Diagnosis accuracy of clinical tests for diagnosis of ankle syndesmosis injury: a systematic review

Amy D Sman, Claire E Hiller, Kathryn M Refshauge

ABSTRACT

Objectives To determine the value of clinical tests for accurate diagnosis of ankle syndesmosis injury.

Design Systematic review.

Data sources An electronic database search was conducted (to 6 August 2012) of databases such as: MEDLINE, CINAHL, EMBASE, PubMed and Cochrane Databases. References from identified articles were examined and seven authors of eligible studies were contacted for additional information.

Study selection Studies of any design, without language restriction, were included; however, systematic reviews were excluded. Eligible studies included participants with a suspected ankle syndesmosis injury but without fracture. Reliability studies compared one or more clinical tests and studies of test accuracy compared the clinical test with a reference standard.

Results The database search resulted in 114 full text articles which were assessed for eligibility. Three studies were included in the review and raw data of these studies were retrieved after contacting the authors. Eight clinical diagnostic tests were investigated; palpation of the tibiofibular ligaments, external rotation stress test, squeeze, Cotton, fibula translation, dorsiflexion range of motion (ROM) and anterior drawer tests. Two studies investigated diagnostic accuracy and both investigated the squeeze test by conflicting results. Likelihood ratios (LR) ranging from LR=1.50 to LR=1.50 were found for other tests. High intra-rater reliability was found for the squeeze, Cotton, dorsiflexion ROM and external rotation tests (83–100% close agreement). Inter-rater reliability was good for the external rotation test (ICC2,1>0.70). Fair-to-poor reliability was found for other tests.

Conclusions This is the first systematic review to investigate the reliability and accuracy of clinical tests for the diagnosis of ankle syndesmosis injury. Few studies were identified and our findings show that clinicians cannot rely on a single test to identify ankle syndesmosis injury with certainty. Additional diagnostic tests, such as MRI, should be considered before making a final diagnosis of syndesmosis injury.

INTRODUCTION

Ankle injuries are the most common injury among sporting populations1 and between 1% and 20%2-7 of these injuries involve the distal tibiofibular joint or ankle syndesmosis. Gerber et al2 suggested that the true incidence in the general population is higher than that reported since syndesmosis injuries are probably under-diagnosed.

The ankle syndesmosis provides integrity to the tibia and fibula and the dense syndesmosis ligaments provide the main stability to the distal tibiofibular joint. The distal tibiofibular syndesmosis was previously regarded as no more than an interossous ligament8 but since 1977 the definition has been extended to include the interossous membrane and ligament, together with the anterior inferior tibiofibular ligament (AITFL) and posterior inferior tibiofibular ligament (PITFL).8-10

Most ankle injury research has focused on the lateral ankle ligaments and inversion injuries, while research on injuries to the ankle syndesmosis, including diagnosis, is rare. However, several authors have reported that syndesmosis injury is difficult to evaluate and diagnose.3 11 12 and is therefore often missed in the clinic.13 but is thought to have a longer recovery time than other ankle injuries.4 14 A lack of consensus on diagnostic criteria, and reliability and accuracy of the clinical diagnostic tests, can result in late or missed diagnosis of injury to the ankle syndesmosis. This could lead to an inappropriate and even harmful choice of treatment and might ultimately lead to chronic symptoms.

Injuries to the ankle syndesmosis are distinguished from other ankle ligamentous injuries based on the history, including mechanism of injury, and physical examination.15 Several mechanisms of injury have been reported; however, dorsiflexion and external rotation with a firmly planted foot is most commonly described.4 7 12 16 Dorsiflexion produces a slight widening between the tibia and fibula, as the fibula undergoes a small degree of lateral rotation because of the varying slope of the lateral surface of the body of the talus.17 The classic feature of syndesmosis injury is palpatory tenderness over the anterior and posterior tibiofibular ligaments and between the tibia and fibula.15 Patients may also have swelling just proximal to the ankle joint,15-17 at or above the anterior tibiofibular ligament18-20 in the acute stage, although ankle syndesmosis injuries are not always prone to severe swelling, potentially because the damaged tissue is extracapsular.15

Several clinical diagnostic tests are used to assess ankle syndesmosis injury; however, their reliability and accuracy have not been extensively researched. Palpation of the tibiofibular ligaments,14 16 21 the dorsiflexion with external rotation stress test9 22 23 and the squeeze test7 24-25 are the most commonly described tests, but the Cotton21 23 26 and fibula translation22 26 tests have also been described.

