Diagnosis, prevention and treatment of common lower extremity muscle injuries in sport – grading the evidence: a statement paper commissioned by the Danish Society of Sports Physical Therapy (DSSF)

Lasse Ishøi,1 Kasper Krommes,1 Rasmus Skov Husted,2,3 Carsten B Juhl,4 Kristian Thorborg1,2

ABSTRACT
This statement summarises and appraises the evidence on diagnosis, prevention and treatment of the most common lower extremity muscle injuries in sport. We systematically searched electronic databases, and included studies based on the highest available evidence. Subsequently, we evaluated the quality of evidence using the Grading of Recommendations Assessment, Development and Evaluation framework, grading the quality of evidence from high to very low. Most clinical tests showed very low to low diagnostic effectiveness. For hamstring injury prevention, programmes that included the Nordic hamstring exercise resulted in a hamstring injury risk reduction when compared with usual care (medium to large effect size; moderate to high quality of evidence). For prevention of groin injuries, both the FIFA 11+ programme and the Copenhagen adductor strengthening programme resulted in a groin injury risk reduction compared with usual care (medium effect size; low to moderate quality of evidence). For the treatment of hamstring injuries, lengthening hamstring exercises showed the fastest return to play with a lower reinjury rate compared with conventional hamstring exercises (large effect size; very low to low quality of evidence). Platelet-rich plasma had no effect on time to return-to-play and reinjury risk (trivial effect size; moderate quality of evidence) after a hamstring injury compared with placebo or rehabilitation. At this point, most outcomes for diagnosis, prevention and treatment were graded as very low to moderate quality of evidence, indicating that further high-quality research is likely to have an important impact on the confidence in the effect estimates.

INTRODUCTION
Lower extremity muscle injuries are frequent in sports involving explosive actions such as high-speed running, jumping, change of direction and kicking. In professional football, muscle injuries constitute up to half of all injuries, and in sports, such as American football, Australian football, rugby, basketball and track and field, the incidence is also high. The majority of muscle injuries in football occur during non-contact situations, classified as ‘indirect muscle injuries’ or ‘muscle strains’. These are typically thought to occur in the muscle-tendon junction when the force applied exceeds the tissue capacity.

In football, muscle strain injuries constitute up to 31% of all injuries, and up to 37% of players experience absence from training and/or match play during a season due to a muscle injury. Besides from significant financial costs at the professional level, the high injury burden of muscle injuries also have substantial implications on player availability, potentially affecting team performance. Up to 92% of all muscle strain injuries encountered in football are located in the hamstrings (37%), adductors (23%), rectus femoris/quadriceps (19%) and calf muscles (13%), with hamstring strain injuries also being the most common diagnosis in sports such as Australian football and track and field.

The fact that incidence of muscle strain injury has remained constant during the last decades in football and Australian football with slightly increased incidence for hamstring strain and calf muscle strains, highlights the ongoing challenge of muscle strain injuries in sport. Thus, to aid management of common muscle injuries in sport, the aim of this statement, commissioned by the Danish Society of Sports Physical Therapy, was to provide an overview of the existing literature. We identify, evaluate and grade the quality of evidence concerning the diagnostic effectiveness of clinical tests and the effect of preventive and treatment strategies for the most common lower limb muscle injuries including hamstring, adductor, rectus femoris/quadriceps and calf muscle injuries.

METHODS
This statement is divided into four sections: (1) hamstring, (2) adductor, (3) rectus femoris/quadriceps and (4) calf muscle strain injuries, and it includes three domains in each section: (1) diagnosis, (2) prevention and (3) treatment. We did not include studies solely reporting on non-acute problems, and/or clear traumatic injuries such as total muscle ruptures, avulsion injuries and muscle contusions. A systematic search was employed to identify literature for the three domains in each section, with inclusion of studies based on the highest level of available evidence. Data were synthesised and the quality of evidence were evaluated using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group.
Search
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) were conducted in MEDLINE (via PubMed), CENTRAL and Embase (via Ovid) during July 2018 and updated in September 2019. 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 (MeSHs)’ OR ‘Reduce’*). If studies from one search was deemed relevant for one of the other 11 searches, the study was included as literature identified from other sources. In addition, reference lists of the included trials and relevant systematic reviews were scanned for relevant references. A flow chart of searches (see online supplementary file 1) and the complete search strategy for all searches and databases (see online supplementary file 2) is available as supplementary.

Selection
Three authors (LI, KK, RSH) screened records ordered from the most recent to the oldest records, using the Rayyan system (http://rayyan.qcri.org). With 12 sets of search results, two authors screened 8 sets each, and thus all searches were screened independently by two authors. For all domains we included studies based on the highest level of available evidence. For the diagnosis domain, we initially screened for systematic reviews and diagnostic cohort studies as these represent the highest starting point for the GRADE assessment. Thus, for diagnosis, we intended to include systematic reviews of diagnostic studies supplemented with additional diagnostic cohort studies. We aimed for studies that compared clinical tests to either ultrasonography or magnetic resonance imaging as the diagnostic reference standard. For the prevention and treatment domains, we initially screened for systematic reviews and randomised controlled trials (RCTs) as these represent the highest starting point for the GRADE assessment. If no systematic reviews and/or RCTs were identified, we screened for observational studies. Thus, for prevention and treatment, we intended to include systematic reviews on RCTs supplemented with additional RCTs, or secondly observation studies if no systematic reviews or RCTs were identified. For the prevention domain, we aimed for studies that assessed the risk of injury using a definition of injury based on time loss, problems or medical attention. For the treatment domain, we aimed for studies that investigated the effect of treatment on risk of reinjuries and/or time to return-to-play, defined as completion of treatment, criteria passed or self-reported completion of a full training session or match play. Studies that investigated the effect of treatment and/or prevention, but did not data extraction of individual muscle injuries were not included.

