

Supplementary Table 5. GRADE Summary of Findings — Cold-water immersion versus passive seated rest for post-exercise recovery

Patient or population: healthy adults and adolescent athletes undergoing post-exercise recovery

Setting: sports and exercise recovery contexts (non-clinical)

Intervention: single-session cold-water immersion (CWI)

Comparison: passive seated rest

Outcome	No of participants (studies)	Certainty of the evidence (GRADE)	Standardized Mean Difference (95% CI)	Comments
Delayed onset muscle soreness (DOMS)	972 (47 RCTs)	⊕⊕⊕○ MODERATE a,b	SMD 0.81 lower (1.10 lower to 0.52 lower)	CWI probably reduces DOMS after exercise.
Creatine kinase (CK)	767 (34 RCTs)	⊕⊕⊕○ MODERATE a,b	SMD 0.64 lower (0.90 lower to 0.30 lower)	CWI probably reduces post-exercise CK concentrations.
Myoglobin (Mb)	169 (7 RCTs)	⊕⊕⊕○ MODERATE c	SMD 0.93 lower (1.40 lower to 0.40 lower)	CWI probably reduces post-exercise Mb concentrations, though based on a small number of studies.

GRADE Working Group grades of evidence

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate.

Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate.

a. Downgraded one level for inconsistency: $I^2 = 73\%$ (DOMS) and 67% (CK), indicating substantial heterogeneity that was not fully explained by temperature-subgroup analysis (between-subgroup $p = 0.79$ for DOMS; $p = 0.72$ for CK).

b. Evidence of possible publication bias was observed based on Egger's test (DOMS, $p = 0.022$; CK, $p = 0.010$). This is noted as a caveat but was not applied as a formal downgrade, as Egger's test alone was not considered sufficient grounds to reduce certainty.

c. Downgraded one level for imprecision because only seven studies contributed evidence for this outcome.