ABSTRACT

Objective To investigate the diagnostic accuracy of MRI for identifying posterior element bone stress injury (PEBSI) in the athletic population with low back pain (LBP).

Study Design A systematic review searched for published sources up until July 2020. Eligibility criteria: prospective cohort design, MRI diagnosis, adolescents/young adults, chief symptoms of LBP, PEBSI as the clinical diagnosis and SPECT-CT as reference standard. Risk of bias and overall quality were assessed using QUADAS-2 and GRADE, respectively. A narrative synthesis was conducted.

Results Four studies were included, with three included in the quantitative synthesis. Compared with SPECT-CT, two studies involving MRI demonstrated sensitivity and specificity of 80% and 100%, and 88% and 97%, respectively. Compared with CT, one study involving MRI demonstrated sensitivity and specificity of 97% and 91%, respectively. Risk of bias was moderate to high although consistency across studies was noted.

Conclusion Findings support further research to consider MRI as the modality of choice for diagnosing PEBSI. MRI was consistent with SPECT-CT for ruling-in PEBSI, but the clinical value of cases where MRI had false negatives remains uncertain due to possible over-sensitivity by SPECT-CT.

INTRODUCTION

Posterior element bone stress injury (PEBSI) is one of the most common reasons for low back pain (LBP) in athletic populations with reported incidence of 14–35%, especially in sports such as gymnastics, diving and throwing sports. LBP may hinder athletic performance, impact on health and contribute to time loss from training and competitions. This may also lead to untimely cessation of professional careers. Diagnosis of PEBSI, however, may help where existing evidence suggests that early diagnosis (ie, prior to an established fracture (spondylolysis)) allows causative factors to be identified and remedied, thus lessening the probability of progression to complete fracture and improving full-union bone healing. Late diagnosis increases the risk of non-union, surgery and lengthy rehabilitation.

Clinically, individuals present with back pain that is exacerbated with extension, side flexion, or a combination of both that increases during sporting activities but decreases with rest, and is usually without neurological signs. The single-leg hyperextension test is commonly used to confirm secondary conditions that may develop as a result of delayed diagnosis of PEBSI (spondylolysis and/or spondylolisthesis); however, a recent systematic review has concluded that neither this test nor the clinical history have the diagnostic utility to confirm any of these conditions.
These recent findings support the importance of diagnostic imaging for this spectrum of conditions.

Existing evidence suggests that early PEBSI is occult on X-rays\(^1\)\(^{17}\) and can be even missed with CT,\(^8\)\(^{10}\) when the cortical bone is still intact. Consequently, MRI and single-photon emission computerised tomography (SPECT) are used for their sensitivity to detect this crucial early phase with bone marrow oedema,\(^8\) and increased tracer uptake (hotspot),\(^9\) respectively. Nevertheless, the gold standard modality for diagnosing PEBSI remains a subject of controversy.\(^1\)\(^1\)\(^1\) Despite this lack of agreement and the exposure to ionising radiation, which frequently includes radioactive tracer injections in adolescents, there is still a tendency by some experts to use X-ray as the first line of investigation followed by SPECT and/or CT.\(^1\)\(^1\)\(^1\)\(^1\) The possible consequences of either late diagnosis and/or over-exposing young athletes to ionising radiation warrant further clarification.

Three recent systematic reviews were identified\(^1\)\(^2\)\(^3\)\(^4\) however none focused on the crucial early phase. Based on the scarcity of literature on PEBSI, there is an urgent need to provide recommendations of the most suitable modality for diagnosing PEBSI, with a particular interest in the early stages of bone stress reactions.

**Objective**

To investigate the diagnostic accuracy of MRI for identifying PEBSI in the athletic population with LBP.

**METHODS**

**Design**

A systematic review was conducted according to a registered protocol (PROSPERO CRD42015023979) and based on the Centre for Reviews and Dissemination.\(^1\)\(^5\) The review is reported in line with PRISMA (online supplemental file 1).\(^1\)\(^6\) See online supplemental file 2 for detailed report of the methodology.

