



Acute caffeine and carbohydrate strategies for fatigue-related sport-specific performance in team-sport athletes: a systematic review and multilevel meta-analysis

Shu Xie & Xueming Wen

To cite this article: Shu Xie & Xueming Wen (2026) Acute caffeine and carbohydrate strategies for fatigue-related sport-specific performance in team-sport athletes: a systematic review and multilevel meta-analysis, *Journal of the International Society of Sports Nutrition*, 23:1, 2731524, DOI: [10.1080/15502783.2026.2731524](https://doi.org/10.1080/15502783.2026.2731524)

To link to this article: <https://doi.org/10.1080/15502783.2026.2731524>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 21 Oct 2026.



[Submit your article to this journal](#)



Article views: 12



[View related articles](#)



[View Crossmark data](#)

Acute caffeine and carbohydrate strategies for fatigue-related sport-specific performance in team-sport athletes: a systematic review and multilevel meta-analysis

Shu Xie  and Xueming Wen 

Shanghai University of Sport, Shanghai, China

ABSTRACT

Background: Although caffeine and carbohydrate are established ergogenic strategies for physical performance, it remains unclear whether these benefits translate to the decision-making and sport-specific skills that determine team-sport performance under fatigue.

Methods: PubMed, Web of Science Core Collection, and SPORTDiscus were searched on 28 July 2026. Thirty-one reports were retained; their reported analyzed sample sizes summed to 441 athletes. Twenty-three independent quantitative study clusters, with reported analyzed sample sizes summing to 341 athletes, contributed 67 effects. Multilevel random-effects models nested effect sizes within studies; primary assumptions were $r = 0.5$ and $\rho = 0.80$. Inference used CR2 robust standard errors and Satterthwaite degrees of freedom.

Results: Soccer supplied 11 of the 23 quantitative clusters; every other sport category supplied four or fewer. Broad caffeine was centered near zero (Hedges' $g = 0.040$, 95% CI -0.211 to 0.291 ; $p = 0.691$; $df = 4.41$). Tier 1a perceptual-motor and Tier 1b technical/tactical estimates were imprecise. Broad carbohydrate was positive but uncertain ($g = 0.256$, 95% CI -0.094 to 0.605 ; $p = 0.118$; $df = 4.96$), and one study cluster materially influenced this estimate. All six GRADE comparisons were rated very low certainty. The two-study mental-fatigue calculation was too fragile for conventional inference.

Conclusions: Current evidence does not support assuming that the established physical ergogenic effects of caffeine translate to fatigue-related decision-making or sport-specific skill. Carbohydrate may provide a small benefit, but the evidence remains sparse and highly uncertain. Sensitivity to study composition was driven primarily by one influential cluster, whose omission increased the pooled estimate. The carbohydrate estimates apply only to heterogeneous ingested-beverage protocols and should not be extrapolated to mouth-rinse or gel strategies. Parallel syntheses do not establish intervention superiority or combined-strategy synergy.

ARTICLE HISTORY

Received 5 August 2026

Accepted 4 September 2026

KEYWORDS

Caffeine; carbohydrate; team sport; fatigue; sport-specific performance; meta-analysis

1. Introduction

Although caffeine and carbohydrate are established ergogenic strategies for physical performance, it remains unclear whether their benefits translate to the decision-making and sport-specific skills that determine team-sport performance under fatigue. The practical question is not simply whether a strategy can improve running capacity or perceived exertion, but whether it produces a favourable intervention-control difference in the information processing and skilled execution required late in play. Existing physical-performance evidence cannot answer that question indirectly.

Team-sport performance under fatigue requires athletes to detect relevant information, select an action, and execute a skilled response under time pressure. In this review, perceptual-motor performance denotes task-relevant information processing expressed through a motor response, whereas sport-specific technical/tactical performance denotes ecologically relevant execution such as passing, shooting, dribbling, serving, or tactical play. These constructs are related but are not interchangeable with laboratory cognition,

CONTACT Xueming Wen  wenxueming0519@163.com  Shanghai University of Sport, 399 Changhai Road, Yangpu District, Shanghai 200438, China

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15502783.2026.2731524>.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

sprint capacity, or general work output. Mental fatigue is also physiologically distinct from exercise- or match-induced fatigue and should not be pooled with it merely to increase the evidence base [1].

Caffeine and carbohydrate are commonly used acute strategies. Caffeine can influence arousal, perceived effort, neuromuscular function, and exercise tolerance, although responses vary by dose, individual, task, and context [2,3]. Carbohydrate may support performance through substrate availability and oral-sensing or central pathways; broader endurance and soccer syntheses report heterogeneous physical, technical, cognitive, and gastrointestinal findings [1,4]. Evidence that caffeine plus carbohydrate improves physical performance cannot be assumed to demonstrate synergy for fatigue-stage perceptual-motor or technical outcomes [5]. Accordingly, caffeine and carbohydrate require parallel, outcome-specific syntheses rather than an indirect ranking of separate pooled estimates.

The evidence also poses an inferential challenge: one trial may contribute several correlated tasks, time points, or scoring variants. Treating those effects as independent can overstate precision, whereas selecting one outcome per study discards information. Multilevel meta-analysis retains the effects while modelling their nesting within studies [6]. CR2 robust variance estimation with Satterthwaite degrees of freedom is particularly relevant when the number of independent clusters is small [7–9]. We therefore conducted a systematic review and multilevel meta-analysis of acute caffeine and carbohydrate strategies for fatigue-related sport-specific performance in team-sport athletes. The prespecified analysis dataset and statistical code were finalised before the formal analyses and were not modified in response to the results. We analysed caffeine and carbohydrate separately, distinguished perceptual-motor from technical/tactical outcomes, retained general cognition/coordination as secondary, and analysed mental fatigue separately. The primary specification used REML, $r = 0.5$, $\rho = 0.80$, and CR2/Satterthwaite inference; sensitivity analyses varied r and ρ , excluded high-risk studies, varied the standardisation scale, and examined influence and model diagnostics.

The primary research question was: among trained team-sport athletes, what are the intervention-control effects of acute caffeine and carbohydrate strategies on perceptual-motor and sport-specific technical/tactical performance during or following physical or match-related fatigue? Secondary questions addressed general cognition/coordination and mental-fatigue evidence separately and examined the robustness of estimates to correlation, risk-of-bias, standardisation, and influence assumptions.

2. Materials and methods

2.1. Reporting framework and protocol status

The review was reported in accordance with PRISMA 2020 [10] and registered in PROSPERO on 28 July 2026 (PROSPERO CRD420261463593). At registration, formal searching and title/abstract screening were recorded as complete, whereas data extraction, risk-of-bias assessment, and synthesis had not been completed. The registration therefore preceded extraction and quantitative synthesis but not study identification. The prespecified analysis dataset and statistical code were finalised before the formal analyses and were not modified in response to the results. Effect sizes, study identifiers, outcome directions, risk-of-bias judgements, and analysis-group assignments were retained unchanged.

2.2. Information sources and searches

PubMed, Web of Science Core Collection, and SPORTDiscus with Full Text (EBSCOhost) were searched on 28 July 2026. Search concepts combined caffeine and carbohydrate interventions with team-sport populations and randomised or controlled trial terminology; the focused PubMed search additionally targeted perception-action, cognitive, technical, tactical, and motor-performance terms. SPORTDiscus was searched using the TX All Text field. No restrictions were imposed by publication year, sex, age, or language. Complete database-specific search strategies, including field codes and the focused PubMed search, are reproduced verbatim in Supplementary Table S2. Records were deduplicated sequentially by exact DOI, exact PMID, and exact normalised title; no records were merged solely on a fuzzy title match. This process removed 642 duplicates and produced 726 unique records, including 56 unique additions from the focused search. The PROSPERO record listed MEDLINE, PubMed, and Scopus as planned main sources; the implemented search instead used

PubMed, Web of Science, and SPORTDiscus, with a focused supplementary PubMed search. The substitution retained biomedical coverage through PubMed, including MEDLINE-indexed records, while adding multi-disciplinary citation coverage through Web of Science and sport-specific indexing through SPORTDiscus. This source change is reported as a protocol deviation rather than being retrospectively harmonised.

2.3. Eligibility criteria

Eligible participants were trained adolescent or adult team-sport athletes or players. Studies of non-athletes, individual endurance- or strength-sport samples, or general student samples without a team-sport background were excluded unless team-sport data were separately extractable. Eligible interventions comprised acute caffeine, coffee or another isolatable caffeine delivery method, carbohydrate ingestion or mouth rinse, and caffeine plus carbohydrate. Chronic dietary programmes, recovery-only protein interventions, and multi-ingredient products in which the target component could not be isolated were excluded. Eligible comparators included placebo, water, non-caloric or electrolyte controls, active caffeine or carbohydrate comparators, and dose comparisons.

