ACCURACY OF EXAMINATION OF THE LONG HEAD OF THE BICEPS TENDON IN THE CLINICAL SETTING: A SYSTEMATIC REVIEW

Valérie BÉLANGER, MD, Frédérique DUPUIS, BSc, Jean LEBLOND, PhD, and Jean-Sébastien ROY, PT, PhD
From the 1Department of Physical Medicine and Rehabilitation (Physiatry), Centre Hospitalier Universitaire de Québec – Université Laval, 2Center for Interdisciplinary Research in Rehabilitation and Social Integration, and 3Department of Rehabilitation, Faculty of Medicine, Université Laval, Quebec City, Canada

Objective: To determine the diagnostic validity of high-resolution ultrasound and orthopaedic special tests in diagnosing long head of the biceps tendon pathologies in patients with shoulder pain.

Design: Systematic review with meta-analysis tools.

Data extraction: Included studies had to report on the diagnostic validity of orthopaedic special tests or high-resolution ultrasound (HRUS) compared with a reference standard for diagnosing long head of the biceps tendon target conditions (superior labrum anterior and posterior lesions, long head of the biceps tendon tendinopathy, dislocation, effusion or rupture). Risk of bias was assessed using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) tool.

Results: Of the 30 included studies, 8 focused on high-resolution ultrasound and 22 on orthopaedic special tests. High-resolution ultrasound proved highly specific for the diagnosis of long head of the biceps tendon pathologies. Pooled positive (LR+) and negative (LR−) likelihood ratios were 38.00 and 0.24 for dislocation, respectively, and 35.50 and 0.30 for complete rupture, respectively. The accuracy of orthopaedic special tests varied greatly across studies. The only test of value was Yergason’s manoeuvre in confirming proximal long head of the biceps tendon pathologies except superior labrum anterior and posterior lesion (high specificity): the summary LR+ and LR− were 2.56 and 0.70, respectively.

Conclusion: High-resolution ultrasound is reliable to confirm suspected long head of the biceps tendon pathologies. There is insufficient evidence to recommend individual orthopaedic special tests.

Key words: shoulder; biceps tendon; glenoid labrum; imaging; diagnostic ultrasound.

Accepted May 3, 2019; Epub ahead of print May 16, 2019

J Rehabil Med 2019; 51: 479–491

Correspondence address: Valérie Bélanger, Centre Hospitalier Universitaire de Québec – Université Laval, Hôpital de l’Enfant-Jésus, 1401, 18e Rue, Quebec City, Canada, G1J 1Z4. E-mail: valerie.belanger.20@ulaval.ca

Shoulde...
The aim of this study was to determine the diagnostic accuracy of: (i) diagnostic HRUS for detecting LHBT tendinopathy, dislocation, rupture (partial or complete) and bicipital recess effusion; and (ii) OSTs for detecting any pathology of the LHBT in patients with shoulder pain. The study determined the accuracy of each OST related to LHBT, for detecting the specific clinical entity for which they were designed (Appendix I) (14).

METHODS

Criteria for considering studies for this review
Included studies were prospective, either delayed cross-sectional or diagnostic case-control studies, which included patients recruited in primary, secondary or tertiary care settings. There was no limit to sample sizes or prevalence in the included studies; however, 100% prevalence studies were eliminated because they do not allow calculation of specificity.

Participants
Any patients with shoulder pain were considered, with no limit on diagnosis or age group. However, studies including exclusively rheumatological or neurological populations were not considered, since these disorders encompass a diverse group of musculoskeletal conditions that differ from those found in the general population.

Index tests
OSTs (Appendix I) and HRUS were the index tests. HRUS methods for examining the LHBT had to be congruent with accepted standards (15, 16).

Target conditions
SLAP lesions, tendinopathy, dislocation, rupture and effusion (bicipital recess) of the LHBT were considered.

Reference standards
HRUS had to be compared with surgery (open or arthroscopy), magnetic resonance (MR) imaging or MR arthrography. OSTs had to be compared with surgery, HRUS, MR imaging or MR arthrography.

