Implementing the 27 PRISMA 2020 Statement items for systematic reviews in the sport and exercise medicine, musculoskeletal rehabilitation and sports science fields: the PERSiST (implementing Prisma in Exercise, Rehabilitation, Sport medicine and SporTScience) guidance

Clare L Ardern,1,2,3 Fionn Büttner,4 Renato Andrade,5,6 Adam Weir,7,8,9 Maureen C Ashe,3,10 Sinead Holden,11,12 Franco M Impellizzeri,13 Eamonn Delahunt,4 H Paul Dijkstra,14,15 Stephanie Mathieson,16 Michael Skovdal Rathleff,11,12 Guus Reurink,17 Catherine Sherrington,16 Emmanuel Stamatakis,18 Bill Vicenzino,19 Jackie L Whittaker,20,21 Alexis A Wright,22 Mike Clarke,23 David Moher,24 Matthew J Page,25 Karim M Khan,3,26,27 Marinus Winters,12

ABSTRACT
Poor reporting of medical and healthcare systematic reviews is a problem from which the sports and exercise medicine, musculoskeletal rehabilitation, and sports science fields are not immune. Transparent, accurate and comprehensive systematic review reporting helps researchers replicate methods, readers understand what was done and why, and clinicians and policy-makers implement results in practice. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement and its accompanying Explanation and Elaboration 11 document provide general reporting examples for systematic reviews of healthcare interventions. However, implementation guidance for sport and exercise medicine, musculoskeletal rehabilitation, and sports science does not exist. The Prisma in Exercise, Rehabilitation, Sport medicine and SporTScience (PERSiST) guidance attempts to address this problem. Nineteen content experts collaborated with three methods experts to identify examples of exemplary reporting in systematic reviews in sport and exercise medicine (including physical activity), musculoskeletal rehabilitation (including physiotherapy), and sports science, for each of the PRISMA 2020 Statement items. PERSiST aims to help: (1) systematic reviewers improve the transparency and reporting of systematic reviews and (2) journal editors and peer reviewers make informed decisions about systematic review reporting quality.

INTRODUCTION
High-quality systematic reviews deliver quality evidence to readers in a timely fashion—supporting informed decision making in practice.1 Poor reporting of medical and healthcare systematic reviews is prevalent, including in sport and exercise medicine, musculoskeletal rehabilitation and sports science.2 3 Systematic reviews must be clearly, transparently, accurately and comprehensively reported. There must be sufficient detail to allow researchers to replicate methods, for readers to understand what the systematic reviewers have done and why, and for clinicians/practitioners and policy-makers to act.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 Statement, published in March 2021, in five leading biomedical journals—a substantially revised and updated version of the PRISMA 2009 Statement—outlines the minimum items that should be reported in a systematic review to improve transparency and interpretation. The PRISMA Statement is recommended by the EQUATOR Network as the reporting guideline for systematic reviews. The PRISMA 2020 Statement and its accompanying Explanation and Elaboration document explain and elaborate general reporting examples for systematic reviews of healthcare interventions, but do not include implementation guidance specific for the context of sport and exercise medicine, musculoskeletal rehabilitation and sports science. Context-specific examples and elaboration are likely to enhance implementation.12 The PRISMA 2020 Statement focuses on systematic reviews of interventions, but it can be used as a basis for reporting systematic reviews of other types of research (eg, aetiology, prevalence, prognosis).10

SCOPE OF THE PRISMA IN EXERCISE, REHABILITATION, SPORT MEDICINE AND SPORTS SCIENCE GUIDANCE
Prisma in Exercise, Rehabilitation, Sport medicine and SporTScience (PERSiST) aims to support systematic reviewers in the sport and exercise medicine (including physical activity), musculoskeletal rehabilitation (including physiotherapy and physical...
Consensus statement

Box 1  Key terms

Systematic review
A systematic review employs specific, systematic methods of searching, selecting, assessing, collating and synthesising evidence to address a clearly formulated review question.44

Systematic review methods aim to minimise bias and maximise the practical relevance of the results to a broader spectrum of end users (eg clinicians, practitioners, patients, athletes, care givers, researchers and policy-makers).

Meta-analysis
Meta-analysis is a statistical technique for synthesising results when appropriate data (eg, effect estimates and their variances) are available, to yield a quantitative summary of the results.30

When making decisions about meta-analysis, consider the available data, methods and clinical differences (heterogeneity) among included studies.46 54 Data pooling using meta-analysis can provide a more precise estimate of treatment effect, diagnostic accuracy or prognosis because of greater statistical power when the results of multiple studies are combined.

Network meta-analysis
Traditional systematic reviews compare two different treatments (or compare one treatment to a control). Network meta-analysis is a statistical technique for synthesising and comparing results from more than two different interventions, even when the interventions have not been directly compared in a trial (the analysis builds a network of interventions based on interventions that have been directly and indirectly compared in different trials).51

Network meta-analysis is a more sophisticated analysis approach because it compares all available interventions, can produce a more precise effect estimate and can rank interventions from most to least effective.48 We recommend collaborating with a statistician to plan and conduct a network meta-analysis.

Individual participant data meta-analysis
Systematic reviewers typically pool aggregate data (eg, group means) extracted from included studies. Individual participant data (IPD) meta-analysis involves systematic reviewers collecting the original data for each participant included in the eligible studies, validating/checking and reanalysing the data.62

Compared with an aggregate data review, synthesising IPD can substantially improve the quantity and quality of data, and ultimately produce a more robust synthesis of the field.62 We recommend collaborating with a statistician to plan and conduct an IPD meta-analysis.

Establishing the PERSiST project contributing authors
Nineteen PERSiST team members (working group; 7 women, 12 men) were recruited from the primary research areas of sports medicine (n=13), exercise medicine (n=6), physiotherapy and physical therapy (n=10), musculoskeletal rehabilitation (n=13), physical activity (n=8) and sports science (n=2) (N.B. numbers sum to >19 because working group members each had relevant experience in more than 1 of the primary research areas). We balanced early career researchers (within 7 years of PhD award) (n=9), clinician researchers (n=7) and senior researchers (n=10). All had systematic review methods knowledge and experience, and had contributed to at least one systematic review in their primary research field. Three methodologists, who were members of the group that developed the PRISMA 2009 and 2020 Statements,13 40 served as advisors and provided feedback on the specific examples we proposed. The working group identified appropriate examples; the advisory panel provided input into the structure and organisation of the PERSiST project.

Identifying and appraising examples
We established five teams (three working group members per team, balancing gender, research experience and field of expertise) and allocated up to six of the PRISMA 2020 Statement4 items to each team. Each team was tasked with collaborating to identify exemplary reporting from systematic reviews in sport and exercise medicine (including physical activity), musculoskeletal rehabilitation (including physiotherapy and physical therapy) or sports science. No systematic search was conducted to identify published examples.

Each team sent the examples for each item they were assigned to the project leader who collated them into a single document and circulated it to the working group members for review and feedback. Over two feedback rounds (of 2 weeks each), the 19 working group members considered the examples and provided written feedback via group email. After each feedback round, the PERSiST project leader collated, summarised and synthesised the feedback, and circulated the draft examples. We used a final consensus meeting round (via group email) to decide on the draft examples for all 27 main items and 12 abstract items. The examples were then reviewed by the advisory panel, approved and finalised. Including an example does not imply anything about the overall quality of the full systematic review from where the example was drawn.

How to use the PERSiST guidance
PERSiST complements the primary PRISMA 2020 Statement4 to help systematic reviewers in sport and exercise medicine, musculoskeletal rehabilitation, and sports science implement the PRISMA 2020 items in their research context. We recommend systematic reviewers use PERSiST alongside the PRISMA 2020 Statement4 and PRISMA 2020 Explanation and Elaboration11 when planning and reporting their systematic review; systematic reviewers may also find the METHODS MATTER statement13 helpful for general guidance on research design, methods and reporting.

PERSiST presents at least 1 exemplar illustrating exemplary reporting for each of the 27 PRISMA 2020 Statement4 items. Some examples were lightly edited for flow, including removing citations or web addresses. In some examples, we highlight additional considerations, share helpful resources and make suggestions for reporting systematic reviews in sport and exercise medicine, musculoskeletal rehabilitation, and sports science—these are boxes headed: Note for systematic reviewers.

METHODS
In this section, we outline the process for developing the PERSiST guidance and explain how the paper is intended to be used. An expert panel collaborated to produce PERSiST.
Where relevant, we retained the exemplar’s citations or hyperlinks. Citations were renumbered to appear in the PERSiST guidance reference list. The examples are intended to guide systematic reviewers in sport and exercise medicine, musculoskeletal rehabilitation, and sports science regarding what to report—systematic reviewers will make decisions about how to present information (eg, whether to use tables, what information is appropriate for appendices/supplementary files) based on the review content and journal requirements.

Terminology
We adopted and followed the terminology of the PRISMA 2020 Explanation and Elaboration box 1. Most systematic reviews use group level (aggregate) data (eg, mean and SD) from included studies. More sophisticated approaches to data analysis and synthesis include network meta-analysis, individual participant data (IPD) meta-analysis, umbrella reviews (sometimes called overviews or review of reviews) and prospective meta-analysis.