Randomised crossover and parallel acute-intervention trials formed the main synthesis; nonrandomized controlled evidence was retained only for narrative or sensitivity purposes. R0042 was appraised with ROBINS-I and excluded from all randomised-trial main models. Outcomes had to assess perception-action performance or sport-specific technical/tactical execution during or after fatiguing exercise, simulated match play, real competition, a late-game period, or another prespecified fatigue exposure. Reactive agility, decision-linked movement, action-linked reaction time, passing, shooting, dribbling, serving, and tactical measures were eligible primary outcomes. Isolated jump, sprint, planned change-of-direction, biochemical, or general physical-output measures without a perception-action or skill component were not included in the primary synthesis. General laboratory cognition and coordination were retained as secondary outcomes. Studies using prolonged cognitive effort to induce mental fatigue were retained in a separate subgroup and were never pooled with physical or match-related fatigue.

2.4. Study selection

Xueming Wen (X.W.) and Shu Xie (S.X.) independently screened every title and abstract against the eligibility criteria. They then compared decisions and resolved differences through discussion until consensus was reached. Of 726 unique records, 692 were excluded at title/abstract screening and 34 reports were sought for full-text assessment. Two reports could not be retrieved, 32 were assessed in full, and one was excluded for an ineligible intervention, leaving 31 reports in the review. The final set comprised 25 physical/match-fatigue reports, three secondary cognition/coordination reports, two mental-fatigue reports, and one nonrandomized report retained only for narrative/ROBINS-I purposes. Final per-record consensus decisions were recorded in a screening ledger and reconciled with the full-text assessment log: 34 records proceeded to full-text retrieval and 692 were excluded at title/abstract screening. The ledger contained no blank final decisions and reproduced the PRISMA flow counts.

2.5. Data extraction and outcome hierarchy

Each retained contrast–outcome–timepoint combination received a unique Effect_ID. Record_ID identified the report-level cluster and Study_ID identified the study or participant sample. In the quantitative dataset, Study_ID and Record_ID mapped one-to-one; clustering by Record_ID was therefore algebraically identical to clustering by Study_ID for the observed data. When several assessments were available, the prespecified rule prioritised late or post-fatigue measurements relevant to performance during or following fatigue. Multiple eligible outcomes and time points were retained rather than selecting a single result, and their dependence was handled in the statistical model.

Primary outcomes were organised into Tier 1a perceptual-motor outcomes and Tier 1b sport-specific technical/tactical outcomes. Tier 1a included reactive-agility tasks with an external stimulus, decision time or accuracy coupled to movement, representative reaction or composure measures, and attack/counterattack decisions. Tier 1b included passing, shooting, dribbling, ball handling, technical-action quality, and tactical

indices. Operationally, Tier 1a required a scored real-time response to an external or changing stimulus, with perception or choice integral to the extracted outcome. Tier 1b captured the quality or success of sport-specific execution or aggregate tactical behaviour when a stimulus-contingent perceptual or decision component was not scored separately. When a task contained both features, classification followed the construct represented by the extracted effect rather than the task label. Laboratory attention, Stroop, simple- or choice-reaction-time, memory, and eye-hand coordination measures that lacked a representative sport action were assigned to Tier 3 and analysed only in the secondary cognition/coordination model. The classification applied to every effect is supplied in Supplementary Table S7. Broad caffeine and carbohydrate sets included all eligible physical/match-fatigue effects within the corresponding intervention family. Strict sets excluded rows designated as exploratory complex-factorial or practical beverage-strategy sensitivity evidence. Mental fatigue constituted a separate subgroup. The single caffeine-plus-carbohydrate add-on contrast was retained descriptively but was insufficient to estimate a pooled interaction or synergy effect.

Within each intervention family, different doses, timings, and delivery formats were eligible as heterogeneous acute strategies; they were not assumed to be pharmacologically identical or dose-equivalent. At review level, the carbohydrate evidence included mouth-rinse and gel reports, but neither format contributed an effect to P6 or P7. The quantitative carbohydrate models therefore combined ingested beverage or solution protocols that varied in formulation, concentration, dose, and timing. These estimates should not be extrapolated to mouth-rinse or gel strategies. Route- or dose-specific subgroup analysis or meta-regression was not fitted because each intervention-specific synthesis contained too few independent clusters for reliable estimation. Study-level formulations remained visible in Supplementary Table S1, and pooled estimates were interpreted as averages across the implementations represented in each model.

2.6. Risk of bias

X.W. and S.X. independently assessed risk of bias, compared their judgements, and resolved differences through discussion. Randomised reports were evaluated using RoB 2, including domains for randomisation, deviations from intended interventions, missing outcome data, outcome measurement, and selection of the reported result; crossover-specific carryover concerns were recorded where applicable. The assessment targeted the eligible late/post-fatigue outcome set. Where a report contributed several eligible outcomes or time points, a report-level summary judgement was used for synthesis and sensitivity analysis rather than a separate judgement for every effect. R0042 was evaluated separately with ROBINS-I. RoB 2 categories (Low risk, Some concerns, and High risk) were not combined with ROBINS-I categories. A prespecified sensitivity analysis excluded randomised studies judged High risk under RoB 2 while retaining the same outcome definitions, effect-size input, covariance assumptions, random structure, and inferential procedure.

2.7. Effect-size calculation

Continuous outcomes measured on different scales were expressed as Hedges' g . The primary effect was a paired standardised mean difference based on the standard deviation of within-participant differences. The direction was harmonised so that positive values favoured the intervention. For higher-is-better outcomes, the benefit-aligned mean difference was the intervention mean minus the control mean; for lower-is-better outcomes, it was the control mean minus the intervention mean. For paired or crossover comparisons, the standard deviation of the within-participant difference was calculated as $SD_{diff} = \sqrt{SD_I^2 + SD_C^2 - 2rSD_{I \times C}}$, where SD_I and SD_C are the intervention and control standard deviations and r is the assumed within-participant correlation. The small-sample correction was $J = 1 - 3/(4N - 5)$, and $g = J(MD/SD_{diff})$. The corresponding sampling variance was $J^2[1/N + (MD/SD_{diff})^2/(2(N - 1))]$. Study-specific paired correlations were not reported in a form that could be recovered consistently for the extracted outcomes. We therefore prespecified $r=0.5$ as a pragmatic mid-range working value rather than borrowing an empirical correlation from non-equivalent tasks, scales, or fatigue protocols. Values calculated with $r=0.3$ and 0.7 tested the sensitivity of the standardised effects and their sampling variances to this choice. In a separate standardisation-scale sensitivity analysis, the same benefit-aligned paired mean difference was divided by the square root of the average condition variance, $\sqrt{(SD_I^2 + SD_C^2)/2}$, with the Hedges correction applied. The assumed paired correlation ($r=0.5$) then entered the sampling variance but not the standardising denominator. The same

model filters, covariance assumption ($\rho^* = 0.80$), nested random structure, and CR2/Satterthwaite procedure were retained.

2.8. Multilevel meta-analysis and robust inference

Sampling-error covariance matrices were constructed with `metafor::vcalc()`, clustering effects by Record_ID and identifying each observation by Effect_ID. Because Study_ID and Record_ID were one-to-one in the quantitative dataset, the independent CR2 clusters corresponded to the observed study/participant-sample units. Because the exact within-study sampling correlations were generally unavailable, the primary working correlation was $\rho = 0.80$. A multilevel random-effects model was then fitted with `metafor::rma.mv()`, using restricted maximum likelihood and random intercepts for effects nested within records (Record_ID/Effect_ID) [6]. The two estimated variance components are reported as study-level tau-squared and effect-level tau-squared.

Inference for each pooled intercept used CR2 cluster-robust standard errors with the record as the independent cluster and Satterthwaite t tests and confidence intervals. Robust variance methods retain internally correlated effects without requiring the working covariance model to be exactly correct [7]. CR2 and Satterthwaite corrections were selected because uncorrected cluster-robust tests can be anticonservative with few clusters [8,9]. Models with fewer than three study clusters were designated “too few studies for CR2” in the prespecified grid. A separate two-study calculation was attempted for the mental-fatigue subgroup only as an explicitly fragile sensitivity result; it did not replace the grid decision or enter any physical/match-fatigue model.

The prespecified models were: any eligible caffeine or carbohydrate strategy versus control (P1); caffeine versus control, broad (P2) and strict (P3); caffeine Tier 1a (P4) and Tier 1b (P5); carbohydrate versus a non-carbohydrate control, broad (P6) and strict (P7); secondary cognition/coordination (P8); and the separate mental-fatigue subgroup (P9). P1 was retained as a prespecified descriptive omnibus check. Because caffeine and carbohydrate differ mechanistically and contextually, P1 was not used as the primary intervention-specific inference or interpreted as a common treatment-class effect. P2-P9 are parallel syntheses. No model directly contrasted caffeine with carbohydrate, and no pooled model estimated a caffeine-by-carbohydrate interaction.