Table I. Quality assessment tool for diagnostic accuracy studies (QUADAS-2) items’ specifications developed by authors of the review

<table>
<thead>
<tr>
<th>QUADAS-2 item</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient selection</td>
<td>Study samples had to include unrestricted population of patients with shoulder pain. If the inclusion criteria were too specific, we considered a high risk of bias.</td>
</tr>
<tr>
<td>Index tests</td>
<td>Had to be described, or at least, referenced. If neither of those, we attributed a high risk of bias.</td>
</tr>
<tr>
<td>Reference standards</td>
<td>The following reference standards were considered adequate for each LHBT pathology (2): SLAP lesion: histopathology, open surgery, arthroscopy and MR arthrography; LHBT tendinopathy: histopathology and open surgery; LHBT rupture or dislocation: histopathology, open surgery, arthroscopy, MR imaging/arthrography and HRUS.</td>
</tr>
<tr>
<td>Flow and timing</td>
<td>A 1-month period in between the index test and reference standard was considered adequate. We chose this cut-off to minimize disease progression bias, since tendon condition can evolve over time. We are aware that some authors (58) consider an 8-month period as a cut-off, but this seems too long.</td>
</tr>
</tbody>
</table>

LHBT: long head of the biceps tendon; SLAP: superior labrum anterior and posterior; MR: magnetic resonance; HRUS: high-resolution ultrasound.
participants they recruit and the test that they evaluate (19). In that respect, data were combined where studies measured the accuracy of the same index test for the diagnosis of the same LHBT pathology; (i) according to the same reference standard; and (ii) according to all reference standards. Meta-analysis tools were used when a minimum of 4 primary studies were identified (Table II) (20). Where a limited number of studies prevented the use of meta-analysis tools, only sensitivity (Sn) and specificity (Sp) estimates are presented from each study, together with forest plots.

Meta-analyses were conducted using the approach developed by Rutter & Gatsonis with the V3.3.3 of R statistical software (http://www.r-project.org/) (21). The HSROC package was used to calculate overall pooled estimates of the included diagnostic studies taking into account the between-study and within-study variability. This routine, based on Bayesian statistics, estimates the overall sensitivity (Sn) and specificity (Sp) for group of studies and produces a receiver operating characteristic (ROC) curve with credible interval and a 95% prediction region. The classical confidence interval (CI) presumes that differences in Sn and Sp between studies are caused only by a statistical instability related to sampling or measurement errors. All estimates would turn around a unique value of Sn and a unique value of Sp. In reality, for the same technique, Sn and Sp may vary in time, with different populations, with different operators or any other relevant conditions that change the nature of the test. Across different conditions, Sn and Sp could fluctuate among a range of values that reflect a change in reality rather than a statistical instability. The credible intervals delimit how Sn and Sp could fluctuate for reasons other than sampling or measurement errors. In this context, the CI adds to the credible interval the uncertainty caused by sampling and measurement errors. The credible intervals are narrower than the CI. The prediction region is defined by pairing the CI with the credible interval. Heterogeneity was explored graphically using forest plots. Positive (LR+) and negative (LR–) likelihood ratios were calculated from the overall Sn and Sp. However, confidence and credible intervals could not be calculated for likelihood ratios. All estimates would turn around a unique value of LR+ and a unique value of LR–. In reality, for the same technique, LR+ and LR– may vary in time, with different populations, with different operators or any other relevant conditions that change the nature of the test. Across different conditions, LR+ and LR– could fluctuate among a range of values that reflect a change in reality rather than a statistical instability. The credible intervals delimit how LR+ and LR– could fluctuate for reasons other than sampling or measurement errors. In this context, the CI adds to the credible interval the uncertainty caused by sampling and measurement errors. The credible intervals are narrower than the CI. The prediction region is defined by pairing the CI with the credible interval.

For SLAP lesions, because the degenerative fraying of the SLAP I lesion is often considered a normal variant and asymptomatic, type II–IV and type I–IV lesions studies were isolated (22). The type II–IV group comprised studies either designed to assess the diagnosis of SLAP II–IV lesions or where only SLAP II–IV lesions were ascertained by the reference standard.

### RESULTS

#### Search results

Searches resulted in 777 citations (duplicates removed). Twenty-eight articles were accepted for the review after full-text screen. Fourteen articles were obtained by scrutiny of the reference lists of reviews and primary studies. Of the 42 eligible studies, 30 were included in the analysis of the review (8 for HRUS, 22 for OSTs; Fig. 1, Table III).

#### Methodological quality of included studies

For the risk of bias assessment, inter-rater agreement was excellent (Gwet’s AC1 of 0.85). The overall

