

SYSTEMATIC REVIEW

Open Access



Cold-Water Immersion for Post-exercise Recovery: An Overlap-Aware Umbrella Review

Weibao Liang^{1*†} , Xiaoge Ma^{2†}, Guanping Sun³, Chuannan Liu⁴ and Xujie Yan⁵

Abstract

Background Cold-water immersion (CWI) is widely used after strenuous exercise, but review-level findings vary by outcome, comparator, recovery window, population, and protocol. Reuse of the same primary studies across reviews can also exaggerate apparent consistency.

Methods We searched PubMed, Web of Science, SPORTDiscus, Cochrane Library, Embase, and Scopus from inception to 30 January 2026 for systematic reviews, meta-analyses, and network meta-analyses of post-exercise CWI. Fourteen reviews were included. Primary-study overlap was quantified using the corrected covered area (CCA) overall and for outcome-specific study-to-review mappings. Because overlap was high and review-level estimates were statistically dependent, we did not perform second-order meta-analysis. Source-verified estimates were synthesized in a structured, overlap-aware format. Methodological quality was assessed with AMSTAR 2.

Results Of the 14 included reviews, 13 provided quantitative estimates, whereas one review provided narrative evidence on heart rate variability (HRV). Overall CCA was 12.68% (high); 99 of 165 primary-study families appeared in at least two reviews. Outcome-specific CCA was 15.03% for delayed onset muscle soreness (DOMS; very high) and ranged from 0.00% for interleukin-6 (IL-6) to 25.00% for C-reactive protein (CRP). It was not calculable for HRV, because one review contributed. AMSTAR 2 confidence was low in 2 reviews and critically low in 12. DOMS showed the most consistent favorable pattern: source-review standardized mean differences ranged from -1.04 to -0.37 , although the largest estimate was soccer-specific and its 95% prediction interval crossed zero. Because DOMS evidence was highly redundant across the contributing reviews, this apparent consistency does not represent independent replications. Findings for objective performance and biochemical outcomes were inconsistent, time-dependent, or comparator-dependent.

Conclusions Within the available review-level evidence, CWI was associated with the most consistent reduction in self-reported short-term post-exercise DOMS, but the clinical importance of the observed effects remains uncertain. Evidence does not support a uniform benefit across objective performance or inflammatory outcomes. Protocol observations, including commonly studied temperatures and durations, are descriptive and should not be interpreted as an optimized prescription.

Trial Registration PROSPERO CRD420261332482

[†]Weibao Liang and Xiaoge Ma have contributed equally to this work.

*Correspondence:

Weibao Liang
2017020800@m.scnu.edu.cn

Full list of author information is available at the end of the article

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Key Points

- The 14 included reviews had a high overall primary-study overlap (corrected covered area [CCA] 12.68%), so dependent review-level estimates were not combined in a second-order meta-analysis.
- Delayed onset muscle soreness (DOMS) showed the most consistent favorable pattern, with overall source-review standardized mean differences (SMDs) from -1.04 to -0.37 , but clinical importance, expectation effects, and future-study reproducibility remain uncertain.
- Evidence on cold-water immersion (CWI) for objective performance, biochemical, inflammatory, and autonomic outcomes was inconsistent or insufficient, and protocol-related findings remain descriptive rather than prescriptive.

Keywords Cold-water immersion, Post-exercise recovery, Delayed onset muscle soreness, Umbrella review, Corrected covered area, Evidence overlap

Background

Post-exercise recovery strategies are used to reduce soreness, restore function, and maintain readiness between exercise bouts [1, 2]. Cold-water immersion (CWI) is common in sport and exercise settings, because it is inexpensive, practical, and readily standardized [1, 2]. Evidence for CWI is not uniform across recovery domains. Earlier reviews generally reported reduced muscle soreness, whereas effects on performance, muscle-damage biomarkers, and inflammatory outcomes were smaller, inconsistent, or dependent on the comparator and recovery window [3–7]. These differences limit broad claims that CWI improves recovery in all settings.

CWI may act through thermal and hydrostatic mechanisms. Hydrostatic pressure shifts blood centrally and increases central blood volume, whereas cooling reduces peripheral skin and muscle blood flow and lowers tissue temperature [1, 2, 8]. Cooling can also alter afferent signaling, reduce nerve conduction velocity, and increase pain threshold or tolerance [2, 9]. These mechanisms may affect perceived soreness more directly than objective performance or circulating biomarkers. Evidence for a reproducible systemic anti-inflammatory effect remains limited [4, 10].

Responses may also vary across recreationally active participants, trained athletes, and sport-specific populations. Training status, sex, age, body composition, exercise mode, and the magnitude of exercise-induced muscle damage can modify thermal strain, symptom reporting, and functional recovery. Review-level findings should be interpreted within their population context and not assumed to apply to all athletes. The evidence base contains many systematic reviews that reuse the same primary trials while applying different comparators, outcome definitions, and recovery windows [6, 7, 11–14]. Treating their pooled estimates as independent observations can overstate precision. An umbrella review should,

therefore, quantify primary-study overlap and preserve the context of each source review.

This umbrella review aimed to (1) synthesize source-verified review-level evidence on the effectiveness of CWI for post-exercise recovery; (2) quantify overall and outcome-specific primary-study overlap; (3) describe how population, comparator, recovery window, and protocol characteristics influence interpretation; and (4) appraise the methodological quality of the included reviews.

Methods

Search Strategy and Selection Criteria

This umbrella review was reported in accordance with PRISMA 2020 and with consideration of the PRIOR statement for overviews of reviews [15, 16]. The protocol was prospectively registered in PROSPERO (CRD420261332482). The systematic review was the unit of searching and inclusion [14, 17].

Six electronic databases (PubMed, Web of Science, SPORTDiscus, Cochrane Library, Embase, and Scopus) were searched from inception to 30 January 2026. Search terms combined CWI and related cold-immersion terms, post-exercise recovery outcomes, and review-level study designs. No language restrictions were applied. Reference lists of included reviews were also screened. Full database-specific strategies are provided in Additional file 1: Table S1.

Eligible reviews included healthy individuals, physically active individuals, or athletes; evaluated post-exercise CWI; compared CWI with passive recovery, active recovery, thermoneutral water immersion, contrast water therapy, body cryotherapy, or another recovery modality; and reported perceptual, functional, biochemical, inflammatory, or autonomic recovery outcomes. These participants, intervention, comparators, outcomes, and study-design criteria constituted the prespecified PICOS

framework. Systematic reviews with or without meta-analysis and network meta-analyses were eligible. Narrative reviews, expert commentaries, and primary studies were excluded.

Study Selection and Data Extraction

Two reviewers independently screened titles, abstracts, and full texts, with disagreements resolved by discussion and, when required, a third reviewer. Re-examination of the screening record identified one full-text report (Veen et al., 2026) that had been misclassified at final eligibility adjudication; it met the prespecified PICOS criteria and was included [18]. Two reviewers independently extracted review design, primary-study count, participant characteristics and training level, exercise context, comparator structure, CWI protocol, outcome definition, effect metric, pooled estimate and 95% confidence interval (CI), heterogeneity, prediction interval where available, and the risk-of-bias method used by the source review. Every retained estimate was rechecked against the source review. Exact source time labels and comparator categories were retained, and only source-verifiable estimates were included. When a source review reported both fixed-effect and random-effects estimates, the model selected by the source-review authors was treated as the primary estimate. Source-review *p* values were retained only as descriptive information; when a reported *p* value was inconsistent with its 95% CI, both were reproduced as reported and interpretation followed the CI. Full characteristics and extraction details are reported in Table 1 and Additional file 1: Tables S2 and S3; complete datasets are provided as Additional files 2 and 3.

For descriptive comparability, “direction-harmonized” means that estimates were sign-inverted where necessary, so that negative values consistently favored CWI. This harmonization did not change effect magnitude or precision. Source time labels were also mapped to broad recovery windows for display only. Overall and time-specific estimates from the same review were treated as dependent and were not counted as independent replications. No primary-study effect estimate was newly calculated from a trial report.