2.9. Sensitivity analyses and diagnostics

The complete model grid crossed $r = 0.3, 0.5$, and 0.7 ; $\rho = 0.50$ and 0.8 ; inclusion and exclusion of high-risk studies; and P1-P9. The primary specification was $r = 0.5$, $\rho = 0.80$, with all eligible randomised studies. We assessed missingness, duplicate effect IDs and duplicate rows, model convergence, optimiser use, and positive definiteness of every working covariance matrix. Effect-level flags were generated for $|g| > 1$, extreme $|g| > 2$, or an absolute sampling-standardised residual above 3. Study influence was examined using cluster Cook's distance with $4/k$ as a heuristic, where k is the number of study clusters, and leave-one-record-out CR2 refits were performed for P2 and P4-P8. Flagged effects or studies were not automatically excluded. Small-study-effect tests were not conducted for intervention-specific models because each contained fewer than 10 independent clusters and the outcome structures were heterogeneous.

Analyses were conducted in R 4.6.1 using `metafor` 5.0-1, `clubSandwich` 0.7.0, `readr` 2.2.0, `dplyr` 1.2.1, `purrr` 1.2.2, `tibble` 3.3.1, and `tidyr` 1.3.2. Optimiser convergence was tracked for every fitted cell. Iteration limits were increased and an `optim/BFGS` fallback was used only when `nlminb` returned a non-convergence error; this numerical safeguard did not alter the likelihood, random structure, covariance assumptions, effect input, or inferential method. Complete scripts, package versions, session information, and diagnostic files are provided as supplementary material.

2.10. Certainty of evidence

Certainty was rated using the GRADE framework for the broad caffeine, caffeine Tier 1a, caffeine Tier 1b, broad carbohydrate, strict carbohydrate, and general cognition/coordination comparisons [11]. Randomised evidence began at High certainty and was evaluated for risk of bias, inconsistency, indirectness, imprecision, and publication

bias. Risk-of-bias judgements, effect-level heterogeneity, cluster counts, Satterthwaite degrees of freedom, confidence-interval width, and influence diagnostics informed downgrading. Publication bias was not assessed statistically because every comparison contained fewer than 10 independent clusters. No automatic downgrading was applied solely because statistical assessment was unavailable. Domain judgements and row-specific downgrading reasons are reported in Supplementary Table S3.

2.11. Generative artificial intelligence use

OpenAI Codex was used after the analytical specification had been finalised to assist with code-execution guidance, language editing, document generation, file organisation, and file-level quality-control checks. It did not independently select studies, adjudicate risk of bias, modify effect sizes or study identifiers, determine outcome direction, or alter analysis groupings. The authors reviewed all generated material and verified every reported quantitative result against the formal R outputs.

3. Results

3.1. Study flow and analytical dataset

The database and focused searches yielded 1,368 records before deduplication; 642 duplicates were removed, leaving 726 unique records for title/abstract screening. X.W. and S.X. excluded 692 records after independent assessment and consensus. Thirty-four reports were sought, two could not be retrieved, and 32 were assessed in full. One report was excluded because the intervention was ineligible, leaving 31 reports in the review: 25 in the physical/match-fatigue evidence stream, three in secondary cognition/coordination, two in mental fatigue, and one nonrandomized report retained only for narrative/ROBINS-I purposes (Figure 1). Study characteristics, analysed sample sizes, intervention delivery, performance setting, eligible outcomes, and quantitative-inclusion status are provided in Supplementary Table S1.

Across the 31 reports, analysed sample sizes ranged from 6 to 24 and summed to 441. The 23 independent clusters entering the quantitative synthesis had reported analysed sample sizes summing to 341. These totals are sums of report- or study-level analysed sample sizes and should not be interpreted as a guaranteed count of unique individuals across all publications, because unreported cross-report overlap cannot be excluded. Soccer dominated the evidence base, with 17 reports overall and 11 quantitative clusters. The remaining evidence included four basketball reports, all quantitative; two volleyball reports, both quantitative; and two rugby-sevens reports, both quantitative. Rugby union contributed two reports, one quantitative. Beach volleyball contributed one quantitative report, whereas the only field-hockey report was narrative. Two additional quantitative reports included mixed team-sport or team-game samples. No non-soccer sport contributed more than four quantitative clusters [12–42].

Three reports were embedded in live match or tournament contexts, six used small-sided-game or team-training settings, 20 used controlled or simulated physical-fatigue or match protocols, and two used mental-fatigue inductions. Two of the three live-context reports contributed quantitative data. These descriptors characterise ecological setting only and were not used as additional model strata.

Caffeine implementations included ingestion of 3–6 mg/kg, a 200-mg gum, other gum, mouth rinse, espresso coffee, an energy drink, capsules, and beverages; reported timing ranged from 60 min before testing to simulated half-time or after 90 min. Carbohydrate implementations included mouth rinse; 6–12% carbohydrate–electrolyte solutions; intakes reported as approximately 52–90 g/h, 60 g, or 1.0 g/kg; gels; a low-glycemic-index food; and isomaltulose or maltodextrin drinks administered before or during testing, including late or extra-time administration. Exact formulations, comparators, and protocols are listed report by report in Supplementary Table S1.

The quantitative synthesis contained 67 unique effect sizes from 23 independent study clusters. Fifty-one effects from 15 clusters belonged to the primary physical/match-fatigue set, 10 effects from four clusters to secondary cognition/coordination, four effects from two clusters to mental fatigue, and two effects from two clusters to active-comparator sensitivity analyses. Ecological classification yielded 13 Tier 1a perceptual-motor effects from five studies, 44 Tier 1b technical/tactical effects from 14 studies, and 10 Tier 3 cognition/coordination effects from four studies. After grouping synonymous labels descriptively, recurring domains