The clinical diagnostic tests aim to reproduce symptoms by applying stress to the syndesmosis. The dorsiflexion with external rotation test is thought to reproduce pain over the ankle syndesmosis ligaments by mimicking the commonly described mechanism of injury. The squeeze test involves mid-calf compression of the tibia and fibula which is thought to cause separation at the distal tibiofibular joint, in turn increasing tension in

the remaining syndesmosis ligament fibres resulting in pain at the ankle.27 Biomechanical analysis has confirmed separation at the distal tibiofibular joint when the calf was compressed.24

The dense syndesmosis ligaments afford only a small degree of movement between the bones in healthy ankles, and the Cotton test and fibula translation test are used for the diagnosis of ankle syndesmosis injury because they are thought to produce increased movement compared to the opposite ankle. With the Cotton test, medial and lateral forces are applied to the talus with the ankle in the neutral position.15 26 For the fibula translation test, the fibula is translated from anterior to posterior on the tibia.15 These tests are positive when increased mediolateral (Cotton) or anteroposterior (fibula translation) movement is felt compared to the opposite ankle.15 26

In addition to clinical examination, several surgical and radiographic tools have been investigated for their diagnostic value for ankle syndesmosis injury. The most valuable technique to accurately diagnose ankle syndesmosis injury is arthroscopy.22 28 29 However, it is expensive and invasive and therefore is usually only performed to diagnose and treat the more severe grades of injury. That is, it is used to enable treatment and not simply as a diagnostic tool.27

Of the radiographic tools, MRI has been found to have the highest specificity and sensitivity, similar to that of arthroscopy.24 32 However, the costs for such examinations are still relatively high. X-ray is widely used, especially to rule out fractures, but the reliability of x-ray to diagnose injury to the ankle syndesmosis is questionable.27 33 CT, however, has been shown to be more sensitive than x-ray for detecting ankle syndesmosis injury,34 but has a high radiation dose. Therefore, although MRI would be the preferred diagnostic tool for this injury, x-ray and CT are also often used in clinical practice.

An accurate and reliable diagnostic protocol for diagnostic tests of ankle syndesmosis injury should facilitate timely and accurate diagnosis, enabling reduced use of expensive diagnostic tools, and an earlier start of appropriate treatment. Therefore, the aim of this systematic review was to determine whether clinical diagnostic tests accurately and reliably diagnose ankle syndesmosis injury.

METHODS
Identification of selected studies
Search strategy
Studies were retrieved from searches of the following databases from earliest record to 6 August 2012: MEDLINE, EMBASE, CINAHL, PsycINFO, PEDro, AMED, SPORTDiscus, Scopus, Web of Science, PubMed and Cochrane Register of Clinical Trials. The search strategy was developed for MEDLINE and modified for use in other databases (figure 1). The reference lists from identified articles were also examined for additional relevant references. Seven authors were contacted to obtain additional information.16 22 25 26 30 35 36 We received five responses16 22 26 30 36 and data from three studies16 26 36 that met the inclusion criteria were provided.

Eligibility criteria
The criteria for the selection of eligible studies were developed in accordance with the handbook for systematic reviews of diagnostic test accuracy from the Cochrane collaboration17 and by Devillé et al.18 Studies were selected based on the following eligibility criteria: any trial design; participants who sustained an ankle injury with suspicion of an ankle syndesmosis injury without fracture. There was no language restriction. Studies examining the accuracy of clinical tests needed to compare the

---

**Figure 1** Keyword search MEDLINE.
were calculated. Diagnostic accuracy was calculated using the quality was assessed by two independent reviewers.

Methodological quality of the included articles was assessed using the Quality Assessment for Diagnostic Accuracy (QUADAS) tool.39 40 The tool consists of 14 items rated as either ‘yes’, ‘no’, ‘unclear’ or ‘N/A’. Nine items relate to bias (item 3–7, 10–12 and 14), three to the quality of reporting (items 8, 9 and 13) and two to variability (items 1 and 2). Methodological quality was assessed by two independent reviewers.