Methodological Quality Assessment and Handling of Risk of Bias

Methodological quality was appraised with A MeaSurement Tool to Assess Systematic Reviews 2 (AMSTAR 2) [19]. Two reviewers assessed each review independently, and disagreements were resolved by consensus or a third reviewer. Items 2, 4, 7, 9, 11, 13, and 15 were treated as critical domains, and overall confidence was rated as high, moderate, low, or critically low according to the number of critical flaws and non-critical weaknesses

[19]. Item-level judgments and rationales are provided in Additional file 1: Table S4.

Risk of bias in the primary trials was not reassessed de novo. We extracted the tool and judgments reported by each source review and examined whether primary-study risk of bias was incorporated into the review conclusions. These data informed interpretation but were not used as an exclusion criterion.

Primary-Study Overlap Assessment

Primary-study overlap was calculated for the included reviews using the corrected covered area (CCA) [11, 12]. Within each review, multiple reports, arms, subgroups, and timepoints from the same primary-study or publication family were collapsed to one unit. CCA was calculated as $(N - r)/[r(c - 1)]$, where *N* is the number of study-review inclusion cells, *r* is the number of unique primary-study families, and *c* is the number of reviews. Values of 0–5%, >5–10%, >10–15%, and >15% were interpreted as slight, moderate, high, and very high overlap, respectively [12]. Outcome-specific cells were included only when the source review directly linked a primary-study family to the corresponding synthesis domain. Broad muscular-power syntheses were coded as power and were not duplicated into jump or sprint domains solely because of the primary test label. Narratively synthesized outcomes were eligible when complete study-level mapping was available; unresolved cells were not coded as absent. CCA was calculated overall and for outcomes represented by at least two reviews. Pairwise Jaccard coefficients, primary-study multiplicity, analyses excluding selected reviews, and leave-one-review-out CCA analyses were also examined. The matrices, outcome tags, and decision rules are provided in Additional file 4; summary results are provided in Additional file 1: Tables S5 and S6 and Figures S1–S3.

Overlap-Aware Evidence Synthesis

Pooled estimates from different reviews were not combined. Overall CCA was high, several outcome-specific CCA values were high or very high, and CCA does not estimate the covariance induced by repeated primary studies. Standard second-order meta-analysis would treat dependent estimates as independent and could produce confidence intervals that are too narrow [12–14]. We presented source-review estimates with population, comparator, metric, recovery window, 95% CI, and available prediction intervals (Additional file 1: Table S7).

Interpretation emphasized effect magnitude, CI compatibility with the null, temporal consistency, and replication across distinct review contexts. No multiplicity adjustment was applied to the many source outcome-window tests, because these tests were not recomputed.

Table 1 Characteristics of included reviews

Study	Review design	Primary studies, n	Total sample, n	Participant characteristics	Age (years)	Sex	RoB tool used in original review	Quantitative contribution
Batista et al. 2023 [27]	SR with MA	44	NR	Adults including young athletes, amateurs, non-athletes and physically active participants	NR at review level	NR at review level	PEDro scale; GRADE used for certainty	MD/WMD
Chen et al. 2024 [21]	NMA	57	1220	Healthy individuals, including athletes and non-athletes	Study samples approximately 18–37 years	Mixed; predominantly male in several trials	Cochrane risk-of-bias tool (v5.1.0)	SMD (NMA relative effects)
Choo et al. 2022 [23]	SR with MA	68	1064	Healthy physically active adults	Eligibility ≥18 years	Crossover: 402 male/25 female; parallel control: 258 male/17 female; parallel CWI: 298 male/19 female; sex unreported in two parallel studies (n=26)	Cochrane risk-of-bias tool	SMD
Galvez-Rodriguez et al. 2025 [20]	SR	12	240	Elite or amateur athletes and physically active adults	Review-reported mean age across included studies: approximately 23.38 years; individual study samples varied	Mixed, with many male-only cohorts	Cochrane risk-of-bias tool	Narrative only
Higgins et al. 2017 [25]	SR with MA	23	606	Well-trained, elite and professional team-sport athletes	NR as a single review-level value	506 men, 47 women, sex not reported for 53 participants	Modified Delphi scale and Jadad scale	SMD
Leeder et al. 2012 [5]	SR with MA	14	239	Untrained healthy adults through elite athletes	23.2 ± 3.5 years	182 men and 57 women	Cochrane Collaboration guidance	Hedges g (SMD)
Liu and Ma 2026 [26]	SR with MA	13	214	Healthy and physically active participants	20–47.2 years	Mixed; predominantly male	Cochrane risk-of-bias tool (v5.1.0)	MD and SMD
Ma et al. 2025 [22]	SR with MA	24	475	Healthy adults, athletes and non-athletes	Approximately 13–50 years across included studies	Mixed, predominantly male	Cochrane risk-of-bias tool (RoB 1)	SMD and temperature-subgroup MD
Machado et al. 2016 [28]	SR with MA	9	169	Physically active individuals and athletes	NR as a single value	141 men and 28 women	PEDro scale	Weighted mean difference (MD), not SMD
Moore et al. 2022 [6]	SR with MA/meta-regression	52	NR	Adults involved in sport or deemed physically active	>18 years eligibility	Mixed, predominantly male	SIGN RCT checklist	SMD and meta-regression coefficients
Moore et al. 2023 [7]	SR with MA/meta-regression	39	NR	Adults involved in sport or deemed physically active	>18 years eligibility	Mixed, predominantly male	SIGN RCT checklist	SMD; comparator-stratified meta-analysis

Table 1 (continued)

Study	Review design	Primary studies, n	Total sample, n	Participant characteristics	Age (years)	Sex	RoB tool used in original review	Quantitative contribution
Veen et al. 2026 [18]	SR with MA	10	210	Competitive trained male and female soccer players aged 15–45 years	Eligibility 15–45 years	201 male and 9 female; only one study included women	PEDro scale; studies scoring <5 excluded from meta-analysis	SMD with 95% prediction intervals
Wang et al. 2025 [29]	NMA	55	1139	Healthy individuals	Approximately 14–47 years	Mixed, predominantly male	Cochrane risk-of-bias tool	SMD NMA node effects
Xiao et al. 2023 [24]	SR with MA	20	357	Healthy participants	Commonly young adults, approximately 19–31 years	Mixed, predominantly male	Cochrane risk-of-bias tool	SMD and MD

CWI cold-water immersion, GRADE Grading of Recommendations Assessment, Development and Evaluation, MA meta-analysis, MD mean difference, NMA network meta-analysis, NR not reported, PEDro Physiotherapy Evidence Database, RCT randomized controlled trial, RoB risk of bias, SIGN Scottish Intercollegiate Guidelines Network, SMD standardized mean difference, SR systematic review, WMD weighted mean difference

Isolated significant timepoints were not interpreted as confirmatory evidence.

Standardized mean differences (SMDs), Hedges g values, mean differences (MDs), weighted mean differences (WMDs), and network relative effects were kept in their source metric and comparator context. SMD-based and MD/WMD-based estimates were not combined. Thirteen reviews provided quantitative evidence, and the HRV-focused systematic review contributed narrative evidence only.

Water temperature, duration, immersion depth, timing, and frequency were summarized descriptively (Additional file 1: Table S8). Protocol ranges were often mixed within reviews, and dose nodes from the same network meta-analysis were dependent. No trial-level dose–response model was fitted. Observations around 5–15 °C and 5–15 min are reported as commonly studied or descriptively favorable patterns, not as an optimal protocol. Evidence for protocols below 5 °C was especially sparse.

Analyses Not Performed Because of Dependence

Second-order random-effects models, review-level funnel plots, Egger-type tests, and inferential leave-one-source-out meta-analyses were not performed, because review-level estimates were not independent. Although CCA quantifies redundancy, it does not supply the covariance matrix required to repair statistical dependence. Sensitivity analyses were limited to overlap metrics, comparator-stratified source results, and transparent inspection of source-review estimates.

Structured Interpretation

No *de novo* credibility grade was assigned. The PICOS framework and eligibility criteria were prospectively registered. The interpretive domains used to structure the synthesis were defined during peer review, after protocol registration, and were not part of the registered protocol: direction and magnitude of source effects; temporal consistency; subjective versus objective measurement; participant and exercise context; comparator structure; AMSTAR 2 confidence; primary-study overlap, quantified from primary-study/publication-family matrices; and prediction intervals where available. These domains organized the presentation of the evidence and were not used as a certainty-grading instrument. Evidence was described as insufficient when too sparse for a conclusion, which was kept distinct from evidence showing no effect.