Appraisal
Two authors assessed risk of bias (LI and KK) of individual studies, as required for the GRADE framework and in line with Cochrane procedures. Therefore, tools assessing risk of bias rather than the study quality was chosen, and thus we used The Cochrane Collaboration’s risk of bias assessment tool (version 1) for RCTs, QUADAS-2 tool for diagnostic studies and Scottish Intercollegiate Guidelines Network Methodology Checklist 3 for cohort studies. Details of each tool can be found in online supplementary file 4. Furthermore, two authors assessed risk of bias (RSH and CBJ) in systematic reviews using the ROBIS assessment tool. Agreement was reached by consensus. If a systematic review included a risk of bias assessment of individual studies using one of the assessment tools stated above, no further risk of bias assessment was conducted for these individual studies. However, if these tools were not used, or if risk of bias was conflicting between two or more systematic reviews, we reassessed all risk of bias domains in the specific individual studies as part of this statement. The risk of bias assessments for systematic reviews and individual studies can be found in supplementary material (see online supplementary file 3; Table I–11).

Data synthesis
Two authors assessed the quality of evidence (LI and KK) 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 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 was 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 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 supplementary material (see online supplementary file 3; Table I–11).

Diagnostic tests
The diagnostic effectiveness of clinical tests was estimated from positive (LR+) and negative (LR−) likelihood ratios, and express the change in probability of the patient having the injury. A LR+ > 1 increases the post-test probability of a diagnosis following a positive test, while a LR− <1 decreases
the post-test probability of a diagnosis following a negative test. The diagnostic effectiveness of a positive and negative test was classified based on current guidelines as: very low (LR+: 1 to 2; LR−: 0.5 to 1), low (LR+: >2 to 5; LR−: 0.2 to <0.5), moderate (LR+: >5 to 10; LR−: 0.1 to <0.2); high (LR+: >10; LR−: <0.1).41

Prevention and treatment
Review Manager V.5.3 (The Nordic Cochrane Centre, Copenhagen) was used to calculate risk ratio and standardised effect sizes (Hedges’ g) if not available through published meta-analyses or individual RCT’s, to allow for consistency of interpretation across studies. Thus, if a meta-analysis did not report standardised effect size for continuous variables or risk ratio for dichotomised variables, we re-ran the analysis, if possible, using a random-effect model, unless otherwise stated in the original meta-analysis.43 Heterogeneity in study results was calculated using the I² statistic, which is a measure to indicate the consistency of results across studies, from 0% (no inconsistency) to 100% (maximal inconsistency).44 For continuous variables, for example, time to return-to-play, the effect of treatment was based on the standardised between-group difference calculated as Hedges’ g and assessed as trivial (g<0.2), small (g≥0.2), medium (g≥0.5) and large (g≥0.8).45 For dichotomised variables (eg, risk of reinjury for treatment or risk of injury for prevention), the effect of treatment and prevention was based on the magnitude of the risk reduction following the intervention calculated as: Risk Ratio (RR) = a/(a+b) / c/(c+d) where a is the number of injuries in the intervention group, b is the number without injuries in the intervention group, c is the number of injuries in the control group and d is the number without injuries in the control group.46 The magnitude of effect was inspired by the risk ratios proposed by Bahr47 for sample size calculations in sports injury prevention, and thus reflects arbitrary cut off values. Since these values were originally presented to study the magnitude of risk factors, that is a higher risk, we calculated the inverse values as 1/RR to determine the preventive effect and assessed the magnitude as trivial (RR >0.78), small (0.78≥ RR >0.61), medium (0.61≥ RR >0.47) and large (RR ≤0.47).45

RESULTS
In total 44 studies were included. For a detailed overview of included studies, including risk of bias assessment, GRADE and which individual studies are contained in systematic reviews, we refer to online supplementary file 3.

Hamstring injuries
Domain 1: diagnostic tests
In total 7081 studies were identified in the literature search. One systematic review59 and three cohort studies were included.60-63 No meta-analysis of diagnostic effectiveness was available from previous literature64 or conducted as part of this statement. Based on one prospective study, where diagnostic effectiveness was calculated by the authors of this statement,61 and two diagnostic studies,68 69 the ‘Taking off shoe test’ showed high diagnostic effectiveness, with perfect agreement between clinical testing and ultrasonography (n=140) (very low quality of evidence; table 1).68 However, since no false positives were observed in Zeren and Oztekin,60 this precludes calculation of LR+ for tests in that study. All other tests displayed very low to low diagnostic effectiveness (LR+: 0.95 to 1.50; LR−: 0.37 to 0.96) in predicting a positive or negative MRI (n=58 to 180) (moderate to high quality of evidence)69 71 (table 1).