Systematic reviewers who are considering these more sophisticated approaches should visit the PRISMA website (www.prisma-statement.org) for guidance on reporting (including PRISMA Statement extensions for network meta-analysis, IPD meta-analysis, scoping reviews, diagnostic test accuracy, reporting of harms and health equity studies, and protocols). Other extensions are in development for: newborn and child health research (PRISMA), rapid reviews (PRISMA), ethically sensitive topics (PRISMA-Ethics), animal research (PRISMA Extension of Preclinical In Vivo Animal Experiments) and outcome measurement instruments (PRISMA-COSMIN).

IMPLEMENTING THE PRISMA 2020 STATEMENT: EXAMPLES FOR SPORT AND EXERCISE MEDICINE, MUSCULOSKELETAL REHABILITATION, AND SPORTS SCIENCE

Title
Item 1: title
Identify the report as a systematic review.

Example: ‘Comparative effectiveness of treatment options for plantar heel pain: a systematic review with network meta-analysis.’

Example: ‘Effectiveness of conservative interventions including exercise, manual therapy and medical management in adults with shoulder impingement: a systematic review and meta-analysis of randomised controlled trials (RCTs).’

Abstract
Item 2
See the PRISMA 2020 for Abstracts checklist (see list and examples in the summary box at the end of this document).

Introduction
Item 3: rationale
Describe the rationale for the review in the context of existing knowledge.

Example: ‘The International Olympic Committee, among others, has called for more diligence to safeguard the physiological development of the paediatric athlete. Performing a cardiac preparticipation evaluation within paediatric populations is controversial due to a lack of international consensus with regard to when, how and who should undertake such examinations. While data from the USA indicate that paediatric black athletes are particularly susceptible to sudden cardiac death (SCD), there is a general lack of understanding as to which factors (eg, physical growth, race and sex) have the potential to increase the likelihood of generating a false-positive diagnosis and unnecessary disqualification from competitive sport. Consequently, the distinction between paediatric athlete’s heart and cardiac pathology associated with SCD is especially important for this population.’

Methods
Item 5: eligibility criteria
Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.

Note for systematic reviewers. Structure the eligibility criteria according to the framework used to define the review objective(s) or question(s) (eg, PICO) to help readers understand the scope of the systematic review.

Example:

‘Types of studies
Eligible RCTs were identified from systematic reviews investigating the effects of exercise therapy and published in the Cochrane database of Systematic Reviews. RCTs, cluster-randomised trials and randomised crossover studies were included if they compared an exercise therapy intervention with a non-exercising control treatment.

Types of participants
Studies that included participants with or without a medical condition were eligible, except for participants receiving chemotherapy, as all or nearly all these participants are anticipated to experience adverse events. Otherwise, no studies were excluded based on specific characteristics of the participants.

Types of intervention
Exercise therapy was the main intervention and each exercise session had to include active exercise therapy for at least 50% of the total time. Furthermore, the exercise could not be combined with any pharmacological, surgical or electrotherapeutic...
intervention. Besides strength/resistance, aerobic and neuromuscular exercise (defined as exercise interventions targeting sensorimotor deficiencies and functional stability), the following active exercise interventions were also included: dancing, running, cycling, QiGong and Tai Chi. However, interventions like whole body vibration, facial exercises, yoga, stretching or range of motion exercises or bladder training were excluded. There were no restrictions on the setting in which the exercise therapy was performed, that is, classes, gymnasium, etc.

Types of control intervention
Studies with comparators such as a non-exercise therapy control group, usual care, attention intervention, etc were included, as were nutraceuticals, placebo and education (eg, back school and similar interventions). However, studies where the control group involved any exercise (including stretching), pharmacological, surgical or electrotherapeutic intervention were excluded.

Type of outcome measures
The outcomes of interest were measures of adverse events. As classifying adverse events as treatment-related is largely subjective, and with unknown validity, the current study was not focused on reported adverse effects, but on adverse events as any undesirable event occurring during the study, divided into serious and non-serious adverse events.36

Example:
‘Electronic searches
We [...] searched the following databases from inception to 11 March 2016 without restrictions to language or publication status:
► Cochrane Central Register of Controlled Trials (CENTRAL, which includes the Cochrane Back and Neck group (CBN trials register) (The Cochrane Library, 2016, Issue 2).
► MEDLINE (OvidSP, 1946 to March week 1 2016; Appendix 1).
► MEDLINE In-Process & Other Non-Indexed Citations (OvidSP, 10 March 2016),
► Embase (OvidSP, 1980 to 2016 week 10),
► Cumulative Index to Nursing and Allied Health Literature (CINAHL) (EBSCO, 1981 to 11 March 2016),
► PsycINFO (OvidSP, 2002 to March week 2 2016),
► Allied and Complementary Medicine Database (AMED) (OvidSP, 1985 to March 2016),
► CBN Trials Register (Cochrane Register of Studies (CRS)),
► Cochrane Complementary Medicine Field Trials Specialized Register (Cochrane Register of Studies Online (CRSO)),
► IndMED,
► PubMed,
► US National Institutes of Health ClinicalTrials.gov,
► World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP).

The searches were previously run in 2013 and 2014. In 2014, the ClinicalTrials.gov, WHO ICTRP and a supplementary search of the CBN Specialised Register in the CRS were added to the search strategy. In 2016, the PubMed search was revised to capture studies not in MEDLINE using the strategy recommended by Duffy 2014. The Information Specialist of the CBN conducted all searches except for the Cochrane Complementary Medicine Field Specialised Register, which we searched through the CRSO.

Searching other resources
We screened the reference lists of included studies and contacted experts in the field (eg, authors of included studies) for information on additional trials, including unpublished or ongoing studies.37

Example: ‘We conducted a comprehensive database search using PubMed, EMBASE, Cochrane Library, SPORT-Discus and PEDro to search clinical practice guidelines (CPGs) that presented rehabilitation of ACL injuries. To search grey literature and CPG repositories we used the OpenGrey, National Guideline Clearinghouse of the Agency for Healthcare Research and Quality, Guidelines International Network and National Institute for Health and Care Excellence (NICE) databases. [...] all searches up to the 30 September of 2018 [...]’.34

Methods
Item 6: information sources
Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.

Note for systematic reviewers. Ensure search decisions are tailored to the systematic review—avoid arbitrary decisions (eg, number of information sources to search).

For more information about grey literature, publication bias, and why it is important to search for grey literature in sport and exercise medicine, musculoskeletal rehabilitation, and sports science, see Winters and Weir36 and chapter 4 of the Cochrane Handbook (searching for and selecting studies).31 Searching grey (unpublished) literature identifies (1) ongoing studies that could be included when a systematic review is updated or that may warrant delaying publishing the review to ensure the study can be included, (2) completed studies that are unpublished and may never get published (eg, injury surveillance conducted by sports federations) and (3) books, monographs, dissertations, policy documents, reports etc that may address the systematic review question and provide relevant data. Reporting sources of grey literature will help readers determine the risk of bias due to missing (unpublished) studies.

Collaborate with a medical/healthcare librarian or information specialist—professionals who have extensive training in literature searching—when planning, developing and executing systematic review searches.32 Quality searching includes choosing appropriate databases and other information sources, designing search strategies for databases and registers, executing searches and saving and collating the results, documenting and reporting the search, and updating the search.31 For more information we recommend chapter 4 of the Cochrane Handbook.31

Item 7: search strategy
Present the full search strategies for all databases, registers and websites, including any filters and limits used.

Example: ‘Twelve systematic searches covering diagnosis, prevention and treatment for each of the four sections (1: hamstring, 2: adductor, 3: rectus femoris/quadriceps and 4: calf) [...]. No restrictions were applied concerning year of publication, however, only publications in English were included. We
searched individual text words in title and abstract supplemented with Medical Subject Headings (MeSH) terms. We combined anatomical region of interest (eg, ‘Groin (MeSH)’ OR ‘adductor’ OR ‘groin’) AND type of injury (eg, ‘Athletic Injury (MeSH)’ OR ‘Strains and Sprains (MeSH)’ OR ‘strain’’ OR ‘injur**’ OR ‘re-injur**’ OR ‘reinjur**’) AND outcome for diagnosis and treatment domains (eg, ‘Diagnosis (MeSH)’ OR ‘exam?’ AND ‘Return To Sport’ OR ‘full training,’ respectively) OR intervention for prevention domains (eg, ‘Primary Prevention (MeSH)’ OR ‘Reeduc*’). […] A flow chart of searches and the complete search strategy for all searches and databases is available as supplemental.38

**METHODS**

**Item 8: selection process**
Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.

Example: ‘The selection of studies was a three-stage process, with the identified citations independently evaluated for inclusion by two reviewers. The first stage was evaluation of titles selected with systematic searches described above. The article was included in this first screen if the title identified athletes and/or lumbar discectomy. We then reviewed the abstracts of all articles identified as meeting the search criteria. Full-text articles meeting criteria were retrieved and read independently by both reviewers and assessed for inclusion in the study. Disagreement was resolved by consensus between the two reviewers and a third reviewer if consensus could not be reached.39

