The effect of post-exercise heat exposure (passive heat acclimation) on endurance exercise performance: a systematic review and meta-analysis.

Thomas P.J. Solomon[1] and Matthew J. Laye[2].

[1] Blazon Scientific, London, UK.
[2] Idaho College of Osteopathic Medicine, Meridian, ID, USA.

Corresponding author: Matthew J Laye (mlaye@icom.edu). Idaho College of Osteopathic Medicine, 1401 E. Central Dr, Meridian, ID 83642, USA.

ORCID: Thomas PJ Solomon 0000-0002-0579-284X.
Matthew J Laye 0000-0003-3214-3113.

Keywords: Heat acclimation; Heat exposure; Hot water immersion; Sauna; Steam bath; Exercise; Exercise performance; Endurance performance; Athlete; Sport.

Suggested Citation:
Solomon TPJ, Laye MJ (2024). The effect of post-exercise heat exposure (passive heat acclimation) on endurance exercise performance: a systematic review and meta-analysis. *SportRχiv*.

Statements and Declarations:

Data availability statement. Data can be downloaded from the Open Science Foundation registry at https://doi.org/10.17605/OSF.IO/6FGC2 [1].

Acknowledgements. The authors thank Janice Thompson (Professor of Public Health Nutrition, University of Birmingham; J.Thompson.1@bham.ac.uk) for peer-reviewing the search strategy.

Funding. No funding was received for this work.

Author contributions. TS and ML contributed equally to conceptualising the idea, the study design, methodology, data extraction, data analysis, data presentation, and the writing and editing of the manuscript. Both authors approved the final version before submission. TS curated the data and takes responsibility for its integrity.

Competing Interests. TS has given invited talks at societal conferences and university/pharmaceutical symposia for which the organisers paid for travel and accommodation; he has also received research money from publicly funded national research councils and medical charities, and private companies, including Novo Nordisk Foundation, AstraZeneca, Amylin, AP Møller Foundation, and Augustinus Foundation; and, he has consulted for Boost Treadmills, GU Energy, and Examine.com, and owns a consulting business, Blazon Scientific, and an endurance athlete education business, Veohtu. These companies have had no control over the research design, data analysis, or publication outcomes of this work. ML has given invited talks at societal conferences and university symposia and meetings for which the organisers paid for travel and accommodation; he has received research money from Augustinus Foundation, American College of Sports Medicine, and national research institutions; and, he has consulted for Zepp Health, Levels Health, GU Energy, and EAB labs, and has coached for Sharman Ultra Coaching. These companies have had no control over the research design, data analysis, or publication outcomes of this work.

Abstract

Background

"Active" heat acclimation (exercise-in-the-heat) can improve exercise performance in the heat. An alternative strategy is "passive" heat acclimation using post-exercise heat exposure, but its efficacy is unclear.

Objectives

To synthesise a systematic review and meta-analysis that answers the question: Does post-exercise heat exposure improve endurance exercise performance?

Methods

MEDLINE, CENTRAL, ClinicalTrials.gov, WHO ICTRP, and EU CTR were searched from inception to February 2024 to identity studies meeting the following inclusion criteria: (i) healthy male or female healthy adults; (ii) an exercise training intervention with post-exercise heat exposure (*treatment* group); (iii) a non-heat exposure *control* group completing the same training; and (iv) outcomes measuring the primary outcome (endurance exercise performance in the heat) or any of the secondary outcomes (performance in cool conditions, VO_2max, lactate threshold, economy, heart rate, RPE, core temperature, sweat rate, and/or thermal sensations during exercise. Study quality was assessed using the Cochrane Risk of Bias 2 tool. To determine the effect of post-exercise heat exposure, between-group ratio of means or standardized mean differences (SMD) were calculated for each outcome and weighted by the inverse of their variance to calculate an overall effect estimate (ratio of mean or Hedges' g) in a random effects meta-analysis, with 95% confidence intervals (CI) and prediction intervals (PI). The quality of evidence was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool.

Results

Ten studies (k=10) including 199 participants (156 males and 43 females, aged 20–32 years) met the inclusion criteria. Interventions lasted 5–21 days with 40–105 minutes/day of exercise and 14–40 minutes/session of post-exercise heat exposure (sauna or hot water immersion). The effect of post-exercise heat exposure on the primary outcome — performance in hot conditions — was trivial with poor precision (ratio of means = 1.04, 95%CI 0.94 to 1.15, P=0.46; k=4, n=60) and had low predictive certainty (95%PI 0.81 to 1.33). There was also a trivial effect on performance in cool conditions and speed at lactate threshold, and a small effect on VO_2max, along with a trivial effect on RPE, small effects on heart rate, core temperature, and sweat rate, and a moderate effect on thermal sensations. However, the certainty of evidence was graded as low to very low across all outcomes due to small sample sizes, high risk of bias, risk of publication bias, imprecision in the effect estimates, and low statistical power.

Conclusions

Given the predominance of low to very low certainty of evidence, the effect of post-exercise heat exposure as a method of heat acclimation to improve endurance exercise performance is uncertain. Further high-quality trials are needed to make conclusions concerning its efficacy.

Protocol registration

OSF https://doi.org/10.17605/OSF.IO/256XZ (June 23, 2022) [2].

Introduction

The current evidence shows that "active" heat acclimation (exercise-in-the-heat) achieved with daily exercise sessions performed in a 30–40°C heat chamber can improve endurance exercise performance in hot conditions [3–9]. However, training in the heat can be unpleasant, increases the risk of heat illness, and requires a reduction in external load (e.g. power output or speed) during sessions. From a practical perspective, "passive" heat acclimation with daily post-exercise heat exposure is favourable because it does not impede an athlete's typical training: desired during-session power output or speed can be maintained. Furthermore, due to logistical or financial constraints, athletes may lack access to the necessary resources — a bicycle or treadmill in a heat chamber — for active heat acclimation, whereas they will likely have access to a bathtub or sauna with which to undertake passive heat acclimation. Consequently, post-exercise heat exposure may be a more practical method of heat acclimation [10,11].

Several existing systematic reviews on heat acclimation and exercise performance have pooled the outcomes from active and passive heat acclimation interventions [3–6,8,9,12], and no systematic review of passive heat acclimation currently exists. Therefore, the specific effect of post-exercise heat exposure on exercise performance is unclear. Consequently, this systematic review aims to determine whether passive heat acclimation using post-exercise heat exposure via sauna bathing or hot water immersion improves exercise performance in healthy adults. To obtain relevant information, endurance exercise performance outcomes from time-to-completion tests (time trials or races) and time-to-exhaustion tests in hot conditions (≥25°C) were chosen as the primary outcome. Secondary outcomes included endurance exercise performance tests in cool conditions (<25°C); as well as VO_2max, economy/efficiency, and lactate threshold; and physiological measurements (heart rate, RPE, core temperature, sweat rate, and thermal sensations) during submaximal exercise in hot or cool conditions.

Methods

Research question.

The purpose of this paper is to answer the question: *Does post-exercise heat exposure improve endurance exercise performance?* To answer this question, a systematic review was synthesised in line with the Cochrane Handbook for Systematic Reviews of Interventions [13]. **Table 1** shows the structure of the research question — the Population, Interventions, Comparisons, and Outcomes (PICO).

Search strategy.

The search strategy was planned in January 2022 and independently peer-reviewed on February 11 2022 by Professor Janice Thompson (University of Birmingham, UK) using the 2015 PRESS (Peer Review of Electronic Search Strategies) Guideline Statement [14] (**Supplemental File 1**). The following databases were searched: MEDLINE, CENTRAL, and clinical trials databases for unpublished data (ClinicalTrials.gov, WHO International Clinical Trials Registry Platform [WHO ICTRP], and EU Clinical Trials Register [EUCTR]). The following MEDLINE search string was used and adapted for other databases (**Supplemental File 2**):

> (heat exposure[Title/Abstract] OR post exercise sauna[Title/Abstract] OR post exercise hot water bath[Title/Abstract] OR heat acclimation OR heat acclimatization OR heat acclimat* OR post exercise bath OR heat adaptation OR heat adapt* OR "hot water immersion"[Title/Abstract] OR Sauna OR

"heat chamber") AND (time trial[Title/Abstract] OR performance[Title/Abstract] OR race[Title/Abstract] OR time to fatigue[Title/Abstract] OR athletic performance[MeSH Terms] OR VO2max OR "aerobic capacity" OR exercise[title/abstract]) AND humans

Note that “heat acclimatisation” describes the adaptations to heat exposure gained naturally through exposure to living in hot conditions whereas “heat acclimation” describes the adaptations gained from purposeful exposure to artificial conditions. These definitions are used throughout this paper, but the two phrases are used interchangeably in the literature. Therefore, both phrases were included in the search strategy.

The initial search was conducted on April 7, 2022. The study protocol was then registered on Open Science Foundation (OSF) before data extraction began [2]. Due to delays in responses to data requests, subsequent searches were completed on Feb 8, 2023 and Feb 7 2024. Other amendments to the original protocol were documented on the OSF registry [2].

Study selection.

The search hits were downloaded into Endnote and the review authors TS and ML independently screened titles and abstracts, coding each study as “include” (eligible or potentially eligible/unclear) or “exclude”. Duplicates were identified and excluded. Studies were only selected if they included: (i) healthy male or female adults (≥18 years of age) of any race/ethnicity, (ii) an exercise training intervention (≥ 30 mins/day) with post-exercise heat exposure for at least 2 consecutive days, (iii) a non-heat-exposure control group completing the same daily exercise, and (iv) outcomes measuring endurance exercise performance during time-to-completion tests (time trials or races) or time-to-exhaustion tests, or exercise capacity (VO_2max, lactate threshold, economy, etc), or heart rate, RPE, core temperature, sweat rate, and/or thermal sensations during exercise. Studies were excluded if they were written in a non-English language, published in non-peer-reviewed journals, or if the full text was unavailable.

Reference lists of included articles were also screened using the same inclusion criteria. Full-text versions of the included articles were retrieved and independently screened by TS and ML to identify the final list of studies meeting the inclusion criteria. Reasons for exclusion were recorded and any disagreements were resolved through discussion or, when required, via the independent search strategy peer-reviewer, JT. The study selection process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [15] and the PRISMA checklist for systematic review protocols (PRISMA-P) was completed (**Supplemental File 3)**.

Data extraction.

TS and ML independently extracted article info (author, year, journal, title), intervention details (duration; heat type, temperature, and duration; exercise type, duration, and intensity), sample sizes (N), mean and standard deviation (SD) values for subject characteristics (age, sex, BMI), and outcome variables of interest. If studies reported standard error (SE), it was converted to SD by multiplying it by √N. When full data were not reported, they were requested from the authors. If authors were unreachable or refused to share data, data were estimated from figures where possible using WebPlotDigitizer software [16]. When data were unobtainable, the study was still described in the qualitative summary but excluded from the meta-analysis. Independent data extraction from TS and ML was cross-checked and disagreements were resolved through discussion. To ensure data accuracy, TS & ML compared the magnitude and direction of effects reported in the included studies with how they appear in the study database.

Risk of bias analysis.

TS and ML independently assessed the quality of the included studies using the Cochrane Risk of Bias 2 (RoB2) Excel tool [17], which assesses bias in five domains: Bias arising from the randomization process (including allocation concealment and blinding of participants and investigators); deviations from intended interventions; missing outcome data; the measurement of the outcome; and the selection of the reported result. The risk of bias for each domain was rated as "high", "some concerns", or "low". A separate risk of bias analysis was conducted for each outcome of interest: exercise performance (time-to-completion and time-to-exhaustion tests, VO_2max, economy, lactate threshold), heart rate, RPE, core temperature, sweat rate, and thermal sensations during exercise. TS and ML's independent RoB2 results were cross-checked, and disagreements were resolved through discussion. The overall risk of bias for each study was determined according to Cochrane guidelines [17]: "Low" if the trial was judged to be at low risk of bias for all domains; "Some concerns" if the trial was judged to raise some concerns in at least one domain but not to be at high risk of bias for any domain; or "High" if the trial was judged to be at high risk of bias in at least one domain or if the trial was judged to have "some concerns" for multiple domains. Risk of bias figures were built using the *robvis* R Shiny app [18].

Data synthesis and statistical analysis.

A qualitative synthesis of the included studies was written, and an overall qualitative assessment was made based on the between-study heterogeneity in settings, participants, sample sizes, study designs, interventions, outcomes, risk of bias, funding sources, and conflicts of interest. Next, a quantitative synthesis was completed to determine the effect of the interventions. Any changes to the planned comparisons or outcomes or any new comparisons that became necessary were documented on the OSF protocol registration record [2].

Measures of treatment effects — meta-analysis.

Different performance tests use different units of measurement — watts, kcals, seconds, metres, etc — and an improvement in exercise performance can be represented by an *increase* in work done (kilojoules) or mean power output (watts), an *increase* in time-to-exhaustion, or a *decrease* in time to complete a specific distance. Therefore, intervention-induced changes for performance outcomes were calculated as a *ratio of means* (post divided by pre) [19,20]. With most types of performance test, a "post ÷ pre" ratio of means greater than 1 indicates an improvement in performance. However, with some performance tests (e.g. time to completion tests), a "post ÷ pre" ratio of means that is greater than 1 inadvertently reflects a decrease in performance. To prevent this issue, an inverse transformation was applied to "post/pre" ratio of means values for time-to-completion tests (i.e. they were raised to the power of -1). Then, the between-group difference was calculated as a *between-group ratio of means*: $(\text{Post}_{\text{Treatment}} \div \text{Pre}_{\text{Treatment}}) \div (\text{Post}_{\text{Control}} \div \text{Pre}_{\text{Control}})$ [19,20] and the standard deviation of the ratio of means was calculated as standard error (SE) × √n, where $SE = \sqrt{(1 / n_{\text{Treatment}} \times (SD_{\text{Treatment}} / \text{mean}_{\text{Treatment}})^2 + 1 / n_{\text{Control}} \times (SD_{\text{Control}} / \text{mean}_{\text{Control}})^2)}$ [19–21]. Because the sampling distribution of a ratio of means is not symmetrical, individual trial ratio of mean estimates must be pooled on a log scale [19–21]. Therefore, the ratio of means values were natural-log-transformed before meta-analytical calculations were made and summary effect estimates were then inverse-log-transformed (e^x) back to ratio of mean (ROM) values, which were interpreted as: trivial (ROM<1.08), small (1.08≤ROM<1.22), moderate (1.22≤ROM<1.37), or large effects (ROM≥1.37) of treatment [19,20]. Ratio of means meta-analyses were completed with RevMan v7.5.0 [22].

For the remaining outcomes, the between-group standardised mean difference (SMD) in each study was calculated as: $((Post_{Treatment} - Pre_{Treatment}) - (Post_{Control} - Pre_{Control})) \div SD_{pooledbaseline}$ [23]. The pooled baseline SD of the treatment and control groups was used rather than the SD of the change score to prevent the error of measurement from introducing bias into the estimate of the SMD. Between-group SMDs were weighted according to the relative sample sizes (N) in the treatment and control groups and reported as Hedges' g effect size estimates, which were interpreted as: trivial (g<0.2), small (0.2≤g<0.5), moderate (0.5≤g<0.8), or large effects (g≥0.8) of treatment. SMD meta-analyses were completed with Meta-Essentials v1.5 [24]

For all outcomes, a combined summary effect estimate was calculated using a DerSimonian-Laird random effects model to account for between-study heterogeneity by weighting each study's effect estimate by its inverse variance (i.e., $1/SE^2$) [24]. Z-tests were used to test for the overall effect and statistical significance was reported when $P \leq 0.05$. The precision of the effect size estimates — SE and 95% confidence intervals (95%CI) — were also calculated.

Statistical heterogeneity.

To examine statistical heterogeneity (inconsistency of results), the Q statistic and its corresponding $\chi 2$ test P-value were calculated to test whether effect sizes departed from homogeneity [25]. I^2 was also calculated to report the proportion of dispersion due to heterogeneity, and was interpreted as follows: 0% to 40% might not be important; 30% to 60% represents moderate heterogeneity; 50% to 90% represents substantial heterogeneity; 75% to 100% represents considerable heterogeneity [25]. τ^2 and τ were calculated to quantify the amount of between-study heterogeneity in the true effects [25]. Prediction intervals, which are derived from τ, were also calculated to visualise the range in which, in 95% of the cases, the outcome of a future study will fall — i.e. the expected range of effects of future studies [26–28]. If the effect sizes of the included and not yet included studies are normally distributed, a large prediction interval indicates low certainty in the effect size estimate.

Statistical power in the meta-analysis.

The *metameta* R package was used to calculate statistical power — the probability of detecting a true effect — for each outcome variable using the effect sizes and SEs measured in individual studies [29].

Assessment of reporting bias.

To help identify the risk of publication bias and small study effects, the funnel plot relationship between the effect estimate (either ratio of means or standardised mean difference) and the SE of the effect estimate for each outcome variable was visually inspected [30]. Contour-enhanced funnel plots were constructed to illustrate conventional levels of statistical significance (e.g., P<0.01, P<0.05, P<0.10) [31] and Egger's test was used to identify funnel plot asymmetry [32]. If statistically significant asymmetry (P≤0.05) was detected, Duval and Tweedie's trim and fill correction was used to simulate a model without publication bias and, therefore, estimate the number of missing studies and the intervention effect adjusted for publication bias [33].

Subgroup analysis.

Subgroup analyses were planned for *training status* (trained athletes vs. non-athletes/untrained people) and *temperature dependency* (performance in hot [≥25°C] vs. cool [<25°C] conditions), and $\chi 2$ tests were planned to test for subgroup interactions [34].

Sensitivity analysis.

For each outcome variable, sensitivity analyses were planned if studies with a high risk of bias were identified and to explore between-study heterogeneity. A sensitivity analysis consisted of repeating the meta-analysis with the high-risk of bias studies removed and, if substantial statistical heterogeneity was identified (I^2>50%), comparing the outcome of fixed-effect and random-effects analyses [35].

Certainty of evidence.

The certainty of evidence was assessed according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidelines [36], which assesses certainty using five domains: risk of bias, inconsistency, indirectness, imprecision and publication bias. Note that imprecision was indicated if the 95%CI of the effect size overlapped zero or if the sample size in the meta-analysis was less than the optimal information size (OIS) criterion. The OIS was calculated for each variable using an online sample size calculator [37] recommended by GRADE, using alpha=0.05 and beta=0.20.

Results

Search results.

The study selection process is presented in **Fig. 1**. The database searches yielded 2516 records. After removal of duplicates, 2458 records were screened and 2444 were excluded. Fourteen records were obtained for full-text review, of which 10 met the inclusion criteria [38–47]. Reasons for exclusion included one or more of the following criteria: unrelated topic, no control group, non-English-language, non-healthy participants, less than 30-mins of exercise with less than 2 consecutive days of post-exercise heat exposure, and no outcomes related to the primary and/or secondary outcomes stated in the PICO. The characteristics of excluded studies are available in **Supplemental Files 4a to 4d**.

Qualitative synthesis of the included studies.

The characteristics of the included studies are described in **Table 2**. The following is a qualitative summary of all evidence:

Settings.