---

**Table II. Possible combinations of index test/reference standard/target condition for meta-analyses**

<table>
<thead>
<tr>
<th>Index test</th>
<th>Reference standard</th>
<th>LHBT pathology identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRUS</td>
<td>Surgery (open or arthroscopy)</td>
<td>One of: Tendinopathy, Dislocation, Rupture – partial, Rupture – total, Effusion (bicipital recess)</td>
</tr>
<tr>
<td>Another diagnostic imaging modality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any reference standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSTs (for each)</td>
<td>Surgery (open or arthroscopy)</td>
<td>One pathology which each OST is designed to detect (see Appendix I): SLAP-lesion, Tendinopathy</td>
</tr>
<tr>
<td>HRUS</td>
<td>MR imaging/arthrography</td>
<td>Proximal LHBT pathology other than SLAP (dislocation, rupture, tendinopathy)</td>
</tr>
<tr>
<td>Any reference standard</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HRUS: high-resolution ultrasound; OSTs: orthopaedic special tests; MR: magnetic resonance; SLAP: superior labrum anterior and posterior; LHBT: long head of the biceps tendon.
<table>
<thead>
<tr>
<th>Study</th>
<th>n of shoulders</th>
<th>Study design</th>
<th>Index test</th>
<th>Reference standard</th>
<th>Target condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRUS accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armstrong, 2006 (23)</td>
<td>71</td>
<td>Prospective</td>
<td>HRUS</td>
<td>Arthroscopy</td>
<td>Dislocation</td>
</tr>
<tr>
<td>Farin, 1995 (44)</td>
<td>266</td>
<td>Prospective</td>
<td>HRUS</td>
<td>Open surgery</td>
<td>Dislocation</td>
</tr>
<tr>
<td>Fischer, 2015 (24)</td>
<td>45</td>
<td>Prospective</td>
<td>HRUS</td>
<td>MRI</td>
<td>Dislocation (complete)</td>
</tr>
<tr>
<td>Moosmayer, 2005 (32)</td>
<td>75</td>
<td>Prospective</td>
<td>HRUS</td>
<td>Arthroscopy or open surgery</td>
<td>Dislocation Rupture (partial) Rupture (total)</td>
</tr>
<tr>
<td>Naredo, 1999 (43)</td>
<td>36</td>
<td>Prospective</td>
<td>HRUS</td>
<td>MRI</td>
<td>Effusion Dislocation Tendinopathy</td>
</tr>
<tr>
<td>Read, 1998 (33)</td>
<td>42</td>
<td>Prospective</td>
<td>HRUS</td>
<td>Arthroscopy</td>
<td>Tendinopathy Dislocation Rupture (total)</td>
</tr>
<tr>
<td>Skendzel, 2011 (34)</td>
<td>67</td>
<td>Prospective</td>
<td>HRUS</td>
<td>Arthroscopy</td>
<td>Tendinopathy Rupture (partial) Rupture (total)</td>
</tr>
<tr>
<td>Teefey, 2000 (42)</td>
<td>100</td>
<td>Prospective</td>
<td>HRUS</td>
<td>Arthroscopy</td>
<td>Dislocation Rupture (total)</td>
</tr>
<tr>
<td>OSTs accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrigoni, 2014 (28)</td>
<td>109</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation</td>
<td>Arthroscopy</td>
</tr>
<tr>
<td>Chen, 2011 (47)</td>
<td>143</td>
<td>Prospective</td>
<td>ACT BLII</td>
<td>Palpation Speed Yergason</td>
<td>SLAP I-IV</td>
</tr>
<tr>
<td>Cook, 2012 (35)</td>
<td>87</td>
<td>Prospective</td>
<td>ACT DLST</td>
<td>LTT Speed</td>
<td>Arthroscopy</td>
</tr>
<tr>
<td>Fowler, 2010 (25)</td>
<td>101</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP I-IV</td>
</tr>
<tr>
<td>Gill, 2007 (29)</td>
<td>847</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP I-IV</td>
</tr>
<tr>
<td>Guanche, 2003 (36)</td>
<td>60</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP I-IV</td>
</tr>
<tr>
<td>Kibler, 2009 (37)</td>
<td>101</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP I-IV</td>
</tr>
<tr>
<td>Kim HA, 2007 (51)</td>
<td>176</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP I-IV</td>
</tr>
<tr>
<td>Kim SH, 2001 (46)</td>
<td>127</td>
<td>Prospective</td>
<td>ACT BLII</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Kim YS, 2007 (45)</td>
<td>61</td>
<td>Prospective</td>
<td>ACT PCT</td>
<td>Arthroscopy</td>
<td>SLAP I-IV</td>
</tr>
<tr>
<td>Lasbleiz, 2014 (48)</td>
<td>39</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>McFarland, 2002 (38)</td>
<td>409</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Michener, 2011 (22)</td>
<td>55</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Micheroli, 2013 (49)</td>
<td>100</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Mimori, 1999 (26)</td>
<td>7</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Nakawaga, 2005 (27)</td>
<td>54</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Oh, 2008 (31)</td>
<td>146</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Parentis, 2006 (39)</td>
<td>132</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Salaffi, 2010 (50)</td>
<td>203</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Schlechter, 2009 (40)</td>
<td>254</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Sodha, 2017 (41)</td>
<td>774</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
<tr>
<td>Toprak, 2013 (30)</td>
<td>69</td>
<td>Prospective</td>
<td>ACT Speed</td>
<td>Palpation Speed Yergason</td>
<td>SLAP II-IV</td>
</tr>
</tbody>
</table>