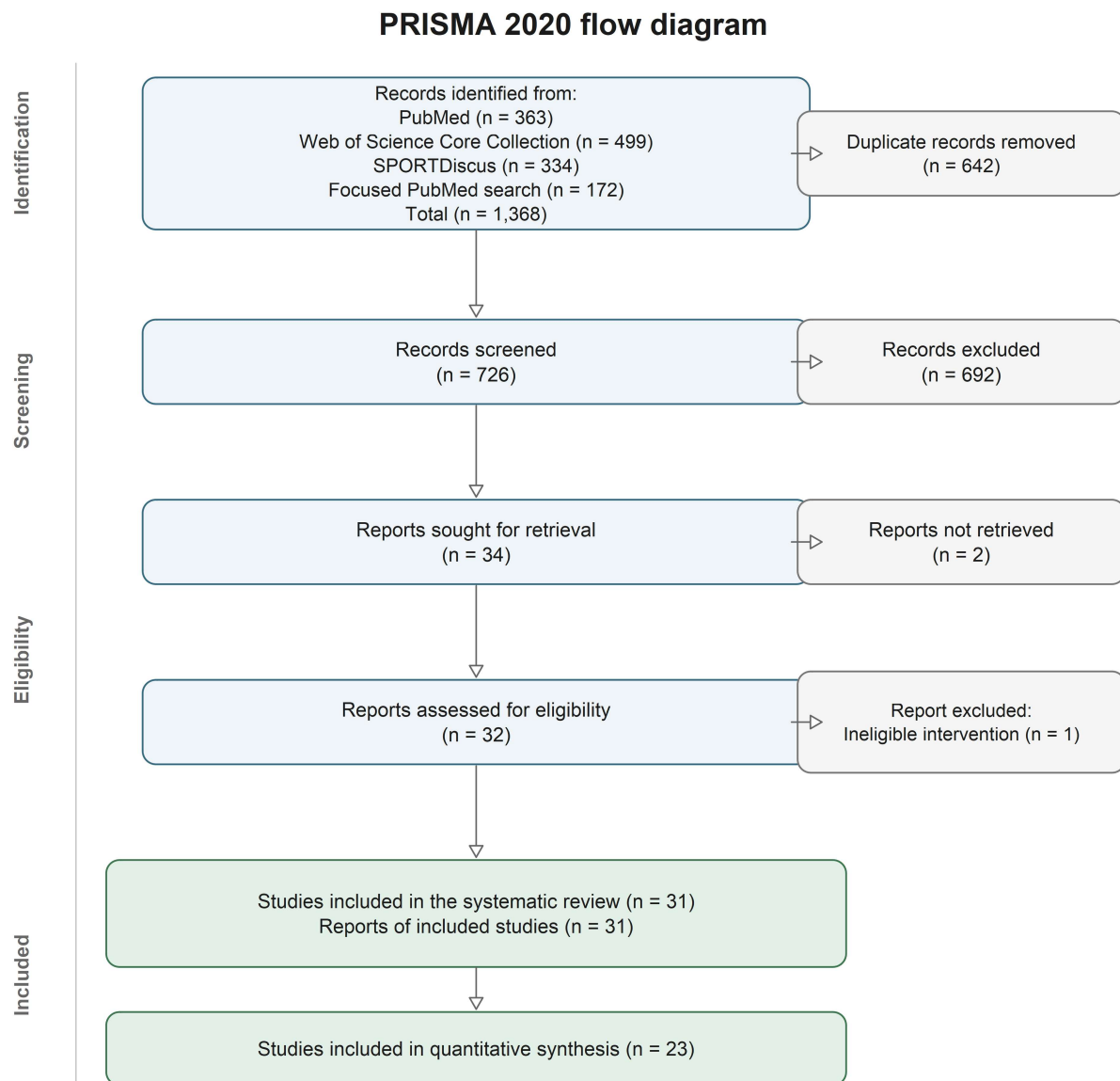


Figure 1. PRISMA 2020 flow diagram. The 31 retained reports comprised 25 physical/match-fatigue reports, three secondary cognition/coordination reports, two mental-fatigue reports, and one nonrandomized report retained for narrative ROBINS-I assessment. The quantitative synthesis included 23 independent study clusters contributing 67 effects.

included shooting (12 effects from five studies), tactical indices (10 effects from one study), passing (nine effects from eight studies), reactive agility (eight effects from three studies), and dribbling (eight effects from three studies). Technical-action quality contributed five effects from one study, and laboratory cognition/coordination contributed 10 effects from four studies. The remaining effects involved attack or counterattack decisions, representative reaction or composure processes, and other task-specific measures. Caffeine-only versus control contributed 32 effects, carbohydrate versus a non-carbohydrate control 26 effects, and the remaining nine represented active-comparator, combined-strategy, or other sensitivity evidence. Supplementary Table S7 provides the effect-by-effect tier, grouped domain, reported measure, selected time point, and analysis-set mapping.

3.2. Data integrity and risk of bias

All 67 effect IDs were unique, no complete duplicate rows were identified, and the variables required by the primary and sensitivity models contained no missing values. R0042 was absent from the quantitative CSV

and therefore from every randomised main model. Among the 30 randomised reports assessed with RoB 2, 28 were judged to have Some concerns and two were judged to be at High risk (R0002 and R0017). The nonrandomized report R0042 was assessed separately with ROBINS-I and was judged to be at Serious risk of bias; it did not enter the randomised quantitative synthesis. Within the 23-study quantitative dataset, 21 clusters had Some concerns and two were at High risk. X.W. and S.X. independently rechecked the judgements and reached consensus after discussion.

3.3. Descriptive omnibus check and caffeine models

Primary estimates for all prespecified models are summarised in Table 1. For transparency, the prespecified P1 model was retained as a descriptive omnibus check across eligible caffeine and carbohydrate strategies (50 effects, 14 clusters; Hedges' $g = 0.117$, 95% CI -0.062 to 0.297 , $p = 0.174$, $df = 9.43$). It is not interpreted as a common treatment-class effect. Under the primary assumptions ($r = 0.5$; $\rho = 0.80$), the broad caffeine model was centred close to zero (P2: 28 effects, seven clusters; $g = 0.040$, 95% CI -0.211 to 0.291 , $p = 0.691$, $df = 4.41$), as was the strict caffeine model (P3: 22 effects, six clusters; $g = 0.049$, 95% CI -0.312 to 0.409 , $p = 0.709$, $df = 3.36$).

The outcome-tier models were less precise. Tier 1a perceptual-motor outcomes yielded $g = 0.180$ (11 effects, four clusters; 95% CI -0.510 to 0.870 , $p = 0.404$, $df = 2.22$), whereas Tier 1b technical/tactical outcomes yielded $g = -0.059$ (17 effects, three clusters; 95% CI -0.708 to 0.590 , $p = 0.699$, $df = 1.73$). Study-level tau-squared estimates near zero did not imply homogeneity: effect-level tau-squared ranged from 0.135 to 0.261 and intervals remained wide. Degrees of freedom were very low for P3-P4 and extremely low for P5.

3.4. Carbohydrate models

Carbohydrate estimates were directionally positive but did not exclude zero. The broad model yielded $g = 0.256$ (22 effects, seven clusters; 95% CI -0.094 to 0.605 , $p = 0.118$, $df = 4.96$), with study-level tau-squared = 0.020 and effect-level tau-squared = 0.070. The strict model yielded $g = 0.298$ (19 effects, six clusters; 95% CI -0.152 to 0.749 , $p = 0.142$, $df = 4.23$), with study-level tau-squared = 0.043 and effect-level tau-squared = 0.078. These estimates were not formally compared with the caffeine estimates.

3.5. Secondary cognition/coordination and mental fatigue

The secondary cognition/coordination model produced a positive but imprecise estimate (P8: 10 effects, four clusters; $g = 0.288$, 95% CI -0.031 to 0.606 , $p = 0.063$, $df = 2.76$). Study-level tau-squared was near zero and effect-level tau-squared was 0.002. The prespecified mental-fatigue grid model was not fitted because only two clusters were available. A separate exploratory descriptive two-study CR2 calculation at the primary correlation assumptions yielded $g = 0.028$ (four effects; 95% CI -0.662 to 0.719 , $p = 0.694$, $df = 1.00$; study-level tau-squared near zero; effect-level tau-squared = 0.001). Across correlation assumptions, this estimate ranged from 0.025 to 0.033; every Satterthwaite df remained approximately 1. It is not conventional pooled inference and was not combined with physical/match-fatigue studies.

Table 1. Primary multilevel estimates with CR2/Satterthwaite inference.

Model	Effects/studies	Hedges' g (95% CI)	p	df	τ^2 , study	τ^2 , effect
P1 Descriptive omnibus check: any strategy vs control	50/14	0.117 (-0.062 , 0.297)	0.174	9.43	<0.001	0.135
P2 Caffeine broad	28/7	0.040 (-0.211 , 0.291)	0.691	4.41	<0.001	0.190
P3 Caffeine strict	22/6	0.049 (-0.312 , 0.409)	0.709	3.36	<0.001	0.261
P4 Caffeine Tier 1a	11/4	0.180 (-0.510 , 0.870)	0.404	2.22	<0.001	0.211
P5 Caffeine Tier 1b	17/3	-0.059 (-0.708 , 0.590)	0.699	1.73	<0.001	0.180
P6 Carbohydrate broad	22/7	0.256 (-0.094 , 0.605)	0.118	4.96	0.020	0.070
P7 Carbohydrate strict	19/6	0.298 (-0.152 , 0.749)	0.142	4.23	0.043	0.078
P8 Cognition/coordination	10/4	0.288 (-0.031 , 0.606)	0.063	2.76	<0.001	0.002
P9 Mental fatigue, exploratory descriptive two-study calculation	4/2	0.028 (-0.662 , 0.719)	0.694	1.00	<0.001	0.001

3.6. Correlation and risk-of-bias sensitivity analyses

Across $r = 0.3, 0.5$, and 0.7 and $\rho = 0.50$ and 0.8 , the direction of each P1-P8 estimate was stable. Estimate ranges were 0.090-0.132 for P1, 0.028-0.049 for P2, 0.035-0.061 for P3, 0.144-0.213 for P4, -0.062 to -0.049 for P5, 0.211-0.334 for P6, 0.247-0.391 for P7, and 0.246-0.360 for P8. Precision varied, particularly in the sparsest models; all corresponding intervals included zero.

Excluding high-risk studies changed P1 from $g = 0.117$ to 0.198 (38 effects, 12 clusters; 95% CI -0.015 to 0.412 , $p = 0.065$, $df = 8.16$). The broad caffeine estimate increased to 0.177 (16 effects, five clusters; 95% CI -0.170 to 0.524 , $p = 0.204$, $df = 3.01$), and the strict caffeine estimate increased to 0.267 (10 effects, four clusters; 95% CI -0.279 to 0.813 , $p = 0.183$, $df = 2.22$). Tier 1b caffeine could not be refitted because exclusion left one cluster. P4 and P6-P8 were unchanged because their contributing clusters did not carry a High-risk judgement.