Results

Study Selection and Characteristics of Included Reviews

The search identified 679 records. After 345 duplicate or other pre-screening removals, 334 records were screened and 79 full-text reports were assessed. Of these, 65 were excluded and 14 reviews were included (Fig. 1).

The 14 included reviews comprised 11 systematic reviews with pairwise meta-analysis, 2 network meta-analyses, and 1 systematic review without meta-analysis. Thirteen reviews contributed quantitative estimates; Galvez-Rodriguez et al. (2025) contributed narrative HRV evidence [20]. Veen et al. (2026) contributed evidence from 10 studies of trained-soccer players, including SMDs and 95% prediction intervals for maximal voluntary contraction, countermovement jump, 20-m sprint, creatine kinase (CK), and delayed onset muscle soreness (DOMS) [18]. Review characteristics are summarized in Table 1 and Additional file 1: Tables S2 and S3. Figure 2 summarizes source-review outcome patterns, and Fig. 3 displays overall source-review SMDs for key outcomes.

Methodological Quality and Primary-Study Overlap

AMSTAR 2 confidence was low in 2 reviews and critically low in 12 (Fig. 4; Additional file 1: Table S4). No review achieved moderate or high confidence. Common critical weaknesses were absence of an excluded-study list, lack of prospective protocol information, inadequate integration of primary-study risk of bias into interpretation, and incomplete investigation of reporting bias.

The overall overlap matrix contained 437 study-review inclusion cells from 165 unique primary-study families across 14 reviews, yielding a CCA of 12.68% (high). Ninety-nine of 165 families (60.0%) appeared in at least two reviews, and 272 of 437 inclusions (62.24%) were redundant. Outcome-specific CCA was very high for DOMS (15.03%), perceived recovery (23.08%), jump performance (17.10%), strength (19.64%), power (21.79%), CK (16.75%), CRP (25.00%), and lactate (15.38%); high for sprint performance (14.29%); moderate for perceived fatigue (6.25%); and slight for IL-6 (0.00%). CCA was not calculable for HRV/autonomic recovery, because one review contributed. The 0% IL-6 value did not establish exchangeability or justify pooling, because only two clinically heterogeneous reviews contributed. Overall CCA remained high in every leave-one-review-out analysis (11.11–14.23%) (Table 2; Additional file 1: Tables S5 and S6 and Figures S1–S3).

Structured Synthesis of Perceptual and Performance Outcomes

DOMS showed the most consistent favorable pattern. Overall SMDs from Chen et al., Leeder et al., Ma et al.,

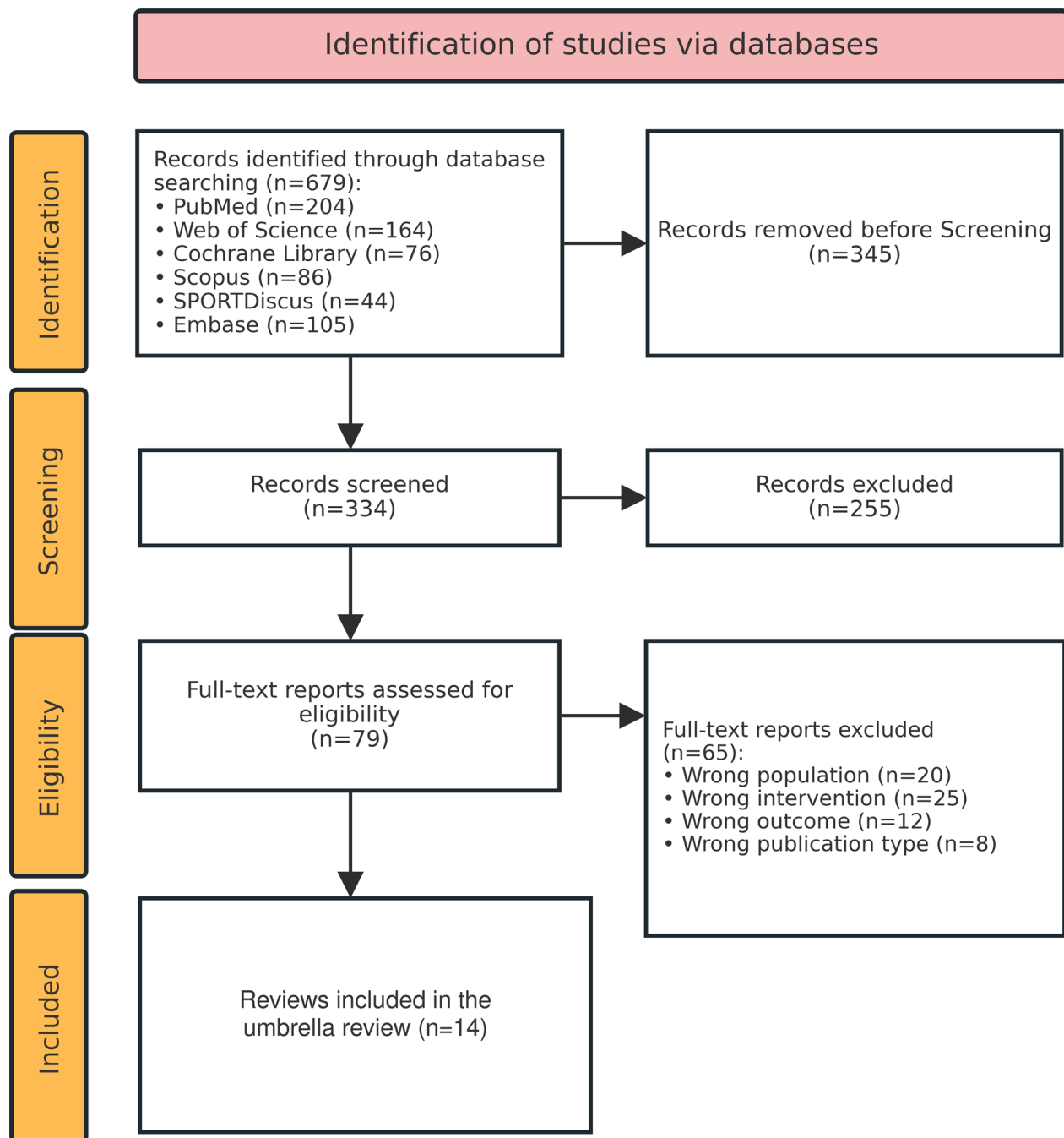


Fig. 1 Study selection flow. The search identified 679 records; 334 records were screened, 79 full-text reports were assessed, 65 reports were excluded, and 14 reviews were included in the umbrella review

and Veen et al. ranged from -1.04 to -0.37 [5, 18, 21, 22]. These four estimates were not independent replications: Chen et al. shared three to six DOMS study families with each of the other three reviews, although Leeder et al., Ma et al., and Veen et al. shared at most one family with one another, and DOMS evidence across the

13 contributing reviews was highly redundant (CCA 15.03%; Additional file 4). Choo et al. reported favorable estimates from <1 h through 72 h but no clear difference at ≥ 96 h; Leeder et al. reported favorable estimates at 24–96 h; and Xiao et al. reported a favorable estimate at 0 h, whereas the 24- and 48-h estimates crossed zero