**RESULTS**

**Study selection**

From 1058 records, 588 studies were included. See the PRISMA flow diagram for study selection process (figure 1). One author failed to respond for missing data to determine eligibility.\(^1\)\(^7\) Four studies were included in the synthesis.\(^1\)\(^8\)\(^2\)\(^1\)\(^2\)

**Study characteristics**

The weighted mean age was 14.9 years,\(^1\)\(^8\)\(^2\)\(^0\)\(^2\)\(^1\) including 62–74% males. Symptom onset to imaging investigation varied but was less than 6 months\(^1\)\(^9\) and 36 days on average.\(^2\)\(^1\) The common objective for all studies was to evaluate the usefulness of MRI in identifying spondylolysis and the use of optimised parameters to diagnose PEBSI. Most studies aimed to evaluate the efficacy of MRI at early and/or acute stages of spondylolysis that is, PEBSI\(^1\)\(^8\)\(^1\)\(^9\)\(^2\)\(^1\); with one study investigating whether MRI could effectively replace SPECT-CT altogether.\(^2\)\(^0\) Two studies considered SPECT-CT and MRI as the reference standard and index test, respectively.\(^1\)\(^9\)\(^2\)\(^0\) The other two studies used only CT in their comparison with MRI\(^1\)\(^8\)\(^2\)\(^1\); and although not explicit, CT was considered as the reference standard. Further, Yamane et al\(^2\)\(^1\) also reported the results of their follow-up CT, which were used in their comparison with MRI findings. Kobayashi et al\(^2\)\(^1\) performed CT scans based only on positive MRI results, and thus did not provide the data to construct a 2×2 table summary. This study was therefore excluded from the quantitative synthesis. See table 1 for details of study characteristics.

---

**Figure 1** PRISMA flow chart.

**Figure 2** Risk of bias (ROB) and applicability concerns.

**Figure 3** Diagnostic accuracy estimates with 95% CI forest plots.
<table>
<thead>
<tr>
<th>Lead author (year of publ.)</th>
<th>Inclusion criteria</th>
<th>Participants</th>
<th>Study design</th>
<th>Objective</th>
<th>Target condition</th>
<th>Reference standard and description of technique</th>
<th>Index test and description of technique</th>
<th>Sen and Spe [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kobayashi (2013)</td>
<td>LBP without neurological Sx</td>
<td>Period of study: N/A; n=200</td>
<td>Age &lt; 18; Mean age: 14.1 years; Male/female: 144/56</td>
<td>Prospective, consecutive, cohort study</td>
<td>Early stage spondylolysis</td>
<td>CT — Reverse gantry angle in the plane parallel to the pars interarticularis only for the vertebral body corresponding to the area of the high intensity change seen on MRI</td>
<td>MRI — Sagittal T2W images</td>
<td>Excluded from quantitative synthesis</td>
</tr>
<tr>
<td>Masci (2006)</td>
<td>LBP Age 10–30 years Engaging in regular activity</td>
<td>Period of study: N/A; n=71</td>
<td>Age &lt; 25 (of those found +ve to SPECT) Male/female: 29/10</td>
<td>Prospective, cohort study</td>
<td>Active spondylolysis</td>
<td>SPECT-CT — A standard dose of 99Tcm-MDP was injected. Planar and SPECT images were obtained about 3 hours after injection — CT — Images were acquired in the reverse gantry axial plane</td>
<td>MRI — Sagittal T1W images</td>
<td>Sen 0.80 [0.66, 0.90]</td>
</tr>
<tr>
<td>Campbell (2005)</td>
<td>Extension LBP Adolescent and young adults</td>
<td>Period of study: N/A; n=72</td>
<td>Mean age: 16 years; Male/female: 45/27</td>
<td>Prospective, cohort study</td>
<td>Pars fracture</td>
<td>SPECT-CT — Procedure was initiated approximately 3 hours after administration of an appropriate paediatric dose of 99Tcm-HDP — CT — Images were acquired in the reverse-angle axial plane</td>
<td>MRI — Sagittal T1W</td>
<td>Sen 0.89 [0.76, 0.96]</td>
</tr>
<tr>
<td>Lead author (year of publ.)</td>
<td>Inclusion criteria</td>
<td>Patients</td>
<td>Study design</td>
<td>Objective</td>
<td>Target condition</td>
<td>Reference standard and description of technique</td>
<td>Index test and description of technique</td>
<td>Sen and Spe [95% CI]</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------</td>
<td>----------</td>
<td>--------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Yamane (1993)</td>
<td>Extension LBP without neurological Sx</td>
<td>Period of study June 1991 to May 1992 n=79 Mean age 14.6 years Male/female 59/20</td>
<td>Prospective, consecutive, cohort study</td>
<td>Report the significance of a hypo-intense signal in the pars-interarticularis in the early diagnosis of spondyloysis</td>
<td>Early stage spondyloysis</td>
<td>CT — Transverse views were obtained. Slices were made parallel to the vertebral arch — Results were compared with MRI based on initial and follow-up CT scan</td>
<td>MRI Imaging was performed in the coronal and sagittal planes using: — T1W images — T2W images</td>
<td>Sen 0.97 [0.87, 1.00] Spe 0.92 [0.85, 0.96]</td>
</tr>
</tbody>
</table>