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Patient-important outcomes identified by the guideline panel aligned to the most corresponding outcome set available in the included trials and studies (reproduced with permission)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel-identified outcomes</td>
<td>Outcomes chosen from included trials and studies</td>
</tr>
<tr>
<td>Pain</td>
<td>Pain (Visual Analogue Scale and Numeric Rating Scale, various scales)</td>
</tr>
<tr>
<td>Physical function</td>
<td>Combined outcomes of physical function, physical capacity and pain items (Constant Score and its modifications, Shoulder Disability Questionnaire, Neer Score)</td>
</tr>
<tr>
<td>Global perceived effect</td>
<td>Global perceived effect. The outcome was derived by subtracting the patients who reported ‘worse’ or ‘much worse’ from the no of patients reporting ‘much better’ or ‘no shoulder problems at all’/’healed completely’ at the relevant time points</td>
</tr>
<tr>
<td>Quality of life</td>
<td>Health-related quality of life (EQ-5D-3L and 15D)</td>
</tr>
<tr>
<td>Participation</td>
<td>Return to leisure activities/sport. Working status (working: yes/no)</td>
</tr>
<tr>
<td>Development of full-thickness rotator cuff tears</td>
<td>Prevalence of full-thickness rotator cuff tears at follow-up</td>
</tr>
<tr>
<td>Harms</td>
<td>Serious harms</td>
</tr>
</tbody>
</table>

**Methods**

**Item 9: data collection process**
Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.

Example: ‘Two reviewers independently extracted data using a specifically designed standardised data extracting form (see study protocol), and afterwards the reviewers compared the extracted data for consistency. All inconsistencies between the two forms were resolved by discussion between the two data extractors. Any disagreement between the data extractors after the initial discussion related to inconsistencies between the two individual data extractions was to be solved by involving a third person. General study information, participants and intervention characteristics, compliance, adverse events, withdrawals and outcome measures were extracted. Where data were not available from tables or the results section, the authors of the study in question were contacted by email, with one reminder after 2 weeks, if they did not respond to the first email.40

**Item 10a: data items**
List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (eg, for all measures, time points, analyses), and if not, the methods used to decide which results to collect.

Example: ‘Outcomes
The guideline panel identified outcomes of importance to patients; we used outcomes from the trials that most closely corresponded to those chosen by the patients (the systematic reviewers worked with 33 patient partners to identify and prioritise the outcomes of interest for the systematic review) (table 1). Because it was the most bothersome symptom for 86% of the patients who completed our survey, we considered pain as the most important outcome in this systematic review.

We defined serious harms as death, bleeding (uncontrolled or requiring transfusion), cardiac arrest requiring cardio-pulmonary resuscitation, myocardial infarction, cerebrovascular accident, acute renal failure, unplanned intubation, requiring ventilator for >48 hours, deep infection (surgical site or organ/space) sepsis, septic shock, pneumonia, wound dehiscence, pulmonary embolism, deep vein thrombosis or peripheral nerve injury. […] We used a priori-defined decision rules for data extraction: 1. When trialists reported final values and change from baseline values for the same outcome, we extracted final values. 2. When trialists reported unadjusted and adjusted values for the same outcome, we extracted the unadjusted values. 3. When trialists reported data based on the intention-to-treat (ITT) sample and another sample (eg, per-protocol, as-treated), we extracted ITT-analysed data.'
Data management
When trials used different outcome measures to evaluate the same construct, we chose the most common outcome measure as the index and transformed mean differences and SD of other outcome measures to the index instrument, and pooled the data using mean difference as the summary estimate. For trials not reporting the index instrument, we followed the data using mean difference as the summary estimate. When trials used different outcome measures to evaluate the same construct, we chose the most common outcome measure as the index and transformed mean differences and SD of other outcome measures to the index instrument, and pooled the data using mean difference as the summary estimate. (If multiple periods during which the pain was evaluated were available, the shortest was chosen).

**Function outcomes and mixed function-capacity-pain scores**
- Oxford Shoulder Score.
- Constant-Murley Score.
- American Shoulder and Elbow Surgeons Standardised Form.
- UCLA Shoulder Score.
- Simple Shoulder Test.
- Shoulder Disability Questionnaire.
- Neer Score.
- Watson-Sonnabend Score.

**Health-related quality of life.**
- EQ-SD.
- 15D.
- EQ-VAS.
- SF-36.

Example of list of other variables for which data were sought:
‘We also extracted the following data: trial characteristics, patient demographic variables, diagnosis, treatment and data about trial methodology. Online supplementary appendix table 1 (box 3) presents a full list of extracted data items.’

Methods
**Item 10b: data items**
List and define all other variables for which data were sought (eg, participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.

Note for systematic reviewers. If you have a long and/or detailed list of variables for which data were sought, consider including a summary of the data items (ie, variables other than outcome(s)) that were extracted as an appendix/supplemental file or as a file uploaded to an online repository (eg, Open Science Framework).

Example of explanation for assumptions made about any missing or unclear information:
‘If authors did not report relevant numeric outcome data in the text, we contacted the authors or, when available, extracted the data from figures and graphs.’

Methods
**Item 11: study risk of bias assessment**
Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.

Example: ‘We used the Risk of Bias 2 tool to assess risk of bias for each trial outcome. We assessed risk of bias on the basis of ‘assignment to intervention’ for all five domains: (1) randomisation process, (2) deviations from intended interventions, (3) missing outcome data, (4) outcome measurement and (5) selection of the reported result. An overall risk of bias judgement was made for each outcome and each time point as either ‘low risk’, ‘some concerns’ or ‘high risk’ of bias.'
Note for systematic reviewers. For information on assessing risk of bias in the context of systematic reviews in the sport medicine, musculoskeletal rehabilitation, and sports science fields, we recommend Büttner et al.42 43 Some systematic reviewers use the terms ‘quality assessment’ and ‘risk of bias assessment’ interchangeably. These terms are not synonymous. Quality is poorly defined, but is often used to convey how well the research was conducted and reported. Bias refers to systematic deviations from the truth, which can occur due to flaws in research design, conduct, analysis and/or reporting (eg, outcome reporting biases, spin, per-protocol analyses).42–44

We recommend systematic reviewers complete and report a risk of bias assessment using a clinimetric tool that is appropriate for the review question (eg, causation, prediction, diagnosis). Use a risk of bias assessment tool that facilitates domain-based assessment wherever possible. Do not calculate and present a numerical ‘methodological quality’ score when you mean to assess risk of bias (including quality assessment scales or composite reporting scales). When choosing a risk of bias assessment tool, carefully consider the ideal study design to answer the review question, and the key sources of bias that may influence the results of the systematic review.

The most relevant sources of bias differ depending on the study design. Assessing risk of bias requires a careful approach, tailored to the key threats to the internal validity (ie, bias) of the research question the systematic review aims to address. Different tools are appropriate for different study designs, and may include:

► The Cochrane ROB 2.0 for assessing bias in randomised controlled trials.
► ROBINS-I for assessing bias in non-randomised intervention studies.
► PROBAST for assessing bias in prediction modelling studies.
► QUIPS for assessing bias in prognostic studies.
► QUADAS-II for assessing bias in diagnostic accuracy studies (see PRISMA-DTA Statement).18

The assessment was performed independently by two reviewers […]. The reviewers did not perform risk of bias assessment or data extraction for publications in which they were involved as an author. Disagreements were resolved via consensus or by a third reviewer […] if necessary.”45

Methods

Item 12: effect measures
Specify for each outcome the effect measure(s) (eg, risk ratio, mean difference) used in the synthesis or presentation of results.

Example:

‘Synthesis of results
For the analysis on benefits, we calculated the effect sizes in the individual studies as standardised mean differences, allowing pooling and comparison of the various outcomes assessed in the individual trials. We estimated the standardised mean difference as the difference between the mean score of the intervention and control groups divided by the pooled SD of the final score. This estimate of the effect size using standardised mean difference has a slight bias overestimating the effect size, and we applied a correction factor to convert the effect size to Hedges’ g.

In the analysis on harms, we transformed the numbers of adverse events into log odds of events, allowing pooling of data from the individual studies. Results are reported as number of adverse events per 1000 procedures with 95% CIs.”46

Example: ‘We conducted meta-analyses, guided by considerations of bias. For the primary comparison, we pooled data from comparisons at low risk of bias. For the secondary comparison, we pooled data from comparisons irrespective of bias.

We assessed outcomes at 3 months, 6 months, 1 year, 2 years (for which we pooled data up to 3 years if no 2-year data were available), 5 years (we prioritised time points closest to 5 years) and >10 years following randomisation.

Example: ‘Studies were stratified by follow-up time categories: <2 years, 2 to 3 years, 5 to 10 years and >10 years […]. Predefined subgroups were (1) patients treated with ACL reconstruction compared with non-operative treatment and (2) skeletally immature patients compared with skeletally mature patients. We accepted skeletal immaturity as defined in the study […]. If skeletal immaturity was not defined in the study, we applied our definition of age under 16 years at injury for all patients.”47

Methods

Item 13a: synthesis methods
Describe the processes used to decide which studies were eligible for each synthesis (eg, tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item 5)).

Methods

Item 13b: synthesis methods
Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics or data conversions.

Example: ‘In studies with no ORs presented, the data were transformed to ORs from standard mean difference of muscle strength between the group of participants who developed osteoarthritis and those who did not. Data from adjusted analyses were extracted if available.”48

Example: ‘When trials used different outcome measures to evaluate the same construct, we chose the most common outcome measure as the index and transformed mean differences and SDs of other outcome measures to the index instrument.”49

Example: ‘When not reported, the SD was estimated from the SE of the mean, 95% CI, p value or other methods suggested in the Cochrane Handbook. Means and SDs were estimated from median and range for two studies. Following Cochrane guidelines, for any study that included two different intervention groups (ie, cycling vs walking) and one control group, the sample size in the control group was evenly divided so a comparison could be made to each intervention. We imputed r=0.5 when the correlation of prescores and postscores was required and we performed sensitivity analyses using r values ranging from 0.1 to 0.9.”49

Methods

Item 13c: synthesis methods
Describe any methods used to tabulate or visually display results of individual studies and syntheses.