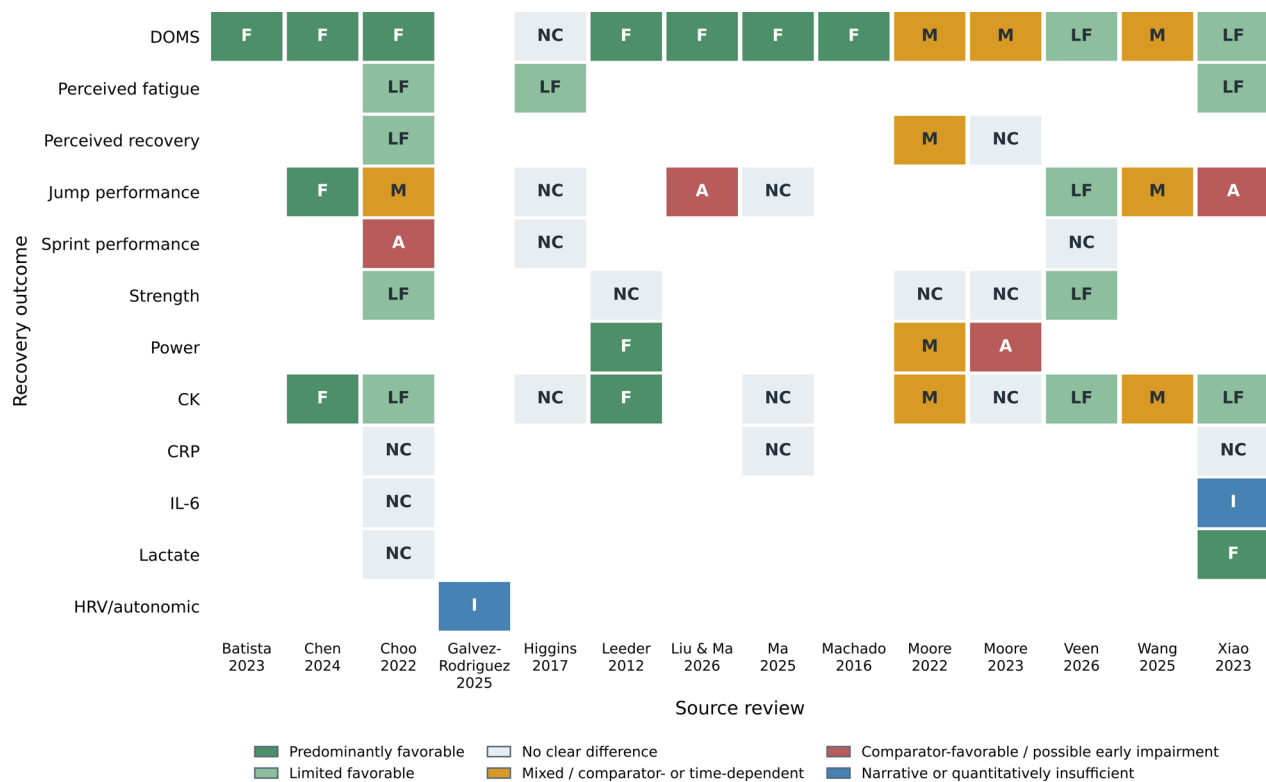


Fig. 2 Overlap-aware source-review evidence map. Cells summarize the direction and consistency of source-review findings without second-order pooling. Blank cells indicate that the review did not report the outcome. A comparator-favorable or possible early impairment, CK creatine kinase, CRP C-reactive protein, DOMS delayed onset muscle soreness, F predominantly favorable, HRV heart rate variability, I narrative or quantitatively insufficient, IL-6 interleukin-6, LF limited favorable, M mixed or comparator/time-dependent, NC no clear difference

under the random-effects model that the source review selected because of heterogeneity [5, 23, 24]. Higgins et al. reported imprecise time-specific estimates with CIs crossing zero [25]. Veen et al. reported the largest overall effect in trained-soccer players (SMD -1.04 , 95% CI -1.91 to -0.17), but its 95% prediction interval (-4.12 to 2.03) crossed zero [18] (Additional file 1: Table S7; Fig. 3).

Perceived fatigue was assessed quantitatively in three reviews. Choo et al. and Xiao et al. reported favorable immediate estimates, whereas later estimates were mixed or imprecise; Higgins et al. reported a favorable estimate only at 72 h [23–25]. Perceived recovery was source-linked in Choo et al., Moore et al. (2022), and Moore et al. (2023), yielding a very high outcome-specific CCA of 23.08% [6, 7, 23]. The estimates nevertheless remained sparse and time- or comparator-dependent and did not show a stable cross-review pattern.

Sprint findings did not show a consistent benefit. Choo et al. reported an immediate estimate favoring the comparator and no clear difference from 1 to 72 h [23]. Higgins et al. reported best- and accumulated-sprint estimates with 95% CIs crossing zero; the 24-h best-sprint estimate was described as beneficial in the source review

($P = 0.02$), but its CI also crossed zero [25]. In trained-soccer players, Veen et al. reported no clear overall 20-m sprint effect (SMD -0.59 , 95% CI -1.41 to 0.23), with a wide prediction interval [18].

Jump, strength, and power outcomes were heterogeneous (Additional file 1: Figure S4). For jump performance, Chen et al. favored CWI in a network comparison; Choo et al. and Xiao et al. reported early impairment (at 1–6 h and 0 h, respectively), with selected later favorable estimates only in Choo et al. Xiao et al. had an imprecise source-selected random-effects estimate crossing zero at 24 h, and no clear difference at 48 h; Higgins et al. reported estimates with CIs crossing zero; and Liu and Ma found no overall advantage over body cryotherapy [21, 23–26]. Ma et al. reported no clear overall counter-movement-jump effect, whereas Veen et al. reported a favorable overall estimate in trained-soccer players but a prediction interval crossing zero [18, 22]. Strength was favorable only at selected timepoints in Choo et al.; Leeder et al. reported no clear overall effect, while Veen et al. reported favorable maximal voluntary contraction estimates with prediction intervals crossing zero [5, 18, 23]. Broad power syntheses in Leeder et al. and the