Domain 2: prevention
In total 2468 studies were identified in the literature search. Three systematic reviews62-64 and 12 RCT’s was included.21-23 65 A systematic review and meta-analysis performed by van Dyk et al (2019) showed a medium and significant risk reduction (RR: 0.55, 95% CI 0.34 to 0.89; I²=67.0%) of hamstring injuries in football players in interventions including the Nordic hamstring exercise versus usual care (n=5362) (moderate quality of evidence).35 A systematic review and meta-analysis by Goode et al (2015) showed a medium but non-significant risk reduction (RR: 0.39, 95% CI 0.24 to 1.44; I²=69.6%) of mixed eccentric hamstring exercises versus usual care on hamstring injuries in football players (n=1229) (low quality of evidence).32 A systematic review and meta-analysis by Thorborg et al (2017) showed a large and significant risk reduction (RR: 0.35, 95% CI 0.22 to 0.54; I²=0.0%) of the FIFA 11+ programme versus usual care on hamstring injuries in football players (n=3417) (moderate quality of evidence).54 Since the meta-analysis from van Dyk et al (2019) only assessed the effect of interventions that included the Nordic hamstring exercise, a post hoc meta-analyses, as part of the present statement paper, was performed to investigate the isolated effect of the Nordic hamstring exercise protocol.66 Thus, data from Petersen et al68 and van der Horst et al77 was pooled. Based on this, the Nordic hamstring exercise protocol showed a large and significant risk reduction (RR: 0.35, 95% CI 0.22 to 0.54; I²=0.0%) compared with usual care on hamstring injuries in football (n=1521) (high quality of evidence).56 57 Data from individual studies62-63 are presented in table 2.

Domain 3: treatment
In total 978 studies were identified in the literature search. One systematic review67 and 11 RCT’s were included.22 60 68 69

Return to play
A systematic review by Pas et al (2015) performed two meta-analyses with time to return to play as outcome.67 First, the effect of lengthening hamstring exercises versus conventional exercises was estimated as HR,67 thus we re-ran the analysis to report Hedges’ g for consistency across studies. Based on this, lengthening hamstring exercises showed a large and significant effect versus conventional hamstring exercises (Hedges’ g=1.23, 95% CI 0.85 to 1.60; I²=0.0%) on return to play in elite football28 and track and field29 (n=131) (low quality of evidence). Second, platelet-rich plasma showed a trivial and non-significant effect versus control interventions (placebo saline27 or rehabilitation25 26) (HR=1.03, 95% CI 0.87 to 1.22; I²=75.0%) on return to play in athletes (n=154) (moderate quality of evidence). No additional RCTs could be added to the meta-analysis by Pas et al67 due to heterogeneous interventions. Data from individual studies22-24 68 69 are presented in table 3.

Reinjuries
A systematic review by Pas et al from 2015 performed two meta-analyses with risk of reinjuries as outcome.67 First, lengthening hamstring exercises showed a large but non-significant risk reduction (RR: 0.25, 95% CI 0.03 to 2.20; I²=0.0%) versus conventional hamstring exercises69 at 12-month follow-up in elite football28 and track and field athletes29 (n=131) (very low quality of evidence). Second, platelet-rich plasma showed a trivial and non-significant risk reduction (RR: 0.88, 95% CI
Table 1: Hamstring injury diagnosis: effectiveness of clinical tests and grading the quality of evidence

<table>
<thead>
<tr>
<th>Clinical tests</th>
<th>Likelihood ratio, (95% CI)</th>
<th>Diagnostic effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain on trunk flexion</td>
<td>LR+=1.48 (1.12 to 1.97)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Pain on active knee flexion</td>
<td>LR+=1.50 (0.91 to 2.49)</td>
<td>High quality of evidence</td>
</tr>
<tr>
<td>Painful passive straight leg raise</td>
<td>LR+=1.33 (1.04 to 1.70)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Painful active knee extension</td>
<td>LR+=1.33 (1.02 to 1.72)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Painful resisted knee flexion 90°</td>
<td>LR+=1.18 (0.99 to 1.41)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Painful resisted knee flexion 30°</td>
<td>LR+=1.13 (0.94 to 1.36)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Active slump</td>
<td>LR+=1.16 (0.59 to 2.28)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Composite test*</td>
<td>LR+=0.95 (0.89 to 1.02)</td>
<td>Moderate quality of evidence</td>
</tr>
</tbody>
</table>

US used as reference standard

| Taking off shoe test                   | LR+=NA | Very low quality of evidence |
| Resisted range of motion test         | LR+=NA | Very low quality of evidence |
| Passive range of motion test          | LR+=NA | Very low quality of evidence |
| Active range of motion test           | LR+=NA | Very low quality of evidence |

*Passive straight leg raise, active knee extension, manual muscle testing, active slump; MRI; The diagnostic effectiveness of the positive (LR+) and negative (LR−) likelihood ratios are classified individually as: very low (LR+: 1 to 2; LR−: 0.5 to 1), low (LR+: >2 to 5; LR−: 0.2 to <0.5), moderate (LR+: >5 to 10; LR−: 0.1 to <0.2); high (LR+: >10; LR−:<0.1). NA (non-applicable); diagnostic effectiveness unknown NA, non-applicable; US, ultrasonography.