SYNTHESIS OF RESULTS
Sensitivity, specificity and likelihood ratios (LR), with 95% CIs, were calculated. Diagnostic accuracy was calculated using the method as described by Portney and Watkins (true positive + true negative)/total number of cases.41 LR are considered to be both clinically useful statistics42 43 and the best indices of diagnostic accuracy.44 Guidelines described by Jaeschke et al45 were used to interpret the LR. A likelihood ratio between 0.5 and 2 is considered a very small or irrelevant shift, 0.2–0.5 or 2–5 is a small shift, 0.1–0.2 or 5–10 is a moderate shift and <1 or >10 is a large, often conclusive shift.45 Intra-rater reliability could not be calculated with ICCs due to insufficient variability between scores. Therefore, intra-reliability results were analysed by percentage agreement between sessions for the same rater. Inter-rater reliability scores were calculated with ICC(2,1) and 95% CIs. Interpretation of the ICC values was according to the guidelines by Fleiss.46

RESULTS
Description of included studies
The initial electronic database search resulted in a total of 7699 articles, leaving 4815 articles after the removal of duplicates. Following the title and abstract screening, 114 were selected for possible inclusion in the review and full text articles were retrieved. Following the screening of the full text, seven articles met the inclusion criteria16 22 25 26 30 35 36 (figure 3). Raw data from three articles were available and included in the systematic review for further analysis.16 26 36 Two studies were included for the analysis of accuracy of the clinical tests.16 26 Nussbaum et al16 compared findings from three clinical tests with the reference standard of radiography in 60 participants and Beumer et al26 compared findings from six clinical tests with arthroscopy in three participants. Two studies were included for the reliability analysis.26 36 Beumer et al26 reviewed six clinical tests on 12 participants with seven examiners and Alonso et al36 reviewed four tests on 53 participants with two examiners. A flow diagram of the search history and selection process is presented in figure 3.

A total of eight clinical diagnostic tests were investigated, of which six tests have previously been described in the literature as commonly used tests for the diagnosis of ankle syndesmosis.

### Table 1 Protocol used for clinical tests in the included studies

<table>
<thead>
<tr>
<th>Test</th>
<th>Authors</th>
<th>Protocol</th>
<th>Positive findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squeeze test</td>
<td>Nussbaum et al16</td>
<td>Patient sitting over side of the bed. Compression of fibula to the tibia above the midpoint of the calf</td>
<td>Pain over area of the syndesmosis ligaments</td>
</tr>
<tr>
<td></td>
<td>Alonso et al16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beumer et al16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dorsiflexion with compression test</td>
<td>Alonso et al26</td>
<td>Patient standing and actively dorsiflexing the ankle. Performed once unassisted and once with the therapist applying a manual compressive force to the malleoli</td>
<td>Significant increase in ankle range of motion when compression added, or decreased pain at end of range when compression added</td>
</tr>
<tr>
<td>Dorsiflexion range of motion</td>
<td>Beumer et al26</td>
<td>Application of passive dorsiflexion</td>
<td>Reduced dorsiflexion range of motion compared with contralateral ankle</td>
</tr>
<tr>
<td>External rotation test</td>
<td>Alonso et al26</td>
<td>Participant sitting over side of the bed. Application of a passive external rotation stress to affected foot and ankle with the knee at 90°and ankle in neutral</td>
<td>Pain over the syndesmosis ligaments</td>
</tr>
<tr>
<td></td>
<td>Beumer et al26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nussbaum et al16</td>
<td>Performed with the ankle passively and maximally dorsiflexed in external rotation on a fixed lower leg</td>
<td>Reproduction of pain along the syndesmosis</td>
</tr>
<tr>
<td>Palpation</td>
<td>Nussbaum et al16</td>
<td>Palpation over anterior tibiofibular ligament</td>
<td>Report of pain after pressing directly over the ligament</td>
</tr>
<tr>
<td></td>
<td>Alonso et al16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cotton test</td>
<td>Beumer et al26</td>
<td>Distal tibia stabilised and lateral force applied to the foot</td>
<td>Increased lateral translation of the talus from medial to lateral compared with contralateral side</td>
</tr>
<tr>
<td>Fibula translation</td>
<td>Beumer et al26</td>
<td>Apply anterior–posterior translation of the fibula</td>
<td>Anteroposterior displacement of the fibula is greater than contralateral side</td>
</tr>
<tr>
<td>Anterior drawer</td>
<td>Beumer et al26</td>
<td>Not described</td>
<td>Used as a differential diagnosis between lateral collateral instability and syndesmosis instability</td>
</tr>
</tbody>
</table>
injury. These are the palpation of the tibiofibular ligaments, the external rotation stress test, the squeeze test, the Cotton test and the fibula translation test. Beumer et al. also investigated the diagnostic accuracy of reduced dorsiflexion range of motion (ROM) and the anterior drawer test. However, although these tests are used as part of clinical examination, their use for the diagnosis of ankle syndesmosis injury has not been previously researched or described.