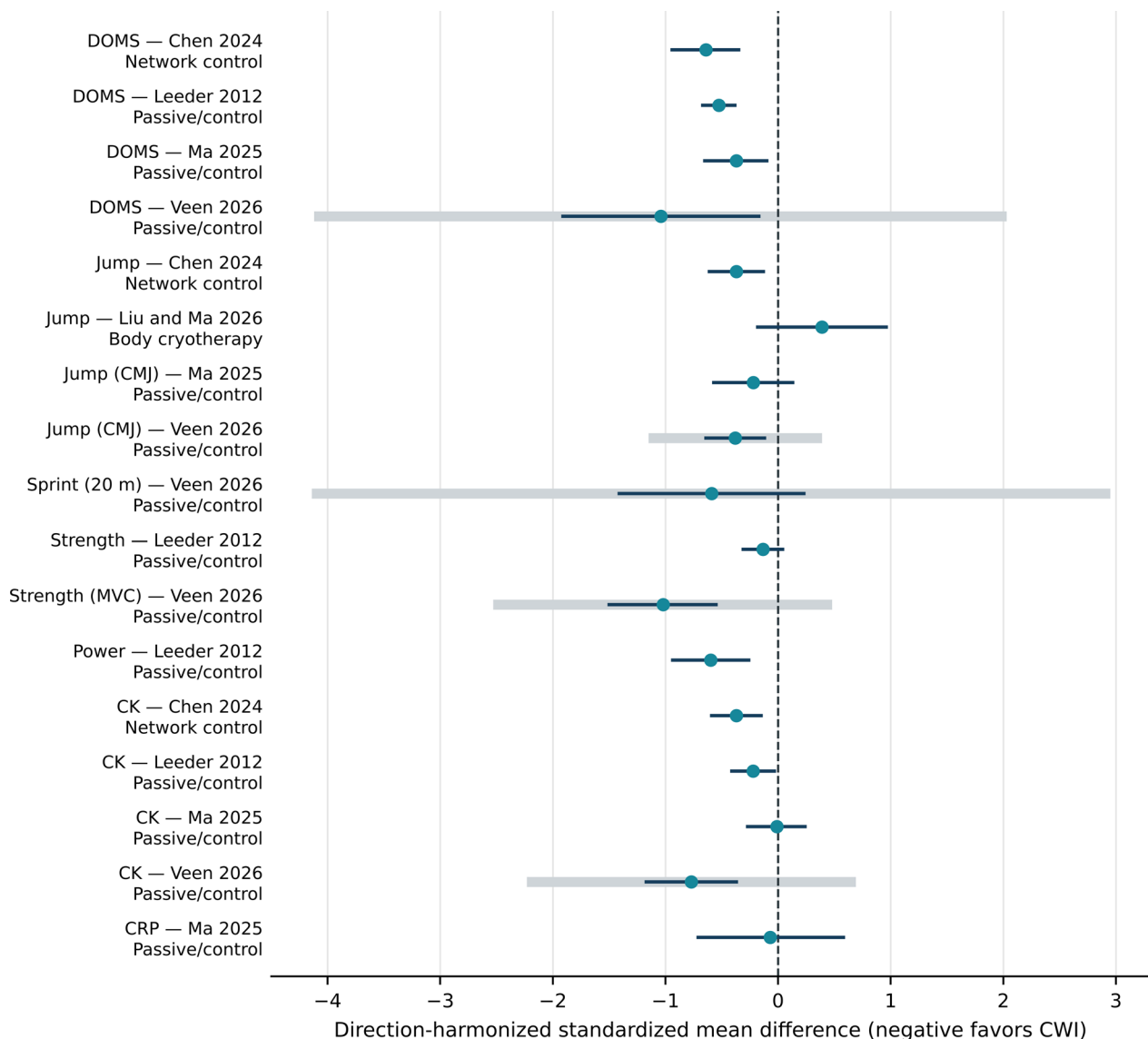


Fig. 3 Overall source-review standardized mean difference estimates for key outcomes. Points and thin lines show direction-harmonized source-review estimates and 95% confidence intervals; pale thick lines show direction-harmonized 95% prediction intervals where reported. Negative values favor CWI. No pooled diamonds are shown, because review-level estimates are dependent. *CK* creatine kinase, *CMJ* countermovement jump, *CRP* C-reactive protein, *CWI* cold-water immersion, *DOMS* delayed onset muscle soreness, *MVC* maximal voluntary contraction

Moore reviews combined heterogeneous tests and were retained as a separate power domain rather than duplicated into jump or sprint [5–7].

Biochemical, Inflammatory, and Autonomic Outcomes

CK findings varied by review and context (Additional file 1: Figure S5). Chen et al. and Leeder et al. reported favorable overall estimates; Choo et al. reported favorable estimates at 24–72 h; Ma et al. reported no clear overall effect; and Veen et al. reported favorable

soccer-specific estimates, although the prediction intervals were compatible with null effects in future settings [5, 18, 21–23]. CRP showed no consistent benefit in Choo et al. or Ma et al. [22, 23]. The IL-6 matrix contained two reviews and eight unique study families with no shared family (CCA 0.00%). Choo et al. reported time-specific estimates with CIs crossing zero, while Xiao et al. provided one source-linked family within otherwise sparse evidence [23, 24]. Lactate was null immediately in Choo et al. and Xiao et al., with

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Overall confidence
Batista 2023	Y	Y	Y	PY	Y	Y	N	Y	PY	N	Y	N	Y	Y	Y	Y	Low
Chen 2024	Y	Y	Y	PY	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	Critically low
Choo 2022	Y	N	Y	PY	Y	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	Critically low
Galvez-Rodriguez 2025	Y	PY	Y	PY	Y	Y	N	Y	Y	N	NA	NA	Y	Y	NA	Y	Low
Higgins 2017	Y	N	Y	PY	Y	N	N	Y	PY	N	Y	N	Y	Y	N	Y	Critically low
Leeder 2012	Y	N	Y	PY	N	N	N	Y	PY	N	Y	N	Y	Y	N	Y	Critically low
Liu and Ma 2026	Y	N	Y	PY	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	Critically low
Ma 2025	Y	Y	Y	PY	Y	Y	N	Y	Y	N	Y	N	N	Y	N	Y	Critically low
Machado 2016	Y	Y	Y	PY	Y	Y	N	Y	PY	N	Y	N	Y	Y	N	Y	Critically low
Moore 2022	Y	Y	Y	PY	Y	Y	N	Y	PY	N	Y	Y	Y	Y	N	Y	Critically low
Moore 2023	Y	Y	Y	PY	Y	Y	N	Y	PY	N	Y	Y	Y	Y	N	Y	Critically low
Veen 2026	Y	Y	Y	PY	Y	Y	N	Y	PY	N	Y	Y	Y	Y	N	Y	Critically low
Wang 2025	Y	Y	Y	PY	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	Critically low
Xiao 2023	Y	N	Y	N	Y	Y	N	Y	Y	N	Y	N	N	Y	N	Y	Critically low

Fig. 4 AMSTAR 2 appraisal of the 14 included reviews. The matrix shows yes, partial yes, no, and not applicable judgments for the 16 AMSTAR 2 items, with the overall confidence rating for each review. AMSTAR 2 A MeaSurement Tool to Assess Systematic Reviews 2, N no, NA not applicable, PY partial yes, Q1–Q16 AMSTAR 2 items 1–16, Y yes

favorable 24- and 48-h estimates reported only by Xiao et al. [23, 24].

The HRV review included 12 studies and did not perform meta-analysis, because protocols and outcomes were heterogeneous [20]. Seven studies reported statistically favorable findings, and the narrative pattern was compatible with acute parasympathetic reactivation. These data are quantitatively insufficient and should not be read as evidence of no effect. The available findings were mostly acute and cannot establish durability or identify an optimal protocol.

Protocol Characteristics: Descriptive Findings

Several reviews reported favorable results in protocol strata around 5–15 °C for 5–15 min, commonly applied immediately after exercise [20, 27–29]. However, these observations came from review-level subgroups, meta-regressions, or dependent network nodes. Colder protocols below 5 °C were rarely studied, protocol ranges were often mixed, and population and comparator distributions differed across strata. Accordingly, Table 3 and Additional file 1: Table S8 describe observed patterns only and do not recommend 11–15 °C or 10–15 min as an optimized prescription.

Comparator Dependence and Metric-Specific Evidence

Comparator choice materially changed interpretation. Passive recovery, active recovery, thermoneutral immersion, contrast water therapy, body cryotherapy, and network controls are not exchangeable. In Moore et al. (2023), comparator-stratified estimates varied by modality and time; CWI reduced DOMS relative to active recovery at 1 h but not at 24 or 48 h [7] (Additional file 1: Table S9 and Figure S6). Liu and Ma reported lower DOMS with CWI than body cryotherapy but no overall jump advantage [26]. MD/WMD findings were retained in source units and comparator strata rather than used as confirmation of SMD results (Additional file 1: Table S10 and Figure S7).

Structured Interpretation by Outcome

Outcomes were not assigned de novo moderate, low, or very low credibility labels. Table 4 instead summarizes each outcome across the interpretive domains described in the Methods and gives an outcome-specific conclusion that distinguishes a consistent favorable pattern, mixed evidence, no clear difference, and insufficient evidence.

Table 2 Quantitative source-review findings and outcome-specific overlap

Outcome	Overlap summary (reviews; unique families)	Quantitative source-review pattern	Outcome CCA	Important qualifiers
DOMS	13 reviews; 107 families	Most overall and delayed estimates favored CWI; overall SMDs -1.04 to -0.37	15.03% Very high	Subjective outcome; heterogeneous scales/comparators; largest soccer estimate had PI crossing zero
Perceived fatigue	3 reviews; 16 families	Immediate estimates were sometimes favorable; later estimates were mixed or imprecise	6.25% Moderate	Few reviews; subjective and expectation-sensitive
Perceived recovery	3 reviews; 13 families	Estimates were sparse and time- or comparator-dependent	23.08% Very high	Three source-linked reviews; subjective and comparator-sensitive
Jump performance	8 reviews; 61 families	Favorable, null, and early comparator-favorable estimates occurred	17.10% Very high	Test protocols, populations, and comparators differed; repeated timepoints were dependent
Sprint performance	3 reviews; 28 families	Immediate or overall effects were generally null; no consistent delayed benefit	14.29% High	Different sprint constructs; trained-soccer PI was wide and crossed zero
Strength	5 reviews; 42 families	Selected favorable timepoints; overall estimates varied by review and population	19.64% Very high	MVC and other strength tests were not interchangeable; soccer PIs crossed zero
Power	3 reviews; 39 families	Broad source-review syntheses were favorable, mixed, or comparator-dependent	21.79% Very high	Heterogeneous tests were retained as a broad power domain rather than split into jump/sprint
CK	10 reviews; 67 families	Several overall or delayed estimates favored CWI, while others were null	16.75% Very high	Indirect, variable biomarker; favorable soccer PIs crossed zero
CRP	3 reviews; 18 families	Available source-review estimates were null or imprecise	25.00% Very high	Sparse evidence with substantial redundancy
IL-6	2 reviews; 8 families	Time-specific estimates crossed zero; additional evidence was quantitatively sparse	0.00% Slight	No shared study family; 0% overlap does not establish exchangeability
Lactate	2 reviews; 13 families	Immediate estimates were null; delayed favorable estimates came from one review	15.38% Very high	Only two reviews; high overlap and time dependence
HRV/autonomic recovery	1 review; 12 families	Narrative findings generally favored acute parasympathetic reactivation	Not calculable (single review)	No pooled effect; protocols and HRV outcomes were heterogeneous