Table 2: Hamstring injury prevention: effect and grading the quality of evidence

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>RR (95% CI)</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of injury</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interventions including the Nordic hamstring exercise versus usual care*; n=5362, male/female football</td>
<td>0.55 (0.34 to 0.89); i²=47.0%</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Mixed eccentric hamstring training versus usual care based*; n=1229, male football</td>
<td>0.59 (0.24 to 1.44); i²=49.6%</td>
<td>Low quality of evidence</td>
</tr>
<tr>
<td>FIFA 11+ programme versus usual care; n=3417, male/female football</td>
<td>0.39 (0.24 to 0.64); i²=0.0%</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Nordic hamstring exercise protocol versus usual care*; n=1521, male football</td>
<td>0.35 (0.22 to 0.54); i²=0.0%</td>
<td>High quality of evidence</td>
</tr>
<tr>
<td>Individual studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bounding exercise programme versus usual care; n=400, male football</td>
<td>0.89 (0.55 to 1.44)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>FIFA 11+ programme performed pre-football and post-football versus FIFA 11+ performed pre-football; n=280, male football</td>
<td>0.21 (0.05 to 0.95)</td>
<td>Very low quality of evidence</td>
</tr>
<tr>
<td>Modified FIFA 11+ with rescheduling of Part 2 versus standard FIFA 11+; n=806, male football</td>
<td>0.86 (0.59 to 1.25)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Balance board training versus usual care; n=140, female football</td>
<td>0.19 (0.02 to 1.42)</td>
<td>Very low quality of evidence</td>
</tr>
</tbody>
</table>

*Based on pooled data from meta-analysis. RR (risk ratio); i² (heterogeneity in study results); The preventive effect is assessed as RR assessed as trivial (RR >0.78), small (0.78≥ RR >0.61), medium (0.61≥ RR >0.47) and large (RR ≤0.47).
<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Effect size</th>
<th>Meta-analyses</th>
<th>Individual studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to return-to-play</td>
<td>Large</td>
<td>Medium</td>
<td>Small</td>
</tr>
<tr>
<td>Hedges’ g (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lengthening hamstring exercises versus conventional hamstring exercises*; n=131, male football and track and field athletes</td>
<td>1.23 (0.85 to 1.60); I²=0.0%</td>
<td>Low quality of evidence</td>
<td></td>
</tr>
<tr>
<td>Platelet-rich plasma versus placebo or rehabilitation*; n=154, various athletes</td>
<td>1.03 (0.87 to 1.22); I²=75.0%</td>
<td>Moderate quality of evidence</td>
<td></td>
</tr>
<tr>
<td>Lengthening hamstring exercises versus a criteria-based algorithm; n=48, male football</td>
<td>0.23 (−0.34 to 0.80)</td>
<td>Low quality of evidence</td>
<td></td>
</tr>
<tr>
<td>Hamstring stretching four times per day versus hamstring stretching one time per day; n=80, track and field athletes</td>
<td>2.31 (1.75 to 2.88)</td>
<td>Very low quality of evidence</td>
<td></td>
</tr>
<tr>
<td>Agility and trunk stabilisation versus hamstring stretching and strengthening; n=24, various athletes</td>
<td>0.75 (−0.08 to 1.58)</td>
<td>Very low quality of evidence</td>
<td></td>
</tr>
<tr>
<td>Running and eccentric hamstring strengthening versus agility and trunk stabilisation; n=29, various athletes</td>
<td>0.39 (−0.42 to 1.20)</td>
<td>Very low quality of evidence</td>
<td></td>
</tr>
<tr>
<td>Pain-threshold (≤4 on the 0 to 10 NRS) versus pain-free (0 on the 0 to 10 NRS) rehabilitation; n=37, male/female</td>
<td>0.75 (0.40 to 1.40)</td>
<td>Low quality of evidence</td>
<td></td>
</tr>
<tr>
<td>Pain-threshold (≤4 on the 0 to 10 NRS) versus pain-free (0 on the 0 to 10 NRS) rehabilitation at 6 month follow-up; n=37, male/female</td>
<td>1.05 (0.14 to 7.47)</td>
<td>Low quality of evidence</td>
<td></td>
</tr>
</tbody>
</table>

*Based on pooled data from meta-analysis. RR (Risk ratio); HR; I² (Heterogeneity in study results); NRS (Numeric Rating Scale); the effect of treatment regarding return to play is assessed by Hedges’ g as trivial (g<0.2), small (g≥0.2), medium (g≥0.5) and large (g≥0.8). The effect of treatment on reinjuries is assessed as risk ratio as trivial (RR >0.78), small (0.78≥ RR >0.61), medium (0.61≥ RR >0.47) and large (RR ≤0.47).
diagnostic effectiveness was included (n=81).\textsuperscript{71} Adductor palpation showed high diagnostic effectiveness in predicting a negative MRI (LR−: 0.08) (low quality of evidence).\textsuperscript{71} For the remaining tests (see table 4), very low to low diagnostic effectiveness (LR+: 1.45 to 3.30; LR−: 0.20 to 0.79) were observed in predicting a positive or negative MRI (low to moderate quality of evidence).