Of the 10 included studies, 8 were conducted in university research labs [38,40,42–47] and 2 were conducted in a Sports Institute research lab [39,41]. Five trials were conducted in the UK [40,44–47], 2 in Australia [39,41], and 1 each in New Zealand [38], Norway [42], and Spain [43]. All studies conducted the heat exposure interventions indoors, while 5 studies conducted the exercise training intervention outdoors [38,41,43,46,47] and 5 were indoors [39,40,42,44,45].

Participants and sample sizes.

The 10 studies included a total of 199 participants enrolled in interventions relevant to this review. Participants included 156 males and 43 females aged 20±2 to 32±4 years. Study sample sizes ranged from N=6 to N=41. Three of the studies included recreationally active participants [40,44,45] while the remaining 7 studies included trained athletes: runners [46,47], runners and triathletes [38], race walkers [41], cyclists [39,42], and soccer players [43]. Only 3 studies included female participants [41,46,47]; these same 3 studies also included male participants, while the other 7 studies included only males [38–40,42–45].

Study designs.

All 10 studies included a post-exercise heat exposure *treatment* group and a non-post-exercise heat exposure *control* group. Note that 2 studies [39,42] included additional intervention groups (see **Table 2**) not relevant to this review and are, therefore, not discussed. Of the included studies, two [38,39] used a crossover design and all but one [38] included a repeated-measures design. Six of the studies randomised participants to the intervention groups while two [46,47] allowed participants to self-select into a treatment group, 1 study [41] divided participants into two groups to match VO_2max and 10,000-m performance time between groups, and 2 studies [38,42] did not describe whether participants were randomised.

Interventions.

The intervention durations ranged from 5 to 21 days. The exercise training interventions included running [38,44–47], race walking [41], cycling [39,42], and soccer [43] with exercise performed daily in 6 studies [39–42,44,45]. It is possible that 3 further studies (7.7±2.3 times per week [38] and 7±2 times per week [46,47]) also used daily exercise but it is not explicitly stated. The remaining study [43] did not explicitly state the exercise frequency, simply that participants engaged in at least 10 hours per week. Exercise sessions lasted between 40 and 105 minutes. A steady-state moderate-intensity (65% of VO_2peak [40,44,45] or 45% to 51% of power at 4 mmol lactate [42]) was used in 4 studies while daily high-intensity intervals were completed in 1 study [39] and the remaining studies used variable intensities [38,41,43,46,47]. The during-exercise conditions were cool (18–19°C, 35–45% relative humidity [RH]) in 4 studies [40,42,44,45], hot (≈33°C, ≈34% RH) in 1 study [41], and not reported in 5 studies [38,39,43,46,47]. The type of post-exercise heat exposure was a sauna in 4 studies [38,43,46,47] and hot water immersion in 6 studies [39–42,44,45]. Heat exposure was delivered for 14–40 mins per session. The hot water immersion studies used a water temperature of ≈40°C, while the sauna studies used air temperatures between 90 and 108°C with relative humidities between 5 and 55%. Five studies described fluid intake during post-exercise heat exposure: 3 [38,46,47] reported that *ab libitum* water intake was permitted, while 1 described that there was no fluid intake [40], and 1 further study [43] described that fluid intake was limited to 250 ml, but none of the studies reported hydration status. Post-exercise heat exposure sessions were daily and followed every exercise training session in 5 studies [39,40,42,44,45]. Three studies [43,46,47] used post-exercise heat exposure on 3 days per week for a total of 9 sessions, while 2 studies did not explicitly state the frequency of heat exposure (12.7±2.1 sessions in 21 days [38], and 8 sessions in 14 days [41]).

Outcomes.

Primary outcomes: Four studies [40,42,44,45] reported endurance exercise performance (time trial, time-to-exhaustion, or a race) in hot conditions (≥25°C).
Secondary outcomes: Five studies [40,41,43,46,47] reported performance (time trial, time-to-exhaustion, or a race) in cool conditions (<25°C). Two further studies [38,39] did not report the temperature during the performance tests but, given the experimental design, they were likely conducted in cool conditions. Three studies reported VO_2max in cool conditions [43,46,47], while one study [41] did not report the conditions for VO_2max. Two studies reported running economy in cool conditions [46,47]. Two studies reported speed at lactate threshold in cool conditions [46,47] while one study [41] did not report the ambient temperature in which lactate threshold was assessed. Heart rate during submaximal exercise was reported by 7 studies [40–42,44–47] in hot conditions and 1 study [40] in cool conditions, while 1 study [39] did not report the conditions. RPE during submaximal exercise was reported by 7 studies [40–42,44–47] in hot conditions and 1 study [40] in cool conditions, while 1 study [39] did not report the conditions. Core temperature during submaximal exercise was reported by 7 studies [40–42,44–47]

in hot conditions and 1 study [40] in cool conditions, while 1 study [39] did not report the conditions. Seven studies [40–42,44–47] reported sweat rate during submaximal exercise in hot conditions and 1 study [40] in cool conditions. Thermal sensation during submaximal exercise was reported by 6 studies [40,41,44–47] in hot conditions and 1 study [40] in cool conditions. No studies compared any of the outcome variables in trained athletes vs. non-athletes.

Risk of bias.

The risk of bias analyses for the exercise performance tests (time trials, races, time-to-exhaustion, VO_2max, and speed lactate threshold) and physiological measures (heart rate, RPE, core temperature, sweat rate, and thermal ratings) are presented in **Fig. 2**. Bias in the different domains is summarised as follows:

- *Bias arising from the randomisation process*: For all outcomes, no study had a low risk of bias, while 2 studies [46,47] had a high risk of bias and the remaining studies [38–45] had some concerns. The most common reasons for a high risk of bias or some concerns were that: the randomisation, allocation, and blinding processes were not described [38–45], participants self-selected their intervention group [46,47], or participants were assigned to groups to match fitness between groups [41]. It is important to note that while it's impossible to blind a participant from exercise, sauna, or hot water immersion, it would have been possible for investigators and/or data analysts to blind themselves to the group allocation.
- *Bias due to deviations from the intended intervention*: For the exercise performance outcomes and most of the physiological outcomes, 4 studies [40,42,44,45] had a low risk of bias, 4 studies [38,39,41,43] had some concerns, and 2 studies [46,47] had a high risk. For RPE during exercise, one additional study [45] had a high risk of bias. The most common reasons for a high risk of bias or some concerns were: not having a pre-registered trial protocol to define the intended intervention.
- *Bias due to missing outcome data*: For the exercise performance outcomes, 6 studies [38,39,41–44] had a low risk of bias, 2 studies [40,46] had some concerns, and 2 studies [45,47] had a high risk. For the physiological outcomes, most studies were either low risk of bias or had some concerns, except for 1 study [47] that had a risk of bias due to missing data for heart rate. The most common reasons for a high risk of bias or some concerns were: not describing how missing data were handled and excluding data from participants who dropped out during the intervention.
- *Bias in the measurement of the outcomes*: For the exercise performance outcomes, 3 studies [40,41,44] had a low risk of bias, 3 studies [39,42,43] had some concerns, and 4 studies [38,45–47] had a high risk. For the secondary outcomes, 2 studies [40,42] had a high risk of bias for the measurement of heart rate, 1 study [43] had a high risk of bias for the measurement of core temperature, 2 studies [40,45] had a high risk of bias for the measurement of thermal ratings, and 1 study [46] had a high risk of bias for the measurement of sweat rate. All other ratings for the secondary outcomes were either low risk of bias or some concerns. The most common reasons for a high risk of bias or some concerns were: providing insufficient details of how the outcome was measured, and not describing the method and/or stating the manufacturer and model of methodological equipment.
- *Bias in the selection of the reported results*: For the exercise performance outcomes, no study had a low risk of bias, while 7 studies [38–44] had some concerns and 3 studies [45–47] had a high risk of bias. For the secondary outcomes, most studies had a high risk of bias, none had a low risk, and the remaining studies [41,44] had some concerns. The most common reasons for a high risk of bias or some concerns were: not having a pre-registered trial protocol to define the planned reported results, and not defining how the time point of a continuous variable was selected (e.g. whether a heart rate, RPE, core temperature, etc, was an average or peak value achieved during exercise).

To summarise the risk of bias, 4 studies [38,45–47] had a high risk of bias for the exercise performance outcomes while the remaining 6 studies [39–44] had some concerns. Meanwhile, 5 studies [40,42,43,46,47] had high risk and 4 studies [39,41,44,45] had some concerns for heart rate; 3 studies [45–47] had high risk and 5 studies [39–42,44] had some concerns for RPE; 6 studies [39,42,43,45–47] had high risk and 3 studies [40,41,44] had some concerns for core temperature; 5 studies [40,42,45–47] had high risk and 2 studies [41,44] had some concerns for sweat rate; and, 4 studies [40,45–47] had high risk and 2 studies [41,44] had some concerns for thermal sensations. No study was rated with an overall low risk of bias.

Funding sources.

Of the 10 included studies, one [43] was funded by the Regional Government of Extremadura, Spain, with a contribution from the European Union; one [42] received funding from the Norwegian Olympic and Paralympic Committee and Confederation of Sports; two [46,47] were funded by internal university funds with additional support from the Canadian Centennial Scholarship Fund; and, two [44,45] received funding from the Ministry of Defence. The remaining four studies [38–41] did not report funding sources.

Conflicts of interest.

Six studies [40,42,44–47] reported that no conflicts of interest existed, while four [38,39,41,43] did not provide this information.

Studies without usable data.

Of the 10 included studies, none of the studies presented all the necessary data to complete a meta-analysis. The authors of 6 studies shared their data upon request [40,41,44–47] but the authors of 2 studies were unreachable [39,43] and the authors of 2 studies refused to share necessary data [38,42]. Fortunately, data from 4 of these studies [38,39,42,43] could be calculated (e.g. post-minus-pre changes, and SD from CI [48]) or extracted from figures [42,43] using WebPlotDigitizer software [16]. Additionally, the two studies from Kirby *et al*. [46,47] re-used several participants' data, but the authors agreed to share their raw data so it could be analysed as a single dataset and thus reduce meta-bias due to data duplication.

Quantitative synthesis — the effect of interventions.

Before analysis, the following data manipulations were necessary: (i) Stevens *et al*. [41] reported the 10 km time trial finish time in mm:ss format, and Scoon *et al*. [38] and Bartolomé *et al*. [43] reported time-to-exhaustion in minutes; these were all converted to seconds. (ii) Because the studies from Kirby *et al.* [46,47] reported total sweat loss (litres) during a 30-minute heat tolerance test, the values were multiplied by 2 to derive a *litre per hour* sweat rate to enable a fair comparison with other studies. (iii) Similarly, Lundby *et al*. [42] reported total sweat loss in litres during a 50-minute exercise test consisting of 15 minutes of submaximal riding in the heat, 5 minutes of riding in ambient conditions, and a 30-minute time trial in the heat. Therefore, the reported sweat loss was divided by 50/60 to derive a sweat rate in *litres per hour*.

Effect of post-exercise heat exposure on performance in hot conditions.

A random effects model showed a trivial effect with poor precision on endurance performance in hot conditions (**Fig. 3**: ratio of means=1.04, 95%CI 0.94 to 1.15, P=0.46; k=4, n=60). Heterogeneity of results across studies was negligible and might not be important (Q=1.22, P=0.75; I^2=0%), but there was low predictive certainty (prediction interval [PI] 0.81 to 1.33) regarding the magnitude of future effects. However, it is important to highlight that the performance outcomes are from highly diverse performance test types (**Table 2**). Egger's regression did not indicate a high risk of publication bias (t=2.27, P=0.15; **Fig. 12A**) but the very small number of studies (*k*=4) makes publication bias likely and decreases the certainty in the validity of the funnel plot. A sensitivity analysis removing the study by McIntyre *et al*. 2022 did not change the effect estimate or its precision (ratio of means=1.03, P=0.53; 95%CI 0.93 to 1.15).

Effect of post-exercise heat exposure on performance in cool conditions.

A random effects model showed a trivial effect with satisfactory precision on endurance performance in cool conditions (**Fig. 4**: ratio of means=1.06, 95%CI 0.99 to 1.12, P=0.08; *k*=6, n=144). There was substantial heterogeneity of results across studies (Q=13.75, P=0.02; I^2=64%), and the predictive certainty about future effects was low (PI 0.85 to 1.33). Furthermore, the performance outcomes are from highly diverse performance test types (**Table 2**). Egger's regression did not indicate a high risk of publication bias (t=1.30, P=0.26; **Fig. 12B**) but the small number of studies does not rule it out and decreases the certainty in the validity of the funnel plot. A sensitivity analysis removing the high risk of bias study by Kirby *et al.* had little influence on the effect estimate (ratio of means=1.05, P=0.18; 95%CI 0.98 to 1.13) and removing the high risk of bias study by Scoon *et al.* further reduced the effect estimate (ratio of means=1.02, P=0.17; 95%CI 0.99 to 1.06). The complex study design and analysis in Vaile *et al*. (see **Table 2**) forced the selection of a single performance outcome (total work done in 9 minutes) among several reported; removing this study slightly increased the effect estimate but worsened its precision (ratio of means=1.07, P=0.09; 95%CI 0.99 to 1.15). Removing all three studies further reduced the trivial effect size (ratio of means=1.01, P=0.50; 95%CI 0.98 to 1.05). Because substantial statistical heterogeneity was identified (I^2>50%), a fixed-effects analysis was also completed but a trivial effect size persisted (ratio of means=1.04, 95%CI 1.00 to 1.07).

Effect of post-exercise heat exposure on VO_2max.

A random effects model showed a small effect for increasing VO_2max (**Fig. 5**: Hedges' g=0.33, 95%CI 0.19 to 0.47, P<0.0001, PI 0.24 to 0.42; *k*=3, n=91) with negligible heterogeneity across studies (Q=0.05, P=0.98; I^2=0.0%). Egger's regression did not indicate a high risk of publication bias (t=-0.26, P=0.84; **Fig. 12C**) but the very small number of studies makes publication bias highly likely and negates the validity of the funnel plot. A sensitivity analysis to remove the high risk of bias study by Kirby *et al* had little impact on the effect size estimate but reduced its precision (g=0.36, P<0.0001; 95%CI -0.16 to 0.88).

Effect of post-exercise heat exposure on running economy.

Two papers [46,47] reported that the between-trial (heat acclimation vs. control) difference in the intervention-induced change in running economy was not statistically significant. A meta-analysis was not possible because these 2 papers are different analyses from the same experimental study and duplicate several participants' data.

Effect of post-exercise heat exposure on speed at lactate threshold.

A random effects model showed a trivial effect on speed at lactate threshold (**Fig. 6**: g=0.19, 95%CI -0.53 to 0.92, P=0.0007, PI 0.05 to 0.33; *k*=2, n=55) with negligible heterogeneity across studies (Q=0.04, P=0.83; I^2=0%). There was an insufficient number of studies to objectively test for publication bias (**Fig. 12D**); the very small number of studies negates the validity of the funnel plot and makes publication bias highly likely. Sensitivity analyses were not possible due to an insufficient number of studies.

Effect of post-exercise heat exposure on heart rate during submaximal exercise.

A random effects model showed a small effect with high precision for reducing heart rate (**Fig. 7**: g=-0.32, 95%CI -0.45 to -0.20, P<0.0001; *k*=6, n=163), with high predictive certainty (PI -0.64 to -0.01) and negligible heterogeneity across studies (Q=0.83, P=1.00; I^2=0%). Egger's regression did not indicate a high risk of publication bias (t=0.64, P=0.55; **Fig. 12E**) but the small number of included studies does not rule it out and decreases the certainty in the validity of the funnel plot. Sensitivity analyses removing individual high risk of bias studies had little impact on the effect size estimate or its precision and predictive certainty (excluding Kirby *et al*: g=-0.32, P<0.0001; 95%CI -0.47 to -0.16; excluding Zurawlew *et al*: g=-0.29, P<0.0001; 95%CI -0.44 to -0.13; excluding Lundby *et al:* g=-0.32, P<0.0001; 95%CI -0.50 to -0.14). However, removing all high risk of bias studies reduced the effect size and worsened its precision (g=-0.20, P=0.13; 95%CI -0.75 to 0.36).

Effect of post-exercise heat exposure on RPE during submaximal exercise.

A random effects model showed a trivial effect with poor precision on RPE (**Fig. 8**: g=-0.07, 95%CI -0.65 to 0.51, P=0.77; *k*=6, n=163), with substantial heterogeneity across studies (Q=16.5, P=0.02; I^2=57.6%) and low predictive certainty (PI -1.36 to 1.22) concerning where future effects will be found. Egger's regression did not indicate a high risk of publication bias (t=-0.79, P=0.46; **Fig. 12F**) but the small number of included studies does not rule it out and decreases the certainty in the validity of the funnel plot. Sensitivity analyses removing individual or all high risk of bias studies had little impact on the original effect estimate or its precision (excluding Kirby *et al:* g=-0.02, P=0.94; 95%CI -0.72 to 0.68; excluding McIntyre *et al*. 2022: g=0.06, P=0.80; 95%CI -0.50 to 0.62; excluding both: g=0.16, P=0.55; 95%CI -0.51 to 0.83). Because substantial statistical heterogeneity was identified (I^2>50%), a fixed-effects analysis was also completed but a trivial effect with poor precision persisted (g=-0.05, 95%CI -0.42 to 0.32).

Effect of post-exercise heat exposure on core temperature during submaximal exercise.

A random effects model showed a small effect with high precision for preventing the rise in core temperature (**Fig. 9**: g=-0.44, 95%CI -0.79 to -0.09, P=0.003; *k*=6, n=163). There was negligible heterogeneity in results across studies (Q=6.41, P=0.49; I^2=0%), but low predictive certainty (PI -1.32 to 0.44). Egger's regression did not indicate a high risk of publication bias (t=0.81, P=0.45; **Fig. 12G**) but the small number of included studies does not rule it out and decreases the certainty in the validity of the funnel plot. Sensitivity analyses removing individual high risk of bias studies had little impact on the original effect estimate or its precision (excluding Lundby *et al:* g=-0.53, P=0.008; 95%CI -1.04 to -0.01; excluding Kirby *et al:* g=-0.35, P=0.04; 95%CI -0.76 to 0.06; excluding McIntyre *et al*. 2022: g=-0.41, P=0.01; 95%CI -0.82 to -0.01). The small effect size persisted when all high risk of bias studies were removed, but its precision was worsened (g=-0.36, P=0.26; 95%CI -1.40 to 0.67). Please note that core temperature data in Vaile *et al*. was not included in the meta-analysis because exact baseline data could not be retrieved for the control and treatment groups. Although their paper states that *"average pre-exercise rectal temperature regardless of intervention group or day of exercise was 37.3 ± 0.28"*, the authors could not be reached to resolve this matter.

Effect of post-exercise heat exposure on sweat rate during submaximal exercise.