Outcome-specific CCA uses study-to-review contribution matrices defined by the source reviews

CCA corrected covered area, CK creatine kinase, CRP C-reactive protein, CWI cold-water immersion, DOMS delayed onset muscle soreness, HRV heart rate variability, IL-6 interleukin-6, MVC maximal voluntary contraction, PI prediction interval, SMD standardized mean difference

Discussion

Principal Findings

This overlap-aware synthesis included 14 reviews and did not perform second-order pooling. Primary-study overlap was high overall and high or very high for most key outcomes, while all included reviews were rated low or critically low by AMSTAR 2. Across these reviews, CWI was associated with the most consistent reduction in self-reported DOMS. Objective performance, biochemical, inflammatory, and autonomic findings were less consistent, more context-dependent, or too sparse for firm conclusions.

This affects how apparent agreement across reviews should be read. Multiple favorable estimates do not represent the same number of independent replications when the underlying trials are repeatedly reused.

Presenting each source-review estimate with its overlap, population, comparator, timepoint, and methodological quality avoids false precision.

Interpretation and Plausible Physiological Mechanisms

The DOMS pattern is biologically plausible, because cooling can reduce tissue temperature and nerve conduction velocity, alter nociceptive input, and increase pain threshold, while immersion changes hydrostatic pressure and fluid distribution [2, 3, 8, 9]. However, these responses may differ with training level, sex, body composition, exercise mode, and the magnitude of muscle damage. Mechanistic plausibility in mixed physically active samples should not be assumed to apply uniformly to elite athletes or to every sport.

Performance recovery is multifactorial. Sprint, jump, strength, and power tests differ in neuromuscular

Table 3 Descriptive protocol, comparator, and population characteristics

Characteristic	Observed review-level pattern	Illustrative source reviews	Principal limitation	Interpretation
Water temperature	Across the included reviews, approximately 5–15 °C was the most commonly reported range; some DOMS subgroups/nodes were favorable at 10–15 °C or 11–15 °C	Batista et al. [27]; Machado et al. [28]; Wang et al. [29]	Mixed ranges within reviews; dependent NMA nodes; very few protocols <5 °C	Descriptive observation only; no optimal temperature established
Immersion duration	Approximately 5–15 min was most commonly studied; some DOMS analyses favored 10–15 min	Batista et al. [27]; Machado et al. [28]; Wang et al. [29]	Duration was confounded with temperature, depth, and exercise model	No independent duration-response inference
Immersion depth	Lower-limb, hip/iliac–spine, and deeper immersion were variably reported	Galvez-Rodriguez et al. [20]	Depth was often unreported or combined across trials	No conclusion on optimal immersion depth
Timing and frequency	Application was usually immediate/soon after exercise and generally involved a single bout	Most included reviews [5–7, 18, 20–29]	Limited variation and inconsistent reporting; repeated use was not consistently separated	Findings primarily concern acute, single-bout recovery
Comparator	Passive rest, active recovery, thermal-neutral/warm-water immersion, contrast water therapy, body cryotherapy, air cryotherapy, massage, and network controls were used	Chen et al. [21]; Liu and Ma [26]; Moore et al. [7]	Comparators are not exchangeable and can change effect direction	Interpret within the source comparator; do not combine indiscriminately
Participant/training status	Samples ranged from non-athletes and recreationally active adults to trained, elite, and professional athletes	All included reviews [5–7, 18, 20–29]; Veen et al. [18]	Most reviews combined training levels	Do not generalize mixed-population findings directly to all athletes
Sex representation	Most samples were predominantly male; women were under-represented	Choo et al. [23]; Higgins et al. [25]; Veen et al. [18]	Sex-specific estimates were rarely available	Applicability to female athletes remains uncertain
Exercise/sport context	Eccentric, resistance, endurance, HIIT, team-sport, and soccer-specific models were represented	All included reviews [5–7, 18, 20–29]	Recovery demands and outcome tests differed materially	Interpret effects within the exercise and sport context

DOMS delayed onset muscle soreness, HIIT high-intensity interval training, NMA network meta-analysis

Table 4 Structured interpretation of the evidence by outcome

Outcome/domain	Direction and magnitude	Temporal consistency	Population/comparator dependence	AMSTAR 2 confidence of contributing reviews	Structured conclusion
DOMS	Mostly favorable; overall SMDs -1.04 to -0.37	Favorable in several immediate/delayed analyses, but not every estimate	Largest effect was soccer-specific; comparator structures varied	13 reviews: 1 low, 12 critically low	Consistent favorable pattern, but clinical importance and expectation bias remain uncertain
Perceived fatigue/recovery	Immediate favorable signals; later estimates mixed or sparse	No stable pattern beyond immediate recovery	Subjective outcomes and comparator choice matter	5 reviews: all critically low	Possible immediate subjective benefit; insufficient for broader claims
Jump	Favorable, null, and early unfavorable estimates	No reproducible benefit across review contexts	Test, population, and comparator dependent	8 reviews: all critically low	Mixed evidence; no uniform objective benefit
Sprint	Mostly null; no clear overall soccer effect	No stable temporal pattern	Sprint construct and training level varied	3 reviews: all critically low	No consistent benefit
Strength	Selected favorable estimates, including soccer-specific MVC	Not reproduced uniformly; source PIs crossed zero	Strength tests and populations differed	5 reviews: all critically low	Possible context-specific benefit; uncertain reproducibility
Power	Favorable, mixed, and comparator-dependent broad syntheses	Sparse	Source constructs combined heterogeneous functional tests	3 reviews: all critically low	Insufficient for a general performance conclusion
CK	Several favorable estimates with null findings in other reviews	Time- and population-dependent	Indirect biomarker with marked inter-individual variability	10 reviews: all critically low	Possible biochemical signal; functional meaning uncertain
CRP	Mostly null/imprecise	No reproducible favorable pattern	Sparse, redundant evidence	3 reviews: all critically low	No consistent benefit
IL-6	No clear source-review effect	Sparse	Two reviews with no shared study family	2 reviews: both critically low	Insufficient evidence, not evidence of no effect
Lactate	Immediate null; delayed favorable estimates from one review	Time-dependent and sparse	Only two overlapping reviews	2 reviews: both critically low	Sparse favorable delayed signal; not confirmatory
HRV/autonomic recovery	Narrative signal compatible with acute parasympathetic reactivation	Mostly acute; durability unclear	One narrative review; heterogeneous protocols/outcomes	1 review: low	Insufficient quantitative evidence with a favorable narrative signal

Contributing reviews are those included in the outcome-specific overlap matrices; outcome-specific CCA is reported in Table 2. Only two reviews were rated low (Batista et al., contributing to DOMS; Galvez-Rodriguez et al., contributing to HRV); all other contributing reviews were rated critically low. Conclusions integrate effect direction and magnitude, temporal consistency, population and comparator context, methodological confidence, overlap, and prediction intervals where available; they are not formal certainty grades

AMSTAR 2 A MeaSurement Tool to Assess Systematic Reviews 2, CCA corrected covered area, CK creatine kinase, CRP C-reactive protein, DOMS delayed onset muscle soreness, HRV heart rate variability, IL-6 interleukin-6, MVC maximal voluntary contraction, PI prediction interval, SMD standardized mean difference

demands and measurement error, and their response can depend on exercise mode, comparator, and assessment time. The trained-soccer review added a sport-specific signal for maximal voluntary contraction and counter-movement jump but not sprint performance; prediction intervals indicated that future settings could produce null effects [18]. This sport-specific evidence should not be generalized to all athletic populations.

