impact of mobile fitness applications on motivational orientation in sustaining exercise behaviors: a qualitative investigation. *Phys. Cult. Sport Stud. Res.* **103**, 1–14 (2023).
30. Burchartz, A. et al. Assessing physical behavior through accelerometry—state of the science, best practices and future directions. *Psychol. Sport Exerc.* **49**, 101703 (2020).
31. Carr, R. M. et al. Dyadic interventions to promote physical activity and reduce sedentary behaviour: systematic review and meta-analysis. *Health Psychol. Rev.* **13**, 91–109 (2019).
32. John, J. C. et al. Dyad and group-based interventions in physical activity, diet, and weight loss: a systematic review of the evidence. *J. Behav. Med.* **47**, 355–373 (2024).
33. Harrer, M., Cuijpers, P., Furukawa, T. & Ebert, D. D. dmetar: companion R package for the guide 'Doing Meta-Analysis in R'. http://dmetar.protectlab.org/ (2019).
34. Hedges, L. V. & Pigott, T. D. The power of statistical tests in meta-analysis. *Psychol. Methods* **6**, 203–217 (2001).
35. Egger, M., Smith, G. D., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.* **315**, 629–634 (1997).
36. Brunner, J. & Schimmack, U. Estimating population mean power under conditions of heterogeneity and selection for significance. *Meta-Psychology* **4** (2020).
37. Lakens, D. Equivalence tests: a practical primer for *t* tests, correlations, and meta-analyses. *Soc. Psychol. Pers. Sci.* **8**, 355–362 (2017).
38. Kritz, M. et al. Data and code for: A systematic review and meta-analysis comparing individual vs. group-based approaches on physical activity, functional, psychosocial, and health outcomes. *Open Science Framework* https://doi.org/10.17605/OSF.IO/XT2G4 (2026).
39. Harden, S. M., Burke, S. M., Haile, A. M. & Estabrooks, P. A. Generalizing the findings from group dynamics-based physical activity research to practice settings: what do we know? *Eval. Health Prof.* **38**, 3–14 (2015).
40. Leach, H. J., Mama, S. K. & Harden, S. M. Group-based exercise interventions for increasing physical activity in cancer survivors: a systematic review of face-to-face randomized and non-randomized trials. *Support. Care Cancer* **27**, 1601–1612 (2019).
41. Higgings, J. et al. *Cochrane Handbook for Systematic Reviews of Interventions* (John Wiley & Sons, 2024).
42. Barrett, E. et al. Exploring patient experiences of participating in a group exercise class for the management of nonspecific shoulder pain. *Physiother. Theory Pract.* **34**, 464–471 (2018).
43. Thomas, E. M. et al. Association of physical function with hospital readmissions among older adults: a systematic review. *J. Hosp. Med.* **20**, 277–287 (2025).
44. Cooney, E., Toomey, E., Ryan, K., Meade, O. & McSharry, J. What's different about digital? A qualitative interview study exploring experiences of adapting in-person behaviour change interventions for digital delivery. *Br. J. Health Psychol.* **29**, 972–999 (2024).
45. Shrier, I. et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am. J. Epidemiol.* **166**, 1203–1209 (2007).
46. Agius, G., Calleja, P. & Muscat-Inglott, M. Examining individual versus group exercise preference among Maltese gym-goers. *J. Sports Phys. Educ. Stud.* **4**, 07–16 (2024).
47. Meira, C. M. et al. Personality traits and preferred exercise environment of fitness club members. *Acta Gymnica* **50**, 172–179 (2021).
48. Middelkamp, J., van Rooijen, M., Wolfhagen, P. & Steenbergen, B. The effects of two self-regulation interventions to increase self-efficacy and group exercise behavior in fitness clubs. *J. Sports Sci. Med.* **15**, 358–364 (2016).
49. Curran, G. M. et al. Reflections on 10 years of effectiveness-implementation hybrid studies. *Front. Health Serv.* **2**, 1053496 (2022).
50. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br. Med. J.* **372**, n71 (2021).
51. Kritz, M. et al. Comparing the effectiveness of individual and group approaches for promoting physical activity, physical health, and psychosocial wellbeing. *PROSPERO* https://www.crd.york.ac.uk/PROSPERO/view/CRD42021271452 (2024).
52. Harari, M. B., Parola, H. R., Hartwell, C. J. & Riegelman, A. Literature searches in systematic reviews and meta-analyses: a review, evaluation, and recommendations. *J. Vocat. Behav.* **118**, 103377 (2020).
53. Chai, K. E. K., Lines, R. L. J., Gucciardi, D. F. & Ng, L. Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews. *Syst. Rev.* **10**, 93 (2021).

54. Campos, D. G. et al. Screening smarter, not harder: a comparative analysis of machine learning screening algorithms and heuristic stopping criteria for systematic reviews in educational research. *Educ. Psychol. Rev.* **36**, 19 (2024).
55. Kmet, L. M., Lee, R. C. & Cook, L. S. *Standard Quality Assessment Criteria for Evaluating Primary Research Papers From a Variety of Fields* (Alberta Heritage Foundation for Medical Research, 2004).
56. Guyatt, G. H. et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Br. Med. J.* **336**, 924–926 (2008).
57. Lakens, D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Front. Psychol.* **4**, 863 (2013).
58. Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. *An Introduction to Meta-Analysis* (Wiley, 2009).
59. Cohen, J. A power primer. *Psychol. Bull.* **112**, 155–159 (1992).
60. Cheung, M. W.-L. A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychol. Rev.* **29**, 387–396 (2019).
61. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
62. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
63. Gucciardi, D. F., Lines, R. L. J. & Ntoumanis, N. Handling effect size dependency in meta-analysis. *Int. Rev. Sport Exerc. Psychol.* **15**, 152–178 (2022).
64. Wright, C. E., Rhodes, R. E., Ruggiero, E. W. & Sheeran, P. Benchmarking the effectiveness of interventions to promote physical activity: a metasynthesis. *Health Psychol.* **40**, 811–821 (2021).
65. Kossmeier, M., Tran, U. S. & Voracek, M. metaviz: forest plots, funnel plots, and visual funnel plot inference for meta-analysis. *Github* https://github.com/mkossmeier/metaviz (2020).