ACT: Active compression test; AST: anterior slide test; BLII: biceps load II; DLST: dynamic labral shear test; LTT: labral tension test; PCT: passive compression test; PDT: passive distraction test; UCT: upper cut test; SLAP: superior labrum anterior and posterior; LHBT: long head of the biceps tendon; HRUS: high-resolution ultrasound; OSTs: orthopaedic special tests.
Examination of the LHBT in the clinical setting

A study assessment shows some risk of bias in 3 of the 4 categories (Fig. 2). For patient selection, 53% of all studies were assessed as low risk. Nine studies were judged at high risk because of restricted population (n = 5) (23–27), inappropriate exclusions (n = 3) (28–30) and case-control study design (n = 1) (31). In addition, three of them did not enrol patients in a consecutive manner (26, 27, 30). For index test, beside inadequate test description (n = 1) (23) and unknown blinding to the reference standard (n = 2) (26, 32), all were assessed as low risk of bias. For reference standard, 33% of studies included had a low risk of bias. All studies judged as high risk had a blinding issue (n = 14) (23, 25, 29, 31–41). For flow and timing, 27% of the eligible studies were deemed to have low risk. All studies considered to have high risk had inadequate interval between index test and reference standard (n = 8) (22, 25, 26, 32–35, 42). Moreover, for 3 of them, the reference standard was not the same for all patients.

Findings

Few studies compared the same index test with the same reference standard for the same target condition. Therefore, meta-analyses could be considered only for the following combinations: diagnosis of (i) LHBT dislocation with HRUS, (ii) LHBT complete rupture with HRUS, (iii) SLAP I–IV lesions with the Speed test, (iv) SLAP II–IV lesions with the active compression test, the anterior slide test and the crank test, (v) any pathology of proximal LHBT except SLAP lesion with the Speed test and the Yergason’s manoeuvre.

HRUS accuracy

Tendinopathy. Three studies evaluated HRUS for diagnosing LHBT tendinopathy, either with surgery or MRI as reference standard (33, 34, 43). While Sn estimates ranged from 0.22 to 1.00, Sp varied from 0.88 to 1.00 (Fig. S11).

Dislocation. Seven studies assessed the accuracy of HRUS for diagnosing LHBT dislocation, comparing with surgery or MRI (23, 24, 32, 33, 42–44). Sn varied from 0.33 to 1.00, while Sp was in the high end of the spectrum, ranging from 0.96 to 1.00 (Fig. S11). Data from the 7 studies were pooled (Table IV, Fig. 3). Point estimates for Sn and Sp are 0.76 (95% CI 0.15–1.00) and 0.98 (95% CI 0.65–1.00), respectively. Results indicate a quite high Sp but more fluctuating Sn.

Effusion. One study evaluated HRUS accuracy in diagnosing LHBT effusion compared with MRI (43). The Sn and Sp estimates were 0.79 and 0.73, respectively (Fig. S11).

Partial rupture. Two studies investigated HRUS accuracy for the diagnosis of LHBT partial tear, and comparison was made with surgery (32, 34). Sn ranged from 0.27 to 1.00 and Sp was 1.00 for both studies (Fig. S11).

Complete rupture. Five studies evaluated HRUS in diagnosing complete LHBT rupture, compared with surgery or MRI (24, 32–34, 42). Sn and Sp ranged from 0.64 to 1.00 and 0.87 to 1.00, respectively (Fig. S11). Data from the 5 studies were pooled (Table IV, Fig. 3): Sn and Sp are 0.71 (95% CI 0.11–1.00) and 0.98 (95% CI 0.61–1.00), respectively. The results indicate a quite high Sp, but more fluctuating Sn.

Orthopaedic special test accuracy

SLAP I–IV lesions. Accuracy for diagnosing SLAP I–IV lesions was assessed for 10 OSTs (Fig. S21). The

Table IV. Overall accuracy of high-resolution ultrasound in characterization of long head of the biceps tendon pathology

<table>
<thead>
<tr>
<th>Pathology</th>
<th>n (studies)</th>
<th>n (shoulders)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>LR+</th>
<th>LR–</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dislocation</td>
<td>7</td>
<td>624</td>
<td>0.76 (0.15–1.00)</td>
<td>0.98 (0.65–1.00)</td>
<td>38.0</td>
<td>0.24</td>
</tr>
<tr>
<td>Complete rupture</td>
<td>5</td>
<td>333</td>
<td>0.71 (0.11–1.00)</td>
<td>0.98 (0.61–1.00)</td>
<td>35.50</td>
<td>0.30</td>
</tr>
</tbody>
</table>

95% CI: 95% confidence interval; LR+: positive likelihood ratio; LR–: negative likelihood ratio.