Quality assessment
Moderate quality was found for the three included studies. All three studies adequately addressed the majority of items on bias, variability and quality of reporting where applicable to the studies. Of the nine items relating to eliminating bias, Beumer et al. addressed six items, whereas Nussbaum et al. addressed five. Variability accounted for two items and was not addressed by Beumer et al.; however, Nussbaum et al. addressed both items. The final three items related to quality of reporting; Beumer et al. addressed two of the three, and Nussbaum et al. addressed all items. For Alonso et al., however, only 6/14 items applied to the study and ‘yes’ was scored on all of these (table 2).

Diagnostic accuracy of clinical tests
Of the two studies that investigated diagnostic accuracy (table 3) only the squeeze test was investigated in both studies; however, conflicting results were found. Nussbaum et al. reported a positive likelihood ratio of 2.68 (95% CI 1.93 to 3.73) whereas Beumer et al. showed a positive likelihood ratio of 0.67 (95% CI 0.39 to 1.15) and a negative likelihood ratio of 3.00 (95% CI 0.44 to 20.32). LR for the anterior drawer, Cotton test, dorsiflexion ROM, external rotation and fibula translation tests ranged from LR+1.50 to LR−1.50. Diagnostic accuracy, LR and sensitivity and specificity values could not be calculated for the external rotation with dorsiflexion test and AITFL tenderness (table 4).

Reliability of clinical tests
Intra-rater reliability
One study was retrieved with data that enabled the calculation of intra-rater reliability (table 5). We were not able to calculate ICCs due to insufficient variability between scores and the results are therefore presented as per cent agreement. Intra-rater reliability for the squeeze, Cotton, dorsiflexion ROM and the external rotation tests ranged from 83% to 100% agreement between the two test occasions. Intra-rater reliability for fibula...
<table>
<thead>
<tr>
<th>Author</th>
<th>Was the spectrum of patients representative of the patients who will receive the test in practice?</th>
<th>Were selection criteria clearly described?</th>
<th>Was the reference standard likely to correctly classify the target condition?</th>
<th>Did the reference standard receive the same reference standard regardless of the index test result?</th>
<th>Was the execution of the index test described in sufficient detail to permit its replication?</th>
<th>Was the execution of the reference test described in sufficient detail to permit its replication?</th>
<th>Were the index test results interpreted without knowledge of the results of the reference standard?</th>
<th>Were the reference standard results interpreted without knowledge of the results of the index test?</th>
<th>Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?</th>
<th>Were un-interpretable/intermediate test results reported?</th>
<th>Were withdrawals from the study explained?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beumer et al</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Unclear</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Nussbaum et al</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unclear</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Alonso et al</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* QUADAS tool appropriate for reliability, not for validity. Therefore no reference test was required (items 3–7, 9–11).

N/A, the item is not applicable to this study.
translation and anterior drawer test showed a greater variability ranging from 46% to 92% agreement between the two test occasions (table 6).

Inter-rater reliability
Inter-rater reliability was calculated from the raw data of two studies (table 5). Good reliability was found for the external rotation test in both studies (ICC$_{2,1}$ 0.732$^{26}$ and ICC$_{2,1}$ 0.744$^{36}$). Fair reliability was found for the squeeze test (ICC$_{2,1}$ 0.461$^{26}$ and 0.490$^{36}$) and ligament palpation (ICC$_{2,1}$ 0.490$^{36}$). Poor reliability was found for the anterior drawer test (ICC$_{2,1}$ 0.059$^{36}$), the Cotton test (ICC$_{2,1}$ 0.156$^{36}$) and the fibula translation test (ICC$_{2,1}$ 0.282$^{36}$) for the single study that evaluated these tests (table 7).