3.7. Standardisation-scale sensitivity

Using the average condition standard deviation instead of the standard deviation of paired differences produced closely aligned estimates. The broad and strict caffeine estimates were $g = 0.034$ (95% CI -0.251 to 0.320) and 0.045 (95% CI -0.351 to 0.441); Tier 1a and Tier 1b estimates were 0.197 (95% CI -0.543 to 0.937) and -0.079 (95% CI -0.826 to 0.668). Broad and strict carbohydrate estimates were 0.275 (95% CI -0.104 to 0.653) and 0.319 (95% CI -0.164 to 0.802), while general cognition/coordination was 0.306 (95% CI -0.024 to 0.636). The exploratory mental-fatigue calculation was 0.033 (95% CI -0.698 to 0.764). Thus, changing the standardising denominator did not change the direction or uncertainty-based interpretation of any model (Supplementary Table S4).

3.8. Certainty of evidence

GRADE certainty was Very low for all six decision-relevant comparisons (Supplementary Table S3). All were downgraded for risk of bias because no contributing quantitative cluster was rated Low risk, and all were downgraded for imprecision because few clusters, low Satterthwaite degrees of freedom, and confidence intervals compatible with benefit and no effect—and in several comparisons harm—limited inference. Caffeine comparisons were additionally downgraded for inconsistency across outcome structures and substantial effect-level heterogeneity. Carbohydrate comparisons were downgraded for inconsistency because one influential cluster materially changed the pooled estimates. General cognition/coordination was additionally downgraded for indirectness because it comprised secondary laboratory outcomes rather than the review's primary perception-action or sport-specific performance constructs.

3.9. Outliers, influence, and model diagnostics

Study-level Cook's distance exceeded the $4/k$ heuristic for R0002 in P1 (0.337 vs 0.286) and R0023 in P6 (0.744 vs 0.571). In leave-one-out analyses, omitting R0023 increased the broad carbohydrate estimate to $g = 0.367$ (95% CI 0.120 to 0.614 , $p = 0.014$, $df = 4.43$) and the strict estimate to $g = 0.422$ (95% CI 0.121 to 0.724 , $p = 0.019$, $df = 3.43$). R0023 was retained because influence alone was not a prespecified exclusion criterion.

All evaluated working covariance matrices were positive definite. Of 108 prespecified model combinations, 90 fitted and 18 were structurally unavailable because too few study clusters remained: 12 P9 grid cells and six P5 cells after high-risk-study exclusion. All primary P1-P8 models converged with nlminb. An optimiser fallback was required only in selected P8 sensitivity cells at $\rho = 0.50$ and did not change the statistical specification. Model-level forest plots summarise P2-P9 in [Figures 2 and 3](#); complete effect-level caffeine and carbohydrate forests are provided as Supplementary Figures S1 and S2.

Caffeine models

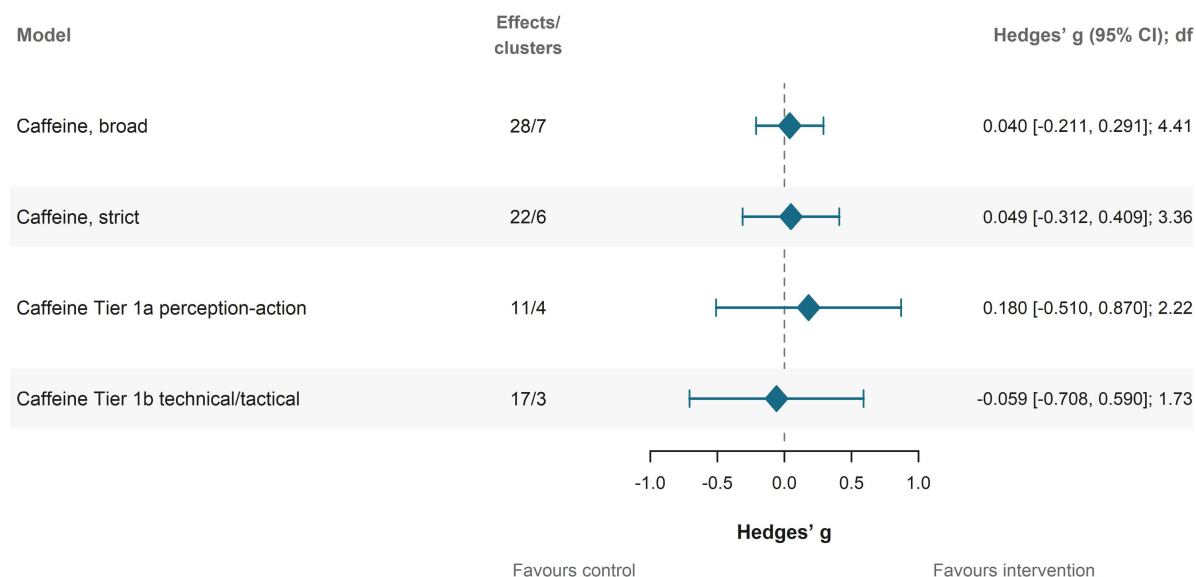


Figure 2. Pooled caffeine models. Diamonds show multilevel REML pooled Hedges' g estimates and horizontal lines show 95% confidence intervals based on CR2 robust standard errors and Satterthwaite degrees of freedom. Positive values favour caffeine, and the dashed vertical line indicates no difference. The Effects/studies column gives the numbers of effect sizes and independent study clusters. Each diamond represents a pooled model and is not scaled by weight.

Carbohydrate and secondary analyses

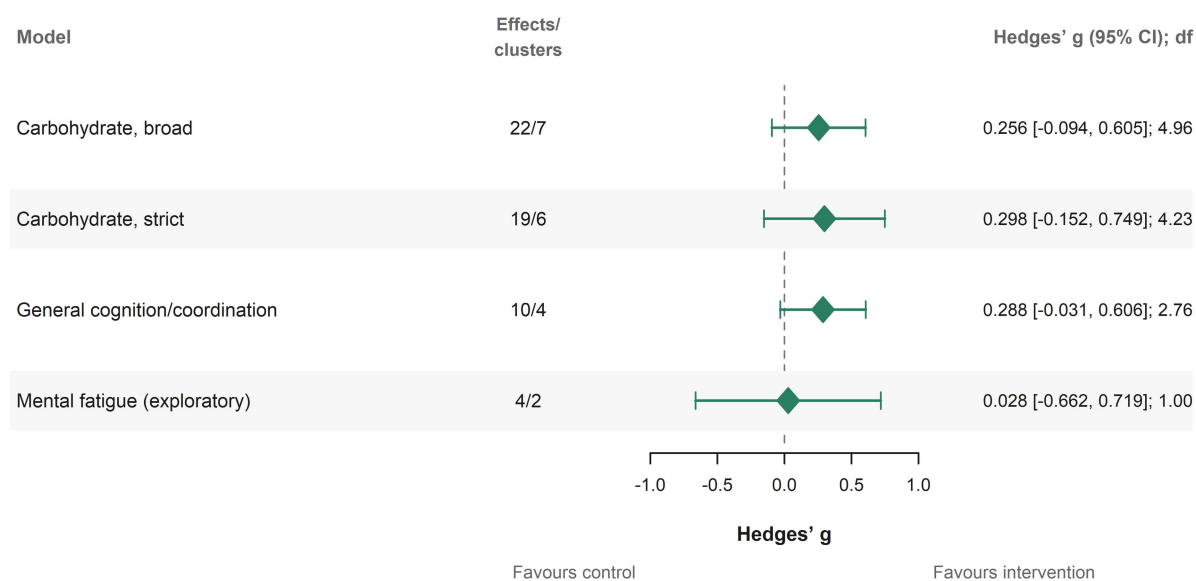


Figure 3. Pooled carbohydrate and secondary analyses. Diamonds show multilevel REML pooled Hedges' g estimates and horizontal lines show 95% confidence intervals based on CR2 robust standard errors and Satterthwaite degrees of freedom. Positive values favour the intervention, and the dashed vertical line indicates no difference. The Effects/studies column gives the numbers of effect sizes and independent study clusters. The mental-fatigue row is an exploratory descriptive two-study calculation (Satterthwaite df approximately 1), not a conventional pooled inference.

4. Discussion

4.1. Principal findings

The central finding is that current evidence does not support routinely assuming that established physical ergogenic effects transfer to fatigue-related decision-making or sport-specific skill. Broad and strict caffeine estimates were centred close to zero, while the carbohydrate estimates were small and positive but compatible with no difference. The tier-specific caffeine estimates were especially imprecise, and the carbohydrate signal was materially influenced by one study cluster. All six decision-relevant comparisons were rated very low certainty under GRADE. General cognition/coordination and mental-fatigue evidence were too sparse for firm conclusions. These results do not support a direct ranking of caffeine and carbohydrate, because the interventions were synthesised in parallel rather than compared within a randomised head-to-head contrast or network. They also do not support a general combined-strategy effect, because only one caffeine add-on-to-carbohydrate contrast was available.