CK should also be interpreted cautiously. It is an indirect marker of muscle membrane disruption with marked inter-individual variability related to training status, muscle mass, sex, ethnicity, and exercise exposure [10, 30]. A lower CK concentration does not necessarily indicate better functional recovery. The absence of consistent CRP or IL-6 effects further argues against using circulating biomarkers alone as evidence of a broad anti-inflammatory recovery mechanism.

Effect Magnitude, Subjective Outcomes, and Clinical Meaning

Overall DOMS SMDs ranged from small-to-moderate-to-large across source reviews, but the scales, populations, exercise models, and comparators differed. No common minimal important difference was available across the review set, so statistical compatibility with benefit cannot be translated directly into clinically or performance-relevant improvement. DOMS is also a subjective outcome that can be influenced by treatment expectations; an experimental placebo-controlled study found that perceived recovery benefits after CWI were not greater than a thermoneutral placebo condition [31]. The DOMS signal should, therefore, be interpreted as a consistent reduction in reported soreness, not proof of accelerated tissue repair.

Population and Contextual Applicability

The included reviews ranged from non-athletes and recreationally active adults to trained, elite, and professional athletes. Most source reviews combined training levels, and only Veen et al. focused specifically on trained-soccer players [18]. Practical conclusions apply to healthy and physically active populations broadly, with sport-specific use considered only when the population, exercise demands, comparator, and recovery window match the underlying evidence. Long-term training adaptation was outside the scope of this review and has been evaluated separately [32–34].

Strengths and Limitations

Strengths include source verification of all retained estimates, explicit comparator and time labels, a primary-study family overlap matrix with overall and outcome-specific CCA, and avoidance of second-order

pooling of statistically dependent estimates. The synthesis also separates standardized and source-unit effects and reports direction-harmonized prediction intervals where available.

Several limitations remain. First, CCA quantifies redundancy but not the covariance among overlapping meta-analytic estimates; it cannot make second-order pooling valid. Second, all reviews had low or critically low AMSTAR 2 confidence, and weaknesses in source reviews carry forward into an umbrella review. Third, participant populations, exercise models, comparators, protocols, outcome scales, and recovery windows were heterogeneous. Sex representation was also imbalanced, with women under-represented and sex-specific estimates rarely available, limiting applicability. Fourth, some outcomes were represented by one or few reviews, and multiple timepoints within a review were dependent. Fifth, outcome-specific CCA depended on the availability of source-linked study-to-outcome contribution mapping. Sixth, the protocol observations remain review-level and cannot establish dose–response. Finally, HRV and several inflammatory outcomes were too sparse for quantitative conclusions.

Practical Implications and Priorities for Future Reviews

CWI may be considered when the immediate objective is short-term soreness management, but the expected magnitude and practical value should be judged against the athlete or participant, exercise context, comparator, and next performance demand. Current evidence does not justify presenting CWI as uniformly effective for performance recovery or systemic inflammation, and it does not establish an optimal temperature-duration prescription. Future systematic reviews should preregister protocols, provide complete excluded-study lists, report duplicate selection and extraction, disclose funding of primary studies, apply an appropriate risk-of-bias tool, and integrate risk-of-bias findings into conclusions. They should also report exact participant training status and sex, exercise model, comparator, immersion depth, temperature, duration, timing, outcome scale, and timepoint; provide study-to-outcome contribution tables; and avoid combining non-exchangeable comparators. These practices would improve AMSTAR 2 confidence and permit more reliable overlap and dose–response analyses.

Conclusions

Within the available review-level evidence, which was highly overlapping and methodologically limited, CWI was associated with the most consistent reduction in self-reported short-term post-exercise DOMS. The clinical importance of this reduction remains

uncertain, subjective expectation effects cannot be excluded, and the apparent consistency partly reflects reuse of the same primary studies. Evidence for objective performance, CK, lactate, CRP, IL-6, and HRV was mixed, context-dependent, or insufficient. CWI should, therefore, be viewed as a targeted option for soreness management rather than a uniformly effective recovery strategy, and current review-level data do not support an optimized protocol.

Abbreviations

AMSTAR 2	A MeaSurement Tool to Assess Systematic Reviews 2
CCA	Corrected covered area
CI	Confidence interval
CK	Creatine kinase
CRP	C-reactive protein
CWI	Cold-water immersion
DOMS	Delayed onset muscle soreness
HRV	Heart rate variability
IL-6	Interleukin-6
MD	Mean difference
NMA	Network meta-analysis
PI	Prediction interval
PICOS	Participants, intervention, comparators, outcomes, and study design
PRIOR	Preferred reporting items for overviews of reviews
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
SMD	Standardized mean difference
WMD	Weighted mean difference

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40798-026-01132-8>.

Additional file 1. Contains Tables S1–S10 and Figures S1–S7, including full search strategies, expanded review and protocol characteristics, AMSTAR 2 appraisal, overall and outcome-specific CCA, CCA sensitivity analyses, source-review SMD and MD/WMD estimates, parameter findings, comparator-stratified evidence, pairwise overlap, and primary-study multiplicity

Additional file 2. Direction-harmonized SMD and Hedges g estimates from the 14 included reviews, with exact source time labels, comparator categories, raw and harmonized prediction intervals, source locations, and dependence metadata

Additional file 3. Metric-specific estimates retained in their source units and comparator contexts

Additional file 4. Contains the review key, overall and outcome-specific CCA calculations, study-review matrices, pairwise overlap, primary-study multiplicity, sensitivity analyses, study-to-outcome mapping, and mapping rules for the 14 included reviews

Acknowledgements

Not applicable.

Author Contributions

XM and WL conceived the review and developed the study design. XM and GS conducted the literature search and initial screening. XM, CL, and XY performed full-text assessment, data extraction, and source verification. CL and XY contributed to AMSTAR 2 appraisal and primary-study mapping. WL reconstructed the primary-study overlap analysis, designed the overlap-aware structured synthesis, and re-derived the AMSTAR 2 appraisal for the revised review set. XM drafted the manuscript, and WL rewrote the Methods, Results, and Discussion sections affected by the revised synthesis. GS, CL, and XY critically revised the manuscript for methodological content, statistical interpretation,

and clinical relevance. WL and XM contributed equally to this work and are joint first authors. All authors read and approved the final manuscript.

Funding

No funds, grants, or other support were received for this study.

Data Availability

The datasets supporting the conclusions of this article are included within the article and its additional files. Additional files 2 and 3 contain the source-verified SMD and MD/WMD extraction datasets, respectively. Additional file 4 contains the primary-study overlap matrices, corrected covered area calculations, sensitivity analyses, study-to-outcome mapping, and mapping rules. No individual participant data were generated or analyzed.

Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Competing Interests

Weibao Liang, Xiaoge Ma, Guanping Sun, Chuannan Liu, and Xujie Yan declare that they have no competing interests.

Author details

¹Faculty of Health Sciences and Sports, Macao Polytechnic University, Macao, SAR, China. ²School of Physical Education and Leisure, Guangdong Ocean University, Zhanjiang, China. ³Department of Physical Education, Kunsan National University, Gunsan-si, South Korea. ⁴School of Physical Education and Health, Zunyi Medical University, Zunyi, China. ⁵School of Physical Education and Sports Science, Hengyang Normal University, Hengyang, China.