Example: ‘Injury risk proportions for individual studies and pooled estimates were summarised in forest plots for the
following subgroups: woman, man and combined. A pooled estimate for the relative risk of ACL injury in women compared with men was calculated and summarised in a forest plot. Raw injury incidence rates for individual studies and pooled estimates were summarised in forest plots for the following groups and subgroups: woman, man and combined. Pooled incidence rate ratios for women compared with men were calculated and summarised in forest plots.\(^5\)

**Methods**

**Item 13d: synthesis methods**

Describe any methods used to synthesise results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.

**Example:** ‘Random effect models were used as large heterogeneity was expected due to the different approaches used to compare (ie, controls or contralateral leg) and assess knee extensor muscle strength (ie, isometric or isokinetic) as well as different pain and function scores. A standard [Cochran] Q-test was used to test the heterogeneity between studies, and the \(I^2\) statistic measuring the proportion of variance attributable to inconsistency was subsequently calculated. […] \(I^2\) each to 100% indicate maximal inconsistency between individual study results. Furthermore, the \(t^2\) value expressing the between study variance was estimated.\(^5\)

**Example:** ‘Owing to expected clinical and methodological heterogeneity between the included studies, we used inverse variance random effects models to estimate relative risks with 95% CIs. […] We dealt with statistical heterogeneity using the \(I^2\) statistic and prediction intervals. […] Analyses were conducted in either RevMan V.5.4 or Stata V.15.\(^3\)

**Example:** ‘We performed meta-analyses using a random-effects model as heterogeneity was expected in participant, intervention and outcome characteristics. […] A random effects meta-analysis was applied to estimate the overall relative risk of adverse events in the exercise therapy groups compared with comparator groups. Heterogeneity was examined as between-study variance and calculated as the \(I^2\) statistic measuring the proportion of variation in the combined estimates due to study variance. An \(I^2\) value of 0% indicates no inconsistency between the results of individual trials, and an \(I^2\) value of 100% indicates maximal inconsistency. Meta-analyses were performed in STATA (V16.1) using the ‘meta’ command.\(^5\)

**Note for systematic reviewers.** There are two common scenarios where systematic reviewers may wish to present a stratified analysis:

1. An analysis based on predefined groups when it is inappropriate for the specific review question to pool the data (eg, present injury incidence separately for women and men).
2. Discovering during the course of the review (eg, after extracting data) the need to stratify/subgroup data for a meaningful analysis (eg, short-term, medium-term and long-term outcomes of treatment for low back pain). This should not be confused with the expected and prespecified scenario for investigating statistical heterogeneity, where systematic reviewers conduct subgroup analyses to assess what is contributing to statistical heterogeneity.

**Methods**

**Item 13e: synthesis methods**

Describe any methods used to explore possible causes of heterogeneity among study results (eg, subgroup analysis, meta-regression).

**Example:** ‘We further explored between-study heterogeneity by comparing results from studies grouped according to several study level characteristics using stratified meta-analysis and meta-regression. Study level characteristics assessed were age, sex, MRI sequences employed, participants in weight-bearing sports, radiographic knee osteoarthritis, sample size and overall risk of bias. The prevalence estimates of primary compartment-specific outcomes (ie, tibiofemoral and patellofemoral cartilage defects, bone marrow lesions, osteophytes; medial and lateral meniscal tears) were pooled wherever reported and differences between compartments assessed with a two-proportion \(z\)-test.\(^5\)

**Example:** ‘Where sufficient trials were identified, meta-regression was undertaken using STATA (metareg command) to explore the impact of the following trial-level characteristics and whether they were associated with greater fall prevention effects:

1. Trial design: sample size: <20% missing outcome data; type of comparator intervention.
2. Participant characteristics: average age ≥75 years; control rate of falls; selected at high risk of falls.
3. Intervention components: included NICE-recommended components; actively provided treatment to address fall-related risk factors; whether adherence was assessed.\(^6\)

**Methods**

**Item 13f: synthesis methods**

Describe any sensitivity analyses conducted to assess robustness of the synthesised results.

**Example:** ‘We planned a sensitivity analysis for pain at 1 year to assess the impact of attrition bias due to missing data […] We also planned to assess small study bias by inspecting the distribution of funnel plots, but there were too few trials.’\(^4\)
Sensitivity analyses are often confused with subgroup analyses, but they differ in two important ways: (1) sensitivity analyses do not report an effect estimate from the studies excluded from the analysis, and (2) sensitivity analyses are different ways of estimating the same effect; with subgroup analyses, systematic reviewers are comparing estimates from different subgroups.57

Example: ‘We controlled for smoking status and pre-existing diseases by performing additional sensitivity analyses. Analyses were restricted to never smokers, healthy participants, and healthy never smokers.’58

Example: ‘We examined the effects of methodological quality [bias] on the pooled estimate by removing studies that were at high or unclear risk of bias for the domains of blinding and incomplete outcome data. We had also intended to examine the effects of measurement device on the pooled estimate by removing studies that used pedometers, as previous studies suggest that these might be less accurate in detecting steps in people with COPD.’59

Example: ‘We planned to carry out sensitivity analyses on pain and function by examining the effects of: […]including trials with unclear allocation concealment (at risk of selection bias); including trials with an incomplete description of mild to moderate knee osteoarthritis; […] including trials at risk of detection bias (ie, unclear or no blinding of participant for participant-reported outcomes).’60

Methods

Item 14: reporting bias assessment
Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).

Example: ‘In 15 of 16 included studies, the effect sizes were not reported; instead, the authors were contacted and asked to provide raw data. Therefore, the risk of publication bias was expected to be minor. Nevertheless, publication bias was assessed visually on a funnel plot (the effect size by the inverse of its standard error) and statistically using Egger’s test, the Begg and Mazumdar rank correlation test and the trim-and-fill method. It should be mentioned, however, that other factors, such as study quality and true heterogeneity, can produce asymmetry in funnel plots.’61

Note for systematic reviewers. Use sensitivity analyses to check the robustness of decisions made when planning the data synthesis. Sensitivity analyses repeat a previous analysis to check whether a decision was reasonable (eg, the first analysis includes all studies; the sensitivity analysis includes only the studies at low risk of attrition bias). Wherever possible, pre-specify sensitivity analyses in the systematic review protocol. Sometimes the need for sensitivity analyses only becomes apparent during the review process—ensure any unplanned sensitivity analyses are clearly labelled in the systematic review report.57

Note for systematic reviewers. There are frameworks (eg, GRADE,62 CINeMA63) that help systematic reviewers provide outcome-based recommendations and judge the certainty (ie, how confident they are in the recommendation) of the body of evidence—supporting transparent critical appraisal and communication. We encourage using the descriptor ‘certainty’ instead of ‘quality’ when describing judgements.

Example: ‘Data were synthesised and the quality of evidence were evaluated using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group. […] Two authors assessed the quality of evidence for each outcome relating to diagnostic tests (eg, effectiveness), prevention (eg, risk of injury) and treatment (eg, time to return-to-play) according to the approach from the GRADE working group. Agreement was reached by consensus. The quality of evidence was graded as: (1) high, indicating that further research is unlikely to change the confidence in the estimate of effect, (2) moderate, indicating that further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate, (3) low, indicating that further research is very likely to have an important impact on the confidence in the estimate of effect and is likely to change the estimate or (4) very low, indicating high uncertainty about the estimate.

The starting quality of evidence was rated as ‘high’ when data were based on either RCTs for treatment and prevention purposes or rated as ‘low’ when based on observational studies. For diagnostic purposes, the starting quality of evidence was rated as high when based on cohort studies (prospective or cross-sectional). Subsequently, the quality of evidence could be downgraded one or two levels (eg, from high to moderate) for each of the following five domains of the GRADE approach: Study limitations (ie, serious risk of bias such as lack of blinding of outcome assessor or other concerns determined to influence the study result), inconsistency (ie, the heterogeneity of the results across studies if more than one study was included for the specific outcome), indirectness (ie, poor generalisability of the findings to the target population, eg, use groin injuries vs acute adductor injuries for prevention, and/or use of a clinically irrelevant outcome in relation to the question, eg, ‘time to end of treatment’ for ‘time to return-to-play’ outcomes), imprecision of the estimates (ie, wide CIs) and the risk of publication bias. Furthermore, the level of evidence for cohort studies could

Note for systematic reviewers. You may choose to pre-define thresholds or criteria for judging each domain tailored to the review context and outcome(s). Specify pre-defined thresholds in a review protocol. We recommend systematic reviewers cautiously select arbitrary thresholds because they may not apply to comparable systematic reviews. One example: a sample size <400 for each meta-analysis (ie, outcome) may be appropriate for consistent continuous patient-reported outcomes (eg, Western Ontario and McMaster Universities Arthritis Index (WOMAC) questionnaire or session rating of perceived exertion), but may be insufficient when pooling data for disease incidence/prevalence or risk differences of complications (eg, incidence of cauda equina syndrome or rotator cuff re-tear), particularly to detect rare events.'
be upgraded due to a large effect, a dose–response relationship or if no effect was found and all plausible confounding factors identified in the study could be expected to increase the effect. An overview of the risk of bias and grading is provided as supplemental material.\(^{38}\)

**Example:** ‘The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach was used to assess the strength of evidence. Studies were downgraded if there were issues with risk of bias, consistency, precision or directness of the outcomes. The reasons for downgrading the evidence are outlined in table 2.’\(^{64}\)

### Table 2 Criteria for downgrading the quality of outcomes using the grade approach (reproduced with permission)

<table>
<thead>
<tr>
<th>Reason to downgrade the level of evidence</th>
<th>Risk of bias</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Indirectness</th>
<th>Publication bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority of studies rated as being at unclear risk of bias.</td>
<td>Large heterogeneity based on the similarity of point estimates, statistical heterogeneity and (I^2 \geq 50%.)</td>
<td>Large CIs when data are presented as standardised mean difference.</td>
<td>Use of surrogate outcomes</td>
<td>Asymmetric funnel plot</td>
<td></td>
</tr>
<tr>
<td>Outcome includes studies that have been rated as being at high risk of bias in two or more categories.</td>
<td>Substantial heterogeneity ((I^2 \geq 50%))</td>
<td>If a recommendation or clinical course of action would differ if the upper versus the lower boundary of the CI represented the truth.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>if no effect was found and all plausible confounding factors identified in the study could be expected to increase the effect.</td>
<td>Sample size &lt;400 within the meta-analysis for each variable.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Figure 1 PRISMA 2020 flow diagram template for systematic reviews (reproduced with permission). PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

### Results

**Item 16a: study selection**

Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram (see figure 1).