A random effects model showed a small effect with high precision for increasing sweat rate (**Fig. 10**: g=0.27, 95%CI 0.06 to 0.47, P=0.002; *k*=6, n=139). There was negligible heterogeneity across studies (Q=1.55, P=0.96; I^2=0%), but low predictive certainty in future effects (PI -0.21 to 0.74). Egger's regression did not indicate a high risk of publication bias (t=0.43, P=0.69; **Fig. 12H**) but the small number of included studies does not rule it out and decreases the certainty in the validity of the funnel plot. Sensitivity analyses removing individual high risk of bias studies had little impact on the original effect estimate or its precision (excluding Lundby *et al:* g=0.28, P=0.006; 95%CI 0.02 to 0.54; excluding Kirby *et al:* g=0.25, P=0.02; 95%CI -0.03 to 0.52; excluding McIntyre *et al*. 2022: g=0.28, P=0.004; 95%CI 0.03 to 0.53; excluding Zurawlew *et al*: g=0.35, P<0.0001; 95%CI 0.12 to 0.59). Removing all high risk of bias studies left only 2 studies in the analysis: the strength of the effect estimate increased but its precision worsened (g=0.62, P<0.0001; 95%CI -0.23 to 1.47).

Effect of post-exercise heat exposure on thermal sensations during submaximal exercise. A random effects model showed a moderate-sized effect with high precision for reducing thermal sensations (**Fig. 11**: g=-0.56, 95%CI -0.94 to -0.17, P=0.0002; *k*=5, n=112). Heterogeneity across studies was negligible (Q=3.13, P=0.68; I^2=0%), but there is low predictive certainty about where future effects will lie (PI -1.23 to 0.12). Egger's regression did not indicate a high risk of publication bias (t=1.75, P=0.15; **Fig. 12I**) but the small number of included studies does not rule it out and decreases the certainty in the validity of the funnel plot. Sensitivity analyses removing individual high risk of bias studies had little impact on the original effect estimate or its precision (excluding Kirby *et al:* g=-0.40, P=0.007; 95%CI -0.81 to 0.02; excluding McIntyre *et al*. 2022: g=-0.50, P=0.002; 95%CI -0.95 to -0.06; excluding Zurawlew *et al*: g=-0.62, P=0.005; 95%CI -1.31 to 0.08). Removing all high risk of bias studies left just 2 studies in the analysis; this reduced the strength of the estimate to a trivial effect with poor precision (g=-0.12, P=0.30; 95%CI -1.60 to 1.36).

Sub-group analysis.

Due to insufficient data, it was not possible to complete the planned sub-group analyses for *training status* (trained athletes vs. non-athletes/untrained people) or *temperature dependency* (performance in hot vs. cool conditions). No other sub-group analyses were planned (e.g. *sex* [male vs. female] or *heat exposure type* [sauna vs. hot water immersion], etc.) and, given the small number of studies, these comparisons are not currently appropriate to add.

Statistical power in the meta-analysis.

The statistical power was very low for each variable (**Figs. 3 to 11**). For example, there was only 5% power to detect a meaningful effect of post-exercise heat exposure on the primary outcome (**Fig. 3**) and between 5 and 21% power to detect meaningful effects across the secondary outcomes (**Figs. 4 to 11**). This suggests that the included body of studies cannot reliably detect true effect sizes of interest and, therefore, may not be very informative.

Certainty of evidence (GRADE).

The certainty of evidence for the primary outcome — exercise performance in hot conditions — was graded as *very low*. For the secondary outcomes, the certainty of evidence for exercise performance in cool conditions, VO_2max, and speed at lactate threshold were also graded as *very low*, and the physiological measurements during submaximal exercise were graded as either *low* (core temperature and thermal ratings) or *very low* (heart rate, RPE, and sweat rate). The GRADE summary of findings and reasons for downgrading the certainty of evidence are presented in **Table 3**.

Data availability

Full outcome variable data extracted from the individual studies (including data extracted using WebPlotDigitizer) are available from the Open Science Foundation data registry at [1].

Discussion

Summary of findings.

This systematic review included 10 studies (199 participants; 156 males and 43 females, aged 20±2 to 32±4 years) investigating the effects of heat acclimation via post-exercise heat exposure. For the primary outcome — endurance exercise performance in hot conditions (as measured by time-to-completion or time-to-exhaustion tests in ≥25°C) — there was a trivial effect of post-exercise heat exposure, with poor precision, low statistical power, and low predictive certainty (*k*=4, n=60: ratio of means=1.04, P=0.46; 95%CI 0.94 to 1.15; PI 0.81 to 1.33; power = 0.05). Similarly, there were trivial effects with poor precision, low power, and low predictive certainty on endurance exercise performance in cool conditions (*k*=6, n=144: ratio of means=1.06, P=0.08; 95%CI 0.99 to 1.12; PI 0.85 to 1.33; power = 0.06), speed at lactate threshold (*k*=2, n=55: g=0.19, P=0.0007; 95%CI -0.53 to 0.92; PI -0.53 to 0.92; power = 0.08), and RPE (k=6, n=163: g=-0.07, P=0.77; 95%CI -0.65 to 0.51, P=0.77; PI -1.44 to 1.29; power = 0.05). For the remaining secondary outcomes, post-exercise heat exposure had small to moderate beneficial effects with better precision, but statistical power and predictive certainty remained low: small effects on VO_2max (*k*=3, n=91: g=0.33, P<0.0001; 95%CI 0.19, to 0.47; PI 0.19, to 0.47; power = 0.15), heart rate (*k*=6, n=163: g=-0.32, P<0.0001; 95%CI -0.45 to -0.20; PI -0.45 to -0.20; power = 0.10), core temperature (*k*=6, n=163: g=-0.44, P=0.003; 95%CI -0.79 to -0.09; PI -0.79 to -0.09; power = 0.15), and sweat rate (*k*=6, n=139: g=0.27, P=0.002; 95%CI 0.06 to 0.47; PI 0.06 to 0.47; power = 0.09), and a moderate-sized effect on thermal sensations (*k*=5, n=112: g=-0.56, P=0.0002; 95%CI -0.94 to -0.17; PI -0.94 to -0.17; power = 0.21). Except for performance in cool conditions and RPE, which both had substantial heterogeneity (I^2 = 64% and 58%, respectively), effects were generally consistent across studies. However, prediction intervals revealed a wide range of possible effects in which future studies would fall.

Quality of the evidence

While there were generally no funding issues, 4/10 studies [38–41] didn't describe the funding sources and 4/10 [38,39,41,43] didn't report whether conflicts of interest existed, or not. Furthermore, all 10 included studies were rated with a *high risk of bias* or *some concerns* for all outcome variables. Bias arose from the allocation and randomisation process, the (participant and investigator/data analyst) blinding process, deviations from the intended interventions, missing outcome data, the measurement of the outcomes, and selective reporting of the outcomes. Across all outcomes, there was also generally poor precision in the effect estimates, low statistical power, and low predictive certainty. Subsequently, the GRADE certainty of evidence, which assesses quality across five domains (risk of bias, inconsistency, indirectness, imprecision and publication bias) and reflects the extent to which there is confidence in the effect estimate, was graded as *very low* for the primary outcome — exercise performance in hot conditions — and as *low* to *very low* for the secondary outcomes (**Table 3**). This means that the true effects may be substantially different from those measured in this analysis.

Sensitivity analyses to remove the high risk of bias studies revealed further information about the quality of the evidence. While sensitivity analyses didn't affect the effect estimates or their precision for either performance in hot conditions or RPE, sensitivity analyses had varying impacts on other secondary outcomes. Removing the high risk of bias studies further reduced the effect on performance in cool conditions; worsened the precision of the effect estimate for VO_2max, sweat rate, and core temperature; and reduced the effect size and its precision, and removed statistical significance for heart rate and thermal sensations. Therefore, effect estimates were sensitive to the high risk of bias studies across most outcomes.

Overall, the low quality of evidence prompts little confidence in the effect estimates. Consequently, firm conclusions cannot currently be made concerning the effect of post-exercise heat exposure interventions to improve endurance exercise performance. To remedy this, further high-quality randomised controlled trials are needed.

Generalisability of the findings

The findings might be generalisable to competitive endurance performance because 7/10 studies [38,39,41–43,46,47] included trained athletes — runners, triathletes, race walkers, cyclists, and soccer players. These 7 studies could also be considered to have a high level of ecological validity because the participants were endurance-trained athletes [38,39,41–43,46,47], athletes continued their normal training habits [38,41–43,46,47], athletes trained in their typical training environment [38,42,43,46,47], the performance tests were designed to simulate the demands of a race [39], or performance was assessed using a race against other athletes [41]. One example from Stevens *et al*. [41] examined world-class race walkers at a training camp and used a 10 km race against other athletes as the performance outcome — this is as ecologically valid as it gets. Meanwhile, the remaining 3/10 studies [40,44,45] have a low-to-moderate level of ecological validity because the participants are recreationally active and not endurance trained, and because the daily exercise sessions were identical (40–60 min/day at the same fixed intensity), which does not represent endurance athletes' habits.

The overall generalisability of the findings is doubtful because there were only 10 studies with small sample sizes (N = 6 to 41) and the studies were only conducted in Europe (Norway, Spain, UK), Australia, and New Zealand. There was also a predominance of young male participants (156 males and 43 females; age 20±2 to 32±4 years) — only 3/10 studies [41,46,47] included female participants — and no study described the race or ethnicity of the participants. Furthermore, only 5/10 studies [38,41,43,46,47] completed the exercise intervention outdoors in the participants' typical training environment, with the remaining 5 studies [39,40,42,44,45] completing the training interventions indoors in a laboratory setting. Additionally, the interventions and performance outcomes were highly heterogeneous across studies, making it difficult to accurately summarise the generalisability.

Strengths and weaknesses in the review process.

Limitations in the body of evidence:

During study selection, several studies were excluded because they lacked a non-heat-exposure control group (e.g. [49–53]) or did not measure performance via time-to-completion (time trials or races) or time-to-exhaustion tests (e.g. [49,52–57]). Of the included studies, the body of evidence is very small: there are few studies (k=10) with small sample sizes (N=6 to N=41 across studies, with 199 participants in total). Larger sample sizes would enable more reliable detection of a wider range of effect sizes. Because of the small sample size, the statistical power was low for every outcome across all studies, indicating that this meta-analysis contains a body of studies that cannot reliably detect effect sizes of interest and, therefore, may not be very informative. All studies were also rated as having a high risk of bias or some concerns across all outcomes; no study was rated with an overall low risk of bias for any outcome. In general, most studies do not describe the randomisation, allocation concealment, or blinding approaches, have poor descriptions of the methods used to measure outcomes, and do not adequately describe the selection of the outcome (e.g. it is often unclear whether a physiological measure during exercise is a mean or peak value or whether the value represents the entire exercise duration, a portion of it, or the endpoint). Specifically, while it is impossible to blind participants to interventions like exercise or heat exposure, it is possible to blind investigators and data analysts to group allocation; no study provided this info.

There is also a sex imbalance in this body of evidence: only 3/10 studies [41,46,47] included female participants and only 43/199 participants in the included studies were female. While a female-specific systematic review published in 2023 [8] concluded that heat acclimation can improve performance in females, it pooled all types of heat acclimation approaches: exercise-in-the-heat and post-exercise heat exposure interventions.

Limitations in this systematic review and meta-analysis:

The small number of included studies and the small sample size created several limitations. For example, it was not possible to conduct planned sub-group analyses to examine whether *training status* (trained athletes vs. non-athletes/untrained people) or *temperature dependency* (performance in hot vs. cool conditions) were influential. It was also not feasible to complete additional unplanned sub-group analyses (*sex*, male vs. female; *heat exposure type*, sauna vs. hot water immersion; etc). The small number of studies and low sample size also create substantial uncertainty in the value of I^2 and the validity of the funnel plots and their associated publication bias metrics, which should be interpreted with caution. This same reason combined with the large heterogeneity of study designs also creates uncertainty in the accuracy of the prediction intervals. There was also an issue with missing data, namely: some of the included papers lacked sufficient data to perform analyses and data requests were problematic: some authors were unreachable, and others refused to share data. This is unfortunate given the current need for open science and transparency. Missing data may influence the conclusions of the meta-analysis; however, this is unlikely given the low number of studies and small sample sizes.

A further limitation arises due to meta-bias. Firstly, one limitation of the meta-analysis is the analysis of study-level rather than individual subject-level data. Secondly, there was potential for meta-bias caused by participant data duplication in the studies by Kirby *et al*. [46,47]; however, their data is open access [58] and the authors kindly agreed to share their raw data to help avoid this issue. Thirdly, due to unobtainable data, the within-group changes for 3/10 studies [39,42,43] were estimated as the change in the post minus pre mean value rather than the mean of the individual participant post minus pre changes. Fourthly, due to unobtainable data, between-group differences and baseline SD values were extracted from figures in 2/10 studies [42,43] using WebPlotDigitizer [16]. Such data estimates reduce the accuracy of the meta-analytical calculations. And, lastly, in the study by Vaile *et al*. [39], participants completed identical training sessions on 5 consecutive days consisting of 105 minutes of cycling with 66 maximal sprints and 9 minutes of sustained time trial effort, with outcomes measured on each of the 5 days. The design and data analysis are complex with multiple time-points measured within multiple days and multiple “performance” outcomes reported; to answer the specific question being asked in the current review, the meta-analysis compared total work done during 9 minutes of time trial on the last (day 5) with measurements made on the first day (day 1). Selecting this specific outcome and these specific time points amongst all combinations introduces meta-bias; a sensitivity analysis showed that this study had a substantial impact on the precision of effect estimate.

Solutions to the limitations:

Following this systematic review, several concepts remain unclear. For example, the current body of evidence cannot conclude whether there is an optimal temperature, duration, or modality (e.g. sauna vs. hot water immersion) of post-exercise heat exposure or an optimal time course of delay between exercise and heat exposure. Furthermore, while most [59–63] but not all [64–66] studies show that hydration status doesn't influence adaptations to active heat acclimation (exercise-in-the-heat), it remains unclear whether hydration status influences the effect of post-exercise heat exposure. Accordingly, none of the studies included in this review measured hydration status and only 5/10 studies described fluid intake: three [38,46,47] reported that *ab libitum* water intake was permitted during post-exercise heat exposure, one [43] described that fluid intake was limited to 250 ml, and one [40] described that there was no fluid intake. Additionally, it is possible that natural heat acclimatisation — caused when athletes' habitual training is conducted outdoors in hot conditions — may mask the effect of post-exercise heat exposure interventions. For example, in the study by Stevens *et al.* [41], participants were likely already heat acclimatized because they trained in the summer heat for 4 weeks before the study and for 15 days during the study. This may explain the trivial effect of post-exercise heat exposure in that study (SMD = 0.15, 95%CI -0.97 to 1.31). The studies by Kirby *et al*. [46,47] attempted to minimise this issue by conducting interventions in the UK between October and March. This type of limitation should be considered in future studies.

To resolve these limitations, future studies should: (i) Pre-register their protocols with methods that clearly describe how outcomes will be compared. (ii) Use a randomised controlled design, ideally with crossover, with large sample sizes that are sufficiently powerful to detect meaningful differences. (iii) Fully describe randomisation, allocation concealment, and (participant and investigator) blinding procedures. (iv) Include female participants. (v) Include endurance-trained athletes, especially elite athletes, if possible. (vi) Measure hydration status and describe fluid intake. (vii) Determine the dose response of post-exercise heat exposure. (viii) Determine the optimal modality of post-exercise heat exposure. And (ix) Publish raw data in line with open science practices, which would improve efficiency in the synthesis of future systematic reviews and enable subject-level meta-analyses.

Strengths of this systematic review:

To minimize reporting bias and increase research transparency, the review protocol was registered on OSF before the literature search commenced and, in line with open science practices, all outcome data is freely-available from the OSF registry [2]. To broaden the coverage of the literature search, an independent scientist followed PRESS guidelines [14] to peer-review the search strategy before the literature search commenced. To obtain relevant information with high ecological validity, performance outcomes from time-to-completion tests (time trials or races) and time-to-exhaustion tests were chosen as the primary outcome, and the Cochrane Handbook [13] was used as a framework to use standardized and repeatable approaches to synthesise the data, assess the risk of bias [17], and GRADE the certainty of the evidence [36]. The authors (TS and ML) independently completed several aspects of the review (literature searching, screening, risk of bias analyses, data extraction) and met to discuss and reach agreements after each step of the process. To minimise the risk of garbage-in-garbage-out, an objective risk of bias analysis [17] was used to assess study quality, and sensitivity analyses were completed to determine the impact of high risk of bias studies on the effect estimates. Furthermore, in addition to traditional estimates of precision (SE and 95%CI), prediction intervals were also calculated to determine the expected range of effects of future studies [26–28].

Comparison to existing systematic reviews.

Several systematic reviews have examined the effect of heat acclimation on exercise performance and physiological measures [3–6,8,9,12]. However, such reviews have either studied the effect of "active" heat acclimation (exercise-in-the-heat) or pooled results from both "active" and "passive" heat acclimation (daily post-exercise heat exposure) protocols. Consequently, the current review is the first to examine "passive" heat acclimation in isolation. Nonetheless, it's important to compare the current findings with those from previous reviews:

Chalmers *et al*. (2014) did not perform a meta-analysis but concluded that heat acclimation generally improves aerobic, not anaerobic, performance [3]. However, the authors warned about the accuracy of their conclusions due to a moderate level of bias in the induced studies. Tyler *et al*. (2016) compared the effect of short, medium, and long-term heat acclimation protocols, finding a moderate to large beneficial effect of all protocols on exercise performance (ES = 0.52, 0.75, and 0.93), along with beneficial effects on heart rate (ES = -0.87), RPE (ES = -0.63), core temperature (ES = -0.51), sweat rate (ES = 0.61), and thermal sensation (ES = -0.68) [4]. However, there was a high risk of bias across several domains and the certainty of the evidence was not assessed. Benjamin *et al*. (2019) also found a positive effect of heat acclimation on performance (time-to-exhaustion: effect size, ES = 0.86); time-to-completion time trials; ES = 0.49) and a small effect on VO_2max (ES = 0.30); however, neither the risk of bias nor the certainty of evidence was assessed [5]. Rahimi *et al*. (2019) found a moderate beneficial effect of heat acclimation on time trial performance (ES = 0.50), a large beneficial effect on heart rate (ES = 1.0), but no benefit for VO_2max, RPE, core temperature, or thermal comfort [6]. However, the study quality assessment found a moderate level of bias in the included studies and the certainty of evidence was not assessed. Waldron *et al*. (2021) found small to moderate benefits on VO_2max in thermoneutral (ES = 0.42) and hot conditions (ES = 0.63), and although the authors concluded that the included studies had a generally low risk of bias [12], the risk of bias was high or unclear across several domains and the certainty of evidence was not assessed. Kelly *et al*. (2023), which studied exclusively female participants, found beneficial effects of heat acclimation on performance (pooled time-to-completion and time-to-exhaustion tests: ES = 1.00), heart rate (ES = -0.60), sweat rate (ES = 0.53), and core temperature (ES = -0.81) [8]. Again, although the authors concluded that included studies mostly had a low risk of bias, their data show a high risk of bias or some concerns across several domains and the certainty of evidence was not assessed. And, lastly, an updated 2024 meta-analysis from Tyler *et al*. confirmed their previous findings (moderate to large beneficial effects across outcomes); however, while the certainty of the evidence was not assessed, there was a high risk of bias across studies, considerable between-study heterogeneity, and wide prediction intervals [9].