1http://www.medicaljournals.se/jrm/content/?doi=10.2340/16501977-2564
Sn and Sp ranged or were for each test, respectively: from 0.60 to 0.91 and from 0.13 to 0.85 for the active compression test (35, 37, 39), from 0.81 to 0.82 for anterior slide test (37, 39), 0.55 and 0.53 for biceps load II test (35), from 0.13 to 0.39 and from 0.67 to 0.83 for crank test (36, 39), from 0.58 to 0.89 and from 0.31 to 0.98 for dynamic labral shear test (35, 37, 41), 0.27 and 0.75 for labral tension test (35), 0.48 and 0.52 for palpation test (36), 0.82 and 0.83 for passive compression test (45), from 0.09 to 0.47 and from 0.56 to 0.74 for Speed test (35–37, 39), and 0.23 and 0.57 for uppercut test (37). Data were pooled from studies assessing the Speed test (Table V, Fig. 4).

The results indicate a widely variable performance. Its point estimates for Sn and Sp are 0.36 (95% CI 0.00–0.82) and 0.71 (95% CI 0.23–1.00), respectively.

SLAP II–IV lesions. Accuracy for diagnosing SLAP II–IV lesions was assessed for 8 OSTs (Fig. S3i). The Sn and Sp for each test were, respectively: from 0.47 to 0.65 and from 0.38 to 0.92 for the active compression test, (22, 25, 27, 31, 38–40), from 0.04 to 0.70 and from 0.69 to 0.98 for anterior slide test (22, 27, 31, 38–40), from 0.29 to 0.90 and from 0.78 to 0.97 for biceps load II test (31, 46), from 0.09 to 0.83 and from 0.42 to 1.00 for crank test (22, 26, 27, 39), from 0.25 to 0.26 and from 0.65 to 0.80 for palpation test (27, 31), 0.89 and 0.82 for passive compression test (45), 0.52 and 0.94 for passive distraction test (40), and from 0.04 to 0.48 and from 0.65 to 1.00 for Speed test (27, 31, 39).

Data were pooled from studies assessing the active compression test, the anterior slide test and the crank test (Table V, Fig. 4). The results indicate a widely variable performance for the 3 tests. The pooled Sn and Sp for the active compression test are 0.59 (95% CI 0.19–0.96) and 0.57 (95% CI 0.18–0.96), respectively, for the anterior slide test 0.21 (95% CI 0.00–0.79) and 0.88 (95% CI 0.35–1.00), respectively, and for the crank test 0.49 (95% CI 0.02–1.00) and 0.70 (95% CI 0.06–1.00), respectively.

Tendinopathy. Accuracy for diagnosing LHBT tendinopathy was assessed for 3 OSTs, and HRUS was the reference standard. The Sn and Sp estimates from each study are shown in forest plots (Fig. S4i). The Sn and Sp were for each test, respectively: from 0.57 to 0.85 and from 0.49 to 0.72 for the palpation test, (30, 47), from 0.47 to 0.83 and from 0.36 to 0.75 for Speed test (47–49), and from 0.32 to 0.86 and from 0.74 to 0.82 for Yergason’s manoeuvre (47–49).

<table>
<thead>
<tr>
<th>Tests</th>
<th>Studies, n</th>
<th>Shoulders, n</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>LR+</th>
<th>LR–</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed test</td>
<td>4</td>
<td>380</td>
<td>0.36 (0.00–0.82)</td>
<td>0.71 (0.23–1.00)</td>
<td>1.24</td>
<td>0.90</td>
</tr>
<tr>
<td>Active compression test</td>
<td>7</td>
<td>1,151</td>
<td>0.59 (0.19–0.96)</td>
<td>0.57 (0.18–0.96)</td>
<td>1.37</td>
<td>0.72</td>
</tr>
<tr>
<td>Anterior slide test</td>
<td>6</td>
<td>1,060</td>
<td>0.21 (0.00–0.79)</td>
<td>0.88 (0.35–1.00)</td>
<td>1.75</td>
<td>0.90</td>
</tr>
<tr>
<td>Crank test</td>
<td>4</td>
<td>248</td>
<td>0.49 (0.02–1.00)</td>
<td>0.70 (0.06–1.00)</td>
<td>1.63</td>
<td>0.73</td>
</tr>
<tr>
<td>Any LHBT pathology except SLAP lesion</td>
<td>7</td>
<td>1,542</td>
<td>0.65 (0.17–1.00)</td>
<td>0.61 (0.15–1.00)</td>
<td>1.67</td>
<td>0.57</td>
</tr>
<tr>
<td>Yergason’s manoeuvre</td>
<td>5</td>
<td>559</td>
<td>0.41 (0.14–0.72)</td>
<td>0.84 (0.65–1.00)</td>
<td>2.56</td>
<td>0.70</td>
</tr>
</tbody>
</table>

SLAP: superior labrum anterior and posterior; 95% CI: 95% confidence interval; LR+: positive likelihood ratio; LR–: negative likelihood ratio.