Domain 2: prevention
In total 1566 studies were identified in the literature search. Two systematic reviews\textsuperscript{54,72} and 13 RCTs were included.\textsuperscript{21,58-60,62,64,65,73-79} None of the studies specifically reported on acute adductor injuries. A systematic review by Esteve \textit{et al} (2015) performed three meta-analyses with risk of groin injuries as outcome.\textsuperscript{72} First, mixed groin prevention programmes showed a trivial and non-significant risk reduction (RR: 0.81, 95% CI 0.60 to 1.09; \(I^2=7.0\%\)) versus usual care on groin injuries in football (n=4191) (low quality of evidence).\textsuperscript{72} Second, specific adductor strength training showed a trivial and non-significant risk reduction (RR: 0.80, 95% CI 0.53 to 1.22; \(I^2=3.0\%\)) versus usual care on groin injuries in football (n=1067) (low quality of evidence).\textsuperscript{72} Third, the FIFA 11+ programme showed a small but non-significant risk reduction (RR: 0.64, 95% CI 0.27 to 1.49; \(I^2=59.0\%\)) versus usual care on groin injuries in football (n=2476) (very low quality of evidence).\textsuperscript{72} Furthermore, a systematic review and meta-analysis by Thorborg \textit{et al} (2017) investigating the effect of the FIFA 11+ programme in football showed a medium and significant risk reduction (RR: 0.58, 95% CI 0.40 to 0.84; \(I^2=8.0\%\)) versus usual care on groin injuries in football (n=3417) (low level of evidence).\textsuperscript{54} No additional RCTs were added to the existing meta-analyses\textsuperscript{44,72} due to heterogeneous interventions, populations and injury definitions (eg, groin injuries vs groin problems as used in Harøy \textit{et al}\textsuperscript{76}). However, two additional studies on the preventive effect of the FIFA 11+ programme in mixed sports and basketball were pooled in a post hoc meta-analysis.\textsuperscript{78,79} Based on this, the FIFA 11+ programme showed a medium but non-significant risk reduction (RR: 0.58, 95% CI 0.06 to 5.93; \(I^2=62.0\%\)) versus usual care on groin injuries in mixed sports (n=3732) (very low quality of evidence). Data from individual studies are presented in table 5.

Domain 3: treatment
In total 217 studies were identified in the literature search; however, no studies on the effect of treatment of groin injury could be included.

Rectus femoris/quadriceps injuries

Domain 1: diagnostic tests
In total 8729 studies were identified in the literature search. Two cohort studies\textsuperscript{60,71} were eligible, however, due to duplication of data between these studies, only the study focusing on diagnostic effectiveness was included (n=81 athletes).\textsuperscript{71} Proximal rectus femoris palpation showed high diagnostic effectiveness when positive (LR+: 11.20) and negative (LR−: 0.00) in predicting a positive or negative rectus femoris injury on MRI (low to moderate quality of evidence).\textsuperscript{71} Furthermore, resisted knee extension showed high diagnostic effectiveness when negative (LR−: 0.00) in ruling out a rectus femoris injury on MRI (moderate quality of evidence). For the remaining tests (see table 6), very low to moderate diagnostic effectiveness (LR+: 1.45 to 5.47; LR−: 0.15 to 0.55) were observed (low to moderate quality of evidence).\textsuperscript{71}

Domain 2: prevention
In total 3002 studies were identified in the literature search. No systematic reviews and five RCTs were included.\textsuperscript{21,58,60,62,64,65} As part of this statement data from Soligard \textit{et al}\textsuperscript{60} and Silvers-Granelli \textit{et al}\textsuperscript{21} was pooled post hoc, as this analysis was not available in the existing literature. Based on this, the FIFA 11+ programme showed a small and not significant effect versus usual care (RR: 0.73, 95% CI 0.48 to 1.12; \(I^2=0.0\%\)) in reducing anterior thigh injuries in football (n=3417) (low quality of evidence).\textsuperscript{60} Data from individual studies\textsuperscript{62,64,65} are presented in table 7.

Domain 3: treatment
In total 484 studies were identified in the search. No systematic reviews or RCT’s were included. One study (case-series) was included with a total of 18 Australian rules football players.\textsuperscript{80} A two-phase criteria-based intervention with increasing running and kicking intensity led to a return to full training at a mean of 13 days (range: 2 to 43) with no reinjuries during an unreported time-frame (very low quality of evidence).\textsuperscript{80}

---

**Table 4** Adductor injury diagnosis: effectiveness of clinical tests and grading the quality of evidence

<table>
<thead>
<tr>
<th>Clinical tests</th>
<th>Likelihood ratio, (95% CI)</th>
<th>Diagnostic effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI used as reference standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adductor palpation (adductor longus, gracilis, pectineus)\textsuperscript{71}</td>
<td>LR+=2.23 (1.51 to 3.29)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td></td>
<td>LR−=0.08 (0.02 to 0.31)</td>
<td>Low quality of evidence</td>
</tr>
<tr>
<td>Squeeze 0°\textsuperscript{71}</td>
<td>LR+=3.13 (1.75 to 5.59)</td>
<td>Low quality of evidence</td>
</tr>
<tr>
<td></td>
<td>LR−=0.26 (0.14 to 0.48)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Squeeze 45°\textsuperscript{71}</td>
<td>LR+=1.81 (1.13 to 2.92)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td></td>
<td>LR−=0.52 (0.33 to 0.81)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Resisted outer range adduction\textsuperscript{71}</td>
<td>LR+=3.30 (1.85 to 5.87)</td>
<td>Low quality of evidence</td>
</tr>
<tr>
<td></td>
<td>LR−=0.20 (0.10 to 0.41)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Passive adductor stretching\textsuperscript{1}</td>
<td>LR+=3.04 (1.51 to 6.14)</td>
<td>Low quality of evidence</td>
</tr>
<tr>
<td></td>
<td>LR−=0.49 (0.34 to 0.71)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Flexion abduction external rotation test\textsuperscript{29}</td>
<td>LR+=1.45 (0.81 to 2.60)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td></td>
<td>LR−=0.79 (0.59 to 1.06)</td>
<td>Moderate quality of evidence</td>
</tr>
</tbody>
</table>