LBP, low back pain; N/A, not available; publ., publication; Sx, symptoms; STIR, short tau inversion recovery, SEN, sensitivity; SPE, specificity; <, below; +ve, Positive; "99Tcm-MDP, Technetium 99 methylene diphosphonate; "99Tcm-HDP, Technetium 99 hydroxymethyl diphosphonate; T1W, T1 weighted; T2W, T2 weighted.
Risk of bias (ROB) within studies and concerns for applicability

A summary of results is presented in Table 2. Strength of agreement for ROB assessment was very good (Kappa=0.9, 95% CI 0.79 to 1). None of the included studies were at low ROB with most domains deemed unclear or at high ROB. Participant recruitment was unclear in most studies, and thus selection bias could not be excluded. Similarly, lack of clarity was also noticeable of the index domain in defining the criteria for a positive result.

Quality of evidence

All included studies started as high quality because of their diagnostic framework. Nevertheless, other factors, primarily the high ROB, decreased the overall quality of evidence. Also, patient important outcomes, such as benefit and/or harm, were not assessed within the included studies. Summary of GRADE can be viewed in Table 3.

Results of individual studies

Most studies showed consistent results. Two studies, investigating the diagnostic accuracy of MRI versus SPECT-CT, demonstrated MRI sensitivity of 80% (95% CI 65 to 89) and MRI specificity of 100% (95% CI 99 to 100), and MRI sensitivity of 88% (95% CI 75 to 95) and MRI specificity of 97% (95% CI 94 to 98). The study by Yamane et al. that investigated the diagnostic accuracy of MRI versus CT, demonstrated MRI sensitivity of 97% (95% CI 85 to 99) and MRI specificity of 91% (95% CI 85 to 95). The sequences used in this study could not distinguish between stages of non-lysis and pseudarthrosis, thus values were calculated based on the presence and absence of abnormal MRI signal. Consequently, pseudarthrosis cases that had no abnormal signal on MRI were considered chronic and, therefore, true negative. If calculated according to CT classification, the outcome would be