Fig. 3. Hierarchical summary receiver operating characteristic (ROC) curve examining the diagnostic value of high-resolution ultrasound (HRUS) for characterization of long head of the biceps tendon (LHBT) (A) dislocation and (B) complete rupture. The 95% prediction region is defined by the blue dotted-curve, while the red dot-dashed-curve marks the boundary of the 95% credible interval of the pooled estimates. Prediction region is defined by pairing the confidence interval with the credible interval.
Any proximal tendon pathology except SLAP lesion.

Accuracy for diagnosing any LHBT pathology except for SLAP lesion was assessed for 5 OSTs. Target conditions included tendinopathy, dislocation, effusion, and rupture. Reference standard varied across studies, including either surgery or HRUS. Sn and Sp estimates from each study are shown in forest plots (Fig. S5). Sn and Sp for each test were, respectively, 0.01 to 1.00 for Heuter’s sign (49), from 0.53 to 0.85 and from 0.49 to 0.72 for palpation test (29, 30, 47), from 0.47 to 0.93 and from 0.27 to 0.81 for Speed test (28, 29, 37, 47–50), 0.72 and 0.78 for upper cut test (37), and
from 0.32 to 0.86 and from 0.78 to 0.88 for Yergason’s manoeuvre (37, 47–49, 51).

Data from studies assessing Speed test and Yergason’s manoeuvre were pooled (Table V, Fig. 4). The results indicate a widely variable performance for the 2 tests, except for Yergason’s manoeuvre. Sn and Sp for the Speed test are 0.65 (95% CI 0.17–1.00) and 0.61 (95% CI 0.15–1.00) and for Yergason’s manoeuvre 0.41 (95% CI 0.14–0.72) and 0.84 (95% CI 0.65–1.00).

DISCUSSION

We identified 30 studies evaluating the accuracy of HRUS or OSTs in diagnosing LHBT pathologies (Table III). The 8 primary studies on HRUS diagnostic accuracy comprised 5 different combinations of target condition/index test. At most, 6 of the studies examined the same combination. The 22 studies assessing OSTs presented 26 such combinations, and no more than 7 research studies tested the same combination. This lack of consistency across studies and the relatively few studies on the subject are a major barrier to the assessment of these clinical tools.

Potential of the tests to inform diagnoses

For a diagnostic test to be useful, it must have the ability to sufficiently revise the pre-test probability of a patient having a disease in order to guide clinical decisions. HRUS for the diagnosis of dislocation and complete rupture had LR+ above 35.5 and LR– below 0.30, indicating a large increase in the post-test probability of dislocation and complete rupture when diagnostic ultrasound is positive, and a moderate decrease in the probability of these diseases when it is negative (23). It should be noted that estimates of Sn of HRUS for diagnosing dislocation and complete rupture had wide confidence intervals (0.15–1.00 and 0.11–1.00), hence their calculated LR– might overlap the evidence. Confidence intervals were narrower for Sp (0.65–1.00 and 0.61–1.00), thus LR+ are probably informative.

OSTs LR+ and LR– demonstrated less compelling evidence. The only test of value was Yergason’s manoeuvre in diagnosing proximal LHBT pathology except SLAP lesion. Its LR+ was 2.56, indicating a slight increase in the probability of the disease. As its Sp confidence interval was 0.65–1.00, we can assume that it is of reasonable value. OSTs LR– varied between 0.57 and 0.90, all indicating no change in the post-test probability of the disease. The current review separated SLAP I–IV and II–IV lesions as 2 target conditions in order to investigate whether the accuracy of each OST changes when SLAP I lesions are considered normal variants. When explored graphically with forest plots, there is no apparent significant difference between the OSTs’ accuracies in diagnosing SLAP I–IV and SLAP II–IV lesions.

Comparison with other systematic reviews

Eight systematic reviews were identified, of which 4 included a meta-analysis that evaluated the diagnostic accuracy of OSTs for diagnosing SLAP lesions. The 4 systematic reviews that did not include a meta-analysis (6, 7, 9, 12) highlighted that OSTs have a wide range of diagnostic accuracy values, with no particular single test appearing to have strong statistical support. This is in line with our conclusions for the accuracy of OSTs.