4.2. Caffeine and fatigue-stage perception–action and sport-specific performance

Broader literature identifies outcome-dependent caffeine effects: caffeine may support physical performance, but technical-skill evidence in live team-sport settings remains mixed [2,3,43]. Our estimates address the missing translation step. A strategy that supports running performance or perceived exertion should not be presumed to improve the tasks that couple stimulus detection, action selection, and skilled execution during or following fatigue.

The point estimates themselves do not suggest a coherent caffeine benefit in the present outcome hierarchy. Broad and strict estimates were 0.040 and 0.049, Tier 1a was 0.180, and Tier 1b was −0.059. However, the tier-specific intervals were extremely wide. The Tier 1a interval ranged from −0.510 to 0.870, spanning a moderate detrimental effect and a large benefit. Effects of potential practical or clinical importance in either direction therefore remain compatible with the data. This analysis cannot conclusively rule out meaningful Tier 1a benefit or harm, despite the small positive point estimate. Tier 1b had only three clusters and $df = 1.73$. Excluding high-risk studies increased the broad and strict caffeine estimates to 0.177 and 0.267, but also reduced the cluster counts and left wide intervals crossing zero. This sensitivity analysis therefore shows that judgements about study quality affect the point estimate, not that risk of bias explains away an otherwise established benefit. Future caffeine trials should designate fatigue-stage perception-action or technical outcomes prospectively, report paired statistics, and avoid presenting numerous correlated task variants without a prespecified primary endpoint.

4.3. Carbohydrate findings and study influence

The carbohydrate estimates were more favourable in direction than the caffeine estimates, with broad and strict pooled effects of 0.256 and 0.298. This pattern is compatible with a small average benefit, but it is not a head-to-head demonstration that carbohydrate outperforms caffeine. The confidence intervals included zero, Satterthwaite df remained between approximately 4 and 5, and both study- and effect-level heterogeneity were present. Broader carbohydrate syntheses report performance benefits that vary by exercise duration and test format [1,4]. The current estimates narrow that literature to performance measured during or after fatigue and show that the available perception-action evidence is substantially less precise than the broader endurance evidence.

Influence analysis is central to this interpretation. R0023 exceeded the Cook's-distance heuristic in the broad carbohydrate model and differed from most other physical/match-fatigue carbohydrate clusters. In this randomised crossover study, 22 university soccer players ingested 8% isomaltulose or maltodextrin carbohydrate–electrolyte drinks, equivalent to approximately 20 g/h, during warm-up and half-time of a 120-min simulation that included extra time [20]. The 8% concentration was within the range of several other included beverage protocols. However, its overall delivery rate was lower and its two scheduled feeding opportunities less frequent than in some other included carbohydrate protocols; these differences may have contributed to its outcome pattern but cannot be isolated in a seven-cluster synthesis. More distinctive features were the two carbohydrate formulations sharing one placebo, late shooting and

dribbling assessments, and 12 of the broad model's 22 effect rows. Each other physical/match-fatigue carbohydrate cluster contributed one to three rows. The effects were mixed, and the isomaltulose comparison for dribbling success at 105–120 min was strongly unfavourable (Hedges' $g = -0.872$). This effect was flagged by a large sampling-standardised residual. Although within-cluster dependence was modelled, this pattern gave R0023 substantial leverage in a seven-cluster evidence base.

Omitting R0023 increased the broad and strict estimates to 0.367 and 0.422 and produced intervals that excluded zero. These leave-one-out results demonstrate sensitivity to evidence-base composition; they do not justify deleting R0023 or treating the favourable leave-one-out estimates as the “true” effects. For practice, the main estimate should be treated as a fragile average across heterogeneous ingestion protocols, not as a formulation or dose recommendation. Neither the main nor leave-one-out analysis establishes an advantage for isomaltulose, maltodextrin, or another delivery strategy. Replication in additional independent clusters using harmonised outcomes is required.

Combined caffeine and carbohydrate ingestion deserves the same restraint. A previous meta-analysis reported additive effects for some physical-performance outcomes when caffeine was added to carbohydrate, but its estimands, participant groups, and performance tests differ from fatigue-stage perception-action outcomes [5]. One add-on contrast in the present dataset cannot identify a stable interaction, separate pharmacological synergy from study-specific features, or support “combined effects” as a headline conclusion. A future synthesis could address this question only after a connected set of comparable trials includes caffeine, carbohydrate, their combination, and an appropriate control under consistent outcome definitions.

4.4. Secondary cognition and mental fatigue

The secondary cognition/coordination estimate was positive ($g = 0.288$) and its interval narrowly included zero, but four clusters and $df = 2.76$ provide little basis for a binary interpretation. General laboratory cognition and coordination can inform potential pathways, yet they are not substitutes for perception-action coupling in a representative sport task. The result is best treated as hypothesis-generating and as support for retaining cognitive measures alongside, rather than in place of, ecologically grounded performance endpoints.

Mental fatigue required a separate interpretation. The exploratory descriptive calculation from two studies was essentially zero, but its confidence interval extended from -0.662 to 0.719 and its Satterthwaite df was approximately 1 across the sensitivity grid. It therefore offers almost no power to distinguish benefit, harm, or no effect and should not be interpreted as a conventional pooled estimate. This uncertainty concerns nutritional countermeasures under mental fatigue, not whether mental fatigue itself can impair sport performance. A domain-specific meta-analysis reported adverse associations of mental fatigue with sport-related reaction time, technical performance, and decision-making [44]. The contrast reinforces why mental fatigue should not be pooled with exercise- or match-induced fatigue merely to increase the number of studies.

4.5. Practical implications

For practitioners, the results do not provide an evidence-based basis to assume that a nutritional strategy with established physical ergogenic effects will preserve fatigue-stage decision quality or technical accuracy. They also do not justify ranking caffeine below carbohydrate or prescribing their combination specifically for sport-specific performance during or following fatigue. Caffeine and carbohydrate decisions for other goals remain dependent on dose, timing, exercise duration, individual response, gastrointestinal tolerance, habitual caffeine exposure, and competition rules, none of which were pooled here [2,4]. The present evidence addresses the narrower translation question: whether these strategies are associated with more favourable perceptual-motor and sport-specific technical/tactical performance during or following fatigue. Caffeine estimates centre near zero. The carbohydrate estimates apply only to the heterogeneous ingested-beverage protocols represented in P6 and P7, not to mouth-rinse or gel strategies. Their sensitivity was driven primarily by R0023, whose inclusion attenuated the pooled estimate and whose omission increased it. This is not a reason to exclude that study or prefer the leave-one-out result. It means that the

average signal is not robust enough to prescribe a formulation, concentration, dose, or timing for sport-specific performance. Practitioners who use either strategy should monitor task-specific outcomes relevant to the athlete's role rather than assume transfer from improvements in physical output.

4.6. Strengths and limitations

The analysis has several strengths. It retained multiple eligible outcomes while explicitly modelling their nesting, used CR2/Satterthwaite inference for small clusters, reported separate study- and effect-level variance components, and preserved the distinction between perceptual-motor, technical/tactical, secondary cognitive, and mental-fatigue outcomes. The complete sensitivity grid varied both the paired correlation and the working within-study correlation, and the conclusions did not depend on one correlation value. High-risk-study exclusion, effect-level flags, Cook's distance, leave-one-out refits, covariance positive-definiteness checks, and optimiser tracking made analytical fragility visible. No effect size, study identifier, outcome direction, risk-of-bias judgement, or analysis group was changed to obtain a preferred result.

The limitations are equally important. RoB 2 was summarised at the report level for the eligible late/post-fatigue outcome set, which may not capture outcome-specific differences in bias within reports. Although Study_ID and Record_ID were one-to-one in the quantitative dataset, unreported participant overlap between separately published reports cannot be excluded with complete certainty. Each intervention-specific model contained seven or fewer independent clusters, and several effective degrees of freedom were below 4. The assumed paired correlations and working sampling correlation were necessary because complete covariance information was unavailable; sensitivity analyses reduce but do not eliminate that uncertainty. Tasks, fatigue protocols, doses, delivery formats, and assessment times were heterogeneous, limiting interpretation of an average standardised effect. No quantitative cluster was rated Low risk, and excluding high-risk studies further reduced already sparse models. Influence analysis showed that a single cluster affected carbohydrate inference. Small-study effects could not be evaluated reliably within intervention-specific models, and the evidence base was too disconnected for a defensible network comparison or combined-strategy interaction.

The review-level evidence included carbohydrate mouth-rinse and gel protocols, but the P6 and P7 models did not. Their estimates therefore apply only to the heterogeneous ingested beverages represented in those models and should not be extrapolated to mouth-rinse or gel strategies. Even among the included beverages, formulations, concentrations, doses, timings, and testing contexts varied. Sparse cluster counts precluded defensible formulation-, dose-, or timing-specific subgroup analysis or meta-regression. Random effects and CR2 inference addressed heterogeneity and dependence but could not identify a protocol-specific effect.