Table 2 QUADAS-2 appraisal form

<table>
<thead>
<tr>
<th>Lead author of included studies</th>
<th>Kobayashi</th>
<th>Masci</th>
<th>Campbell</th>
<th>Yamane</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1: patient selection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was a consecutive or random sample of patients enrolled?</td>
<td>Yes</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Was a case-control design avoided?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Did the study avoid inappropriate exclusions?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Could the selection of patients have introduced bias?</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Are there concerns that the included patients do not match the review question?</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Domain 2: index test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td>Did the study provide a clear definition of what was considered to be a ‘positive’ result?</td>
<td>Yes</td>
<td>Unclear</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Could the conduct or interpretation of the index test have introduced bias?</td>
<td>Low</td>
<td>Unclear</td>
<td>Low</td>
<td>Unclear</td>
</tr>
<tr>
<td>Are there concerns that the index test, its conduct, or interpretation differ from the review question?</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Domain 3: reference standard</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the reference standard likely to correctly classify the target condition?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Were the reference standard results interpreted without knowledge of the results of the index tests?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td>Did the study provide a clear definition of what was considered to be a ‘positive’ result?</td>
<td>Yes</td>
<td>Unclear</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Could the reference standard, its conduct, or its interpretation have introduced bias?</td>
<td>High</td>
<td>Unclear</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Are there concerns that the target condition as defined by the reference standard does not match the question?</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Domain 4: flow and timing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was there an appropriate interval between index test and reference standard?</td>
<td>Unclear</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Did all patients receive a reference standard?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Did patients receive the same reference standard (protocol)?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Did patients receive the same index test (protocol)?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Were all patients included in the analysis?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Could the patient flow have introduced bias?</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

In each domain: signalling questions (white background) are followed by summarising questions of ROB and applicability concerns (light grey background).
incorrect, and misleading. Additionally, calculating the acute stage alone, that is, without pseudarthrosis, maintained a high level of accuracy.

Overall, the diagnostic value of MRI for ruling PEBSI in was conclusive, and moderate to conclusive for ruling it out.

Synthesis of results
Meta-analysis was inappropriate due to the low number of studies (2), quality and clinical heterogeneity.

DISCUSSION
This is the first systematic review to investigate the diagnostic accuracy of MRI for identifying PEBSI in the lumbar spine in young athletes with LBP. MRI’s accuracy was found high and consistent throughout in comparison with SPECT-CT.

The diagnostic value of MRI for ruling PEBSI in was conclusive across all studies, and moderate to conclusive for ruling it out. Accuracy estimates were calculated based on normal versus abnormal scans for clarifying the ability of MRI in distinguishing between the presence and absence of posterior element pathology. The confidence in the overall quality of evidence is low to moderate. MRI was as accurate as CT in detecting fully formed fractures. With respect to follow-ups, where cortical bone disruption is not evident on CT, MRI offers insights to stages of healing owing to its sensitivity to bone marrow oedema. Kobayashi et al demonstrated that 43% of participants with positive findings on MRI were occult on reverse gantry-CT even with prior knowledge of MRI detected changes.

With respect to the early stages of fracture development, high levels of false negative were found for MRI compared with SPECT. Reasons for this are two-fold. First, as opposed to CT and MRI, there is no established grading system defining SPECT abnormalities in the lumbar spine. In the absence of a rigorous grading system to a particular diagnosis, reliability findings of observers are of limited value. Masci et al, for example, not only lacked a clear classification system for SPECT, but also modified a validated classification system for MRI. Second, SPECT is highly sensitivity to ongoing bone turnover activity. Essentially, in the absence of a true reference standard, caution should be taken as to the clinical value of positive SPECT scans. Scintigraphy uptake occurs frequently in athletes (34–45.2%) in non-painful sites. Such, false positive cases are commonly regarded as adaptive changes and are perceived normal. With the evidence supporting SPECT over planar bone scans for its enhanced sensitivity, it may be even more difficult distinguishing what ‘normal’ uptake is.