**Note for systematic reviewers.** When presenting the results of the search, direct readers to the PRISMA flowchart, so the yield at each step is transparent.

**Example:** ‘Following deletion of duplicates, the literature search yielded a total of 616 abstracts. A total of 557 abstracts were immediately excluded based on the title and abstract screen; 59 articles were obtained in full text and the selection criteria applied. Fifty-one articles were excluded, as they did not compare outcomes between patients who had had revision and primary ACL reconstruction. Finally, eight studies were included for meta-analysis.’\(^{65}\)

**Results**

**Item 16b: study selection**

Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.

**Example:** ‘[...] the remaining 1378 were evaluated in full text. Of these, 161 were excluded due to exercise included in the control group, 139 for reporting results from the same population in another study, 79 for an additional...’
therapy being delivered together with exercise or less than 50% of the intervention being exercise, 69 for the intervention involving vibration therapy, bladder training or other interventions not meeting the inclusion criteria for interventions. Further, 59 were excluded as they were reported in a language other than English, 45 because the full text was not accessible. Lastly, 8 were excluded because data were not extractable, 5 because the thesis was not available and 40 for other reasons.45

## Results

### Item 17: study characteristics

Cite each included study and present its characteristics.

**Note for systematic reviewers.** Report the key characteristics of the included studies to help readers understand how the included studies address the review question(s). The Template for Intervention Description and Replication (TIDieR) framework might help systematic reviewers summarise intervention details in intervention systematic reviews.

**Example:** ‘The final analysis included a total of 555 youth (283 and 272 in concurrent exercise and aerobic exercise training group, respectively). Three studies included the same population, but analysed different parameters. Nine studies recruited obese youth exclusively; whereas the rest targeted both overweight and obese children. Most studies (n=8) included adolescents (aged 13–18 years), one included only children, and one enrolled both children and adolescents. All studies included boys and girls. Sample sizes across studies ranged from 30 to 150, with a mean of 55 participants. The primary mode of the aerobic exercise training programmes were based on treadmills and cycle ergometers, elliptical trainers, walking and running programmes, and sports participation. The exercise intensity was monitored using either maximum heart rate, or peak oxygen uptake. For resistance exercise, studies used body weight exercise, free weights, selectorised machines or circuit training. Interventions duration varied from 10 to 48 weeks, with a mean of 30 weeks (figure 2 (labelled table 2 in the original publication) is an example of how one might present study characteristics).

### Item 18: risk of bias in studies

Present assessments of risk of bias for each included study.

**Note for systematic reviewers.** Please read in conjunction with item 11 explanation text.

**Example:** ‘Twenty-two trials (76%) were at high risk of bias. We had some concerns about bias in seven trials (24%). No trials were at low risk of bias. In studies that used both the VISA-A score and return to sport as outcome measures, there was no difference in risk of bias between the outcomes. In 48% of the trials, outcome measurement was a source of bias. All other sources of bias were also commonly judged as high risk: the randomisation procedure (21%), deviations from the intended intervention (28%), missing outcome data (28%) and selection of reported results (24%).43

### Item 19: results of individual studies

For all outcomes, present, for each study: (1) summary statistics for each group (where appropriate) and (2) an effect estimate and its precision (eg, CI/credible interval), ideally using structured tables or plots.

**Example included in figure format:** ‘One low quality study used x-ray to examine the pubic symphyses of athletes with hip/groin pain and those without pain. A reliable grading system quantified the abnormalities, which were present in all the hip/groin pain subjects (9/20 slight, 9/20 intermediate, 2/20 advanced). In contrast, the athletic control subjects had either no (3/20) or slight (17/20) abnormalities seen on x-ray. Another moderate quality study only reported the radiographic findings in the hip/groin pain subjects and not the control subjects. Synthesising the data on x-ray investigations of the pubic symphysis, there is currently limited evidence that x-ray findings differentiate athletes with hip/groin pain from athletes without pain.

Three moderate quality studies examined pubic bone oedema using MRI in athletes with hip/groin pain and controls. Dichotomous data for the presence or absence of bone oedema were extracted and pooled from these studies. The results indicated that there were high odds that participants with bone oedema on MRI would be in the hip/groin pain group with a large effect size; OR=41.63 (95% CI 1.6 to 1096.6). However, there was high heterogeneity demonstrated by this pooled result (I²=88%), and significant sensitivity to the data from the study of Cunningham et al (figure 3 (labelled figures 8 and 9 in the original publication)). The removal of this study data resulted in an OR=8.1 (95% CI 3.1 to 21.2) for the presence of bone oedema in subjects with hip/groin pain, representing moderate evidence, with a large effect size, that bone oedema in the pubic symphysis differentiates athletes with hip/groin pain from those without this pain.48

### Item 20a: results of syntheses

For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.

**Example:** ‘Risk of bias assessment For the comparisons of subacromial decompression surgery versus placebo surgery, the risk of bias was low for all outcomes (figure 4). Due to detection bias (all studies), selection bias, attrition bias and selective reporting, for the comparisons of subacromial decompression surgery vs non-surgical treatment, the risk of bias was high for all outcomes except rotator cuff tears.

**Primary comparisons**

Two trials were at low risk of bias and sufficiently clinically and methodologically homogeneous to allow pooling for the comparison of subacromial decompression surgery plus postoperative rehabilitation versus placebo surgery plus postoperative rehabilitation.41

---

**Consensus statement**

**Table 2** Characteristics of the studies included in the meta-analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants (n)/BMI (kg/m² or percentile)</th>
<th>Groups size</th>
<th>Modality</th>
<th>Intervention duration (weeks)</th>
<th>Frequency (days/week)</th>
<th>Session (minutes)</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ackel-D’Elia et al, 2014</td>
<td>48 (16 boys, 32 girls, 15–18 years of age)/≥95</td>
<td>CE 24</td>
<td>Treadmill+free weights</td>
<td>24</td>
<td>3</td>
<td>60</td>
<td>BM, BMI, BF (% and kg), LBM (% and kg), glucose, insulin, HOMA-IR, and leptin</td>
</tr>
<tr>
<td>Alberga et al, 2015</td>
<td>150 (44 boys, 106 girls, 14–18 years of age)/≥90</td>
<td>CE 75</td>
<td>Treadmill, cycle ergometer or elliptical + free weights and machines Treadmill, cycle ergometer or elliptical</td>
<td>22</td>
<td>2</td>
<td>20–40</td>
<td>VO2max, CRP, VAT, and SAT BM, BMI, BF (% and kg), LBM (kg), WC, TC, HDL, LDL, TG, glucose, insulin, SBP, DBP</td>
</tr>
<tr>
<td>Alberga et al, 2016</td>
<td>42 (28 boys, 14 girls, 14–18 years of age)/≥95</td>
<td>CE 21</td>
<td>Treadmill+free weights</td>
<td>48</td>
<td>3</td>
<td>60</td>
<td>BM, BMI, BF (% and kg), LBM (kg), VAT, SAT, glucose, insulin, HOMA-IR, and leptin</td>
</tr>
<tr>
<td>Carmer et al, 2013</td>
<td>26 (10 boys, 16 girls, 14–18 years of age)/≥95</td>
<td>CE 13</td>
<td>Treadmill+free weights</td>
<td>48</td>
<td>3</td>
<td>60</td>
<td>BM, BMI, BF (% and kg), LBM (kg), VAT, SAT, glucose, insulin, HOMA-IR, and leptin</td>
</tr>
<tr>
<td>Dâmana et al, 2014</td>
<td>116 (57 boys, 82 girls, 14–18 years of age)/≥95</td>
<td>CE 61</td>
<td>Treadmill+free weights</td>
<td>48</td>
<td>3</td>
<td>60</td>
<td>BM, BMI, BF (% and kg), LBM (kg), VAT, SAT, glucose, insulin, HOMA-IR, and leptin</td>
</tr>
<tr>
<td>de Mello et al, 2011</td>
<td>30 (20 boys, 10 girls, 14–18 years of age)/≥95</td>
<td>CE 15</td>
<td>Treadmill+free weights</td>
<td>48</td>
<td>3</td>
<td>60</td>
<td>BM, BMI, BF (% and kg), LBM (kg), VAT, SAT, WC, TC, HDL, LDL, TG, glucose, insulin, HOMA-IR, and leptin</td>
</tr>
<tr>
<td>de Piano et al, 2012</td>
<td>30 (20 boys, 10 girls, 14–18 years of age)/≥95</td>
<td>CE 15</td>
<td>Treadmill+free weights</td>
<td>48</td>
<td>3</td>
<td>60</td>
<td>BM, BMI, BF (% and kg), LBM (kg), VAT, SAT, TC, HDL, LDL, TG, glucose, insulin, HOMA-IR, and leptin, and adiponectin</td>
</tr>
<tr>
<td>Inoue et al, 2015</td>
<td>45 (17 boys, 28 girls, 14–18 years of age)/≥95</td>
<td>CE 13</td>
<td>Treadmill+free weights*</td>
<td>48</td>
<td>3</td>
<td>60</td>
<td>BM, BMI, BF (% and kg), LBM (kg), VAT, SAT, TC, HDL, LDL, TG, glucose, insulin, HOMA-IR, and leptin, and adiponectin</td>
</tr>
<tr>
<td>Lee et al, 2010</td>
<td>36 (12–14 years of age)/≥95</td>
<td>CE 20</td>
<td>Circuit weight training</td>
<td>10</td>
<td>2</td>
<td>60</td>
<td>BM, BMI, BF (% and kg), LBM (kg), VAT, SAT, TC, HDL, LDL, TG, glucose, insulin, HOMA-IR, and leptin, and adiponectin</td>
</tr>
<tr>
<td>Monteiro et al, 2015</td>
<td>32 (19 boys, 13 girls, 11–17 years of age)/sex and-age BMI IOTF</td>
<td>CE 14</td>
<td>Walking and running + own body weight Walking and running</td>
<td>20</td>
<td>3</td>
<td>50</td>
<td>BM, BMI, BF (% and kg), LBM (kg), VAT, SAT, WC, HDL, LDL, TG, glucose, insulin, HOMA-IR, and leptin, and adiponectin</td>
</tr>
</tbody>
</table>