MRI: the diagnostic effectiveness of the positive (LR+) and negative (LR−) likelihood ratios are classified individually as: very low (LR+: 1 to 2; LR−: 0.5 to 1), low (LR+: >2 to 5; LR−: 0.2 to <0.5), moderate (LR+: >5 to 10; LR−: 0.1 to <0.2), high (LR+: >10; LR−: <0.1).\textsuperscript{37}
Table 5  Groin injury prevention: effect and grading the quality of evidence

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Effect size</th>
<th>Large</th>
<th>Medium</th>
<th>Small</th>
<th>Trivial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of injury</td>
<td>RR (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analyses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed groin prevention programmes versus usual care*; n=4191, male/female football</td>
<td>0.81 (0.60 to 1.09); I²=7.0%</td>
<td>Low quality of evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific adductor strength training versus usual care*; n=1067, male football</td>
<td>0.80 (0.53 to 1.22); I²=3.0%</td>
<td>Low quality of evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIFA 11+ programme versus usual care*; n=2476, male/ female football</td>
<td>0.64 (0.27 to 1.49); I²=59.0%</td>
<td>Very low quality of evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIFA 11+ programme versus usual care*; n=3417, male/ female football</td>
<td>0.58 (0.40 to 0.84); I²=8.0%</td>
<td>Low quality of evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIFA 11+ programme versus usual care*; n=3732, male/ female from mixed sports</td>
<td>0.58 (0.06 to 5.93); I²=62.0%</td>
<td>Very low quality of evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual studies</td>
<td>OR (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adductor strengthening programme versus usual care; n=486, male football</td>
<td>0.59 (0.40 to 0.86)</td>
<td>Moderate quality of evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIFA 11+ programme performed pre- and post- football versus FIFA 11+ performed pre- football; n=280, male football</td>
<td>0.16 (0.02 to 1.29)</td>
<td>Very low quality of evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified FIFA 11+ with rescheduling of Part 2 versus standard FIFA 11+; n=806, male football</td>
<td>1.19 (0.81 to 1.76)</td>
<td>Moderate quality of evidence</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Based on pooled data from meta-analysis. RR (risk ratio) ORs; I² (heterogeneity in study results); the preventive effect is assessed as RR assessed as trivial (RR >0.78), small (0.78≥ RR >0.61), medium (0.61≥ RR >0.47) and large (RR ≤0.47).

Calf injuries

Domain 1: diagnostic tests

In total 5410 studies were identified in the search; however, no studies were included.

Domain 2: prevention

In total 2944 studies were identified in the search. No systematic reviews or RCTs were identified. One prospective cohort study could be included (n=24). A soccer-specific balance programme performed during five half-seasons were superior to a 6-month control period with a medium effect size (rate ratio: 0.57; 95% CI not reported) on reducing gastrocnemius injuries in football (very low quality of evidence).

Domain 3: treatment

In total 91 studies were identified in the search. No systematic reviews or RCTs were identified. Three studies (two case-series and one retrospective observational) were included with a total number of 825 subjects. Very low quality of evidence was observed for a multimodal treatment consisting of passive treatment modalities and progressive

Table 6  Rectus femoris/quadriceps injury diagnosis: effectiveness of clinical tests and grading the quality of evidence

<table>
<thead>
<tr>
<th>Clinical tests</th>
<th>Likelihood ratio (95% CI)</th>
<th>Diagnostic effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI used as reference standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectus femoris palpation</td>
<td>LR+=11.20 (4.85 to 25.86) LR=0</td>
<td>Low quality of evidence</td>
</tr>
<tr>
<td>Resisted hip flexion at 0°</td>
<td>LR+=1.45 (0.90 to 2.32) LR=0.55 (0.15 to 1.79)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Resisted hip flexion at 90°</td>
<td>LR+=2.47 (1.41 to 4.34) LR=0.36 (0.11 to 1.21)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Resisted hip flexion (modified Thomas test position)</td>
<td>LR+=2.36 (1.53 to 3.66) LR=0.20 (0.03 to 1.27)</td>
<td>Low quality of evidence</td>
</tr>
<tr>
<td>Resisted knee extension (modified Thomas test position)</td>
<td>LR+=4.17 (2.54 to 6.82) LR=0</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Passive hip extension (modified Thomas test position)</td>
<td>LR+=2.70 (1.50 to 4.86) LR=0.35 (0.10 to 1.17)</td>
<td>Moderate quality of evidence</td>
</tr>
<tr>
<td>Passive knee flexion (modified Thomas test position)</td>
<td>LR+=5.47 (2.75 to 10.87) LR=0.15 (0.02 to 0.94)</td>
<td>Low quality of evidence</td>
</tr>
</tbody>
</table>