Table 3
Summary of included studies investigating diagnostic accuracy

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Inclusion–exclusion criteria</th>
<th>Procedures</th>
<th>Tests investigated</th>
<th>Reference standard used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nussbaum et al$^{16}$</td>
<td>N=60 (51M, 9F) NCAA Division I-A athletes with syndesmosis sprain 34 of 60 injuries in football, 9 lacrosse, 8 soccer, 2 rowing and 1 each in wrestling, gymnastics, swimming, track, field hockey, cheerleading, basketball Right ankle 28 Left ankle 32</td>
<td>History and physical examination consistent with an acute ankle sprain Functional disability by failure on a single leg hop test Evaluated and treated within 24 h of injury No history of prior syndesmosis ankle sprain</td>
<td>Physical examination and standard radiographs including AP, internal rotation mortise and lateral views Examined within 24 h of injury by certified athletic trainer. Physician reported on radiographs</td>
<td>Radiographs</td>
<td></td>
</tr>
<tr>
<td>Beumer et al$^{26}$</td>
<td>N=12 3 Participants with suspected chronic syndesmosis injury underwent arthroscopy 9 healthy subjects with asymptomatic ankles</td>
<td>Not defined</td>
<td>Participants sat behind a curtain with exposure only of the lower legs Both legs were examined twice in a different order by 7 examiners (4 orthopaedic surgeons, 3 orthopaedic registrars) Participants did not speak and indicated pain by tapping on a wooden board then pointing to the place where pain was felt Following day, 3 patients with suspected chronic syndesmosis ankle underwent arthroscopy</td>
<td>Squeeze test Cotton test Fibula translation External rotation stress test Anterior drawer Dorsiflexion range of motion</td>
<td>Arthroscopy</td>
</tr>
</tbody>
</table>

Table 4
Diagnostic accuracy of clinical tests, diagnostic accuracy, sensitivity, specificity, likelihood ratios (LR) (95% CI), $\chi^2$ and p value

<table>
<thead>
<tr>
<th>Author</th>
<th>N</th>
<th>Test</th>
<th>True positive</th>
<th>False positive</th>
<th>False negative</th>
<th>True negative</th>
<th>Diagnostic accuracy (%)$^*$</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>LR+</th>
<th>LR−</th>
<th>$\chi^2$</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beumer et al$^{26}$†</td>
<td>211</td>
<td>Anterior drawer</td>
<td>5 4 9 3</td>
<td>38.1</td>
<td>0.36 (0.16 to 0.61)</td>
<td>0.43 (0.16 to 0.75)</td>
<td>0.63 (0.24 to 1.62)</td>
<td>1.50 (0.59 to 3.84)</td>
<td>0.219 (0.16 to 0.61)</td>
<td>0.640</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cotton test</td>
<td>4 2 10 5</td>
<td>42.9</td>
<td>0.29 (0.12 to 0.55)</td>
<td>0.71 (0.36 to 0.92)</td>
<td>1.00 (0.24 to 4.20)</td>
<td>1.00 (0.56 to 1.78)</td>
<td>0.000 (0.12 to 0.55)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dorsiflexion</td>
<td>7 3 7 4</td>
<td>52.4</td>
<td>0.50 (0.27 to 0.73)</td>
<td>0.57 (0.25 to 0.84)</td>
<td>1.17 (0.43 to 3.18)</td>
<td>0.88 (0.38 to 2.00)</td>
<td>0.000 (0.27 to 0.73)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>External rotation</td>
<td>7 7 7 0</td>
<td>33.3</td>
<td>0.50 (0.27 to 0.73)</td>
<td>0.00 (0 to 0.35)</td>
<td>0.50 (0.30 to 0.84)</td>
<td>N/A</td>
<td>3.241 (0.27 to 0.73)</td>
<td>0.072</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fibula translation</td>
<td>9 3 5 4</td>
<td>61.9</td>
<td>0.64 (0.39 to 0.84)</td>
<td>0.57 (0.25 to 0.84)</td>
<td>1.50 (0.59 to 3.84)</td>
<td>0.63 (0.25 to 1.62)</td>
<td>0.219 (0.39 to 0.84)</td>
<td>0.640</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Squeeze test</td>
<td>8 6 6 1</td>
<td>42.9</td>
<td>0.57 (0.33 to 0.79)</td>
<td>0.14 (0.03 to 0.51)</td>
<td>0.67 (0.39 to 1.15)</td>
<td>3.00 (0.44 to 20.32)</td>
<td>0.670 (0.33 to 0.79)</td>
<td>0.413</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nussbaum et al$^{16}$§</td>
<td>60</td>
<td>Squeeze test</td>
<td>1 22 0 37</td>
<td>63.3</td>
<td>1.00 (0.21 to 1)</td>
<td>0.63 (0.50 to 0.74)</td>
<td>2.68 (1.93 to 3.73)</td>
<td>N/A</td>
<td>0.059 (0.21 to 1)</td>
<td>0.809</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>External rotation$^\ddagger$</td>
<td>1 59 0 0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AITFL tenderness</td>
<td>1 59 0 0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^*$ (True positive + true negative)/number of cases, as described by Portney and Watkins.$^{41}$
†Arthroscopy was used as the reference standard.
‡Three people had arthroscopy and were clinically diagnosed by seven (blinded) assessors.
§X-Ray was used as the reference standard.
$^\ddagger$External rotation test performed in maximal passive dorsiflexion.
AITFL, anterior inferior tibiofibular ligament N/A, results were not able to be calculated.
DISCUSSION