In general, existing systematic reviews [4–6,8,9] find a beneficial effect of heat acclimation (active alone or pooled active and passive) on performance, but they also have a small sample size, a small number of included studies, and concern with the risk of bias. Plus, there is a lack of assessment for the certainty of evidence. This systematic review concludes that there are trivial effects of post-exercise heat exposure (passive heat acclimation) on exercise performance in hot conditions, performance in cool conditions, speed at lactate threshold, and RPE, small effects on VO_2max, heart rate, core temperature, and sweat rate, and a moderate-sized effect on thermal sensations. However, there is a low to very low certainty of evidence across all outcomes — this is the primary reason for the difference in the effect of heat acclimation on performance in hot conditions between this review and existing reviews. However, other reasons include: (i) the possibility that "active" heat acclimation is superior to "passive" heat acclimation, but this remains to be tested, and (ii) the dose of heat exposure in the current review is relatively short (mean of 287 minutes over 8.5 sessions) compared to most of the existing reviews (e.g., Chalmers *et al*.

reported a mean of 419 minutes over 5.8 sessions [3]), while the reviews by Benjamin *et al*., Waldron *et al*., and Kelly *et al*. found that the number of heat exposure days or total heat dose significantly influenced the performance effects [5,8,12].

While this is the first systematic review to examine the specific effect of post-exercise heat exposure, the practicalities of such an approach have been articulated elsewhere. For example, Heathcote and colleagues recommended 6–7 heat sessions on consecutive days for at least 30 minutes as soon as possible after exercise to improve performance but, in agreement with the current review, cautioned the need for more studies to fully understand the effect [11]. Casadio and colleagues agree that heat acclimation using post-exercise heat exposure could improve performance, but pose additional considerations [10]: namely, they ponder whether heat stress alters total training stress and the quality of athletes' subsequent sessions, and how much between-athlete variability in performance outcomes exists between active and passive heat acclimation approaches. Such questions must be answered by future studies.

Relevance of findings to coaching practice and athlete performance.

The current evidence suggests small to moderate-sized beneficial effects of post-exercise heat exposure on VO_2max and some physiological measures (heart rate, core temperature, sweat rate, and thermal sensations). Such effects have the potential to extrapolate to performance benefits because, in elite sports, even small-to-moderate effect sizes can translate to meaningfully large, perhaps unrealistic, improvements in endurance performance (e.g., 2–3 minutes in a marathon). Nonetheless, the current evidence shows only a trivial effect on performance, and the low to very low certainty of evidence prevents firm conclusions about the efficacy of this type of heat acclimation until further high-quality trials are published. That said, because post-exercise heat exposure doesn't harm endurance performance or related physiological variables, coaches and athletes *could* consider its use. However, it is important to consider whether the additional 30–40 minutes a day required for this strategy could be better used in other areas of training and recovery — training load optimization, sleep, nutrition, rest, etc. Furthermore, coaches and athletes should consider the practicality of different heat acclimation approaches. For example, if an athlete travels to a race or lives in an Olympic village before a race, post-exercise heat exposure in a hot bath or sauna (passive acclimation) allows an athlete to train and taper as usual without having to find exercise equipment located in a heat chamber (active acclimation).

Conclusions

The current evidence shows that heat acclimation using post-exercise heat exposure might improve physiological responses during submaximal exercise (increased sweat rate and decreased heart rate, core temperature, and thermal sensations). However, given the predominance of low to very low certainty evidence, the effect of this method of heat acclimation on endurance exercise performance is uncertain. Further high-quality trials are needed to bolster the evidence and to enable conclusions concerning the efficacy of post-exercise heat exposure for improving endurance exercise performance.

References

1. Solomon TP, Laye MJ. Open Science Framework - Data registry: A systematic review of the effect of passive heat acclimation on exercise performance. 2024; Available from: https://doi.org/10.17605/OSF.IO/6FGC2

2. Solomon TP. Open Science Framework - Registered protocol: A systematic review of the effect of passive heat acclimation on exercise performance. 2022; Available from: https://doi.org/10.17605/OSF.IO/256XZ

3. Chalmers S, Esterman A, Eston R, Bowering KJ, Norton K. Short-Term Heat Acclimation Training Improves Physical Performance: A Systematic Review, and Exploration of Physiological Adaptations and Application for Team Sports. Sports Med [Internet]. 2014 [cited 2023 Oct 4];44:971–88. Available from: https://doi.org/10.1007/s40279-014-0178-6

4. Tyler CJ, Reeve T, Hodges GJ, Cheung SS. The Effects of Heat Adaptation on Physiology, Perception and Exercise Performance in the Heat: A Meta-Analysis. Sports Med [Internet]. 2016 [cited 2023 Oct 4];46:1699–724. Available from: https://doi.org/10.1007/s40279-016-0538-5

5. Benjamin CL, Sekiguchi Y, Fry LA, Casa DJ. Performance Changes Following Heat Acclimation and the Factors That Influence These Changes: Meta-Analysis and Meta-Regression. Front Physiol [Internet]. 2019 [cited 2023 Oct 4];10:1448. Available from: https://doi.org/10.3389/fphys.2019.01448

6. Rahimi GRM, Albanaqi AL, Van der Touw T, Smart NA. Physiological Responses to Heat Acclimation: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. J Sports Sci Med [Internet]. 2019;18:316–26. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6543994/

7. Daanen HAM, Racinais S, Périard JD. Heat Acclimation Decay and Re-Induction: A Systematic Review and Meta-Analysis. Sports Med [Internet]. 2018 [cited 2023 Oct 4];48:409–30. Available from: https://doi.org/10.1007/s40279-017-0808-x

8. Kelly MK, Bowe SJ, Jardine WT, Condo D, Guy JH, Snow RJ, et al. Heat Adaptation for Females: A Systematic Review and Meta-Analysis of Physiological Adaptations and Exercise Performance in the Heat. Sports Med [Internet]. 2023 [cited 2023 Oct 4];53:1395–421. Available from: https://doi.org/10.1007/s40279-023-01831-2

9. Tyler CJ, Reeve T, Sieh N, Cheung SS. Effects of Heat Adaptation on Physiology, Perception, and Exercise Performance in the Heat: An Updated Meta-Analysis. J Sci Sport Exerc [Internet]. 2024; Available from: https://doi.org/10.1007/s42978-023-00263-8

10. Casadio JR, Kilding AE, Cotter JD, Laursen PB. From Lab to Real World: Heat Acclimation Considerations for Elite Athletes. Sports Med Auckl NZ [Internet]. 2017;47:1467–76. Available from: https://doi.org/10.1007/s40279-016-0668-9

11. Heathcote SL, Hassmén P, Zhou S, Stevens CJ. Passive Heating: Reviewing Practical Heat Acclimation Strategies for Endurance Athletes. Front Physiol [Internet]. 2018;9:1851. Available from: https://doi.org/10.3389/fphys.2018.01851

12. Waldron M, Fowler R, Heffernan S, Tallent J, Kilduff L, Jeffries O. Effects of Heat Acclimation and Acclimatisation on Maximal Aerobic Capacity Compared to Exercise Alone in Both Thermoneutral and Hot Environments: A Meta-Analysis and Meta-Regression. Sports Med Auckl NZ [Internet]. 2021;51:1509–25. Available from: https://doi.org/10.1007/s40279-021-01445-6

13. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [Internet]. Available from: https://training.cochrane.org/handbook/current

14. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol [Internet]. 2016;75:40–6. Available from: https://doi.org/10.1016/j.jclinepi.2016.01.021

15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ [Internet]. 2021 [cited 2023 Oct 4];n71. Available from: https://doi.org/10.1136/bmj.n71

16. Rohatgi A. Webplotdigitizer: Version 4.6 [Internet]. 2022. Available from: https://automeris.io/WebPlotDigitizer

17. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ [Internet]. 2019 [cited 2023 Oct 4];l4898. Available from: https://doi.org/10.1136/bmj.l4898

18. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods [Internet]. 2021 [cited 2023 Oct 4];12:55–61. Available from: https://doi.org/10.1002/jrsm.1411

19. Friedrich JO, Adhikari NK, Beyene J. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: A simulation study. BMC Med Res Methodol [Internet]. 2008 [cited 2024 Mar 6];8:32. Available from: https://doi.org/10.1186/1471-2288-8-32

20. Friedrich JO, Adhikari NKJ, Beyene J. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. J Clin Epidemiol [Internet]. 2011 [cited 2024 Mar 6];64:556–64. Available from: https://doi.org/10.1016/j.jclinepi.2010.09.016

21. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis—a tutorial and review of methods for enhancing interpretability. Res Synth Methods [Internet]. 2011 [cited 2024 Mar 6];2:188–203. Available from: https://doi.org/10.1002/jrsm.46

22. Review Manager (RevMan) version: 7.5.0. The Cochrane Collaboration (2024). Available at revman.cochrane.org.

23. Morris SB. Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. Organ Res Methods [Internet]. 2008 [cited 2023 Oct 4];11:364–86. Available from: https://doi.org/10.1177/1094428106291059

24. Suurmond R, Van Rhee H, Hak T. Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. Res Synth Methods [Internet]. 2017 [cited 2023 Oct 4];8:537–53. Available from: https://doi.org/10.1002/jrsm.1260

25. Deeks J, Higgins J, Altman D. Chapter 10.10 Heterogeneity. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [Internet]. Available from: https://training.cochrane.org/handbook/current/chapter-10#section-10-10

26. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc [Internet]. 2009;172:137–59. Available from: https://doi.org/10.1111/j.1467-985X.2008.00552.x

27. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ [Internet]. 2011 [cited 2023 Oct 4];342:d549–d549. Available from: https://doi.org/10.1136/bmj.d549

28. Deeks J, Higgins J, Altman D. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [Internet]. Available from: https://training.cochrane.org/handbook/current/chapter-10

29. Quintana DS. A Guide for Calculating Study-Level Statistical Power for Meta-Analyses. Adv Methods Pract Psychol Sci [Internet]. 2023 [cited 2023 Aug 30];6:25152459221147260. Available from: https://doi.org/10.1177/25152459221147260

30. Page M, Higgins J, Sterne J. Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [Internet]. Available from: https://training.cochrane.org/handbook/current/chapter-13

31. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol [Internet]. 2008 [cited 2023 Oct 5];61:991–6. Available from: https://doi.org/10.1016/j.jclinepi.2007.11.010

32. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ [Internet]. 1997;315:629–34. Available from: https://doi.org/10.1136/bmj.315.7109.629

33. Duval S, Tweedie R. A Nonparametric “Trim and Fill” Method of Accounting for Publication Bias in Meta-Analysis. J Am Stat Assoc [Internet]. 2000 [cited 2023 Oct 4];95:89–98. Available from: https://doi.org/10.1080/01621459.2000.10473905

34. Deeks J, Higgins J, Altman D. Chapter 10.11.3 Undertaking subgroup analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [Internet]. Available from: https://training.cochrane.org/handbook/current/chapter-10#section-10-11-3

35. Deeks J, Higgins J, Altman D. Chapter 10.14 Sensitivity analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [Internet]. Available from: https://training.cochrane.org/handbook/current/chapter-10#section-10-14

36. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. [Internet]. Updated October 2013. The GRADE Working Group.; 2013. Available from: https://gdt.gradepro.org/app/handbook/handbook.html

37. Brant R. Sample size calculator. Inference for Means: Comparing Two Independent Samples [Internet]. Department of Statistics, University of British Columbia; Available from: https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html

38. Scoon GSM, Hopkins WG, Mayhew S, Cotter JD. Effect of post-exercise sauna bathing on the endurance performance of competitive male runners. J Sci Med Sport [Internet]. 2007;10:259–62. Available from: https://doi.org/10.1016/j.jsams.2006.06.009

39. Vaile J, Halson S, Gill N, Dawson B. Effect of hydrotherapy on recovery from fatigue. Int J Sports Med [Internet]. 2008;29:539–44. Available from: https://doi.org/10.1055/s-2007-989267

40. Zurawlew MJ, Walsh NP, Fortes MB, Potter C. Post-exercise hot water immersion induces heat acclimation and improves endurance exercise performance in the heat. Scand J Med Sci Sports [Internet]. 2016;26:745–54. Available from: https://doi.org/10.1111/sms.12638

41. Stevens CJ, Ross MLR, Carr AJ, Vallance B, Best R, Urwin C, et al. Postexercise Hot-Water Immersion Does Not Further Enhance Heat Adaptation or Performance in Endurance Athletes Training in a Hot Environment. Int J Sports Physiol Perform [Internet]. 2021;16:480–8. Available from: https://doi.org/10.1123/ijspp.2020-0114

42. Lundby C, Svendsen IS, Urianstad T, Hansen J, Rønnestad BR. Training wearing thermal clothing and training in hot ambient conditions are equally effective methods of heat acclimation. J Sci Med Sport [Internet]. 2021;24:763–7. Available from: https://doi.org/10.1016/j.jsams.2021.06.005

43. Bartolomé I, Siquier-Coll J, Pérez-Quintero M, Robles-Gil MC, Grijota FJ, Muñoz D, et al. 3-Week passive acclimation to extreme environmental heat (100± 3 °C) in dry sauna increases physical and physiological performance among young semi-professional football players. J Therm Biol [Internet]. 2021;100:103048. Available from: https://doi.org/10.1016/j.jtherbio.2021.103048

44. McIntyre RD, Zurawlew MJ, Oliver SJ, Cox AT, Mee JA, Walsh NP. A comparison of heat acclimation by post-exercise hot water immersion and exercise in the heat. J Sci Med Sport [Internet]. 2021;24:729–34. Available from: https://doi.org/10.1016/j.jsams.2021.05.008

45. McIntyre RD, Zurawlew MJ, Mee JA, Walsh NP, Oliver SJ. A comparison of medium-term heat acclimation by post-exercise hot water immersion or exercise in the heat: adaptations, overreaching, and thyroid hormones. Am J Physiol Regul Integr Comp Physiol [Internet]. 2022;323:R601–15. Available from: https://doi.org/10.1152/ajpregu.00315.2021

46. Kirby NV, Lucas SJE, Cable TG, Armstrong OJ, Weaver SR, Lucas RAI. Sex differences in adaptation to intermittent post-exercise sauna bathing in trained middle-distance runners. Sports Med - Open [Internet]. 2021;7:51. Available from: https://doi.org/10.1186/s40798-021-00342-6

47. Kirby NV, Lucas SJE, Armstrong OJ, Weaver SR, Lucas RAI. Intermittent post-exercise sauna bathing improves markers of exercise capacity in hot and temperate conditions in trained middle-distance runners. Eur J Appl Physiol [Internet]. 2021;121:621–35. Available from: https://doi.org/10.1007/s00421-020-04541-z

48. Higgins J, Li T, Deeks J. Chapter 6.5.2.2: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [Internet]. Available from: https://training.cochrane.org/handbook/current/chapter-06#section-6-5-2-2

49. Kissling LS, Akerman AP, Campbell HA, Prout JR, Gibbons TD, Thomas KN, et al. A crossover control study of three methods of heat acclimation on the magnitude and kinetics of adaptation. Exp Physiol [Internet]. 2022;107:337–49. Available from: https://doi.org/10.1113/ep089993

50. Ashworth E, Cotter J, Kilding A. Post-exercise, passive heat acclimation with sauna or hot-water immersion provide comparable adaptations to performance in the heat in a military context. Ergonomics [Internet]. 2023;66:49–60. Available from: https://doi.org/10.1080/00140139.2022.2058096

51. Pokora I, Sadowska-Krępa E, Wolowski Ł, Wyderka P, Michnik A, Drzazga Z. The Effect of Medium-Term Sauna-Based Heat Acclimation (MPHA) on Thermophysiological and Plasma Volume Responses to Exercise Performed under Temperate Conditions in Elite Cross-Country Skiers. Int J Environ Res Public Health [Internet]. 2021;18:6906. Available from: https://doi.org/10.3390/ijerph18136906

52. Zurawlew MJ, Mee JA, Walsh NP. Post-exercise Hot Water Immersion Elicits Heat Acclimation Adaptations in Endurance Trained and Recreationally Active Individuals. Front Physiol [Internet]. 2018;9:1824. Available from: https://doi.org/10.3389/fphys.2018.01824

53. Zurawlew MJ, Mee JA, Walsh NP. Post-exercise Hot Water Immersion Elicits Heat Acclimation Adaptations That Are Retained for at Least Two Weeks. Front Physiol [Internet]. 2019;10:1080. Available from: https://doi.org/10.3389/fphys.2019.01080

54. Barry H, Gendron P, Gagnon C, Bherer L, Gagnon D. Passive heat acclimation does not modulate processing speed and executive functions during cognitive tasks performed at fixed levels of thermal strain. Appl Physiol Nutr Metab Physiol Appl Nutr Metab [Internet]. 2022;47:261–8. Available from: https://doi.org/10.1139/apnm-2021-0243

55. Ko Y, Kang J, Seol S-H, Lee J-Y. Effectiveness of skin-heating using a water-perfused suit as passive and post-exercise heat acclimation strategies. J Therm Biol [Internet]. 2020;93:102703. Available from: https://doi.org/10.1016/j.jtherbio.2020.102703

56. Vesic Z, Jakovljevic V, Nikolic Turnic T, Vukasinovic-Vesic M, Bolevich S, Radakovic S. The influence of acclimatization on stress hormone concentration in serum during heat stress. Mol Cell Biochem [Internet]. 2021;476:3229–39. Available from: https://doi.org/10.1007/s11010-021-04153-x

57. Karolkiewicz J, Nieman DC, Cisoń T, Szurkowska J, Gałęcka M, Sitkowski D, et al. No effects of a 4-week post-exercise sauna bathing on targeted gut microbiota and intestinal barrier function, and hsCRP in healthy men: a pilot randomized controlled trial. BMC Sports Sci Med Rehabil [Internet]. 2022;14:107. Available from: https://doi.org/10.1186/s13102-022-00497-z

58. Kirby NV, Lucas SJE, Cable TG, Armstrong OJ, Weaver SR, Lucas RAI. Additional file 1 of Sex differences in adaptation to intermittent post-exercise sauna bathing in trained middle-distance runners. Available at https://doi.org/10.6084/m9.figshare.15047670.v1. 2021 [cited 2023 Oct 5]; Available from: https://springernature.figshare.com/articles/journal_contribution/Additional_file_1_of_Sex_differences_in_adaptation_to_intermittent_post-exercise_sauna_bathing_in_trained_middle-distance_runners/15047670/1

59. Travers G, González-Alonso J, Riding N, Nichols D, Shaw A, Périard JD. Exercise Heat Acclimation With Dehydration Does Not Affect Vascular and Cardiac Volumes or Systemic Hemodynamics During Endurance Exercise. Front Physiol [Internet]. 2021;12:740121. Available from: https://doi.org/10.3389/fphys.2021.740121

60. Haroutounian A, Amorim FT, Astorino TA, Khodiguian N, Curtiss KM, Matthews ARD, et al. Change in Exercise Performance and Markers of Acute Kidney Injury Following Heat Acclimation with Permissive Dehydration. Nutrients [Internet]. 2021;13:841. Available from: https://doi.org/10.3390/nu13030841

61. Sekiguchi Y, Filep EM, Benjamin CL, Casa DJ, DiStefano LJ. Does Dehydration Affect the Adaptations of Plasma Volume, Heart Rate, Internal Body Temperature, and Sweat Rate During the Induction Phase of Heat Acclimation? J Sport Rehabil [Internet]. 2020;29:847–50. Available from: https://doi.org/10.1123/jsr.2019-0174

62. Schleh MW, Ruby BC, Dumke CL. Short term heat acclimation reduces heat stress, but is not augmented by dehydration. J Therm Biol [Internet]. 2018;78:227–34. Available from: https://doi.org/10.1016/j.jtherbio.2018.10.004

63. Neal RA, Massey HC, Tipton MJ, Young JS, Corbett J. Effect of Permissive Dehydration on Induction and Decay of Heat Acclimation, and Temperate Exercise Performance. Front Physiol [Internet]. 2016;7:564. Available from: https://doi.org/10.3389/fphys.2016.00564

64. Travers G, Nichols D, Riding N, González-Alonso J, Périard JD. Heat Acclimation with Controlled Heart Rate: Influence of Hydration Status. Med Sci Sports Exerc [Internet]. 2020;52:1815–24. Available from: https://doi.org/10.1249/MSS.0000000000002320

65. Pethick WA, Murray HJ, McFadyen P, Brodie R, Gaul CA, Stellingwerff T. Effects of hydration status during heat acclimation on plasma volume and performance. Scand J Med Sci Sports [Internet]. 2019;29:189–99. Available from: https://doi.org/10.1111/sms.13319

66. Garrett AT, Goosens NG, Rehrer NJ, Patterson MJ, Harrison J, Sammut I, et al. Short-term heat acclimation is effective and may be enhanced rather than impaired by dehydration. Am J Hum Biol Off J Hum Biol Counc [Internet]. 2014;26:311–20. Available from: https://doi.org/10.1002/ajhb.22509

67. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I-M, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc [Internet]. 2011;43:1334–59. Available from: https://doi.org/10.1249/MSS.0b013e318213fefb

Tables

Table 1. PICO — Population, Interventions, Comparisons, Outcomes.