MRI; the diagnostic effectiveness of the positive (LR+) and negative (LR−) likelihood ratios are classified individually as: very low (LR+: 1 to 2; LR−: 0.5 to 1), low (LR+: >2 to 5; LR−: 0.2 to <0.5), moderate (LR+: >5 to 10; LR−: 0.1 to <0.2); high (LR+: >10; LR−: <0.1).
exercises, which led to an average treatment time of 9 days (range: not reported) in one study and a 1 year reinjury rate of 6.8% (95% CI not reported) in another study. Furthermore, platelet-rich plasma and rehabilitation showed a large and significant effect versus rehabilitation alone (Hedges’ $g = 1.78$, 95% CI 1.19 to 2.37) following a negative test (moderate to high quality of evidence). The limited diagnostic effectiveness of most tests may be explained by inclusion of low compliance, could likely explain the continuous rise in hamstring strain injuries in elite football. We also found moderate quality of evidence for a trivial non-significant effect of a comprehensive bounding programme involving plyometric and running drills directed towards the injury mechanisms during high-speed running.

In summary, hamstring training, in form of the Nordic hamstring exercise protocol, may be essential for prevention of hamstring injuries, where improvements in eccentric strength, fatigue resistance, and alterations in muscle morphology and architecture, collagen expression at the myotendinous junctions and angle of peak torque are suggested mechanisms of effect.

**Domain 3: treatment**

Several interventions for treatment of hamstring injuries have been investigated, including targeted hamstring exercises and running, stretching, agility and trunk exercises and injection therapy. The majority of included subjects were male football players or track and field athletes. For the outcome measure ‘time to return-to-play’,
large heterogeneity in the outcome definitions (eg, criteria-based vs self-reported vs medical clearance) preclude any clear recommendations of a superior rehabilitation strategy for return to play. However, rehabilitation programmes with a focus on progressive targeted eccentric hamstring exercises supplemented with progressive running drills seems to result in the shortest return-to-play times with an associated low reinjury risk.38–40

Thus, lengthening hamstring exercises showed significant faster return to play and no reinjuries compared with conventional exercises in elite football players28 and track and field athletes.29 However, in a recent study, lengthening hamstring exercises showed a substantial higher reinjury rate of 25% compared with a multifactorial criteria-based programme with a reinjury rate of 4% with only small, and non-significant, differences in return-to-play time.30 Interestingly, the return-to-play times were markedly faster in the latter study, suggesting that longer return-to-play times may decrease risk of reinjury.10 41 46 It should however be noted, that due to the small absolute number of reinjuries in the above studies the effect of the interventions on reinjury risk is associated with substantial imprecision of the estimate leading to a downgrade of the evidence.39

Finally, moderate quality of evidence was observed for a trivial, non-significant, effect of platelet-rich plasma on time to return-to-play and reinjury risk.47

Adductor injuries

Domain 1: diagnostic tests

Based on one study concerning primarily male football players,71 high diagnostic effectiveness was observed for a negative adductor palpation test in ruling out a diagnosis of a MRI-verified adductor injury. Accordingly, Serner et al (2016) observed that 57% of athletes presented with an MRI-defined acute adductor injury (pre-test probability) but after negative palpation test the probability of injury decreased to 9%.71 Thus, the clinician can be fairly certain that no adductor injury is present if the athlete reports no pain on palpation. The test was downgraded to low quality of evidence due to indirectness (clinical value of a negative or positive test unknown)38 and imprecision (wide CIs).39 Additionally, potential important clinical implications for diagnosis were also observed for pain during the squeeze test, outer range isometric adduction and passive adductor stretch, with shifts in pre-test to post-test probability from 57% to 80% to 81% for a positive test.71 Thus, these tests could be used to rule-in an adductor injury, although a positive test is still associated with uncertainty; that is one out of five athletes with a positive test do not have an adductor injury on MRI. Currently, no evidence is available concerning potential implications for return-to-play prognosis, and therefore all tests were downgraded due to indirectness (clinical value of a negative or positive test unknown).38

Domain 2: prevention

Several interventions for prevention of groin injuries have been investigated primarily in football including both males and females,54 72 71 however, no RCT’s have specifically reported on adductor strain injuries. Thus, the quality of evidence for all studies/outcomes have been downgraded due to indirectness.38 The highest quality of evidence for a preventive effect was observed for an adductor strengthening programme consisting of the Copenhagen adduction exercise100 specifically performed the Copenhagen adduction exercise.100

Domain 3: treatment

No studies were identified for treatment of acute adductor injuries, which is remarkable since acute adductor injuries are considered a common muscle injury in football.7 However, based on current literature concerning long-standing adductor-related pain, an exercise-based approach with a large focus on strengthening the adductor muscles has been recommended.101