This systematic review found that the diagnostic sensitivity and specificity, and inter-rater reliability, of clinical tests to identify ankle syndesmosis injury were very low although the intra-rater reliability was adequate. The clinical tests reviewed were the anterior drawer, Cotton, dorsiflexion, external rotation, fibula translation and squeeze tests. However, the findings are based on two articles that investigated diagnostic accuracy and two articles that investigated the reliability of clinical tests used. Overall eight clinical tests and two reference standards were used in the diagnosis of a total number of 123 patients.

Diagnostic accuracy of clinical tests

Diagnostic accuracy for all evaluated clinical tests was very low. LR were all close to 1, indicating that the tests are unhelpful in the clinic. The squeeze test was the only test with a clinically important result.\cite{46} The conclusion that clinical tests are not accurate should be carefully interpreted. The aim of the study conducted by Nussbaum et al\cite{46} was not to investigate diagnostic accuracy, but to evaluate time missed from sport in an attempt to better define the injury. Therefore, the use of x-rays was in relation to clinical pragmatism not for use as a reference standard which is considered as a very valuable tool to accurately diagnose ankle syndesmosis injury. Unfortunately arthroscopy was only performed in three participants because the remaining nine participants were healthy controls.

Despite our extensive search of the literature we did not find any diagnostic accuracy studies using MRI as a reference standard, even though the ability of MRI to accurately detect ankle syndesmosis injuries is similar to that of arthroscopy.\cite{9,15} This could be due to the fact that most studies were conducted prior to knowledge of the excellent diagnostic accuracy of MRI. Future investigation of diagnostic accuracy of clinical tests with the use of MRI as a reference standard would be recommended.

Reliability of the clinical tests

The reliability of the clinical tests included in this systematic review varied among tests, but were consistent between the two studies included in our review. Intra-rater reliability was investigated in only one study\cite{26} which had such limited

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Inclusion–exclusion criteria</th>
<th>Study design</th>
<th>Tests investigated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alonso et al\cite{13}</td>
<td>N=53 (3BM, 15F) Age: 24.3±8.5 years (12–52) Time between injury and test: 34.2±125 days, median 60.4% tested within 7 days of injury All involved in a ball sport</td>
<td>Excluded if:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beumer et al\cite{52}</td>
<td>N=12 3 with suspected chronic syndesmosis injury 9 healthy subjects with asymptomatic ankles</td>
<td>Not defined</td>
<td>Participants sat behind a curtain with exposure only of the lower legs Both legs were examined twice in a different order by 7 examiners (4 orthopaedic surgeons, 3 orthopaedic registrars) Participants did not speak and indicated pain by tapping on a wooden board then pointing to the place where pain was felt</td>
<td>Squeeze test Cotton test Fibula translation External rotation test Ligament palpation</td>
</tr>
</tbody>
</table>

Additionally, the external rotation test in passive dorsiflexion and palpatory tenderness along the AITFL as investigated by Nussbaum et al\cite{46} were excluded from the analysis because there was no variability among the measures. The x-rays, which were used as the reference standard by Nussbaum et al\cite{46} were negative in 59 of the 60 cases, while AITFL tenderness and the external rotation in passive dorsiflexion test was positive for all 60 cases.

Beumer et al,\cite{26} however, used arthroscopy as the reference standard which is considered as a very valuable tool to accurately diagnose ankle syndesmosis injury. Unfortunately arthroscopy was only performed in three participants because the remaining nine participants were healthy controls.