Hanchard et al. (9) conducted a Cochrane systematic review on shoulder impingements and local lesions of tendons and labrum that may accompany impingement. Their review comprised several individual studies that were included in our analysis for the accuracy of OSTs. For these analyses, Sn and Sp were obtained in agreement with Hanchard et al.’s study. For these same combinations of index test/target condition, 8 new studies issued after completion of their review were identified and included (22, 24–30). In addition, we classified the target conditions slightly differently. In the current review, we grouped together studies examining the diagnosis of SLAP II–IV and SLAP II lesions (our SLAP II–IV group) while Hanchard et al. kept them separated.

Four previous meta-analyses (8, 10, 11, 13) have reported pooled accuracy estimates for the active compression test, anterior slide test, crank test and Speed test in diagnosing SLAP lesions. Hegedus et al. (10) and Gismervik et al. (8) reviewed the literature on the accuracy of OSTs of the shoulder. For SLAP lesions, there were some discrepancies between the values obtained by these authors and our estimates for the active compression test and Speed test. These discrepancies may arise from the fact that we separated SLAP I–IV from II–IV studies. Our higher Sp for active compression test could suggest that it has a better profile for confirming a SLAP II–IV than a SLAP I–IV lesion. In addition, Gismervik et al. incorporated Holtby & Razmjou’s study (31) when combining data for the Speed test, while we did not. It should be noted that Holtby & Razmjou’s study was not included in our analysis for the combination Speed test/SLAP I–IV lesions because this study evaluates Speed test’s accuracy in diagnosing not only SLAP lesions, but any proximal LHBT pathology including SLAP lesions.

Meserve et al. (11) conducted a meta-analysis examining the accuracy of OSTs for assessing SLAP lesions
(active compression test, anterior slide test, crank test, and Speed test). They found that the anterior slide test was statistically inferior to the 3 other tests; this can be appreciated when looking at their ROC curves. In our review, the curve for the anterior slide test resembles the 3 others. This inconsistency may be explained by the 3 studies included in our analysis that were published after their review (22, 32, 33). After reviewing the literature on the same research question, Walton et al. (13) performed a meta-analysis for the OSTs that have been evaluated at least 3 times in the literature. They provided estimates of the pooled LR+ for, among others, the active compression test (1.07), crank test (1.51), and Speed test (1.12). Our pooled LR+ estimates were 1.37 for the active compression test, 1.63 for the crank test, and 1.24 for the Speed test. Our values are slightly higher for the active compression test because we included 3 studies that have been published after their work (26, 34, 35). Also, for the Speed test, they incorporated Holtby & Razmjou’s (31) as well as Bennet’s (36) studies in their analysis, which evaluates not only SLAP lesions, but any LHB pathology.

Strengths and weaknesses of the review

Strengths. First, this systematic review was based on a rigorous search of the literature, which resulted in the inclusion of 30 articles. Secondly, a recommended appraisal tool was used to determine the risk of bias of included studies. In addition, the statistics presented in the included studies were double-checked by back-calculating 2×2 tables. Where we observed discrepancy between text and tables, or when values presented had arithmetical errors, the study was excluded. Finally, judicious use was made of meta-analysis’ tools: they were used when there was a minimum of 4 primary studies identified, as suggested by Sotiriadis et al. (20).

Weaknesses. In our protocol design, we chose to exclude non-English or French studies, which may have led to selection bias. There was one study in Persian and one in Turkish languages that could have been eligible. We also recognize the possibility of information bias in the studies included. More specifically, as appraised with QUADAS-2 instrument, there is a possibility of misclassification due to spontaneous recovery or progression of disease. Of the 30 included studies, 9 had an inadequate interval between index test and reference standard. In the same vein, misclassification in the primary studies due to inaccurate reference standard is another possibility to consider. It was “unclear” if the reference standard was likely to correctly classify the target condition in 8 of the 30 included studies. For instance, in order to assess the accuracy of OSTs in diagnosing tendinopathy, HRUS was the reference standard in the only individual studies identified in the literature (Fig. S4). Since the role of ultrasound in the diagnosis of biceps tendinopathy is still poorly understood, this area of uncertainty would need to be addressed before a more definitive conclusion can be drawn (2).

Applicability of findings to the review question

From the findings of this systematic review, HRUS had variable Sn and thus would be of lower interest as a screening test. Nevertheless, it can be considered a highly specific clinical tool for the diagnosis of dislocation, rupture and tendinopathy of the LHB; it can be useful in ruling-in disease. Besides its effectiveness, HRUS has several advantages over other imaging modalities: there is no contraindication, it has high spatial resolution, dynamic assessment is possible as well as correlation of findings with patients’ symptoms. Furthermore, it has been shown to be cost-effective in specific situations, such as in the context of rotator cuff disease (37), and proved to be a reliable method for the measurement of the LHB in healthy shoulders (38).