Acknowledgements
The authors received no specific funding for this work.

Author contributions
Conceptualization: C.T.-N., N.N., M.K., S.M.B. and S.M.H. conceived the meta-analysis, including the research question and objectives. Methodology: the search strategy, literature search and data extraction protocols were led by M.K. with input from N.N. and C.T.-N. Abstract screening was conducted by M.K., D.O. and H.R., with C.T.-N. and N.N. serving as third reviewers. Full-text screening was conducted by M.K., H.R. and D.O., with input from C.T.-N. and N.N. Data coding and extraction were performed by M.K., H.R. and D.O., with guidance from N.N. and C.T.-N. Statistical analysis was led by H.R., with input from N.N., C.T.-N. and M.K. Formal analysis and investigation: H.R. conducted the statistical analysis, while N.N., C.T.-N. and M.K. contributed to the interpretation and presentation of findings. Writing (original draft preparation): the initial paper draft was prepared by N.N., C.T.-N., H.R. and M.K. Writing (review and editing): all authors (N.N., C.T.-N., M.K., H.R., D.O., S.M.H. and S.M.B.) critically reviewed and revised the paper.

Funding
Open access funding provided by Danube University Krems University for Continuing Education.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-026-02429-0.

Correspondence and requests for materials should be addressed to Marlene Kritz.

Peer review information *Nature Human Behaviour* thanks Fiona Gillison, Jongho Moon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

[1]Department for Economy and Health, University for Continuing Education, Krems, Austria. [2]Curtin School of Population Health, Faculty of Health Sciences, Curtin University, Perth, Western Australia, Australia. [3]Department of Human Nutrition, Foods, and Exercise, Virginia Tech, Blacksburg, VA, USA. [4]School of Health Studies, Faculty of Health Sciences, Western University, London, Ontario, Canada. [5]Children's Health Research Institute, London, Ontario, Canada. [6]Danish Centre for Motivation and Behaviour Science, University of Southern Denmark, Odense, Denmark. [7]School of Sport, Exercise, and Rehabilitation Sciences, University of Birmingham, Birmingham, UK. ✉e-mail: marlene.kritz@donau-uni.ac.at

nature portfolio

Corresponding author(s): Marlene Kritz

Last updated by author(s): 29.1.2026 by Marlene Kritz

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	An automated research screening tool (Research Screener) was used to assist with title and abstract screening, alongside manual screening and full-text assessment by the research team. Excel was used to assist with data extraction.
Data analysis	All analyses were conducted in R (version 4.3.1) using established meta-analytic packages (metafor, metaSEM, and metaviz packages). Scripts/code used for the analysis are available on the Open Science Framework at https://doi.org/10.17605/OSF.IO/XT2G4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The dataset generated and analyzed during this meta-analysis and additional materials, are publicly available via the Open Science Framework (OSF) at https://doi.org/10.17605/OSF.IO/XT2G4

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This meta-analysis synthesizes data from previously published studies. Where reported by the original authors, sex and gender were extracted and considered as potential moderators. However, not all included studies disaggregated results by sex or gender. No new participant data were collected
Reporting on race, ethnicity, or other socially relevant groupings	Race, ethnicity, or other socially constructed groupings were not consistently reported across included studies. Where relevant and available, this information was extracted, but no new data were collected for this meta-analysis.
Population characteristics	Population characteristics (e.g., clinical vs. non-clinical, age group, delivery mode) were extracted based on the information reported in the original studies and analyzed as potential moderators.
Recruitment	No participants were recruited for this study. All data were extracted from previously published research. Recruitment procedures were those of the original studies and are reported therein.
Ethics oversight	As this is a meta-analysis of published data, no new ethical approval was required.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This was a quantitative meta-analysis of published studies comparing individual versus group-based physical activity interventions. A three-level random-effects model was used.
Research sample	The sample consisted of 71 published studies reporting on adult populations from various countries. Sample characteristics (e.g., age, sex, delivery mode) were extracted where available and analyzed as moderators. No new data were collected.
Sampling strategy	Studies were selected through systematic searches in peer-reviewed databases using predefined inclusion criteria. No additional sampling was conducted. A sensitivity analysis was conducted to address potential bias.
Data collection	Research Screener was used for abstract screening. All data were extracted manually from published studies and recorded in structured Excel spreadsheets. Data were manually extracted by two independent reviewers using structured coding sheets. Discrepancies were resolved through discussion. Study characteristics and outcome variables to calculate effect sizes were coded.
Timing	The final literature search was completed on 19 March 2024.
Data exclusions	Outlier studies were excluded in sensitivity analyses based on pre-specified criteria (e.g., extreme effect sizes with large standard errors). All decisions were transparently reported.
Non-participation	Not applicable. No new participants were recruited; all data were drawn from published sources.
Randomization	Not applicable. This was a meta-analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Not applicable
Novel plant genotypes	Not applicable
Authentication	Not applicable