Two review-process limitations warrant explicit consideration. PROSPERO registration occurred after formal searching and title/abstract screening but before data extraction, risk-of-bias assessment, and synthesis; it therefore did not prospectively constrain study identification. In addition, the implemented sources differed from the registered plan: Web of Science and SPORTDiscus replaced the planned Scopus/MEDLINE configuration, while PubMed and supplementary searching were retained. The implemented search and deviation are reported transparently; the final consensus ledger reconciled exactly to the PRISMA and full-text counts.

4.7. Research priorities

The next generation of trials should prespecify a small number of fatigue-stage outcomes that distinguish perceptual-motor processing from technical execution, report paired means, standard deviations, and correlations, and use protocols representative of late-match demands. Multi-arm trials that include placebo, caffeine, carbohydrate, and their combination would permit direct and interaction estimates. Adequate independent replication is more valuable than adding many correlated outcomes within one small cross-over sample. Mental-fatigue trials should remain a distinct evidence stream and use validated induction checks, sport-representative tasks, and sufficient reporting to support dependence-aware synthesis. These design improvements would allow future reviews to move from describing wide compatible ranges to estimating which strategy works, for which task, and under which fatigue conditions.

5. Conclusions

Current evidence does not support assuming that the established physical ergogenic effects of caffeine translate to fatigue-related decision-making or sport-specific skill in team-sport athletes. In multilevel models with CR2/Satterthwaite inference, caffeine estimates were centred near zero, but wide tier-specific intervals could not rule out meaningful benefit or harm. Carbohydrate may provide a small benefit, but the evidence remains sparse and highly uncertain. Sensitivity to study composition was driven primarily by one influential cluster, whose omission increased the pooled estimate. The carbohydrate models represented heterogeneous ingested-beverage protocols and do not support extrapolation to mouth-rinse or gel strategies. The exploratory mental-fatigue calculation was too fragile for conventional inference. These parallel syntheses cannot establish intervention superiority or combined-strategy synergy. Better-powered trials should prespecify representative fatigue-stage outcomes, distinguish perceptual-motor from technical/tactical performance, report paired covariance information, and provide enough independent study clusters for reliable small-sample inference.

Acknowledgements

During the preparation of this manuscript, the authors used the OpenAI Codex desktop application (OpenAI; accessed 28 July–1 August 2026) for code-execution guidance, language editing, document preparation, file organisation, and quality-control checks. The authors reviewed and edited the output and take full responsibility for the content of this publication. No generative AI tool was used as an author or as an independent source of scientific evidence.

Both authors have read and approved the final manuscript.

Author contributions

CRedit: **Shu Xie:** Conceptualization, Methodology, Validation, Investigation, Writing – review & editing, Supervision, Project administration; **Xueming Wen:** Writing – review & editing, Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research received no external funding.

ORCID

Shu Xie  <http://orcid.org/0009-0007-1142-2953>

Xueming Wen  <http://orcid.org/0009-0006-2256-1599>

Data availability statement

The effect-size dataset, statistical code, complete model outputs, diagnostic files, package versions, and session information are provided in the Supplementary Materials.

Institutional review board statement

Not applicable. This systematic review synthesised previously published reports and did not enrol participants or collect identifiable individual-level data.

Informed consent statement

Not applicable.

References

- [1] Pueyo M, Llodio I, Cámara J, et al. Influence of carbohydrate intake on different parameters of soccer players' performance: systematic review. *Nutrients*. 2024;16(21):3731. doi: [10.3390/nu16213731](https://doi.org/10.3390/nu16213731)
- [2] Guest NS, VanDusseldorp TA, Nelson MT, et al. International society of sports nutrition position stand: caffeine and exercise performance. *J Int Soc Sports Nutr*. 2021;18(1):1. doi: [10.1186/s12970-020-00383-4](https://doi.org/10.1186/s12970-020-00383-4)
- [3] Diaz-Lara J, Nieto-Acevedo R, Abian-Vicen J, et al. Can caffeine change the game? Effects of acute caffeine intake on specific performance in intermittent sports during competition: a systematic review and meta-analysis. *Int J Sports Physiol Perform*. 2024;19(11):1180–1196. doi: [10.1123/ijsspp.2023-0232](https://doi.org/10.1123/ijsspp.2023-0232)
- [4] Ramos-Campo DJ, Clemente-Suárez VJ, Cupeiro R, et al. The ergogenic effects of acute carbohydrate feeding on endurance performance: a systematic review, meta-analysis and meta-regression. *Crit Rev Food Sci Nutr*. 2024;64(30):11196–11205. doi: [10.1080/10408398.2023.2233633](https://doi.org/10.1080/10408398.2023.2233633)
- [5] López-Seoane J, Buitrago-Morales M, Jiménez SL, et al. Synergy of carbohydrate and caffeine ingestion on physical performance and metabolic responses to exercise: a systematic review with meta-analysis. *Crit Rev Food Sci Nutr*. 2024;64(10):2941–2959. doi: [10.1080/10408398.2022.2128298](https://doi.org/10.1080/10408398.2022.2128298)
- [6] Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1–48. doi: [10.18637/jss.v036.i03](https://doi.org/10.18637/jss.v036.i03)
- [7] Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. *Res Synth Methods*. 2010;1(1):39–65. doi: [10.1002/jrsm.5](https://doi.org/10.1002/jrsm.5)
- [8] Tipton E. Small sample adjustments for robust variance estimation with meta-regression. *Psychol Methods*. 2015;20(3):375–393. doi: [10.1037/met0000011](https://doi.org/10.1037/met0000011)
- [9] Pustejovsky JE, Tipton E. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *J Bus Econ Stat*. 2018;36(4):672–683. doi: [10.1080/07350015.2016.1247004](https://doi.org/10.1080/07350015.2016.1247004)
- [10] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Brit Med J*. 2021;372:n71. doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)
- [11] Guyatt GH, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction—GRADE Evidence profiles and summary of findings tables. *JCE*. 2011;64(4):383–394. doi: [10.1016/j.jclinepi.2010.04.026](https://doi.org/10.1016/j.jclinepi.2010.04.026)
- [12] Seddik M, Dhahbi W, Bessifi M, et al. Acute caffeine ingestion, calendar-based menstrual-cycle window, time of day, and match-induced fatigue independently and interactively influence psychophysiological, cognitive, and physical performance in elite female volleyball players: a randomized double-blind placebo-controlled crossover design study. *Life (Basel)*. 2026;16(6):922. doi: [10.3390/life16060922](https://doi.org/10.3390/life16060922)
- [13] Yüce G, Soylu Y. Caffeine mouth rinsing in small-sided volleyball games: differential effects of continuous versus intermittent formats on psychophysiological, mood, and technical responses in young female players. *Turk J Kinesiol*. 2026;12(2):144–155. doi: [10.31459/turkjin.1822496](https://doi.org/10.31459/turkjin.1822496)
- [14] Gök U, Soylu Y, Arslan E, et al. A comparative analysis of carbohydrate and caffeine mouth rinsing strategies on cognitive, technical and kinematic performance in young soccer players during 4-a-Side small-sided game. *Turk J Sports Med*. 2025;60(4):143–150. doi: [10.47447/tjsm.0878](https://doi.org/10.47447/tjsm.0878)
- [15] Hsueh CL, Sun CY, Chang CK. Caffeine supplementation improved movement patterns and reactive agility in rugby sevens matches in Male collegiate players. *J Int Soc Sports Nutr*. 2025;22(1):2441763. doi: [10.1080/15502783.2024.2441763](https://doi.org/10.1080/15502783.2024.2441763)
- [16] Fischer A, Corr L, Birdsey LP, et al. Negligible effects of Fructose–Glucose composite carbohydrates on performance across a prolonged soccer match simulation compared with a glucose-only control in semiprofessional soccer players. *Int J Sport Nutr Exerc Metab*. 2025;35(5):404–415. doi: [10.1123/ijsnem.2024-0253](https://doi.org/10.1123/ijsnem.2024-0253)
- [17] Field A, Corr L, Birdsey L, et al. Caffeine gum improves reaction time but reduces composure versus placebo during the extra-time period of simulated soccer match-play in Male semiprofessional players. *Int J Sport Nutr Exerc Metab*. 2024;34(5):286–297. doi: [10.1123/ijsnem.2023-0220](https://doi.org/10.1123/ijsnem.2023-0220)
- [18] de Almeida RF, da Costa IT, Machado G, et al. The effect of acute caffeine ingestion on tactical performance of professional soccer players. *Nutrients*. 2022;14(7):1466. doi: [10.3390/nu14071466](https://doi.org/10.3390/nu14071466)
- [19] Quinones MD, Lemon PW. Hydrothermally modified corn starch ingestion attenuates soccer skill performance decrements in the second half of a simulated soccer match. *Int J Sport Nutr Exerc Metab*. 2019;29(5):498–504. doi: [10.1123/ijsnem.2018-0217](https://doi.org/10.1123/ijsnem.2018-0217)
- [20] Stevenson EJ, Watson A, Theis S, et al. A comparison of isomaltulose versus maltodextrin ingestion during soccer-specific exercise. *Eur J Appl Physiol*. 2017;117(11):2321–2333. doi: [10.1007/s00421-017-3719-5](https://doi.org/10.1007/s00421-017-3719-5)
- [21] Portillo J, Del Coso J, Abián-Vicén J. Effects of caffeine ingestion on skill performance during an international female rugby sevens competition. *J Strength Cond Res*. 2017;31(12):3351–3357. doi: [10.1519/jsc.0000000000001763](https://doi.org/10.1519/jsc.0000000000001763)
- [22] Harper LD, Stevenson EJ, Rollo I, et al. The influence of a 12% carbohydrate-electrolyte beverage on self-paced soccer-specific exercise performance. *J Sci Med Sport*. 2017;20(12):1123–1129. doi: [10.1016/j.jsams.2017.04.015](https://doi.org/10.1016/j.jsams.2017.04.015)
- [23] Harper LD, Briggs MA, McNamee G, et al. Physiological and performance effects of carbohydrate gels consumed prior to the extra-time period of prolonged simulated soccer match-play. *J Sci Med Sport*. 2016;19(6):509–514. doi: [10.1016/j.jsams.2015.06.009](https://doi.org/10.1016/j.jsams.2015.06.009)
- [24] O'Reilly J, Wong SH. Effect of a carbohydrate drink on soccer skill performance following a sport-specific training program. *J Exerc Sci Fit*. 2013;11(2):95–101. doi: [10.1016/j.jesf.2013.11.001](https://doi.org/10.1016/j.jesf.2013.11.001)