Table 5 Summary of included studies investigating reliability

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Inclusion–exclusion criteria</th>
<th>Study design</th>
<th>Tests investigated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alonso et al\cite{13}</td>
<td>N=53 (3BM, 15F) Age: 24.3±8.5 years (12–52) Time between injury and test: 34.2±125 days, median 60.4% tested within 7 days of injury All involved in a ball sport</td>
<td>Excluded if:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beumer et al\cite{52}</td>
<td>N=12 3 with suspected chronic syndesmosis injury 9 healthy subjects with asymptomatic ankles</td>
<td>Not defined</td>
<td>Participants sat behind a curtain with exposure only of the lower legs Both legs were examined twice in a different order by 7 examiners (4 orthopaedic surgeons, 3 orthopaedic registrars) Participants did not speak and indicated pain by tapping on a wooden board then pointing to the place where pain was felt</td>
<td>Squeeze test Cotton test Fibula translation External rotation test Ligament palpation</td>
</tr>
</tbody>
</table>

Diagnostic accuracy of clinical tests

Diagnostic accuracy for all evaluated clinical tests was very low. LR were all close to 1, indicating that the tests are unhelpful in the clinic. The squeeze test was the only test with a clinically important result.\cite{46} The conclusion that clinical tests are not accurate should be carefully interpreted. The aim of the study conducted by Nussbaum et al\cite{46} was not to investigate diagnostic accuracy, but to evaluate time missed from sport in an attempt to better define the injury. Therefore, the use of x-rays was in relation to clinical pragmatism not for use as a reference standard, Nussbaum et al\cite{46} did report a significant relationship (p=0.03) between a positive squeeze test and a longer time lost from competition. This indicates that although the squeeze test might not be very useful as a diagnostic tool, it could be useful as a prognostic tool.

Table 6 Intra-rater reliability in percentages*\cite{26}

<table>
<thead>
<tr>
<th>Rater / test →</th>
<th>Squeeze test</th>
<th>Cotton test</th>
<th>Fibula translation</th>
<th>External rotation</th>
<th>Anterior drawer</th>
<th>Dorsiflexion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>91.7</td>
<td>91.7</td>
<td>87.5</td>
<td>91.7</td>
<td>70.8</td>
<td>95.8</td>
</tr>
<tr>
<td>B</td>
<td>91.7</td>
<td>87.5</td>
<td>87.5</td>
<td>100</td>
<td>91.7</td>
<td>100</td>
</tr>
<tr>
<td>C</td>
<td>91.7</td>
<td>83.3</td>
<td>87.5</td>
<td>95.8</td>
<td>87.5</td>
<td>95.8</td>
</tr>
<tr>
<td>D</td>
<td>91.7</td>
<td>95.8</td>
<td>70.8</td>
<td>95.8</td>
<td>45.8</td>
<td>100</td>
</tr>
<tr>
<td>E</td>
<td>87.5</td>
<td>100</td>
<td>87.5</td>
<td>95.8</td>
<td>87.5</td>
<td>87.5</td>
</tr>
<tr>
<td>F</td>
<td>87.5</td>
<td>91.7†</td>
<td>91.7</td>
<td>95.8</td>
<td>87.5</td>
<td>83.3</td>
</tr>
<tr>
<td>G</td>
<td>87.5</td>
<td>87.5</td>
<td>75.0</td>
<td>87.5</td>
<td>66.7</td>
<td>87.5</td>
</tr>
</tbody>
</table>

*Percentage of agreement between session 1 and session 2 measured in 12 patients (24 legs).
†Data from one leg in a single participant from one of the raters during one session was missing.
variability between test results that ICCs could not be calculated. However, the per cent close agreement was high for the squeeze, Cotton, external rotation and dorsiflexion tests (83–100% agreement).

Inter-rater reliability was investigated in two studies, finding good reliability for the external rotation stress test and fair-to-poor reliability for the squeeze, dorsiflexion, Cotton, anterior drawer and fibula translation tests. Both studies used similar clinical test examination protocols with experienced examiners who were appropriately blinded to the results. The main limiting factor was that only data from the three tests described were available from the two studies. In addition, of these tests, only the squeeze and external rotation test had common protocols. The protocol differed between studies for the dorsiflexion test: Beumer et al interpreted findings based on limited ROM whereas Alonso et al performed a dorsiflexion lunge with compression test. As seen in the results of this systematic review, there are many different ways to perform these clinical tests and few have been validated. We recommend that studies use trained examiners to ensure standardised performance of the investigated tests and consistent interpretation of findings.