Received: 18 April 2026 Accepted: 13 October 2026

Published online: 23 October 2026

References

1. Versey NG, Halson SL, Dawson BT. Water immersion recovery for athletes: effect on exercise performance and practical recommendations. *Sports Med.* 2013;43:1101–30. <https://doi.org/10.1007/s40279-013-0063-8>.
2. Ihsan M, Watson G, Abbiss CR. What are the physiological mechanisms for post-exercise cold water immersion in the recovery from prolonged endurance and intermittent exercise? *Sports Med.* 2016;46:1095–109. <https://doi.org/10.1007/s40279-016-0483-3>.
3. Bleakley CM, Davison GW. What is the biochemical and physiological rationale for using cold-water immersion in sports recovery? A systematic review. *Br J Sports Med.* 2010;44:179–87. <https://doi.org/10.1136/bjsm.2009.065565>.
4. Dupuy O, Douzi WF, Theurot D, Bosquet L, Dugue B. An evidence-based approach for choosing post-exercise recovery techniques to reduce markers of muscle damage, soreness, fatigue, and inflammation: a systematic review with meta-analysis. *Front Physiol.* 2018;9:403. <https://doi.org/10.3389/fphys.2018.00403>.
5. Leeder J, Gissane C, van Someren K, Gregson W, Howatson G. Cold water immersion and recovery from strenuous exercise: a meta-analysis. *Br J Sports Med.* 2012;46:233–40. <https://doi.org/10.1136/bjsports-2011-090061>.
6. Moore E, Fuller JT, Buckley JD, Saunders S, Halson SL, Broatch JR, et al. Impact of cold-water immersion compared with passive recovery following a single bout of strenuous exercise on athletic performance in physically active participants: a systematic review with meta-analysis and meta-regression. *Sports Med.* 2022;52:1667–88. <https://doi.org/10.1007/s40279-022-01644-9>.
7. Moore E, Fuller JT, Bellenger CR, Saunders S, Halson SL, Broatch JR, et al. Effects of cold-water immersion compared with other recovery modalities on athletic performance following acute strenuous exercise

- in physically active participants: a systematic review, meta-analysis, and meta-regression. *Sports Med.* 2023;53:687–705. <https://doi.org/10.1007/s40279-022-01800-1>.
8. Roberts LA, Muthalib M, Stanley J, Lichtwark G, Nosaka K, Coombes JS, et al. Effects of cold water immersion and active recovery on hemodynamics and recovery of muscle strength following resistance exercise. *Am J Physiol Regul Integr Comp Physiol.* 2015;309:R389–98. <https://doi.org/10.1152/ajpregu.00151.2015>.
 9. Algaflly AA, George KP. The effect of cryotherapy on nerve conduction velocity, pain threshold and pain tolerance. *Br J Sports Med.* 2007;41:365–9. <https://doi.org/10.1136/bjsm.2006.031237>.
 10. Peake JM, Neubauer O, Della Gatta PA, Nosaka K. Muscle damage and inflammation during recovery from exercise. *J Appl Physiol.* 2017;122:559–70. <https://doi.org/10.1152/jappphysiol.00971.2016>.
 11. Pieper D, Antoine S-L, Mathes T, Neugebauer EAM, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. *J Clin Epidemiol.* 2014;67:368–75. <https://doi.org/10.1016/j.jclinepi.2013.11.007>.
 12. Hennessy EA, Johnson BT. Examining overlap of included studies in meta-reviews: guidance for using the corrected covered area index. *Res Synth Methods.* 2020;11:134–45. <https://doi.org/10.1002/jrsm.1390>.
 13. Lunny C, Pieper D, Thabet P, Kanji S. Managing overlap of primary study results across systematic reviews: practical considerations for authors of overviews of reviews. *BMC Med Res Methodol.* 2021;21:140. <https://doi.org/10.1186/s12874-021-01269-y>.
 14. Belbasis L, Bellou V, Ioannidis JPA. Conducting umbrella reviews. *BMJ Med.* 2022;1:e000071. <https://doi.org/10.1136/bmjmed-2021-000071>.
 15. Gates M, Gates A, Pieper D, Fernandes RM, Tricco AC, Moher D, et al. Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. *BMJ.* 2022;378:e070849. <https://doi.org/10.1136/bmj-2022-070849>.
 16. 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.* 2020;372:n71. <https://doi.org/10.1136/bmj.n71>.
 17. Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: overviews of reviews. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane handbook for systematic reviews of interventions*. Version 6.5. Cochrane; 2024. <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-v>. Accessed 18 Apr 2026.
 18. Veen J, Bergh C, Cao Y, Randers MB, Krstrup P, Edholm P. The impact of cold-water immersion on post-match recovery in trained soccer players: a systematic review and meta-analysis. *Scand J Med Sci Sports.* 2026;36:e70202. <https://doi.org/10.1111/sms.70202>.
 19. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008. <https://doi.org/10.1136/bmj.j4008>.
 20. Galvez-Rodriguez C, Valenzuela-Reyes P, Fuentealba-Sepulveda S, Farias-Valenzuela C, Salinas AE. Cold water immersion, heart rate variability and post-exercise recovery: a systematic review. *Physiother Res Int.* 2025;30:e70033. <https://doi.org/10.1002/pri.70033>.
 21. Chen R, Ma X, Ma X, Cui C. The effects of hydrotherapy and cryotherapy on recovery from acute post-exercise induced muscle damage: a network meta-analysis. *BMC Musculoskelet Disord.* 2024;25:749. <https://doi.org/10.1186/s12891-024-07315-2>.
 22. Ma J, Guo C, Luo L, Chen X, Zhang K, Liang D, et al. Comparison of the effects of cold-water immersion applied alone and combined therapy on the recovery of muscle fatigue after exercise: a systematic review and meta-analysis. *Life (Basel).* 2025;15:1205. <https://doi.org/10.3390/life15081205>.
 23. Choo HC, Lee M, Yeo V, Poon W, Ihsan M. The effect of cold water immersion on the recovery of physical performance revisited: a systematic review with meta-analysis. *J Sports Sci.* 2022;40:2608–38. <https://doi.org/10.1080/02640414.2023.2178872>.
 24. Xiao FY, Kabachkova AV, Jiao L, Zhao H, Kapilevich LV. Effects of cold water immersion after exercise on fatigue recovery and exercise performance: meta-analysis. *Front Physiol.* 2023;14:1006512. <https://doi.org/10.3389/fphys.2023.1006512>.
 25. Higgins TR, Greene DA, Baker MK. Effects of cold water immersion and contrast water therapy for recovery from team sport: a systematic review and meta-analysis. *J Strength Cond Res.* 2017;31:1443–60. <https://doi.org/10.1519/JSC.0000000000001559>.
 26. Liu S, Ma Y. Effects of cold water immersion vs body cryotherapy on delayed onset muscle soreness and jump performance following acute strenuous exercise in physically active participants: a systematic review and meta-analysis. *Medicine (Baltimore).* 2026;105:e46781. <https://doi.org/10.1097/MD.00000000000046781>.
 27. Batista NP, de Carvalho FA, Machado AF, Micheletti JK, Pastre CM. What parameters influence the effect of cold-water immersion on muscle soreness? An updated systematic review and meta-analysis. *Clin J Sport Med.* 2023;33:13–25. <https://doi.org/10.1097/JSM.0000000000001081>.
 28. Machado AF, Ferreira PH, Micheletti JK, de Almeida AC, Lemes IR, Vanderlei FM, et al. Can water temperature and immersion time influence the effect of cold water immersion on muscle soreness? A systematic review and meta-analysis. *Sports Med.* 2016;46:503–14. <https://doi.org/10.1007/s40279-015-0431-7>.
 29. Wang H, Wang L, Pan Y. Impact of different doses of cold water immersion (duration and temperature variations) on recovery from acute exercise-induced muscle damage: a network meta-analysis. *Front Physiol.* 2025;16:1525726. <https://doi.org/10.3389/fphys.2025.1525726>.
 30. Brancaccio P, Maffulli N, Limongelli FM. Creatine kinase monitoring in sport medicine. *Br Med Bull.* 2007;81:209–30. <https://doi.org/10.1093/bmb/ldm014>.
 31. Broatch JR, Petersen A, Bishop DJ. Postexercise cold water immersion benefits are not greater than the placebo effect. *Med Sci Sports Exerc.* 2014;46:2139–47. <https://doi.org/10.1249/MSS.0000000000000348>.
 32. Roberts LA, Raastad T, Markworth JF, Figueiredo VC, Egner IM, Shield A, et al. Post-exercise cold water immersion attenuates acute anabolic signalling and long-term adaptations in muscle to strength training. *J Physiol.* 2015;593:4285–301. <https://doi.org/10.1113/JP270570>.
 33. Ihsan M, Abbiss CR, Allan R. Adaptations to post-exercise cold water immersion: friend, foe, or futile? *Front Sports Act Living.* 2021;3:714148. <https://doi.org/10.3389/fspor.2021.714148>.
 34. Malta ES, Dutra YM, Broatch JR, Bishop DJ, Zagatto AM. The effects of regular cold-water immersion use on training-induced changes in strength and endurance performance: a systematic review with meta-analysis. *Sports Med.* 2021;51:161–74. <https://doi.org/10.1007/s40279-020-01362-0>.

Publisher's Note

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

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com