*Applied different periodisation.
AE, aerobic training; BF, body fat; BM, body mass; BMI, body mass index; CE, concurrent exercise; CRP, C reactive protein; DBP, diastolic blood pressure; HDL, high-density lipoprotein; HOMA-IR, homeostatic model assessment; IOTF, International Obesity Task Force; LBM, lean body mass; LDL, low-density lipoprotein; SAT, subcutaneous fat; SBP, systolic blood pressure; TC, total cholesterol; TG, triglycerides; VAT, visceral fat; VO2max, maximal oxygen consumption; WC, waist circumference.

**Figure 2** Example of study characteristics presented in a summary table (reproduced with permission).

**Note for systematic reviewers.** Present the study and population characteristics (eg, age, sex, n, etc.) for each synthesis, separately. Below is an example of how to report study and population characteristics of contributing studies for each data synthesis.

**Example:**

‘Mental health symptoms and disorders among current elite athletes
Among those, 11 studies reported prevalence data on distress symptoms among 3335 male and female elite athletes (age ranging from 16 to 29 years) from team sports (eg, cricket, football, handball, ice hockey, rugby) and combined Olympic sports (eg, boxing, gymnastics, judo, rowing, swimming).


Mental health symptoms and disorders among former elite athletes
Among those, eight studies reported prevalence data on distress symptoms among 1686 former male and female elite athletes (age ranging from 34 to 62 years) from team sports (American football, cricket, football, ice hockey, rugby) and combined Olympic sports.’

**Results**

**Item 20b: results of syntheses**
Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (eg, CI/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.

**Example:** ‘Three studies reported on previous history of stress fracture and its association with increased risk of future stress fractures. [...]’
fracture. All three studies had similar findings in that athletes with a previous history of fracture were at increased risk of developing a future stress fracture with ORs ranging from 2.90 to 6.36. An exploratory meta-analysis confirmed the individual study results with runners with a previous history of stress fracture at five times higher risk of a future stress fracture (OR 4.99; 95% CI 2.91 to 8.56; p < 0.001; I² = 0%). Differences in risk between men and women did not reach statistical significance (p=0.87).48

### Results

#### Item 20c: results of syntheses

Present results of all investigations of possible causes of heterogeneity among study results.

Example: ‘Stratified analyses for sex showed an increased risk in both men (OR 1.68, 95% CI 1.10, 2.58; I² = 55.5%) and women (OR 1.59, 95% CI 0.94, 2.68; I² = 54.3%). Differences in risk between men and women did not reach statistical significance (p=0.87).48

#### Item 20d: results of syntheses

Present results of all sensitivity analyses conducted to assess the robustness of the synthesised results.

Example: ‘[…] The effect remained similar across all strata included in our prespecified…sensitivity analyses (Table 3) with the exception of the type of comparator intervention. […] The pooled estimates of effect remained similar in all prespecified…sensitivity analyses (figure 5 (labelled Table 3 in the original publication)).56

#### Item 21: reporting biases

Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.

Example: ‘We assessed publication bias for this comparison. The funnel plot (figure 6) shows that small studies with negative effects are missing in the right lower quadrant, which indicates possible publication bias.’51

### Example:

‘Publication bias

The funnel plot indicated that almost all studies fell within the expected parameters, most with low SE indicating that most studies were large. A majority of studies reported that women had greater incidence proportion than men. The funnel plot for incidence rate ratio indicated that most studies fell within the expected parameters. Standard error was relatively low, indicating that studies were large, and a majority of studies reported that women were at increased risk of ACL injuries relative to men. The studies are not evenly distributed in the funnel, with studies missing from the lower left quadrant. Studies in the lower left quadrant would represent smaller studies that report a greater incidence proportion or incidence rate of ACL injuries in men compared with women.’50

### Results

#### Item 22: certainty of evidence

Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.

Example: ‘We assessed publication bias for this comparison. The funnel plot (figure 6) shows that small studies with negative effects are missing in the right lower quadrant, which indicates possible publication bias.’51

Example:

‘Publication bias

The funnel plot indicated that almost all studies fell within the expected parameters, most with low SE indicating that most studies were large. A majority of studies reported that women had greater incidence proportion than men. The funnel plot for incidence rate ratio indicated that most studies fell within the expected parameters. Standard error was relatively low, indicating that studies were large, and a majority of studies reported that women were at increased risk of ACL injuries relative to men. The studies are not evenly distributed in the funnel, with studies missing from the lower left quadrant. Studies in the lower left quadrant would represent smaller studies that report a greater incidence proportion or incidence rate of ACL injuries in men compared with women.’50
Table 3  Sensitivity analyses for the effect of multifactorial intervention on rate of falls and risk of sustaining one or more falls and recurrent falls

<table>
<thead>
<tr>
<th>Sensitivity analysis</th>
<th>Trials, n</th>
<th>Participants, n</th>
<th>Effect estimate (random effects) 95% CI</th>
<th>I² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of falls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate concealment of allocation</td>
<td>10</td>
<td>4442</td>
<td>0.82 (0.68 to 0.97)</td>
<td>94</td>
</tr>
<tr>
<td>Adequate blinding of outcome assessors</td>
<td>15</td>
<td>8062</td>
<td>0.81 (0.71 to 0.92)</td>
<td>92</td>
</tr>
<tr>
<td>Adequate handling of incomplete outcome data*</td>
<td>12</td>
<td>5096</td>
<td>0.78 (0.68 to 0.89)</td>
<td>90</td>
</tr>
<tr>
<td>Individual randomisation</td>
<td>18</td>
<td>6643</td>
<td>0.80 (0.70 to 0.91)</td>
<td>91</td>
</tr>
<tr>
<td>Risk of sustaining one or more falls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate concealment of allocation</td>
<td>13</td>
<td>5296</td>
<td>0.95 (0.86 to 1.04)</td>
<td>69</td>
</tr>
<tr>
<td>Adequate blinding of outcome assessors</td>
<td>18</td>
<td>8481</td>
<td>0.95 (0.89 to 1.02)</td>
<td>54</td>
</tr>
<tr>
<td>Adequate handling of incomplete outcome data*</td>
<td>15</td>
<td>5732</td>
<td>0.96 (0.90 to 1.02)</td>
<td>31</td>
</tr>
<tr>
<td>Individual randomisation</td>
<td>26</td>
<td>9839</td>
<td>0.95 (0.89 to 1.02)</td>
<td>59</td>
</tr>
<tr>
<td>Risk of sustaining recurrent falls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate concealment of allocation</td>
<td>8</td>
<td>2589</td>
<td>0.89 (0.71 to 1.10)</td>
<td>67</td>
</tr>
<tr>
<td>Adequate blinding of outcome assessors</td>
<td>12</td>
<td>6668</td>
<td>0.89 (0.77 to 1.03)</td>
<td>56</td>
</tr>
<tr>
<td>Adequate handling of incomplete outcome data*</td>
<td>7</td>
<td>2129</td>
<td>0.96 (0.84 to 1.10)</td>
<td>0</td>
</tr>
<tr>
<td>Individual randomisation</td>
<td>14</td>
<td>4095</td>
<td>0.89 (0.78 to 1.03)</td>
<td>46</td>
</tr>
</tbody>
</table>

* Judgement of ‘low risk’ if there are no missing outcome data, or less than 20% of missing outcome data are missing and losses are balanced in numbers across intervention groups with similar reasons for missing data across groups or missing data have been imputed using appropriate methods.