Population	Healthy male or female adults ≥18 years of age of any race or ethnicity, in any research setting.
Interventions	Daily exercise* under thermoneutral (<25°C) environmental conditions followed by post-exercise heat exposure** for at least 2 consecutive days, with baseline and post-intervention exercise performance tests completed in either hot (≥25°C) or thermoneutral (<25°C) conditions.
Comparisons	The control group must complete the same daily exercise intervention but without post-exercise heat exposure.***
Outcomes	Primary outcome: Endurance exercise performance measured by time-to-completion (time trial or race) or time-to-exhaustion tests in hot conditions (≥25°C). Secondary outcomes: (i) Endurance exercise performance measured by time-to-completion (time trial or race) or time-to-exhaustion tests in cool conditions (<25°C). (ii) VO_2max, economy/efficiency, and/or lactate threshold in hot or cool conditions. (iii) RPE, heart rate, core temperature, sweat rate, and/or thermal sensations during exercise in hot or cool conditions. (iv) All outcome variables in trained athletes vs. non-athletes.

* Daily exercise must include at least 30-mins of aerobic activities (walking, hiking, running, cycling, rowing, xc skiing, swimming, stair-stepping etc) at a moderate intensity or higher (≥40% heart rate reserve, ≥64% heart rate max, ≥46% VO_2max, or RPE≥12) per ACSM activity guidelines [67].
** Heat exposure via any method, including but not limited to a sauna or hot water immersion. Post-exercise heat exposure must commence within 30-mins of exercise cessation.
*** Studies including additional comparator groups were included since some studies in this field compare post-exercise heat exposure (passive acclimation) vs. during-exercise heat exposure (active acclimation) vs. no heat exposure.

Table 2. Characteristics of the included studies.
Data represent mean ± SD where relevant.

Scoon *et al*. (2007) *[38]*	
Methods	Location: New Zealand. Setting: University research lab; heat exposure indoors, exercise training outdoors. Study design: Controlled crossover without repeated measures (post-intervention treatment vs control comparison). Method of randomisation: Not described.
Participants	Healthy and endurance-trained runners/triathletes (baseline VO_2max not reported). Total sample size: N = 6 (2 triathletes and 4 runners). Loss to follow-up: Not reported; assumed N = 0. Sex: 6 male, 0 female. Age = 23±3 years.
Interventions	Duration = 21 days. Control intervention = exercise training without post-exercise heat exposure; N = 6 (6 male, 0 female). - Exercise frequency: 7.7±2.3 sessions per week. - Exercise intensity: 53% and 45% of total training time was spent at "hard or very hard intensity" in the control and treatment groups, respectively. - Exercise duration: 53±8 minutes in the control group and 52±7 minutes in the treatment group. - Exercise type: Not specified. Subjects continued their "normal" training; which was likely running or cycling outdoors because the participants were runners and triathletes. - Exercise conditions: Not reported. - Post-exercise exposure control: Not described. Treatment intervention = exercise training with post-exercise heat exposure; N = 6 (6 male, 0 female). - Exercise training same as the control group. - Heat exposure frequency: 12.7±2.1 sessions in 21 days. - Heat exposure duration: 31±5 minutes per session. - Heat exposure timing: Immediately post-exercise. - Heat exposure type: Sauna, with *ad libitum* fluid intake. - Heat exposure conditions: Air temperature 89.9±2.0°C; relative humidity not reported. Washout period between intervention crossover = 14 days.
Outcomes	The following outcome was measured following each intervention: - Treadmill running time-to-exhaustion at 5 km race pace (test conditions not reported).
Notes	The treadmill run to exhaustion was tested twice following each intervention on back-to-back days and the run time-to-exhaustion is reported as the mean of these two measurements.

Vaile *et al*. (2008) *[39]*	
Methods	Location: Australia. Setting: Australian Institute of Sport research lab; heat exposure indoors, exercise training setting not described. Study design: Randomised controlled trial, repeated measures with crossover. Method of randomisation: Not described.
Participants	Healthy and endurance-trained cyclists (baseline VO_2max 68.8 ml/kg/min). Total sample size: N = 12. Loss to follow-up: Not reported; assumed N = 0. Sex: 12 male, 0 female. Age = 32.2±4.3 years.
Interventions	Duration = 5 days. Control intervention = exercise training without post-exercise heat exposure; N = 12 (12 male, 0 female). - Exercise frequency: Daily (5 sessions in total). - Exercise intensity: Intermittent mix of maximal sprints (66 sprints) and 9 minutes of sustained hard effort. - Exercise duration: 105 minutes per session. - Exercise type: Cycling, possibly indoors but not clearly stated. - Exercise conditions: Not reported. - Post-exercise exposure control: Subjects remained seated with minimal movement for 14 minutes, assumably in cool conditions but this is not specified. Treatment intervention = exercise training with post-exercise heat exposure; N = 12 (12 male, 0 female). - Exercise training same as the control group. - Heat exposure frequency: Daily (5 sessions in total). - Heat exposure duration: 14 minutes per session. - Heat exposure timing: Immediately post-exercise, following a 5-minute cycling warm-down. - Heat exposure type: Hot water immersion (fluid intake not reported). - Heat exposure conditions: Water temperature 38°C; relative humidity not relevant because participants were immersed in water. Washout period between intervention crossover = 9 days. There were also two further intervention groups not relevant to this systematic review in which participants completed post-exercise cold water immersion or post-exercise contrast water therapy (see *notes*). Data from these arms were not included or discussed.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Mean sprint cycling power (watts) on a bicycle ergometer (test conditions not reported). - Total cycling time-trial work (kilojoules) in 9 minutes on a bicycle ergometer (test conditions not reported). - Heart rate and RPE at the end of the sprints and time trials (test conditions not reported). - Core temperature at the end of the cycling training sessions (test conditions not reported).

	These outcomes were also measured every day during the intervention. However, a time course was not relevant to this systematic review, so these interim measurements were not included and only the baseline and post-intervention (day 5) measurements were discussed
Notes	Participants completed four 5-day training interventions with identical training sessions on 5 consecutive days containing 105 minutes of cycling with 66 maximal sprints and 9 minutes of sustained time trial effort, followed by 14 minutes of either passive recovery (the *Control* intervention in this systematic review), hot water immersion (the *Treatment* intervention in this systematic review), cold water immersion, or contrast water therapy (alternating hot and cold immersion). The above-described outcomes were measured on each of the 5 days of the interventions. The study met the inclusion criteria for this review but the design and data analysis is very complex with multiple time-points measured within multiple days, and several variables were not reported with sufficient detail to include in a meta-analysis. To answer the specific question, this review planned to compare measurements made on the last (Day 5) with measurements made on the first day (Day 1) of the intervention in the hot water immersion and passive recovery groups. Cherry-picking these time points introduces meta-bias and exclusion of this study was planned in a sensitivity analysis. Although the corresponding author responded, attempts to obtain the data from the first author were unsuccessful, so within-group changes were estimated as the post minus pre mean.
Zurawlew *et al.* (2016) *[40]*	
Methods	Location: UK. Setting: University research lab; heat exposure indoors, exercise training indoors. Study design: Randomised controlled trial, parallel-group repeated measures without crossover. Method of randomisation: Not described.
Participants	Healthy and recreationally active (baseline VO_2max 60.1±8.9 ml/kg/min). Total sample size: N = 17. Loss to follow-up: Not reported; assumed 0. Sex: 17 male, 0 female. Age = 23±3 years.
Interventions	Duration = 6 days. Control intervention = exercise training without post-exercise heat exposure; N = 7 (7 male, 0 female). - Exercise frequency: Daily (6 sessions in total). - Exercise intensity: 65% of VO_2max. - Exercise duration: 40 minutes per session. - Exercise type: Running indoors on a treadmill. - Exercise conditions: Air temperature 18°C; relative humidity. - Post-exercise exposure control: Cool water immersion (34°C), without fluid intake. Treatment intervention = exercise training with post-exercise heat exposure; N = 10 (10 male, 0 female). - Exercise training same as the control group. - Heat exposure frequency: Daily (6 sessions in total). - Heat exposure duration: 40 minutes per session.

	- Heat exposure timing: Immediately post-exercise. - Heat exposure type: Hot water immersion, without fluid intake. - Heat exposure conditions: Water temperature 40°C; relative humidity not relevant because participants were immersed in water.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Treadmill running 5km time trial performance in cool conditions (18°C) and in hot conditions (33°C). - Heart rate, RPE, core temperature, and thermal sensation at the end of submaximal treadmill running (40 minutes at 65% VO_2max) in cool conditions (18°C) and in hot conditions (33°C). - Whole-body sweat rate during submaximal treadmill running in cool conditions (18°C) and in hot conditions (33°C). On testing days, the 40-minute submaximal run was completed, followed by 60 minutes of rest then the 5 km time trial.
Notes	None.

Stevens *et al.* (2020) *[41]*	
Methods	Location: Australia. Setting: Elite race walker training camp at the Australian Institute of Sport; heat exposure indoors, exercise training outdoors. Study design: Non-randomised controlled trial, parallel-group repeated measures without crossover. Method of randomisation: Athletes were divided into two groups to match VO_2max and 10,000-meter performance time between groups.
Participants	Healthy and endurance-trained elite race walkers (baseline VO_2max 55.5±5.2 ml/kg/min for females, 63.2±2.8 for males). Total sample size: N = 13. Loss to follow-up: Not reported but there are 13 subject data points in figure 2, so N = 0. Sex: 4 male, 9 female. Age = 25.4±5.2 years in females and 20.9±6.2 years in males.
Interventions	Duration = 15 days. Control intervention = exercise training without post-exercise heat exposure; N = 6 (2 male, 4 female). - Exercise frequency: Daily (15 sessions in total). - Exercise intensity: Variable. - Exercise duration: 37.9 to 41.1 minutes per session. - Exercise type: Race-walking outdoors. - Exercise conditions: Air temperature 33.4±6.7°C; relative humidity 34.4±26.3%. - Post-exercise exposure control: Seated rest in 21°C, relative humidity 50%. Treatment intervention = exercise training with post-exercise heat exposure; N = 7 (2 male, 5 female). - Exercise training same as the control group. - Heat exposure frequency: 8 sessions in 15 days. - Heat exposure duration: 30 minutes per session in the first week, 40 minutes per session in the second week.

	- Heat exposure timing: Immediately post-exercise. - Heat exposure type: Hot water immersion, access to fluid intake not described. - Heat exposure conditions: Water temperature 40°C; relative humidity not relevant because participants were immersed in water.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Race-walking 10 km outdoor time trial race in cool conditions (wet-bulb globe temperature was 24.4±0.4°C pre and 20.5±0.2°C post-intervention). - Treadmill race-walking VO_2max and race-walking speed at lactate threshold (test conditions not reported). - Heart rate, RPE, core temperature, whole-body sweat rate, and thermal sensation during a submaximal treadmill race-walk (30 minutes at 80% of 10km race pace) in hot conditions (40°C).
Notes	The study included elite/pre-elite racewalkers completing a 15-day training camp at the Australian Institute of Sport and the 10 km performance time trial was a race against other athletes. The temperature and humidity during this race was different between pre and post-intervention time-points because the race was completed outdoors — actually, conditions were cooler at the post-intervention time point (22.8°C and 72.1% humidity) versus the pre-intervention timepoint (28.7°C and 47.1% humidity). Because the racewalkers trained in the summer heat for approximately 4 weeks prior to the study and for 15 days during the study, it is likely that they were already heat acclimatized. The study reported the 10 km race finish time in mm:ss format; this was converted to seconds for this meta-analysis.
Lundby *et al*. (2021) *[42]*	
Methods	Location: Norway. Setting: University research lab; heat exposure indoors, exercise training indoors. Study design: Controlled trial, parallel-group repeated measures without crossover. Method of randomisation: Not described.
Participants	Healthy and endurance-trained cyclists (baseline VO_2max 77.4±4.3 ml/kg/min in the control group, 77.4±8.5 in the treatment group). Total sample size: N = 34 but N = 24 in the intervention groups relevant to this systematic review (see *notes*). Loss to follow-up: Not reported but there are only 32 subject data points in figure 1, therefore loss is N = 3. Sex: 30 male, 4 female; but 24 male and 0 female in the intervention groups relevant to this systematic review. Age = 21±4 years in the control group and 22±7 years in the treatment group.
Interventions	Duration = 10 days. Control intervention = exercise training without post-exercise heat exposure; N = 11 (11 male, 0 female) but only N = 10 for 30 min time trial. - Exercise frequency: Daily (10 sessions in total). - Exercise intensity: 45% to 51% of power output (watts) at 4 mM lactate. - Exercise duration: 50 minutes per session.

	- Exercise type: Cycling indoors on participants' own bikes connected to a Tacx Neo smart trainer. - Exercise conditions: Air temperature 18.4±1.3°C; relative humidity 35.4±1.3% - Post-exercise exposure control: Not described. Treatment intervention = exercise training with post-exercise heat exposure; N = 13 (13 male, 0 female) but only N = 11 for 30 min time trial. - Exercise training same as the control group. - Heat exposure frequency: Daily (10 sessions in total). - Heat exposure duration: 25 to 30 minutes per session. - Heat exposure timing: Within 5 to 10 minutes of finishing exercise. - Heat exposure type: Hot water immersion, access to fluid intake not described. - Heat exposure conditions: Water temperature 40.4±0.2°C; relative humidity not relevant because participants were immersed in water. There was also a third intervention group not relevant to this systematic review in which participants exercised in a heat chamber without post-exercise heat exposure (see *notes*). Data from this arm were not included or discussed.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Power output (W), heart rate, RPE, and lactate during an all-out 30-minute cycling time trial on participants' own bikes connected to a Tacx Neo smart trainer in hot conditions (35°C). - Heart rate, and RPE during a submaximal ride (15 minutes at 60% of power at 4 mM lactate) on participants' own bikes connected to a Tacx Neo smart trainer in hot conditions (35°C).
Notes	This study assigned participants to one of three groups: HEAT (exercise in 35°C), SUIT (exercise in 18°C while wearing clothes that reduced heat loss), and $SUIT_{HWI}$ (exercise in 18°C while wearing clothes that reduced heat loss followed by hot water immersion). For this systematic review, the SUIT group represented the control group and the $SUIT_{HWI}$ group represented the treatment group. The data from the HEAT group were not relevant to the aim of this systematic review and are not included or discussed. The study reported total sweat loss in litres during a 50-minute exercise test (consisting of 15 minutes submaximal riding in the heat, 5 minutes riding in ambient conditions, and a 30 minute time trial in the heat). Therefore, the reported sweat loss was divided by 50/60 to derive a sweat rate in litres per hour. The authors refused to share their intervention-induced change data, so the within-group mean changes for heart rate, RPE, core temperature, and sweat rate were estimated from post minus pre changes in mean values. The individual subject pre and post power output values for the 30-minute time trial were presented in Fig 1 of the paper and were extracted using WebPlotDigitizer [16]. This allowed between-group differences and baseline SD to be calculated.
Bartolomé *et al.* (2021) *[43]*	
Methods	Location: Spain. Setting: University research lab; heat exposure indoors, exercise training outdoors. Study design: Randomised controlled trial, parallel-group repeated measures without crossover. Method of randomisation: Not described.

Participants	Healthy and semi-professional football (soccer) players (baseline VO_2max = 55±7 ml/kg/min). Total sample size: N = 36. Loss to follow-up: N = 2 due to muscular injuries. Sex: 36 male, 0 female. Age = 20.23±1.98 years in the control group and 20.69±2.09 years in the treatment group.
Interventions	Duration = 21 days. Control intervention = exercise training without post-exercise heat exposure; N = 18 (18 male, 0 female) with N = 2 drop-outs due to muscular injuries, leaving N = 16 completing the intervention. - Exercise frequency: Not reported; but participants did 10 hours of training per week. - Exercise intensity: Intensity not stated; subjects continued their "normal" training. - Exercise duration: Not reported; but participants did 10 hours of training per week. - Exercise type: Usual weekly training designed to maintain fitness under direction of their club coach. Included light continuous runs, fartleks, accelerations and changes of pace, light core and physical fitness exercises, short distance sprints, and soccer-specific skills. - Exercise conditions: Not reported. - Post-exercise exposure control: Not described. Treatment intervention = exercise training with post-exercise heat exposure; N = 18 (18 male, 0 female). - Exercise training same as the control group. - Heat exposure frequency: 3 sessions per week (9 sessions in total). - Heat exposure duration: 4 × 10 minutes per session. - Heat exposure timing: Not described. - Heat exposure type: Sauna, with fluid intake limited to 250 ml. - Heat exposure conditions: Air temperature 100±3°C; relative humidity 50–55%.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Treadmill running time-to-exhaustion during a VO_2max test in cool conditions (18°C). - Treadmill running VO_2max in cool conditions (18°C).
Notes	The authors reported pre- and post-intervention mean and SD values for time-to-exhaustion in Figure 1 of their paper. These values were extracted using WebPlotDigitizer [16]. Time-to-exhaustion was reported in minutes and was converted to seconds to enable comparison with other studies in the meta-analysis. No within-group change scores were presented, and the authors did not respond to multiple attempts to request the data, so change scores were estimated from post minus pre changes in mean values. Within-group Cohen's R effect sizes were reported, but only for the sauna group not the control group.
McIntyre *et al.* (2021) *[44]*	
Methods	Location: UK. Setting: University research lab; heat exposure indoors, exercise training indoors. Study design: Randomised controlled trial, parallel-group repeated measures without crossover. Method of randomisation: Used randomizer.org but details not described.