- [25] Russell M, Benton D, Kingsley M. Influence of carbohydrate supplementation on skill performance during a soccer match simulation. *J Sci Med Sport*. 2012;15(4):348–354. doi: [10.1016/j.jsams.2011.12.006](https://doi.org/10.1016/j.jsams.2011.12.006)
- [26] Duncan MJ, Taylor S, Lyons M. The effect of caffeine ingestion on field hockey skill performance following physical fatigue. *Res Sports Med*. 2012;20(1):25–36. doi: [10.1080/15438627.2012.634686](https://doi.org/10.1080/15438627.2012.634686)
- [27] Duvnjak-Zaknich DM, Dawson BT, Wallman KE, et al. Effect of caffeine on reactive agility time when fresh and fatigued. *Med Sci Sports Exerc*. 2011;43(8):1523–1530. doi: [10.1249/mss.0b013e31821048ab](https://doi.org/10.1249/mss.0b013e31821048ab)
- [28] Carvalho P, Oliveira B, Barros R, et al. Impact of fluid restriction and ad libitum water intake or an 8% carbohydrate-electrolyte beverage on skill performance of elite adolescent basketball players. *Int J Sport Nutr Exerc Metab*. 2011;21(3):214–221. doi: [10.1123/ijsnem.21.3.214](https://doi.org/10.1123/ijsnem.21.3.214)
- [29] Gant N, Ali A, Foskett A. The influence of caffeine and carbohydrate coingestion on simulated soccer performance. *Int J Sport Nutr Exerc Metab*. 2010;20(3):191–197. doi: [10.1123/ijsnem.20.3.191](https://doi.org/10.1123/ijsnem.20.3.191)
- [30] Ali A, Williams C. Carbohydrate ingestion and soccer skill performance during prolonged intermittent exercise. *J Sports Sci*. 2009;27(14):1499–1508. doi: [10.1080/02640410903334772](https://doi.org/10.1080/02640410903334772)
- [31] Currell K, Conway S, Jeukendrup AE. Carbohydrate ingestion improves performance of a new reliable test of soccer performance. *Int J Sport Nutr Exerc Metab*. 2009;19(1):34–46. doi: [10.1123/ijsnem.19.1.34](https://doi.org/10.1123/ijsnem.19.1.34)
- [32] Abbey EL, Rankin JW. Effect of ingesting a honey-sweetened beverage on soccer performance and exercise-induced cytokine response. *Int J Sport Nutr Exerc Metab*. 2009;19(6):659–672. doi: [10.1123/ijsnem.19.6.659](https://doi.org/10.1123/ijsnem.19.6.659)
- [33] Ali A, Williams C, Nicholas CW, et al. The influence of carbohydrate-electrolyte ingestion on soccer skill performance. *Med Sci Sports Exerc*. 2007;39(11):1969–1976. doi: [10.1249/mss.0b013e31814fb3e3](https://doi.org/10.1249/mss.0b013e31814fb3e3)
- [34] Dougherty KA, Baker LB, Chow M, et al. Two percent dehydration impairs and six percent carbohydrate drink improves boys basketball skills. *Med Sci Sports Exerc*. 2006;38(9):1650–1658. doi: [10.1249/01.mss.0000227640.60736.8e](https://doi.org/10.1249/01.mss.0000227640.60736.8e)
- [35] Stuart GR, Hopkins WG, Cook C, et al. Multiple effects of caffeine on simulated high-intensity team-sport performance. *Med Sci Sports Exerc*. 2005;37(11):1998–2005. doi: [10.1249/01.mss.0000177216.21847.8a](https://doi.org/10.1249/01.mss.0000177216.21847.8a)
- [36] Zuo Y, Poon ETC, Zhang X, et al. Effects of pre-exercise snack bars with low- and high-glycemic index on soccer-specific performance: an application of continuous glucose monitoring. *J Sports Sci*. 2025;43(14):1397–1405. doi: [10.1080/02640414.2025.2497672](https://doi.org/10.1080/02640414.2025.2497672)
- [37] Niknam A, Abdullahi MH, Hemmatinafar M, et al. Low and high doses of espresso coffee improve repeated sprint performance and Eye–Hand coordination following fatigue status in Male basketball players. *Curr Dev Nutr*. 2024;8(9):104427. doi: [10.1016/j.cdnut.2024.104427](https://doi.org/10.1016/j.cdnut.2024.104427)
- [38] Russell M, Reynolds NA, Crewther BT, et al. Physiological and performance effects of caffeine gum consumed during a simulated half-time by professional academy rugby union players. *J Strength Cond Res*. 2020;34(1):145–151. doi: [10.1519/jsc.0000000000002185](https://doi.org/10.1519/jsc.0000000000002185)
- [39] Ali A, O'Donnell J, Von Hurst P, et al. Caffeine ingestion enhances perceptual responses during intermittent exercise in female team-game players. *J Sports Sci*. 2016;34(4):330–341. doi: [10.1080/02640414.2015.1052746](https://doi.org/10.1080/02640414.2015.1052746)
- [40] Costa YP, Lopes-da-Silva BS, Fonseca FS, et al. Effect of caffeine on attack decision-making in trained beach volleyball players following prolonged cognitive effort induced by social media use: a randomized, single-blind, crossover pilot study. *BMC Sports Sci Med Rehabil*. 2025;17(1):145. doi: [10.1186/s13102-024-01052-8](https://doi.org/10.1186/s13102-024-01052-8)
- [41] Hemmat M, Nazemzadegan G, Niknam A, et al. The effects of mental fatigue and caffeine on basketball performance: insights from a crossover study. *J Int Soc Sports Nutr*. 2025;22(1):2541010. doi: [10.1080/15502783.2025.2541010](https://doi.org/10.1080/15502783.2025.2541010)
- [42] Ostojic SM, Mazic S. Effects of a carbohydrate-electrolyte drink on specific soccer tests and performance. *J Sports Sci Med*. 2002;1(2):47–53. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3963242/>
- [43] Arguedas-Soley A, Townsend I, Hengist A, et al. Acute caffeine supplementation and live match-play performance in team-sports: a systematic review (2000–2021). *J Sports Sci*. 2022;40:717–732. doi: [10.1080/02640414.2021.2003559](https://doi.org/10.1080/02640414.2021.2003559)
- [44] Lu Q, Liu D, Chen X, et al. Effects of mental fatigue on sport-related reaction-time, technical, and decision-making performance: a systematic review and meta-analysis. *BMC Psychol*. 2026. doi: [10.1186/s40359-026-04900-z](https://doi.org/10.1186/s40359-026-04900-z)