

Supplementary File 1 - Risk of Bias Education Review

RESULTS

No disagreement existed amongst the two authors regarding the selection of eligible systematic reviews. Disagreement occurred between these two authors for 3.9% of the extracted data (i.e., 41 of 1056 data points).

Review selection:

136 review articles were identified from the *BJSM* journal archive. Following title and abstract screening, 66 systematic reviews (SRs) (24 with and 42 without meta-analysis) were identified for full-text review and subsequent data abstraction.

Domain Classification

Included articles comprised reviews of healthcare interventions (k = 32), observational epidemiology (k = 26), diagnostic accuracy (k = 2), and 'other' (k = 6) [Supplementary Table 1].

Supplementary Table 1 – Domain classification of included systematic reviews, as determined by pre-specified review question.

Review Domain
Healthcare intervention (k = 32)
Observational Epidemiology (k = 26)
<ul style="list-style-type: none"> • Etiology (k = 11) • Prognosis (k = 10) • Assessment (k = 5)
Diagnosis (k = 2)
Other (k = 6)
Total (K = 66)

Risk of bias tool utilized:

Reviews varied widely in their implementation of a risk of bias tool [Table 2]. Across sixty-five systematic reviews, twenty-five discrete risk of bias tools were implemented to critically appraise included studies. The most frequently utilized risk of bias tools, be it in original form or as modified by authors, included PEDro (k = 12; 18.5%), Downs & Black Quality Assessment Scale (k = 12; 18.5%), and the Cochrane Risk of Bias tool for Randomized Controlled Trials (k = 10; 15%). Other study design-specific tools, implemented within systematic reviews, comprised the Newcastle-Ottawa Scale, to assess the methodological quality of observational studies (k = 4; 6%), QUADAS-II, to evaluate the internal validity of diagnostic accuracy studies (k = 2; 3%), and QAREL, to perform a quality assessment of reliability studies (k = 2; 3%). Systematic review authorship teams developed separate invalidated quality assessment tools to critically appraise included studies within eight systematic reviews (k = 8; 12%). The remaining risk of bias tools identified across the included sports and exercise medicine systematic reviews (k = 16; 25%), were implemented in a systematic review only once [Supplementary Table 2].

Assessment domain (internal validity, methodological quality, reporting, combined, other).

Eleven (17%) systematic reviews implemented a risk of bias tool that evaluated the internal validity of included studies. To assess risk of bias, remaining systematic reviews implemented tools that evaluated methodological quality (50%), reporting (1.5%), a combination of internal validity, external validity, methodological quality, and reporting (18.5%), or 'other' (15%).

Supplementary Table 2 – Frequency of risk of bias tools implemented across BJSM Sports and Exercise Medicine Systematic Reviews.

Critical Appraisal Tool	K = 65
PEDro	k = 12
(modified) Downs & Black	k = 12
Cochrane Risk of Bias Tool for RCTs	k = 10
(modified) Newcastle Ottawa Scale	k = 4
QUADAS-II	k = 2
QAREL	k = 2
Modified QATQS	k = 1
Quality Assessment Tool for Observational & Cross-sectional Studies	k = 1
ROBINS-I	k = 1
ACROBAT-NRSI	k = 1
RoBANS	k = 1
MINORS	k = 1
COSMIN	k = 1
CONSORT	k = 1
AMSTAR	k = 1
Jadad Scale	k = 1
Risk of Bias Tool for Case-Series	k = 1
CASP	k = 1
SIGN	k = 1
CMS	k = 1
EAI	k = 1
MMAT	k = 1
Drummond Checklist	k = 1
Delphi Checklist	k = 1
Tool developed by authors	k = 8
Total	(K = 68)*

*Three SRs implemented two risk of bias tools to account for included studies with different designs.

(modified) = authors omitted/added tool components to assess additional areas of internal validity or methodological quality.

PEDro = Physiotherapy Evidence Database; QUADAS-II = Quality Assessment of Diagnostic Accuracy Studies version II; QAREL = Quality Assessment of Reliability Studies; ROBINS-I = Risk Of Bias In Non-randomized Studies – of Interventions; ACROBAT-NRSI = A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions; RoBANS = Risk of Bias Assessment tool for Non-randomized Studies; Methodological Index for Non-Randomized Studies; COSMIN = Checklist for assessing the methodological quality Of Studies on Measurement properties of health status measurement INSTRUMENTS; CONSORT = CONSolidated Standards Of Reporting Trials; AMSTAR = Assessing the Methodological Quality of Systematic Reviews; CASP = Critical Appraisal Skills Program; SIGN = Scottish Intercollegiate Guidelines Network; CMS = Coleman Methodology Score; EAI = Epidemiological Appraisal Instrument; MMAT = Mixed Methods Appraisal Tool.

CRITICAL RE-ASSESSMENT –**Methods of risk of bias re-assessment:****Systematic review selection criteria**

From the included sample of systematic reviews described above, we sought a systematic review of therapeutic interventions that, intended to assess the risk of bias of included studies but rather evaluated features other than risk of bias. Thus, we sought a systematic review that;

- (1) conducted a critical appraisal of included studies intended to assess risk for bias;
- (2) implemented a standard, established tool;
- (3) used a quality scale;
- (4) generated a summary score, with or without arbitrary cut-off thresholds, to rank included studies;
- (5) did not assess risk of bias for separate outcome types;
- (6) did not conduct a domain-level risk of bias assessment, and;
- (7) included RCTs in the sample of included studies.

Our eligibility criteria identified a systematic review of therapeutic interventions, conducted by Kosik et al., investigating the efficacy of therapeutic interventions to improve patient-reported function in participants with chronic ankle instability (CAI).

Results of risk of bias re-assessment:

Supplementary Table 3 – Frequency (n) and proportion (%) of randomized controlled trials at low risk, some concerns, and high risk of bias.

Risk of Bias Domains	Low Risk	Some Concerns	High Risk
Bias arising from the randomization process	1 (9.1%)	8 (72.7%)	2 (18.2%)
Bias due to deviation from intended interventions	2 (18.2%)	2 (18.2%)	7 (54.5%)
Bias due to missing outcome data	9 (81.8%)	0 (0.0%)	2 (18.2%)
Bias in measurement of the outcome	2 (18.2%)	1 (9.1%)	8 (72.7%)
Bias in selection of the reported result	0 (0.0%)	10 (90.9%)	1 (9.1%)
Overall risk of bias judgement	0 (0.0%)	0 (0.0%)	11 (100%)