Participants	Healthy and recreationally active (baseline VO_2max = 53.4 ml/kg/min in the control group, 53.4 in the treatment group). Total sample size: N= 27 but N = 18 in the intervention groups relevant to this systematic review (see *notes*). Loss to follow-up: N =7 participants were removed from the time-to-exhaustion tests due to reaching the core temperature ethical cut-off (HWI, n = 2), toilet break (EHA, n = 1), lower limb discomfort (TNE, n = 1), exercise-induced bronchoconstriction (TNE, n = 1), nausea (TNE, n = 1), and lack of effort without markers of overreaching at rest (TNE, n = 1). Sex: 27 male, 0 female. Age = 21±2 years in the control group and 22±3 years in the treatment group.
Interventions	Duration = 8 days (2 × 3-day blocks separated by 2 days without intervention). Control intervention = exercise training without post-exercise heat exposure; N = 9 (9 male, 0 female) with N = 4 drop-outs due to lower limb discomfort (N = 1), exercise-induced bronchoconstriction (N = 1), nausea (N = 1), and lack of effort (N = 1), leaving N = 5 completing the intervention. - Exercise frequency: Daily in each 3-day block (6 sessions in total). - Exercise intensity: 65% of VO_2peak. - Exercise duration: 60 minutes per session (40 minutes per session in the treatment group). - Exercise type: Running indoors on a treadmill. - Exercise conditions: Air temperature 19°C; relative humidity 45%. - Post-exercise exposure control: 20 minutes additional exercise compared to treatment group. Treatment intervention = exercise training with post-exercise heat exposure; N = 9 (9 male, 0 female) with N=2 drop-outs due to reaching the core temperature ethical cut-off, leaving N = 7 completing the intervention. - Exercise training same as the control group. - Heat exposure frequency: Daily in each 3-day block (6 sessions in total). - Heat exposure duration: 40 minutes per session. - Heat exposure timing: Not described but assumed to be immediately post-exercise because study design based on that of Zurawlew *et al.* [40]. - Heat exposure type: Hot water immersion; fluid intake not described but possibly without because design based on Zurawlew *et al.* [40]. - Heat exposure conditions: Water temperature 40°C; relative humidity not relevant because participants were immersed in water. There was also a third intervention group not relevant to this systematic review in which participants exercised in a heat chamber without post-exercise heat exposure (see *notes*). Data from this arm was not included or discussed.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Treadmill running time-to-exhaustion at 65% of VO_2max in hot conditions (33°C). - Heart rate, RPE, and thermal sensation at the end of submaximal treadmill running (40 minutes at 65% of VO_2peak) in hot conditions (33°C). - Whole-body sweat rate during submaximal treadmill running (40 minutes at 65% of VO_2peak) in hot conditions (33°C). These outcomes were also measured after the first 3-day block of the intervention. However, a time course was not relevant to this systematic review, so this interim measurement was not included and only the baseline and post-intervention (after the 2 ×

	3-day blocks) measurements were discussed.
Notes	This study randomised participants to one of three groups: EHA (60 minutes of exercise in 33°C), TNE (60 minutes of exercise in 19°C), or HWI (40 minutes of exercise in 19°C followed by 20 minutes of hot water immersion). The HWI group did 20 minutes less exercise per session and, therefore, had a lower training load than the other groups. For this systematic review, the TNE group represented the control group and the HWI group represented the treatment group. The data from the EHA group were not relevant to the aim of this systematic review and were not included or discussed.
McIntyre *et al*. (2022) *[45]*	
Methods	Location: UK. Setting: University research lab; heat exposure indoors, exercise training indoors. Study design: Randomised controlled trial, parallel-group repeated measures without crossover. Method of randomisation: Used randomizer.org but details not described.
Participants	Healthy and recreationally active (baseline VO_2max = 53.4 ml/kg/min in the control group, 53.7 in the treatment group). Total sample size: N = 43 but N = 27 in the intervention groups relevant to this systematic review (see *notes*). Loss to follow-up: N = 22, of which N = 13 were in the intervention groups relevant to this systematic review (see *interventions*). Sex: 14 male, 0 female. Age = 22±2 years in the control group and 22±3 years in the treatment group.
Interventions	Duration = 12 days (2 × 3-day blocks followed by a 6-day block, each block separated by 2 days without intervention). Control intervention = exercise training without post-exercise heat exposure; N = 12 (12 male; 0 female) but N = 5 drop-outs due to lower limb discomfort (N = 3), scheduling conflict (N = 1), failure to comply with protocol (N = 1), leaving N = 7 completing the intervention. Furthermore, only N = 5 completed the time-to-exhaustion tests. - Exercise frequency: Daily in each block (12 sessions in total). - Exercise intensity: 65% of VO_2peak. - Exercise duration: 60 minutes per session (40 minutes per session in the treatment group). - Exercise type: Running indoors on a treadmill. - Exercise conditions: Air temperature 19°C; relative humidity 45%. - Post-exercise exposure control: 20 minutes of additional exercise compared to the treatment group. Treatment intervention = exercise training with post-exercise heat exposure; N = 15 (15 male; 0 female) but N = 8 drop-outs due to lower limb discomfort (N = 1), scheduling conflict (N = 4), failure to comply with protocol (N = 2) and injury (N = 1), leaving N = 7 completing the intervention. Furthermore, only N = 5 completed the time-to-exhaustion tests. - Exercise training same as the control group. - Heat exposure frequency: Daily in each block (12 sessions in total). - Heat exposure duration: 40 minutes per session. - Heat exposure timing: Within 2 to 3 minutes after exercise. - Heat exposure type: Hot water immersion; fluid intake not described but possibly absent because the design is based on

	Zurawlew *et al*. [40] and McIntyre *et al*. (2021) [44]. - Heat exposure conditions: Water temperature 40°C; relative humidity not relevant because participants were immersed in water. There was also a third intervention group not relevant to this systematic review in which participants exercised in a heat chamber without post-exercise heat exposure (see *notes*). Data from this arm was not included or discussed.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Treadmill running time-to-exhaustion at 65% of VO_2max in hot conditions (33°C). - Heart rate, RPE, and thermal sensation at the end of submaximal treadmill running (40 minutes at 65% of VO_2peak) in hot conditions (33°C). - Whole-body sweat rate during submaximal treadmill running (40 minutes at 65% of VO_2peak) in hot conditions (33°C). These outcomes were also measured after the first and second 3-day blocks of the intervention. However, a time course was not relevant to this systematic review, so these interim measurements were not included and only the baseline and post-intervention (after all blocks) measurements were discussed.
Notes	This study randomised participants to one of three groups: EHA (60 minutes of exercise in 33°C), CON (60 minutes of exercise in 19°C), or HWI (40 minutes of exercise in 19°C followed by 20 minutes of hot water immersion). The HWI group did 20 minutes less exercise per session and, therefore, has a lower training load than the other groups. For this systematic review, the CON group represented the control group and the HWI group represented the treatment group. The data from the EHA group were not relevant to the aim of this systematic review and were not included or discussed. This study also measured mood, sleep, and executive function but these variables are not included or discussed because they are not relevant to the aim of this review.
Kirby *et al*. (2021) *[47]*	
Methods	Location: UK. Setting: University research lab; heat exposure indoors, exercise training outdoors. Study design: Non-randomised controlled trial, parallel-group repeated measures without crossover. Method of randomisation: Participants self-selected into their preferred intervention group.
Participants	Healthy and endurance-trained runners (baseline VO_2max = 59.3±7.2 ml/kg/min in the control group, 58.4±7.2 in the treatment group). Total sample size: N = 28 Loss to follow-up: N = 8. Sex: Not specified for the original N = 28 but 7 males and 13 females completed the study. Age = 19±1 years in the control group and 20±2 years in the treatment group.
Interventions	Duration = 21 days. Control intervention = exercise training without post-exercise heat exposure; N = 12 but N = 4 drop-outs due to injury (N = 3) and time commitment (N = 1), leaving N = 8 (4 male, 4 female) completing the intervention. - Exercise frequency: 7±2 sessions per week.

	- Exercise intensity: RPE 4 to 8 out of 10, depending on the day. - Exercise duration: 53.3 minutes per session in the control group and 54.5 minutes per session in the treatment group. - Exercise type: Running outdoors. - Exercise conditions: Not reported. - Post-exercise exposure control: Not described. Treatment intervention = exercise training with post-exercise heat exposure; N = 16 but N = 4 drop-outs due to injury, leaving N = 12 (3 male, 9 female) completing the intervention. - Exercise training same as the control group. - Heat exposure frequency: 3±1 sessions per week (9±1 sessions in total). - Heat exposure duration: 28±2 minutes per session. - Heat exposure timing: Within approximately 5 minutes of cessation of exercise. - Heat exposure type: Sauna, with *ad libitum* fluid intake. - Heat exposure conditions: Air temperature 101–108°C; relative humidity 5–10%.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Treadmill running time-to-exhaustion during a VO_2max test in cool conditions (18°C). - Treadmill running VO_2max, running economy, and running speed at 4 mM lactate in cool conditions (18°C). - Heart rate, RPE, core temperature, sweat rate, and thermal sensation during submaximal treadmill running (30 minutes at 9 kph at a 2% grade) in hot conditions (40°C).
Notes	Participants were given an iron supplement (65 mg/day ferrous sulfate) to take for 2 weeks prior to and during the interventions but the reasons were not given. The study was conducted in the UK between the months of October and March to minimise any natural heat acclimatisation caused by training outdoors in hot conditions. All participants completed a submaximal 9 kph treadmill run at baseline and after the interventions, which means that different participants exercised at different relative intensities. VO_2max was presented in absolute values (L/min). To allow comparisons with the other included studies in this systematic review, the authors agreed to share the relative VO_2max data (mL/kg/min). Total sweat loss was reported during a 30-minute exercise bout; therefore, to derive a sweat rate in L/h and allow comparison to the other included studies, the data were multiplied by 2. Sweat gland activity was also measured during exercise but this outcome was not relevant to the aim of this systematic review so the data were not included or discussed. This study [47] and Kirby *et al.* (2021) *Sports Med Open* [46] (details below) used identical interventions and several participants' data is used in both papers. The main difference between the two publications is that this study aimed to determine the efficacy of regular post-exercise heat exposure on exercise performance whereas Kirby *et al.* (2021) *Sports Med Open* [46] aimed to determine the effect of sex on the efficacy of regular post-exercise heat exposure to induce heat acclimation and improve exercise performance. However, due to large subject drop-out, Kirby *et al.* (2021) *Sports Med Open* [46] only presents results from the sauna group. To avoid duplication of data and prevent errors in this systematic review, the authors agreed to share their raw data. Therefore, the data analysis in the meta-analysis includes the data from both Kirby studies [46,47] as a single data set.

Kirby *et al.* (2021a) *[46]*	
Methods	Location: UK. Setting: University research lab; heat exposure indoors, exercise training outdoors. Study design: Non-randomised controlled trial, parallel-group repeated measures without crossover. Method of randomisation: Participants self-selected into their preferred intervention group.
Participants	Healthy and endurance-trained runners (baseline VO_2max = 52.6±6.9 to 64.6±6.9 ml/kg/min). Total sample size: N = 41. Loss to follow-up: N = 6, specific reasons not described. Sex: 18 male, 23 female. Age = 20±2 years.
Interventions	Duration = 21 days. Control intervention = exercise training without post-exercise heat exposure; N = 15 (8 male, 7 female) but N = 3 (1 male, 2 female) dropouts (specific reasons not described), leaving N = 12 (7 male, 5 female) completing the intervention. - Exercise frequency: 7±2 sessions per week. - Exercise intensity: RPE 4 to 8 out of 10, depending on the day. - Exercise duration: 69.5 minutes per session for males, 42.8 minutes per session for females. - Exercise type: Running outdoors. - Exercise conditions: Not reported. - Post-exercise exposure control: Not described. Treatment intervention = exercise training with post-exercise heat exposure; N = 26 (10 male, 16 female) but N = 3 (1 male, 2 female) dropouts (specific reasons not described), leaving N = 23 (9 male, 14 female) completing the intervention. - Exercise training same as the control group. - Heat exposure frequency: 3±1 sessions per week (9±1 sessions in total). - Heat exposure duration: 29±3 minutes per session. - Heat exposure timing: Within approximately 5 minutes of cessation of exercise. - Heat exposure type: Sauna, with *ad libitum* fluid intake. - Heat exposure conditions: Air temperature 104°C; relative humidity 5–9%.
Outcomes	The following outcomes were measured at baseline and following the intervention in each group: - Treadmill running time-to-exhaustion during a VO_2max test in cool conditions (18°C). - Treadmill running VO_2max, running economy, and running speed at 4 mM lactate in cool conditions (18°C). - Heart rate, RPE, core temperature, sweat rate, and thermal sensation during submaximal treadmill running (30 minutes at 9 kph at a 2% grade) in hot conditions (40°C).
Notes	Participants were given an iron supplement (65 mg/day ferrous sulfate) to take for 2-weeks prior to and during the interventions. Authors stated that iron supplements were intended to ensure equal efficacy of normal training in both sexes because iron deficiency is common in highly trained endurance athletes, particularly female endurance athletes, and because iron deficiency is known to impair training adaptations.

	The study was conducted in the UK between the months of October and March to minimise any natural heat acclimatisation caused by training outdoors in hot conditions. All participants completed a submaximal 9 kph treadmill run at baseline and after the interventions, which means that different participants exercised at different relative intensities. Total sweat loss was reported during a 30-minute exercise bout; therefore, to derive a sweat rate in L/h and allow comparison to the other included studies, the data were multiplied by 2. Sweat gland activity was also measured during exercise but this outcome was not relevant to the aim of this systematic review, so the data were not included or discussed. This study and Kirby *et al.* (2021) *Eur J Appl Physiol* [47] (details above) used identical interventions and several participants' data is used in both papers. The main difference between the two publications is that Kirby *et al.* (2021) *Eur J Appl Physiol* [47] aimed to determine the efficacy of regular post-exercise heat exposure on exercise performance whereas this study aimed to determine the effect of sex on the efficacy of regular post-exercise heat exposure to induce heat acclimation and improve exercise performance. However, due to large subject drop-out, this study only presents results from the sauna group. To avoid duplication of data and prevent errors in this systematic review, the authors agreed to share their raw data. Therefore, the data analysis in the meta-analysis includes the data from both Kirby studies [46,47] as a single data set.

Table 3. GRADE summary of findings.

Population: Healthy male or female adults ≥18 years of age of any race or ethnicity, in any research setting.

Interventions: Exercise followed by post-exercise heat exposure (sauna or hot water immersion) on at least 2 consecutive days.

Comparisons: The control group must complete the same exercise intervention but without post-exercise heat exposure.

Outcomes	Effect of treatment (ratio of mean, or standardised mean difference)	Number of participants and studies in meta-analysis.	Optimal Information Size criterion (OIS).	Certainty of the evidence.
Endurance performance in hot conditions (primary outcome).	ROM (95% CI) = 1.04 (0.94, 1.15) A ratio>1 means treatment is better than control.	Participants = 60 Studies = 4	OIS = 298	⊖ ⊕ ⊖ ⊖ ⊖ VERY LOW due to Risk of bias[1], Indirectness[10], Imprecision[12], and Publication bias[13].
Endurance performance in cool conditions.	ROM (95% CI) = 1.06 (0.99, 1.12) A ratio>1 means treatment is better than control.	Participants = 144 Studies = 6	OIS = 664	⊖ ⊕ ⊖ ⊖ ⊖ VERY LOW due to Risk of bias[2], Indirectness[10], and Imprecision[12].
VO_2max.	SMD (95% CI) = 0.33 (0.19, 0.47) A positive value means treatment is better than control.	Participants = 91 Studies = 3	OIS = 144	⊖ ⊕ ⊖ ⊖ ⊖ VERY LOW due to Risk of bias[3], Indirectness[11], Imprecision[12], and Publication bias[13].
Speed at lactate threshold.	SMD (95% CI) = 0.19 (-0.53, 0.92) A positive value means treatment is better than control.	Participants = 55 Studies = 3	OIS = 393	⊖ ⊕ ⊖ ⊖ ⊖ VERY LOW due to Risk of bias[4], Indirectness[11], Imprecision[12], and Publication bias[13].
Heart rate.	SMD (95% CI) = -0.32 (-0.45, -0.20) A negative value means treatment is better than control.	Participants = 163 Studies = 6	OIS = 173	⊖ ⊕ ⊖ ⊕ ⊖ VERY LOW due to Risk of bias[5], and Indirectness[11], and Imprecision[12].
RPE.	SMD (95% CI) = -0.07 (-0.65, 0.51) A negative value means treatment is better than control.	Participants = 163 Studies = 6	OIS = 432	⊖ ⊕ ⊖ ⊖ ⊕ VERY LOW due to Risk of bias[6], Indirectness[11], and Imprecision[12].
Core temperature.	SMD (95% CI) = -0.44 (-0.79, -0.09) A negative value means treatment is better than control.	Participants = 163 Studies = 6	OIS = 34	⊖ ⊕ ⊖ ⊕ ⊕ LOW due to Risk of bias[7] and Indirectness[11].

Sweat rate.	SMD (95% CI) = 0.27 (0.06, 0.47) A positive value means treatment is better than control.	Participants = 139 OIS = 159 Studies = 6	⊖ ⊕ ⊖ ⊖ ⊕ VERY LOW due to Risk of bias[8], Indirectness[11], and Imprecision[12].
Thermal rating.	SMD (95% CI) = -0.56 (-0.94, -0.17) A negative value means treatment is better than control.	Participants = 112 OIS = 56 Studies = 5	⊖ ⊕ ⊖ ⊕ ⊕ LOW due to Risk of bias[9] and Indirectness[11].

ROM = Ratio of mean between the effect of treatment and the effect of no treatment (control), interpreted as: trivial (ROM<1.08), small (1.08≤ROM<1.22), moderate (1.22≤ROM<1.37), or large effects (ROM≥1.37) of treatment. SE = standard error of ratio of mean. SMD = standardised mean difference (Hedges' g) between the effect of treatment and the effect of no treatment (control), interpreted as: trivial (g<0.2), small (0.2≤g<0.5), moderate (0.5≤g<0.8), or large effects (g≥0.8) of treatment. CI = confidence interval.

GRADE Working Group grades of evidence:
⊕ = no downgrade in the quality of evidence. ⊖ = downgrade quality of evidence.
High = very confident that the true effect lies close to the effect estimate. *Moderate* = moderately confident in the effect estimate; the true effect is likely to be close to the effect estimate but there is a possibility that it is substantially different. *Low* = low confidence in the effect estimate; the true effect may be substantially different. *Very Low* = very little confidence in the effect estimate; the true effect is likely to be substantially different.