Figure 5  Example of how to present results of sensitivity analyses (reproduced with permission).

Note for systematic reviewers. Critical appraisal frameworks (eg, GRADE62 CINeMA63) facilitate transparent critical appraisal and help systematic reviewers clearly communicate information. We encourage using the descriptor ‘certainty’ instead of ‘quality’ when describing judgements.

Example of GRADE summary reporting and outcome reporting in table format: A summary of the quality of evidence, based on risk of bias, study design, CIs and variability in results, has been collated using the GRADE approach (table 3).

All outcomes were rated as very low or low quality [we recommend systematic reviewers use the descriptor ‘certainty’ not ‘quality’] evidence demonstrating that the estimate of effect for those outcomes is uncertain.64

Discussion

Item 23a: discussion

Provide a general interpretation of the results in the context of other evidence.

Example: ‘There exist several other reviews, although previous reviews have focused on fewer interventions. The most important difference between our systematic review, and previously published reviews is that we have a more stringent assessment of the risk of bias and quality of included trials. This is important because the strength of recommendations (eg, in future guidelines) will be based on the quality of the evidence.

For exercise, our results are in line with the other reviews, with the exception that we concluded that there is only very low-quality evidence where other studies reported moderate or even high or strong evidence. Two reviews evaluated scapula-focused treatments, reporting moderate evidence, and significant but clinically not relevant effects; whereas we did not separately analyse the scapula-focused treatments.25

Discussion

Item 23b: discussion

Discuss any limitations of the evidence included in the review.

Example: ‘Overall, there was a lack of consistent high-quality evidence to support nominating any particular movement quality outcome as a lower extremity injury risk factor due to inadequate reporting of concepts essential to establishing internal (how well an experiment was carried out) and external (can the results be applied to people and situations beyond the experiment) validity. The biggest threats to internal validity were related to the possibility of selection bias, and the reporting of, and adjustment for, potential influence of factors such as sex, injury history and training exposure.

Specifically, due to the lack of participant characteristic reporting, it was often difficult to determine if the individuals selected for a study differed systematically from those in...
the source population (selection bias). Equally important was the consistent omission of the characteristics of those lost to follow-up, which made it impossible to determine if participants lost to follow-up were systematically different from those retained in a study. The inability to determine selection bias not only questions the internal validity of several included studies, it impacts the degree to which the findings of these studies can be generalised to the larger population from which the samples were drawn (external validity).²⁷²

**Discussion**

**Item 23c: discussion**

Discuss any limitations of the review processes used; comment on the potential impact of each limitation.

**Example:** ‘We planned threshold analysis as a quantitative means to assess the robustness of network meta-analysis recommendations to potential limitations in the evidence. We were unable to use this approach because of substantial overlap in credible intervals from the network meta-analysis. Due to overlap in the intervals, no recommendations could be made, which is a fundamental prerequisite to performing a valid threshold analysis. To comply with our protocol, we report threshold results in (supplementary Web appendix), but chose to use GRADE to interpret the evidence. We were not able to evaluate small study bias due to too low number of trials. We found three completed trials in trial registers; two are under review, and the publication status of one trial is unknown.’⁴⁵

**Discussion**

**Item 23d: discussion**

Discuss implications of the results for practice, policy, and make recommendations for future research.

**Example:** ‘On average, patients with subacromial pain syndrome reported reduced pain, and improved physical function and quality of life following both surgical and non-surgical treatment. However, at up to 5 years, irrespective of treatment, patients continued to report pain of an average of 1.5–3 on a scale of 0–10 points on a Visual Analogue Scale. Clinicians working in primary care who are treating patients with subacromial pain syndrome should be aware that some patients experience prolonged symptoms and consider care strategies to support coping.

A placebo control helps answer the research question ‘is there a benefit of subacromial decompression surgery?’ because it minimises the risk of detection and performance biases. Both sources of bias contribute to overestimation of treatment effects by up to 20%. The largely consistent findings of the unblinded studies leave little doubt of the inference that subacromial decompression surgery provides no important benefit to patients. The current evidence provides no support for subacromial decompression surgery as an intervention providing important benefit for patients with subacromial pain syndrome. High-quality evidence indicates that surgery versus placebo surgery confers no important benefit on pain and function—the outcomes most important to patients. Considering the body of evidence, further head-to-head comparisons of subacromial decompression surgery compared with placebo surgery or non-surgical management with the same population are unlikely to change the results. Policymakers, funders and clinicians should consider these results in their funding and clinical decisions regarding the management of patients with shoulder pain.

Our review was designed to assess the benefits and harms of subacromial decompression surgery for managing subacromial pain syndrome. To date, no trial has demonstrated benefit of surgery for any clinical subgroup. In the future, subgroup claims should be supported by data from well-conducted trials at low risk of bias and the use of established criteria for credibility of subgroup effects, ideally enhanced by IPD meta-analysis.

For harms we welcome well-performed observational studies that specifically report harms after subacromial decompression surgery separately from harms following other types of shoulder surgery. Although the finding of no important benefit of surgery is robust, the root cause of subacromial pain and the underlying pathological process remain uncertain, as does the possible best treatment—if such exists—for subacromial pain syndrome. Network meta-analysis might provide information on this question and provide hypotheses to be tested in future methodologically sound trials. Future triallists investigating any treatment for subacromial pain syndrome should adopt a
common set of outcomes, and the outcome measures should be standardised.

Note for systematic reviewers. A systematic review protocol is an important way for readers to check whether the systematic review questions and methods have been pre-specified and followed. The protocol guards against bias introduced when too many decisions are made after the data are analysed (eg, selective reporting and outcome switching). Protocols may be (appropriately) modified during the conduct of a systematic review (ensure the registry record is updated when the protocol is modified). The PRISMA-P checklist guides systematic reviewers on developing a review protocol.

Example: ‘The study was registered at PROSPERO (ID CRD42015024120).’

Note for systematic reviewers. The PROSPERO database (www.crd.york.ac.uk/PROSPERO) is an international database of prospectively registered systematic reviews with a health outcome. In addition to the PROSPERO database, systematic reviewers have a range of options to consider when deciding where to register a systematic review. See box 4 for an overview.

Other information
Item 24a: registration and protocol
Provide registration information for the review, including register name and registration number, or state that the review was not registered.

Box 4 Options for prospectively registering systematic review protocols

<table>
<thead>
<tr>
<th>Preprint servers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprint servers are repositories of preliminary reports of scientific work that are yet to complete a full peer review process. The content of preprint servers precedes the formal peer-review, typesetting, copyediting and publishing processes of scientific journals. Currently, there is no cost to post a manuscript on a preprint server.</td>
</tr>
<tr>
<td>Preprints are available open access and are typically assigned a digital object identifier (DOI) by the server’s administrator. The DOI protects authors’ intellectual property and allows researchers to disseminate and cite work in manuscripts, grant applications and curricula vitae. medRxiv (<a href="http://www.medrxiv.org">www.medrxiv.org</a>) accepts work in the medical, clinical and related health science fields. SportRxiv (<a href="http://www.sportrxiv.org">www.sportrxiv.org</a>) accepts work in the sport, exercise, performance and health research fields. OSF Preprints is another alternative for systematic reviewers wishing to obtain a DOI.</td>
</tr>
</tbody>
</table>

Open Science Framework
The Open Science Framework (OSF) is an online community, curated by the Center for Open Science, that supports researchers to collaborate and communicate. The platform allows researchers (individuals and teams) to collate all resources, files, data, statistical code and study protocols for a project. Edits to a project are timestamped, and public projects are searchable. Registering the work creates a time-stamped, read-only version of a project, which can serve as a pre-registered protocol because the files in a registration cannot be modified (the files in a project can be modified).

Registered reports
Registered Reports are an alternative publishing format that elevates the research question and study methods. Approximately 275 biomedical journals offer Registered Reports as a manuscript submission type. Registered Reports facilitate peer review of study protocols, and the work is provisionally accepted based on the research protocol (ie, before the results are known).

The peer review process for Registered Reports occurs in two stages: after the study is designed (protocol) and (for protocols that are ‘accepted in principle’) after the data are collected, analysed and interpreted (full manuscript). The purpose of Registered Reports is to ensure authors’ reporting of their work is consistent with the registered protocol. Registered Reports ensures quality methods (via protocol peer review) and a guarantee that the journal will publish the final version of the systematic review, provided the authors remain true to the protocol.

Institution repositories
Many university libraries offer repositories with the capacity to assign a DOI to authors’ work at the preprint and postprint stages.

Publishing the protocol
Some journals publish systematic review protocols (eg, BMJ Open, Systematic Reviews). Journals typically do not consider protocols for publication once data extraction has commenced. Published protocols usually require prospective registration of the systematic review with an appropriate platform (eg, PROSPERO database or OSF).
the majority of studies reported either 6-month or 12-month outcome data (not both). Therefore, to provide a more complete picture of early, medium and long-term outcomes, we modified the time point at which outcomes were report to: under 6 months, 6–12 months and over 12 months, respectively. One included study did not perform MRI in all patients prior to randomisation.75

Example:

‘Deviations from study registration and study protocol
‘Agreement by raters on risk of bias decisions for the included randomised controlled studies was calculated as a percentage of agreement and K values, and included in the results. Since the secondary analysis concerning type of programme showed that only the FIFA 11+ prevention programme was effective in reducing injuries, all secondary outcomes concerning lower limb, hamstring, knee and ankle injuries were only analysed in relation to this programme.