Rectus femoris/quadriceps injuries

Domain 1: diagnostic tests

One study including primarily male football players was included.71 However, it should be noted that this study only included athletes with acute onset of groin pain, and thus only encompass proximal rectus femoris injuries.71 High diagnostic effectiveness was observed for proximal rectus femoris palpation in predicting a rectus femoris injury in MRI-positive cases, that is cases are only included if they had positive MRI. The test showed a substantial shift in pre-test to post-test probability from 13% to 62% for a positive test and from 13% to 0% for a negative test.71 Although the test showed high diagnostic effectiveness, the post-test probability of 62% suggest that a positive test is still associated with large uncertainties in the diagnosis; that is one out of three athletes with a positive palpation test do not have an injury.13 Conversely, the clinician can be fairly certain that no injury is present if the athlete report no pain on palpation. Currently, no evidence is available concerning potential implications for return-to-play prognosis, and thus all tests were downgraded due to indirectness (clinical value of a negative or positive test unknown).38

Domain 2: prevention

The highest quality of evidence for a preventive effect was observed for the FIFA 11+ programme compared with usual care, which resulted in a 27% lower risk (non-significant) of quadriceps/anterior thigh injuries (low quality of evidence).21 60 Since the FIFA 11+ programme improve hip and knee strength,102 the preventive effect could result from this. It should, however, be noted that anterior thigh injuries may encompass different injury locations, and thus the effect on rectus femoris injuries is uncertain. Furthermore, none of the included studies investigated the preventive effect on rectus femoris/quadriceps injuries as the primary outcome measure.

Domain 3: treatment

Limited literature was found to guide treatment of rectus femoris/quadriceps injuries. Thus, a single small case-series in male Australian rules football observed no reinjuries after a two-phase criteria-based intervention with increasing running and kicking intensity (very low quality of evidence).30 It should,
however, be noted that assessment of reinjury rate was not the purpose of the study, and thus these findings should be treated with caution.

Calf injuries

Despite calf muscle injuries being prevalent in football, very limited literature was found to guide diagnosis, treatment and prevention. Thus, very low quality of evidence was observed for both a multimodal treatment approach, consisting of passive therapy, stretching and strengthening or platelet-rich plasma for treatment and a soccer-specific balance programme in elite female football players for prevention.

Limitations and methodological considerations

The current statement is not without limitations. First, risk of bias assessments of individual studies were not conducted by the authors of this statement, if an assessment was already available as part of a systematic review. This was chosen a priori in acknowledgement of the original study findings; however, this introduces several risk of bias rater teams, and potential rating discrepancies may exist due to differences in interpretation of risk of bias domains. However, since GRADE represents an overall assessment encompassing several domains, and thus does not rely only on risk of bias, potential minor discrepancies in the risk of bias assessments are unlikely to change the quality of evidence provided in the present study. Second, for evaluating the effect of treatment and prevention, we have reported estimates and compared interventions using effect sizes, rather than using absolute measures (eg, days to return to play). This approach was chosen to allow standardised comparisons across studies and to account for the different definitions of return to play (eg, criteria-based, self-reported or medical clearance) or number reinjuries (self-reported, imaging-defined, length of follow-up). Third, the cut off values used to estimate the magnitude of a preventive effect are based on arbitrary values proposed to aid sample size calculations for risk factor studies in sports injury research. Finally, post hoc meta-analyses were performed on existing literature if the results were not available in published meta-analyses, and if the post hoc analysis was deemed to add clinical value. Since post hoc analyses increase the risk of false positive findings these should be interpreted with caution.

CONCLUSION

We have graded the quality of evidence concerning diagnosis, prevention and treatment of the most common muscle injuries and provide a comprehensive and up-to-date summary of the best available evidence. Most clinical tests showed very low to low diagnostic effectiveness. For hamstring injury prevention, programmes that included the Nordic hamstring exercise resulted in a risk reduction of 45% to 65% when compared with usual care. For prevention of groin injuries, both the FIFA 11+ programme and the Copenhagen adductor strengthening programme resulted in risk reduction of 41% when compared with usual care. For treatment of hamstring injuries, lengthening hamstring exercises showed fastest return to play with a lower reinjury rate when compared with conventional hamstring exercises. Platelet-rich plasma had no effect on time to return-to-play and reinjury risk after a hamstring injury when compared with placebo or rehabilitation. Most outcomes for all muscle injuries and domains were graded as very low to moderate quality of evidence, indicating that further high-quality research is likely to have an important impact on the confidence in the effect estimates. At this point, research on diagnosis, prevention and treatment of muscle injuries primarily concerns hamstring muscle injuries, with only limited research on quadriceps, adductor and calf muscle injuries. Furthermore, muscle injury prevention research is mainly conducted in football, whereas muscle injury treatment research is conducted across different sports.

What is already known

- Lower extremity muscle strain injuries are very common in multidirectional sports.
- The quality of evidence related to diagnosis, prevention and treatment for the most common muscle injuries (hamstring, adductor, calf, rectus femoris/quadriceps) have not been investigated.

What are the new findings

- Most clinical tests for acute muscle injuries show very low to low diagnostic effectiveness.
- Hamstring and groin injuries can be reduced 40% to 65% using specific exercise interventions.
- Lengthening hamstring exercises show fastest return to play and lower reinjury rate compared with conventional exercises, and platelet-rich plasma offers no additional effect to current rehabilitation.
- Most outcomes were graded as very low to moderate quality of evidence, thus further high-quality research is likely to have an important impact on the confidence in the effect estimates.

REFERENCES


10. Review


17. Review


33. Review


37. Review


43. Review


45. Review


47. Review


49. Review