Specific reasons for downgrading the quality of evidence:
[1] Risk of bias for performance in hot conditions: 1/4 studies had a high risk of bias due to missing outcome data, 1/4 studies had a high risk of bias due to measurement of the outcome; 1/4 studies had a high risk of bias due to the selection of the reported result.
[2] Risk of bias for performance in cool conditions: 1/6 studies had a high risk of bias due to the randomization process; 1/6 studies had a high risk of bias due to the deviations from intended interventions; 1/6 studies had a high risk of bias due to the missing outcome data; 2/6 studies had a high risk of bias due to the measurement of the outcome; 1/6 studies had a high risk of bias due to the selection of the reported result.
[3] Risk of bias for VO_2max: 1/3 studies had a high risk of bias in all bias domains.
[4] Risk of bias for speed at lactate threshold: 1/2 studies had a high risk of bias in all bias domains.
[5] Risk of bias for heart rate: 1/6 studies had a high risk of bias due to the randomization process, deviations from intended interventions, and missing outcome data; 2/6 studies had a high risk of bias due to the measurement of the outcome; 3/6 studies had a high risk of bias due to the selection of the reported result.
[6] Risk of bias for RPE: 1/6 studies had a high risk of bias due to the randomization process; 2/6 studies had a high risk of bias due to deviations from intended interventions and the selection of the reported result.
[7] Risk of bias for core temperature: 1/6 studies had a high risk of bias due to the randomization process and deviations from intended interventions; 3/6 studies had a high risk of bias due to the selection of the reported result.
[8] Risk of bias for sweat rate: 1/6 studies had a high risk of bias due to the randomization process, deviations from intended interventions, and the measurement of the outcome; 4/6 studies had a high risk of bias due to the selection of the reported result.
[9] Risk of bias for thermal rating: 1/5 studies had a high risk of bias due to the randomization process and deviations from intended interventions; 2/5 studies had a high risk of bias due to the measurement of the outcome and the selection of the reported result.
[10] Indirectness due to substantial differences in interventions and outcome measures between studies.
[11] Indirectness due to substantial differences in interventions between studies.

[12] Imprecision due to the sample size in the meta-analysis being lower than the optimal information size criterion and/or the 95% CI of the effect size overlapping zero effect.
[13] Evidence for publication bias due to a very small number of studies (k<5).

Figures

Fig. 1. PRISMA 2020 flow diagram.

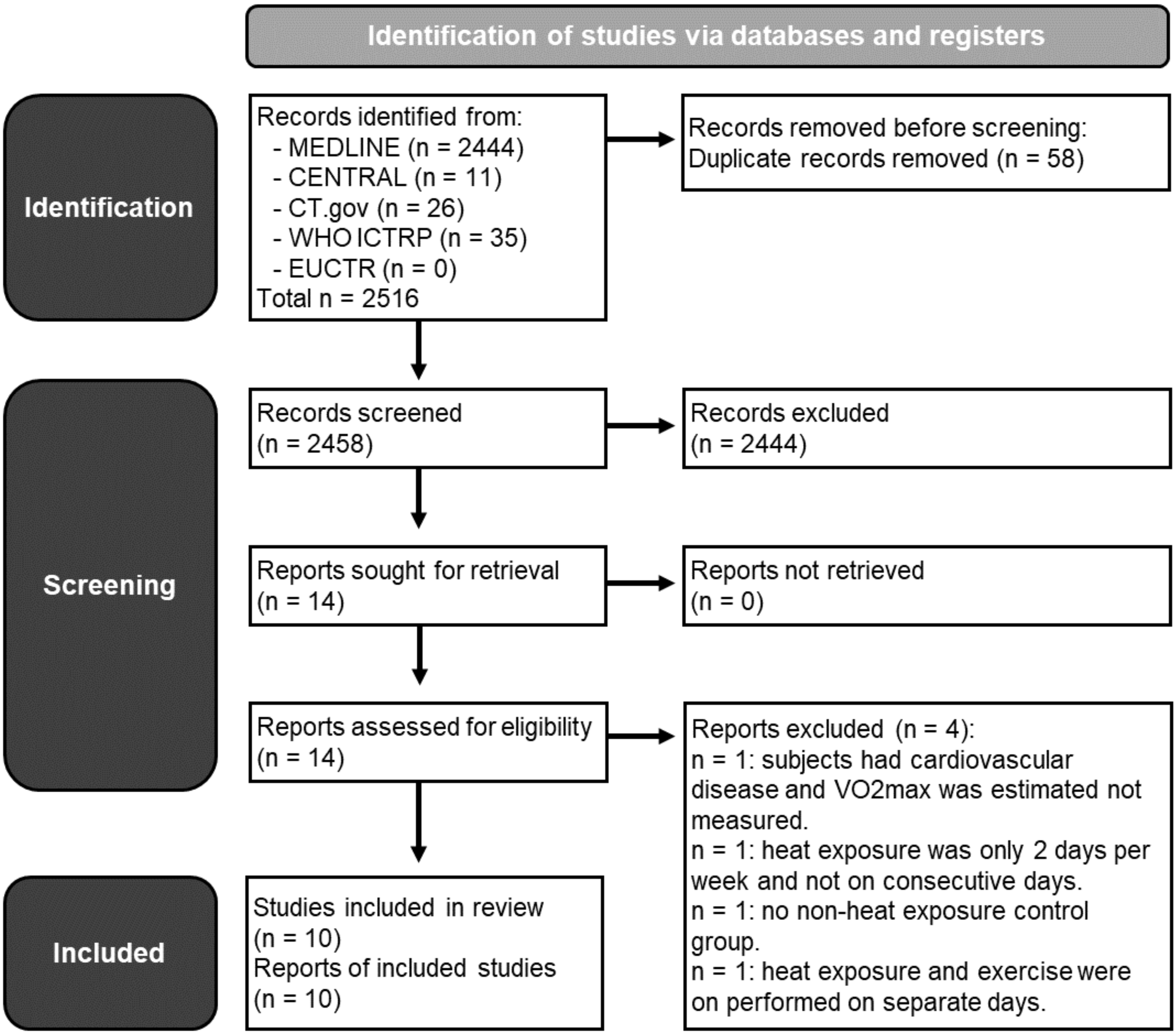


Fig. 2. Risk of bias analysis.

Risk of bias for each outcome variable was rated as high (×), low (+), or some concerns (-) across five domains — randomisation, protocol deviation, missing data, measurement, and reporting — using the Cochrane Risk of Bias 2 (RoB2) tool. The overall risk of bias for each study was determined as: "Low" if the trial was judged to be at low risk of bias for all domains; "Some concerns" if the trial was judged to raise some concerns in at least one domain but not to be at high risk of bias for any domain; or "High" if the trial was judged to be at high risk of bias in at least one domain or if the trial was judged to have "some concerns" for multiple domains. The figure was built using the *robvis* R package [18].

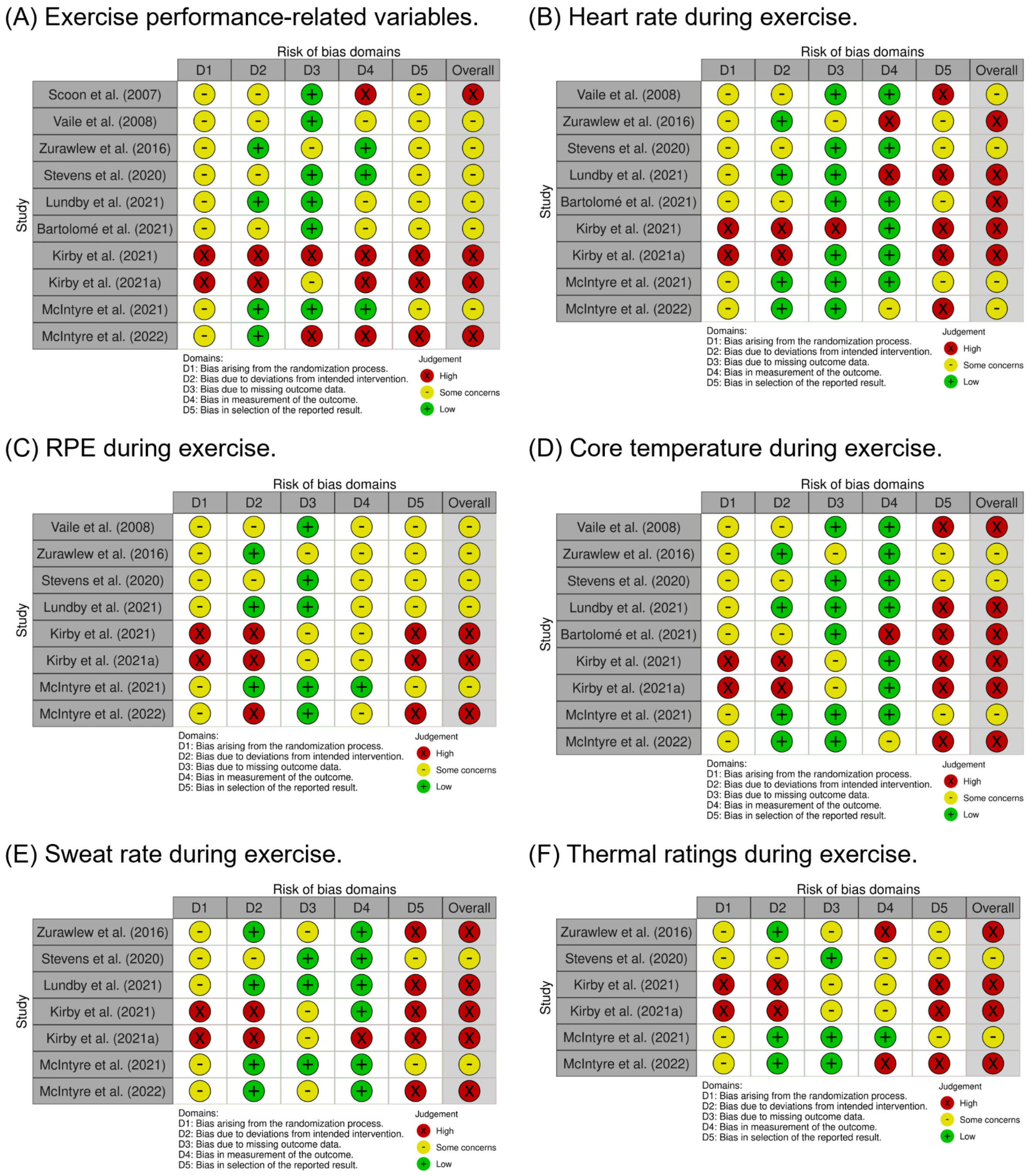


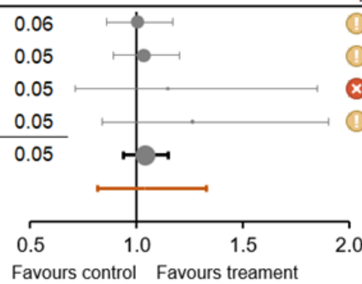
Fig. 3. Meta-analysis of endurance exercise performance in hot (≥25°C) conditions (the primary outcome).

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on endurance exercise performance (time trial, time-to-exhaustion, or a race) in hot conditions (≥25°C). The ln(Ratio of means) values in the control and treatment groups represent intervention-induced changes for performance outcomes calculated as the natural log-transformed ratio of means (post divided by pre). The between-group ln(Ratio of means) values represent the relative difference in the change scores between the treatment and control groups, calculated as the natural log-transformed ratio of means, i.e. $\log_e((Post_{Treatment} \div Pre_{Treatment}) \div (Post_{Control} \div Pre_{Control}))$. Meta-analytical calculations were performed on the ln(Ratio of means) and its corresponding standard error (SE). The effect estimates are presented as the reverse-transformed ratio of means values and the forest plot bubbles represent the ratio of means for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). N = sample size; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I^2 = percentage variability in effect sizes across studies; Tau2 and Tau are measures of the dispersion of true effect sizes between studies.

Endurance exercise performance in hot conditions (≥25°C).

Study	Control N	Control ln(Ratio of means)	Treatment N	Treatment ln(Ratio of means)	Between-group ln(Ratio of means)	Between-group SE	Effect estimate Weight	Ratio of means	95%CI	Power	Overall Risk of bias
Lundby et al. (2021)	10	0.090	11	0.094	0.004	0.079	43.0%	1.00	[0.86, 1.17]	0.06	Some concerns
Zurawlew et al. (2016)	7	0.017	10	0.051	0.034	0.076	46.4%	1.03	[0.89, 1.20]	0.05	Some concerns
McIntyre et al. (2022)	5	0.205	5	0.342	0.138	0.244	4.5%	1.15	[0.71, 1.85]	0.05	High
McIntyre et al. (2021)	5	0.071	7	0.303	0.233	0.210	6.1%	1.26	[0.84, 1.90]	0.05	Some concerns
Number of studies = 4 Total participants = 60					Combined effect:		100%	1.04	[0.94, 1.15]	0.05	
							Prediction interval:		[0.81, 1.33]		

Ratio of means forest plot:



Test for overall effect: Z = 0.73 (P = 0.46)

Heterogeneity: Q = 1.22 (P = 0.75); I^2 = 0%; Tau^2 = 0.02, Tau = 0.12

Statistical power = 0.05

GRADE certainty of the evidence = Very low

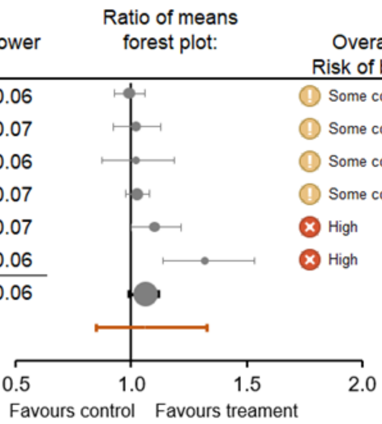
Fig. 4. Meta-analysis of endurance exercise performance in cool (<25°C) conditions.

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on endurance exercise performance (time trial, time-to-exhaustion, or a race) in cool conditions (<25°C). The ln(Ratio of means) values in the control and treatment groups represent intervention-induced changes for performance outcomes calculated as the natural log-transformed ratio of means (post divided by pre). The between-group ln(Ratio of means) values represent the relative difference in the change scores between the treatment and control groups, calculated as the natural log-transformed ratio of means, i.e. $\log_e((Post_{Treatment} \div Pre_{Treatment}) \div (Post_{Control} \div Pre_{Control}))$. Meta-analytical calculations were performed on the ln(Ratio of means) and its corresponding standard error (SE). The effect estimates are presented as the reverse-transformed ratio of means values and the forest plot bubbles represent the ratio of means for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). N = sample size; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I^2 = percentage variability in effect sizes across studies; Tau2 and Tau are measures of the dispersion of true effect sizes between studies.

Endurance exercise performance in cool conditions (<25°C).

Study	Control		Treatment		Between-group		Effect estimate				Overall Risk of bias
	N	ln(Ratio of means)	N	ln(Ratio of means)	ln(Ratio of means)	SE	Weight	Ratio of means	95%CI	Power	
Stevens et al. (2020)	6	0.063	7	0.053	-0.010	0.034	21.6%	0.99	[0.93, 1.06]	0.06	Some concerns
Vaile et al. (2008)	12	-0.038	12	-0.019	0.019	0.052	16.1%	1.02	[0.92, 1.13]	0.07	Some concerns
Zurawlew et al. (2016)	7	-0.007	10	0.013	0.020	0.078	10.4%	1.02	[0.88, 1.19]	0.06	Some concerns
Bartolomé et al. (2021)	18	0.001	18	0.025	0.025	0.025	24.4%	1.03	[0.98, 1.08]	0.07	Some concerns
Kirby et al. (2021; 2021a)	15	-0.025	27	0.072	0.096	0.050	16.7%	1.10	[1.00, 1.21]	0.07	High
Scoon et al. (2007)	6	n/a	6	n/a	0.278	0.076	10.8%	1.32	[1.14, 1.53]	0.06	High
Number of studies = 6 Total participants = 144					Combined effect:		100%	1.06	[0.99, 1.12]	0.06	
							Prediction interval:		[0.85, 1.33]		

Note: Scoon et al. (2007) is a between-group post-intervention comparison.



Test for overall effect: Z = 1.75 (P = 0.08)

Heterogeneity: Q = 13.75 (P = 0.02); I^2 = 64%; Tau^2 = 0.01, Tau = 0.11

Statistical power = 0.06

GRADE certainty of the evidence = Very low

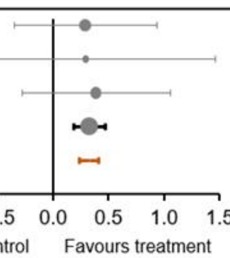
Fig. 5. Meta-analysis of VO_2max.

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on VO_2max. The meanΔ values in the control and treatment groups represent the post-intervention minus pre-intervention change scores. The mean difference values represent the difference in the change scores between the treatment and control groups, i.e. ($Post_{Treatment}$ - $Pre_{Treatment}$) - ($Post_{Control}$ - $Pre_{Control}$). The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). The forest plot bubbles represent the standardised mean difference (SMD; Hedges' g) for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. N = sample size; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I^2 = percentage variability in effect sizes across studies; Tau^2 and Tau are measures of the dispersion of true effect sizes between studies.

VO_2max.

Study	Control N	Control MeanΔ (L/min)	Treatment N	Treatment MeanΔ (L/min)	Mean difference (L/min)	Pooled baseline SD	Effect estimate SMD	95%CI	SE	Weight	Power	Overall Risk of bias
Kirby et al. (2021; 2021a)	15	0.16	27	2.43	2.27	7.69	0.29	[-0.35, 0.94]	0.32	43.4%	0.16	High
Stevens et al. (2020)	6	1.10	7	3.00	1.90	5.99	0.29	[-0.83, 1.46]	0.52	16.2%	0.09	Some concerns
Bartolomé et al. (2021)	18	-1.11	18	1.14	2.25	5.70	0.39	[-0.27, 1.06]	0.33	40.4%	0.15	Some concerns
Number of studies = 3 Total participants = 91					Combined effect:		0.33	[0.19, 0.47]	0.03	100%	0.15	
					Prediction interval:			[0.24, 0.42]				

SMD forest plot:



Test for overall effect: Z = 10.0 (P<0.0001)

Heterogeneity: Q = 0.05 (P = 0.98); I^2 = 0%; Tau^2 = 0.002, Tau = 0.04

Statistical power = 0.15

GRADE certainty of the evidence = Very low

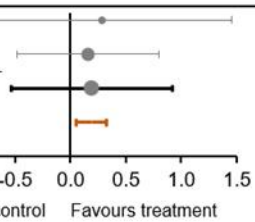
Fig. 6. Meta-analysis of speed at lactate threshold.

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on speed at lactate threshold. The meanΔ values in the control and treatment groups represent the post-intervention minus pre-intervention change scores. The mean difference values represent the difference in the change scores between the treatment and control groups, i.e. $(Post_{Treatment} - Pre_{Treatment}) - (Post_{Control} - Pre_{Control})$. The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). The forest plot bubbles represent the standardised mean difference (SMD; Hedges' g) for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. N = sample size; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I^2 = percentage variability in effect sizes across studies; Tau^2 and Tau are measures of the dispersion of true effect sizes between studies.