Furthermore, a post hoc analysis on hip/groin injury in relation to this programme was also included. Preplanned secondary analyses on the incidence rate ratio in the following subgroups: gender (male and female), and mean age groups (youth (<19 years), seniors (19–30 years), old girls/boys (31–39 years) and veterans (>39 years)) were not conducted, as the included studies did not allow making meaningful comparisons with only six studies, where studies with male (n = 3) and female participants (n = 3) significantly differed in the age group they targeted. The predefined secondary analysis of compliance at team level was not performed as all team-level data could not be obtained from the corresponding authors of the included studies. Instead, the preplanned analysis of the association between prevention programme compliance and injury incidence was further supported by a post hoc analysis of the association between prevention programme compliance and the overall injury incidence rate ratio from each study to accommodate for the risk of substantial variance in injury incidence between studies due to other factors than the FIFA injury prevention programmes.40

Other information

Item 25: support
Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.

Example: ‘This research received no grant from any funding agency in the public, commercial or not-for-profit sectors. […] is supported and funded by the National Osteoporosis Society via the Linda Edwards Memorial PhD Studentship.’76

Item 26: competing interests
Declare any competing interests of review authors.

Example: ‘All authors have completed the ICMJE uniform disclosure form at (available on request from the corresponding author) and declare: […] has received personal fees from Össur, Flexion Therapeutics, Medivir, Teijin, MerckSerono, Allergan, and Galapagos and is editor-in-chief of Osteoarthritis and Cartilage; […] has received personal fees for lectures and royalties for books from Össur, Finnish Orthopedic Society, Studentlitteratur, and Munksgaard and is an associate editor of Osteoarthritis and Cartilage; no other relationships or activities that may appear to have influenced the submitted work.’46

Other information

Item 27: availability of data, code and other materials
Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.

Example: ‘...a publicly available comprehensive study protocol including data extraction forms was uploaded at the following website: http://vbn.aau.dk/files/229186677/The_effect_of_the_FIFA_11+_prevention programmes_on_the_overall_injury_rate_in_football_a_systematic_review_and_meta_analysis_version1_1.pdf’.40

Example: ‘Data availability statement. Data are available in a public, open access repository: https://osf.io/q24xh’77 [the authors provide the dataset of included studies (with reasons for excluding studies from meta-analysis), and a list of excluded studies with reasons].

Note for systematic reviewers. We could not find any examples in the sports and exercise medicine literature of authors reporting the statistical code used for analyses. For an example of how to report statistical code, please refer to Supplementary Material 1) of Weideman et al.49

PERSPECTIVES

All PRISMA 20201 items apply to systematic reviews in the sport and exercise medicine, musculoskeletal rehabilitation, and sports science fields. However, some items require more elaboration to promote implementation. PERSIST is intended to support systematic reviewers to implement PRISMA 2020 in their systematic reviews in the sport and exercise medicine, musculoskeletal rehabilitation, and sports science fields, and to promote transparent reporting. Use PERSIST in parallel with the PRISMA 2020 Checklist4 and Explanation and Elaboration,11 and with the METHODS MATTER statement17 (as appropriate). We encourage journal editors and reviewers in the relevant fields to use PERSIST to help them make informed judgements about the quality and transparency of systematic review reporting.

Researchers make many decisions at the various stages of conducting a systematic review—from planning the review, completing the search and analysis, and writing the report.78 Decisions include choosing search terms, which data to extract, data pooling policies, choice of statistical tests, among others.79 Poor reporting makes it impossible for other researchers to replicate a systematic review, and a systematic review that cannot be replicated has little value. PERSIST and the PRISMA 2020 Statement and Checklist1 will help systematic reviewers deliver quality systematic reviews in sport and exercise medicine, musculoskeletal rehabilitation, and sports science.'
Effect of soft braces on pain and physical function in patients with knee osteoarthritis: systematic review with meta-analyses

To systematically review and synthesize the effect of soft braces on pain, and self-reported and performance-based physical function in patients with knee osteoarthritis.

Item 2: Provide an explicit statement of the main objective(s) or question(s) the review addresses.

Example: ‘To systematically review and synthesize the effect of soft braces on pain, and self-reported and performance-based physical function in patients with knee osteoarthritis.’

Item 3: Specify the characteristics used as criteria for eligibility.

Example: ‘Randomised controlled trials examining the effectiveness of any treatment in patients with both insertional and/or midportion Achilles tendinopathy. We excluded trials with N ≤10 per treatment arm or investigating tendon ruptures.’

Item 4: Specify the information sources (eg, databases, registers) used to identify studies and the date when each was last searched.

Example: ‘The following electronic databases were searched from inception to April 20, 2016: The Cochrane Central Registry for Controlled Trials (CENTRAL), PubMed, EMBASE, Cumulative Index to Nursing and Allied Health Literature, SPORTDiscus, Web of Science and PEDro.’

Item 5: Specify the methods used to assess risk of bias in the included studies.

Example: ‘Two reviewers independently extracted data and assessed risk of bias with Risk of Bias Tool V2. We used Grading of Recommendations, Assessment, Development and Evaluation to appraise the strength of the evidence.’

Item 6: Specify the methods used to present and synthesise results.

Example: ‘We pooled data for meta-analyses by length of follow-up, reported as mean differences or standardised mean differences using random-effects wherever possible, or the fixed-effect model, where appropriate. If a meta-analysis was not possible, we synthesised studies narratively.’

Example: ‘Where the same outcome was assessed across different intervention types, we reported standardised effect sizes for findings from single-study and multiple-study analyses to allow comparison of intervention effects across intervention types. To ease interpretation of the effect size, we also reported the mean difference of effect sizes for single-study outcomes.’

Item 7: Give the total number of included studies and participants, and summarise relevant characteristics of studies.

Example: ‘Twenty studies comprising 2375 injuries from 1234 athletes (all males and mean age of 24 years) from different sports were included. Internal (65%) and external loads (70%) were collected in more than half of the studies and the session-rating of perceived exertion and total distance were the most commonly collected metrics. The acute chronic workload ratio was commonly calculated using the coupled method (95%), 1:4 weekly blocks (95%) and subsequent week injury lag (80%). There were 14 different binning methods with almost none of the studies using the same binning categories.’

Item 8: Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (ie, which group is favoured).

Example: ‘Overall, 42.7% (95% confidence interval (CI): 18%, 69%) of patients passed return to sport (RTS) criteria, and 14.4% (95% CI: 8%, 21%) of those who passed experienced a second ACL injury (graft rupture or contralateral ACL injury). There was a nonsignificant 3% reduced risk of a second ACL injury after passing RTS criteria (risk difference, −3%; 95% CI: −16%, 10%; I^2=74%, p=0.61).’

Item 9: Provide a brief summary of the limitations of the evidence included in the review (eg, study risk of bias, inconsistency and imprecision).

Example: ‘The evidence rating of the Grading of Recommendations Assessment, Development and Evaluation scale was ‘very low certainty’ due to imprecision and heterogeneity of the pooled risk difference estimate.”

Item 10: Provide a general interpretation of the results and important implications.

Example: ‘In our living network meta-analysis no trials were at low risk of bias and there was large uncertainty in the comparative estimates. For midportion Achilles tendinopathy, wait-and-see is not recommended as all active treatments seemed superior at 3-month follow-up. There seems to be no clinically relevant difference in effectiveness between different active treatments at either 3-month or 12-month follow-up. As exercise therapy is easy to prescribe, can be of low cost and has few harms, clinicians could consider starting treatment with a calf-muscle exercise programme.’

Item 11: Specify the primary source of funding for the review.

Example: ‘Funding: This research received a grant from the Dutch Association of Medical Specialists to develop a clinical guideline for the treatment of patients with Achilles tendinopathy. The Dutch Patient Federation is involved in this guideline development and assisted in sending out patient surveys.’

Example: ‘Funding: The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.’

Item 12: Provide the register name and registration number.

Example: ‘Prospero registration: CRD42018086467.’
in-chief of BJM Open Sport & Exercise Medicine from 2019 to 2020. He is a senior adviser to BJSM (September 2021).

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Claire L Ardem http://orcid.org/0000-0001-8102-3631
Finn Büttnér http://orcid.org/0000-0002-5987-3063
Renato Andrade http://orcid.org/0000-0002-7636-7816
Adam Weir http://orcid.org/0000-0003-0861-662X
Maureen C Ashe http://orcid.org/0000-0002-6820-4425
Sinead Holden http://orcid.org/0000-0002-7314-2152
Franco M Impellizzeri http://orcid.org/0000-0002-1703-2573
Eamonn Delahunt http://orcid.org/0000-0001-5449-5932
H Paul Dijkstra http://orcid.org/0000-0003-3166-1357
Stephanie Mathiessen http://orcid.org/0000-0002-7335-8842
Catherine Sherrington http://orcid.org/0000-0001-8934-4368
Emmanuel Stamatakis http://orcid.org/0000-0001-7323-3225
Bill Vicenzino http://orcid.org/0000-0003-0253-5933
Jackie L Whittaker http://orcid.org/0000-0002-6591-4976
Alexis A Wright http://orcid.org/0000-0002-8429-0775
Mike Clarke http://orcid.org/0000-0002-2926-7257
Karim M Khan http://orcid.org/0000-0002-9976-0258
Marinus Winters http://orcid.org/0000-0001-5742-7441

REFERENCES
Consensus statement