Clinical implications
Although specific clinical tests have formed a critical part of the triage for diagnosis of ankle syndesmosis injury, the usefulness of these tests has not been adequately researched. Studies that reported the reliability or accuracy of clinical tests to diagnose ankle syndesmosis injury were limited, and therefore the diagnostic accuracy of the clinical diagnostic tests remains unclear. It is important to note that each test in this systematic review has been investigated and analysed for its individual ability to accurately diagnose ankle syndesmosis injury. Our findings show that clinicians cannot rely on a single test in isolation as no test has been found to be sufficient to identify ankle syndesmosis injury with certainty. It is likely that a combination of tests and inclusion of other elements such as symptoms and the patient’s history might further assist in diagnosis. However, the studies included in this review had several limiting factors in methodology, including the choice of a reference standard and the results should therefore be interpreted with caution. None of the sample sizes of the three articles we retrieved, after contacting the authors, were appropriately powered.

The ability of clinical tests to diagnose ankle syndesmosis injury accurately without the necessity of imaging investigations would enable early start of appropriate management and potentially decrease the risk of persistence of problems. Importantly, it would limit costs for patients and the healthcare system.

CONCLUSION
This is the first systematic review that has investigated reliability and accuracy of clinical tests for the diagnosis of ankle syndesmosis injury. There are clear gaps in the research about the reliability and diagnostic accuracy of the clinical tests and the usefulness of these clinical diagnostic tests remains uncertain. We recommend that future studies concentrate on using MRI or arthroscopy as a reference standard, recruit appropriate sample sizes and combine clinical tests with factors such as the patient’s history. Clinicians should be aware of the limitations of the current clinical tests and conduct additional diagnostic tests, such as MRI or arthroscopy to make a final diagnosis of syndesmosis injury.
What are the new findings

> This is the first systematic review that has investigated reliability and accuracy of clinical tests for the diagnosis of ankle syndesmosis injury. The clinical tests reviewed were the anterior drawer, Cotton, dorsiflexion, external rotation, fibula translation and squeeze tests.

> The findings showed that the diagnostic sensitivity and specificity were very low and only the squeeze test showed a clinically important result. Inter-rater reliability of clinical tests to identify ankle syndesmosis injury was investigated in two studies, finding only good reliability for the external rotation stress test. The intra-rater reliability was investigated in one study and was found to be adequate (83–100% agreement). However, these findings are based on two articles that investigated diagnostic accuracy and two articles that investigated the reliability of clinical tests used.

> Although specific clinical tests have formed a critical part of the triage for diagnosis of ankle syndesmosis injury, the usefulness of these tests has not been adequately researched. Studies that reported the reliability or accuracy of clinical tests to diagnose ankle syndesmosis injury were limited and had several limiting factors in methodology, including the choice of a reference standard and the results should therefore be interpreted with caution. None of the sample sizes of the three articles we retrieved, after contacting the authors, were appropriately powered.

> There are clear gaps in the research about the reliability and diagnostic accuracy of the clinical tests and the usefulness of these clinical diagnostic tests remains uncertain. We recommend that future studies concentrate on using MRI or arthroscopy as a reference standard, recruit appropriate sample sizes and combine clinical tests with factors such as the patient’s history.

How might it impact on clinical practice in the near future

> Our findings show that clinicians cannot rely on a single test in isolation as no test has been found to be sufficient to identify ankle syndesmosis injury with certainty. It is likely that a combination of tests and inclusion of other elements such as symptoms and the patient’s history might further assist in diagnosis.

> Clinicians should be aware of the limitations of the current clinical tests and conduct additional diagnostic tests, such as MRI or arthroscopy to make a final diagnosis of syndesmosis injury.

> The ability of clinical tests to diagnose ankle syndesmosis injury accurately without the necessity of imaging investigations would enable early start of appropriate management and potentially decrease the risk of persistence of problems. Importantly, it would limit costs for patients and the healthcare system.

Contributors ADS contributed to the conception and design of the study, the collection, extraction and analysis of the data and interpretation of the findings, the drafting and reviewing of the article critically for important intellectual content and the final approval of the version to be published. CEH contributed to the conception and design of the study, the collection and extraction of the data, the interpretation of the findings, the revising of the article critically for important intellectual content and the final approval of the version to be published. KMR contributed to the conception and design of the study, the interpretation of the findings, the revising of the article critically for important intellectual content and the final approval of the version to be published.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interest None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

19. Karl EL, Wrazidlo WJ, Peschke WH, et al. Anterior drawer, Cotton, dorsi-translation and squeeze tests. The intra-rater reliability was investigated in one study and was found to be adequate (83–100% agreement). However, these findings are based on two articles that investigated diagnostic accuracy and two articles that investigated the reliability of clinical tests used.