Lastly, patient-important outcomes such as exposure to ionising radiation and the associated risks in the athletic population should not be overlooked. The effective dose from a single course of X-ray and SPECT-CT scans is 10X

<table>
<thead>
<tr>
<th>True positive</th>
<th>Number of studies</th>
<th>Study design</th>
<th>ROB</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other considerations</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 studies</td>
<td>Cohort</td>
<td>Very serious*</td>
<td>Not serious</td>
<td>Serious†</td>
<td>Not serious</td>
<td>Publication bias‡ Strong association§</td>
<td>++++</td>
<td></td>
</tr>
<tr>
<td>True negative</td>
<td>3 studies</td>
<td>Cohort</td>
<td>Very serious*</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Publication bias‡ Strong association§</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>False positive</td>
<td>4 studies</td>
<td>Cohort</td>
<td>Very serious*</td>
<td>Not serious</td>
<td>Serious†</td>
<td>Publication bias‡ Strong association§</td>
<td>++++</td>
<td></td>
</tr>
<tr>
<td>False negative</td>
<td>3 studies</td>
<td>Cohort</td>
<td>Very serious*</td>
<td>Not serious</td>
<td>Serious¶</td>
<td>Publication bias‡ Strong association§</td>
<td>+++</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 GRADE quality assessment of the body of evidence

*Refers to the ROB within studies.
†50% of included studies did not completely adhere to the interventions of interest (lack of SPECT) to answer the research question, hence risking the external validity of findings.
‡Publication bias could not be fully excluded, but it was not deemed sufficient to downgrade the overall quality of evidence either as the search strategy was extensive and up to date overall.
§Quality of evidence was rated up for magnitude of effect as indirect evidence has shown that early diagnosis increases the probability of full bony healing, which may result in shorter rehabilitation period, but mainly for the lack of exposure to ionising radiation as opposed to other modalities.
¶False negatives present the uncertainty linking to patient-important outcomes, for example, the possible deleterious effects of delayed diagnosis.
ROB, risk of bias.
more, than what UK dwellings get in a year from natural background radiation exposure.

In contrast to previous reviews,12–14 our findings recommend seeking consensus about shifting the diagnostic focus to the earliest, potentially reversible, PEBSI stage. Accordingly, use of radiographs should be discouraged7 and with advancements in MRI, the latter should be considered as the first-line investigation in all circumstances. Further, volumetric interpolated breath-hold examination (VIBE) MRI scan was found accurate in characterising incomplete pars fractures in comparison with CT.8 These sequences, however, are not sensitive to bone marrow oedema.31 Consequently, VIBE sequences should be incorporated in a PEBSI MRI protocol with other highly sensitive sequences to bone marrow oedema. This takes into account the continuum of this condition, thus avoiding complementary diagnostic imaging and hence saving time, money and avoiding radiation exposure.

Strengths and limitations
This is the first review to consider patient-important factors to inform evidence-based decisions regarding the use of imaging for identifying PEBSI, especially in the absence of a gold-standard consensus. Publication bias was minimised as the search excluded two possible non-English studies, which nonetheless were not discussed in similar studies.

CONCLUSION
Findings suggest MRI having an important role in the diagnosis of PEBSI, with consistency between MRI and SPECT-CT in ruling-in PEBSI but importantly without the exposure to ionising radiation. Further research is required to consider MRI as an alternative to SPECT-CT and to balance benefits versus risks for the appropriate investigation approach.

Author affiliations
1School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, UK
2The Norwegian Sport Medicine Clinic, Oslo, Norway
3Radiology, Queen’s Medical Centre Nottingham University Hospital NHS Trust, Nottingham, UK
4Loughborough Performance Centre, English Institute of Sport, Loughborough University, Loughborough, UK
5Physiotherapy, English Institute of Sport, Manchester, UK
6British Gymnastics, English Institute of Sport, Newport, UK
7Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, UK

Twitter Nicola Heneghan @Nicola Heneghan.

Contributors The initial protocol and design were developed by RE, RK, KS, SS, LF, AR and NRH; RE, LHJ, RK, KS and NRH were involved in the completion of searches, data extraction, quality assessment and analysis; all authors contributed to the interpretation and write-up of the study.

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.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID ids Alison Rushton http://orcid.org/0000-0001-8114-7669 Nicola R Heneghan http://orcid.org/0000-0001-7599-3674

REFERENCES