Speed at lactate threshold.

Study	Control N	Control MeanΔ (kph)	Treatment N	Treatment MeanΔ (kph)	Mean difference (kph)	Pooled baseline SD	Effect estimate SMD	95%CI	SE	Weight	Power	Overall Risk of bias
Stevens et al. (2020)	6	0.20	7	0.40	0.20	0.65	0.29	[-0.84, 1.45]	0.52	27.0%	0.07	Some concerns
Kirby et al. (2021; 2021a)	15	0.04	27	0.36	0.32	1.99	0.16	[-0.48, 0.80]	0.32	73.0%	0.09	High
Number of studies = 2 Total participants = 55						Combined effect:	0.19	[-0.53, 0.92]	0.06	100%	0.08	
						Prediction interval:		[0.05, 0.33]				

SMD forest plot:



Test for overall effect: Z = 3.38 (P = 0.0007)

Heterogeneity: Q = 0.04 (P = 0.83); I^2 = 0%; Tau^2 = 0.005, Tau = 0.07

Statistical power = 0.08

GRADE certainty of the evidence = Very low

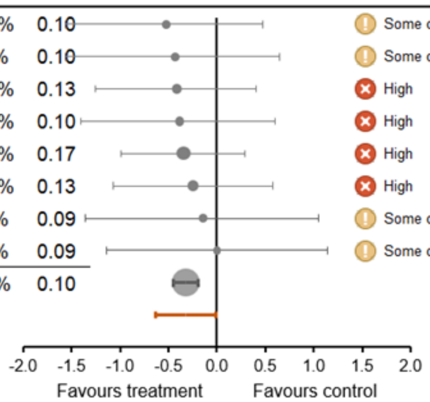
Fig. 7. Meta-analysis of heart rate during submaximal exercise.

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on heart rate during submaximal exercise. The meanΔ values in the control and treatment groups represent the post-intervention minus pre-intervention change scores. The mean difference values represent the difference in the change scores between the treatment and control groups, i.e. $(Post_{Treatment} - Pre_{Treatment}) - (Post_{Control} - Pre_{Control})$. The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). The forest plot bubbles represent the standardised mean difference (SMD; Hedges' g) for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. N = sample size; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I² = percentage variability in effect sizes across studies; Tau^2 and Tau are measures of the dispersion of true effect sizes between studies.

Heart rate during submaximal exercise.

Study	Control N	Control MeanΔ (bpm)	Treatment N	Treatment MeanΔ (bpm)	Mean difference (bpm)	Pooled baseline SD	Effect estimate SMD	95%CI	SE	Weight	Power	Overall Risk of bias
Zurawlew et al. (2016)	7	-1	10	-7	-6	11	-0.52	[-1.56, 0.47]	0.48	10.4%	0.10	Some concerns
McIntyre et al. (2022)	7	-14	7	-20	-6	13	-0.43	[-1.57, 0.64]	0.51	9.2%	0.10	Some concerns
Lundby et al. (2021)	11	3	13	0	-3	7	-0.41	[-1.26, 0.40]	0.40	14.7%	0.13	High
Zurawlew et al. (2016)a	7	0	10	-6	-6	15	-0.38	[-1.41, 0.61]	0.47	10.6%	0.10	High
Kirby et al. (2021; 2021a)	15	-2	27	-9	-7	20	-0.34	[-0.99, 0.29]	0.32	23.3%	0.17	High
Lundby et al. (2021)a	11	-3	13	-5	-2	8	-0.24	[-1.07, 0.57]	0.40	15.0%	0.13	High
McIntyre et al. (2021)	5	-9	7	-11	-2	13	-0.14	[-1.36, 1.05]	0.54	8.1%	0.09	Some concerns
Stevens et al. (2020)	6	-11	7	-11	0	12	0.00	[-1.14, 1.14]	0.52	8.8%	0.09	Some concerns
Number of studies = 6 Total participants = 163					Combined effect:		-0.32	[-0.45, -0.20]	0.05	100%	0.10	
					Prediction interval:			[-0.64, -0.01]				

SMD forest plot:



Note: Zurawlew et al. (2016)a is a 2nd test within Zurawlew et al. (2016).
Note: Lundby et al. (2021)a is a 2nd test within Lundby et al. (2021).

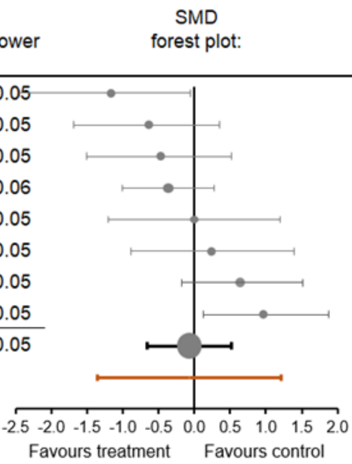
Test for overall effect: Z = -6.10 (P<0.0001)
Heterogeneity: Q = 0.83 (P = 1.00); I² = 0%; Tau² = 0.03, Tau = 0.16
Statistical power = 0.10
GRADE certainty of the evidence = Very low

Fig. 8. Meta-analysis of RPE during submaximal exercise.

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on ratings of perceived exertion (RPE) during submaximal exercise. The meanΔ values in the control and treatment groups represent the post-intervention minus pre-intervention change scores. The mean difference values represent the difference in the change scores between the treatment and control groups, i.e. $(Post_{Treatment} - Pre_{Treatment}) - (Post_{Control} - Pre_{Control})$. The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). The forest plot bubbles represent the standardised mean difference (SMD; Hedges' g) for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. N = sample size; a.u. = arbitrary units; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I^2 = percentage variability in effect sizes across studies; Tau^2 and Tau are measures of the dispersion of true effect sizes between studies.

RPE during submaximal exercise.

Study	Control N	Control MeanΔ (a.u.)	Treatment N	Treatment MeanΔ (a.u.)	Mean difference (a.u.)	Pooled baseline SD	SMD	95%CI	SE	Weight	Power	Overall Risk of bias
McIntyre et al. (2022)	7	0.0	7	-2.0	-2.0	1.6	-1.17	[-2.44, -0.06]	0.55	10.6%	0.05	High
Zurawlew et al. (2016)	7	0.0	10	-1.0	-1.0	1.5	-0.63	[-1.69, 0.36]	0.48	12.0%	0.05	Some concerns
Zurawlew et al. (2016)a	7	0.0	10	-1.0	-1.0	2.0	-0.47	[-1.51, 0.51]	0.47	12.1%	0.05	Some concerns
Kirby et al. (2021; 2021a)	15	0.0	27	-1.0	-1.0	2.7	-0.36	[-1.01, 0.27]	0.32	16.2%	0.06	High
McIntyre et al. (2021)	5	-1.0	7	-1.0	0.0	2.0	0.00	[-1.20, 1.20]	0.54	10.7%	0.05	Some concerns
Stevens et al. (2020)	6	-1.3	7	-0.9	0.4	1.6	0.23	[-0.89, 1.39]	0.52	11.1%	0.05	Some concerns
Lundby et al. (2021)	11	-1.0	13	0.0	1.0	1.5	0.64	[-0.18, 1.51]	0.41	13.8%	0.05	Some concerns
Lundby et al. (2021)a	11	0.0	13	1.0	1.0	1.0	0.97	[0.13, 1.87]	0.42	13.5%	0.05	Some concerns
Number of studies = 6 Total participants = 163					Combined effect:		-0.07	[-0.65, 0.51]	0.25	100%	0.05	
					Prediction interval:			[-1.36, 1.22]				



Note: Zurawlew et al. (2016)a is a 2nd test within Zurawlew et al. (2016).
Note: Lundby et al. (2021)a is a 2nd test within Lundby et al. (2021).

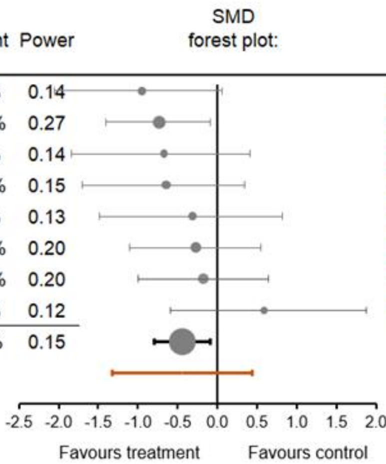
Test for overall effect: Z = -0.29 (P = 0.77)
Heterogeneity: Q = 16.52 (P = 0.02); I^2 = 57.6%; Tau^2 = 0.43, Tau = 0.66
Statistical power = 0.05
GRADE certainty of the evidence = Very low

Fig. 9. Meta-analysis of core temperature during submaximal exercise.

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on core temperature during submaximal exercise. The meanΔ values in the control and treatment groups represent the post-intervention minus pre-intervention change scores. The mean difference values represent the difference in the change scores between the treatment and control groups, i.e. ($Post_{Treatment}$ - $Pre_{Treatment}$) - ($Post_{Control}$ - $Pre_{Control}$). The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). The forest plot bubbles represent the standardised mean difference (SMD; Hedges' g) for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. N = sample size; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I^2 = percentage variability in effect sizes across studies; Tau^2 and Tau are measures of the dispersion of true effect sizes between studies.

Core temperature during submaximal exercise.

Study	Control N	Control MeanΔ (°C)	Treatment N	Treatment MeanΔ (°C)	Mean difference (°C)	Pooled baseline SD	Effect estimate SMD	95%CI	SE	Weight	Power	Overall Risk of bias
Zurawlew et al. (2016)	7	0.04	10	-0.28	-0.32	0.32	-0.95	[-2.06, 0.06]	0.50	9.9%	0.14	Some concerns
Kirby et al. (2021; 2021a)	15	0.00	27	-0.30	-0.30	0.40	-0.74	[-1.41, -0.09]	0.33	22.8%	0.27	High
McIntyre et al. (2022)	7	-0.41	7	-0.64	-0.23	0.32	-0.67	[-1.84, 0.41]	0.52	9.1%	0.14	High
Zurawlew et al. (2016)a	7	0.00	10	-0.36	-0.36	0.53	-0.64	[-1.70, 0.35]	0.48	10.5%	0.15	Some concerns
Stevens et al. (2020)	6	0.00	7	-0.10	-0.10	0.30	-0.31	[-1.48, 0.81]	0.52	8.9%	0.13	Some concerns
Lundby et al. (2021)	11	-0.20	13	-0.30	-0.10	0.36	-0.27	[-1.10, 0.55]	0.40	15.4%	0.20	High
Lundby et al. (2021)a	11	0.00	13	-0.10	-0.10	0.56	-0.17	[-1.00, 0.64]	0.40	15.4%	0.20	High
McIntyre et al. (2021)	5	-0.24	7	-0.10	0.14	0.24	0.59	[-0.60, 1.87]	0.55	7.9%	0.12	Some concerns
Number of studies = 6 Total participants = 163					Combined effect:		-0.44	[-0.79, -0.09]	0.15	100%	0.15	
					Prediction interval:			[-1.32, 0.44]				



Note: Zurawlew et al. (2016)a is a 2nd test within Zurawlew et al. (2016).

Note: Lundby et al. (2021)a is a 2nd test within Lundby et al. (2021).

Test for overall effect: Z = -2.95 (P = 0.003)

Heterogeneity: Q = 6.41 (P = 0.49); I^2 = 0%; Tau^2 = 0.20, Tau = 0.45

Statistical power = 0.15

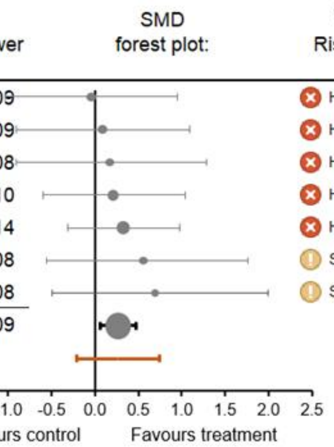
GRADE certainty of the evidence = Low

Fig. 10. Meta-analysis of sweat rate during submaximal exercise.

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on sweat rate during submaximal exercise. The meanΔ values in the control and treatment groups represent the post-intervention minus pre-intervention change scores. The mean difference values represent the difference in the change scores between the treatment and control groups, i.e. $(Post_{Treatment} - Pre_{Treatment}) - (Post_{Control} - Pre_{Control})$. The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). The forest plot bubbles represent the standardised mean difference (SMD; Hedges' g) for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. N = sample size; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I^2 = percentage variability in effect sizes across studies; Tau^2 and Tau are measures of the dispersion of true effect sizes between studies.

Sweat rate during submaximal exercise.

Study	Control N	Control MeanΔ (L/h)	Treatment N	Treatment MeanΔ (L/h)	Mean difference (L/h)	Pooled baseline SD	SMD	95%CI	SE	Weight	Power	Overall Risk of bias
Zurawlew et al. (2016)a	7	0.01	10	0.00	-0.01	0.23	-0.04	[-1.04, 0.95]	0.47	12.6%	0.09	High
Zurawlew et al. (2016)	7	0.00	10	0.01	0.01	0.11	0.09	[-0.91, 1.09]	0.47	12.6%	0.09	High
McIntyre et al. (2022)	7	0.07	7	0.10	0.03	0.16	0.18	[-0.90, 1.28]	0.50	11.0%	0.08	High
Lundby et al. (2021)	11	0.24	13	0.36	0.12	0.54	0.21	[-0.60, 1.04]	0.40	17.6%	0.10	High
Kirby et al. (2021; 2021a)	15	-0.22	27	0.02	0.24	0.73	0.32	[-0.31, 0.97]	0.32	27.4%	0.14	High
Stevens et al. (2020)	6	0.34	7	0.55	0.21	0.35	0.56	[-0.56, 1.76]	0.53	9.9%	0.08	Some concerns
McIntyre et al. (2021)	5	-0.03	7	0.12	0.15	0.22	0.69	[-0.49, 1.99]	0.56	8.9%	0.08	Some concerns
Number of studies = 6 Total participants = 139					Combined effect:		0.27	[0.06, 0.47]	0.08	100%	0.09	
Note: Zurawlew et al. (2016)a is a 2nd test within Zurawlew et al. (2016).					Prediction interval:			[-0.21, 0.74]				



Test for overall effect: Z = 3.17 (P = 0.002)

Heterogeneity: Q = 1.55 (P = 0.96); I^2 = 0%; Tau^2 = 0.06, Tau = 0.24

Statistical power = 0.09

GRADE certainty of the evidence = Very low

Fig. 11. Meta-analysis of thermal sensations during submaximal exercise.

A random effects model was used to determine the effect of post-exercise heat exposure (heat acclimation) on thermal sensations during submaximal exercise. The meanΔ values in the control and treatment groups represent the post-intervention minus pre-intervention change scores. The mean difference values represent the difference in the change scores between the treatment and control groups, i.e. ($Post_{Treatment}$ - $Pre_{Treatment}$) - ($Post_{Control}$ - $Pre_{Control}$). The overall risk of bias for this outcome in each study is rated as high (×), low (✓), or some concerns (!). The forest plot bubbles represent the standardised mean difference (SMD; Hedges' g) for each study. The size of the bubbles represents the weights of each study in the inverse variance model. Error bars represent the 95% confidence interval (CI). The 95% prediction interval is shown in orange. N = sample size; a.u. = arbitrary units; SE = standard error; Z = z-test statistic and the corresponding p-value; Q = Cochrane's Q; I^2 = percentage variability in effect sizes across studies; Tau^2 and Tau are measures of the dispersion of true effect sizes between studies.

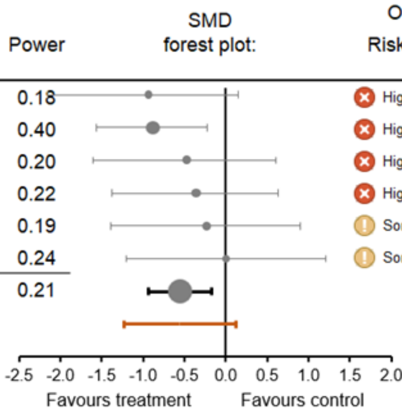
Thermal sensations during submaximal exercise.

Study	Control N	Control MeanΔ (a.u.)	Treatment N	Treatment MeanΔ (a.u.)	Mean difference (a.u.)	Pooled baseline SD	Effect estimate SMD	95%CI	SE	Weight	Power	Overall Risk of bias
McIntyre et al. (2022)	7	0.00	7	-1.00	-1.00	1.00	-0.94	[-2.16, 0.16]	0.53	12.6%	0.18	High
Kirby et al. (2021; 2021a)	15	0.04	27	-0.83	-0.87	0.97	-0.88	[-1.56, -0.23]	0.33	32.5%	0.40	High
Zurawlew et al. (2016)	7	0.10	7	-0.20	-0.30	0.60	-0.47	[-1.61, 0.61]	0.51	13.7%	0.20	High
Zurawlew et al. (2016)a	7	-0.40	10	-0.80	-0.40	1.06	-0.36	[-1.38, 0.63]	0.47	15.9%	0.22	High
Stevens et al. (2020)	6	-0.30	7	-0.40	-0.10	0.40	-0.23	[-1.39, 0.89]	0.52	13.1%	0.19	Some concerns
McIntyre et al. (2021)	5	0.00	7	0.00	0.00	1.00	0.00	[-1.20, 1.20]	0.54	12.1%	0.24	Some concerns
Combined effect:							-0.56	[-0.94, -0.17]	0.15	100%	0.21	

Number of studies = 5   Total participants = 112

Note: Zurawlew et al. (2016)a is a 2nd test within Zurawlew et al. (2016).

Prediction interval: [-1.23, 0.12]



Test for overall effect: Z = -3.73 (P = 0.0002)

Heterogeneity: Q = 3.13 (P = 0.68); I^2 = 0%; Tau^2 = 0.12, Tau = 0.34

Statistical power = 0.21

GRADE certainty of the evidence = Low

Fig. 12. Funnel plots to represent potential publication bias.

Contour-enhanced funnel plots between the natural log of the ratio of means (the effect size estimate) and the standard error of the ratio of means were constructed for: (A) performance in hot conditions, and (B) performance in cool conditions. Contour-enhanced funnel plots between the standardised mean difference (Hedges' g; the effect size estimate) and the standard error of the standardised mean difference were constructed for: (C) VO_2max and (D) speed at lactate threshold, as well as (E) heart rate, (F) RPE, (G) core temperature, (H) sweat rate, and (I) thermal sensations during submaximal exercise. Grey-filled circles represent the individual studies. The vertical light grey line represents the combined effect estimate: ln(ratio of mean) in panels A and B, and Hedges' g in panels C to I. The t-statistics and P-values from Egger's regression are shown. The regions inside the solid orange funnels represent P>0.10; the regions between the solid orange lines and the dashed orange lines represent 0.10>P>0.05; the region between the dashed orange lines and the dotted orange lines represent 0.05>P>0.01; and the regions outside the dotted orange funnels represent P<0.01. IMPORTANT: Because the included studies have a small sample size and are few in number, there is low certainty in the validity of the funnel plots and they should be interpreted with caution.