With regard to OSTs, the evidence was more limited by the variability of the test accuracies across different study settings. A promising screening test (high Sn) for SLAP II–IV lesions is passive compression test, but the test has been evaluated only by its originators. No other test demonstrated high Sn. For ruling-in specific diagnosis, several tests seem to be valuable. The anterior slide test and biceps load II test had high Sp for diagnosing SLAP I–IV lesions. Passive compression test and passive distraction test were highly specific for SLAP II–IV lesions, but only the test’s originators assessed their accuracies. For LHB tendinopathy, Yergason’s manoeuvre proved highly specific. For proximal LHB pathology except SLAP lesions, Heuter’s sign (one study) and Yergason’s manoeuvre had high Sp.

Whereas no single clinical finding, either OSTs or HRUS, is accurate enough to confirm diagnosis and guide subsequent clinical decisions, it is appealing for clinicians and researchers to improve diagnostic accuracy by clustering clinical information. Furthermore, combining clinical findings more closely reflects how clinicians make decisions in practice. Combining the more sensitive clinical information with the more specific data could be quite helpful in improving our ability to diagnose LHB pathology. Future research on the subject should focus on the development of such clusters.

Conclusion

In order to rule in LHB pathology, HRUS has proven its diagnostic efficacy. However, evidence is lacking to recommend its use for the purpose of ruling out pathology. There is insufficient evidence to recommend individual OSTs. In the future, rigour in diagnostic test
accuracy research is of paramount importance. Researchers should minimize bias by using prospective cohort-type study designs, index test in accordance with the original description, adequate reference standards and adequate interval between index test and reference standard. Finally, investigators should consider improving accuracy by clustering OSTs with or without HRUS and information about current or past medical history (39).

The authors have no conflicts of interest to declare.

REFERENCES


## Appendix I. Description of orthopaedic special tests (OSTs)

<table>
<thead>
<tr>
<th>OST</th>
<th>LHBT injury</th>
<th>Description of OST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active compression test (O'Brien)</td>
<td>SLAP (Type 2)</td>
<td>Patient with shoulder 90° forward flexed, 10–15° adducted and in full internal rotation (palm pronated), elbow fully extended. Examiner applies a downward eccentric force to the arm. The manoeuvre is repeated in the same position except with the shoulder in full external rotation (palm supinated). Test considered positive if pain or clicking is produced inside the shoulder in the first part of the test and eliminated or decreased in the second part.</td>
</tr>
<tr>
<td>Anterior slide test (Kibler)</td>
<td>SLAP</td>
<td>Patient has hands on the waist, thumbs posterior. Examiner stabilizes the scapula and clavicle with one hand. The other applies an anterosuperior force at the elbow. Test considered positive if produces pain or click deep in the shoulder.</td>
</tr>
<tr>
<td>Biceps load II (Kim II)</td>
<td>SLAP</td>
<td>Patient sitting or in supine position, 120° shoulder abduction with lateral rotation, 90° elbow flexion, forearm supinated. Examiner fully externally rotates shoulder. If apprehension appears, position is maintained and patient is asked to flex elbow against resistance. Test considered positive if apprehension increases or stays the same with elbow flexion.</td>
</tr>
<tr>
<td>Speed test</td>
<td>SLAP</td>
<td>Test is considered positive if pain or clicking is induced between 90° and 120° abduction.</td>
</tr>
<tr>
<td>Groove palpation</td>
<td>SLAP, rupture, tendinopathy</td>
<td>Patien...</td>
</tr>
<tr>
<td>Heuter's sign</td>
<td>Rupture</td>
<td>Patien...</td>
</tr>
<tr>
<td>Yergason's manoeuvre</td>
<td>Dislocation, tendinopathy</td>
<td>Patien...</td>
</tr>
</tbody>
</table>

SLAP: superior labrum anterior and posterior; LHBT: long head of the biceps tendon.
### Appendix II. Search strategies for MEDLINE and CINAHL

**Search strategy for MEDLINE**

- "Ultrasonography"[Mesh] OR "ultrasonography"[Subheading] OR "Diagnostic Imaging"[Mesh] OR "Physical Examination"[Mesh]) OR Index tests set
  - "CLINICAL TEST" OR "CLINICAL MEASUREMENT" OR "CLINICAL EXAMINATION" OR "PHYSICAL EXAMINATION") OR
  - (ULTRASOUND* OR ULTRASONOGRAPH*) AND
  - ("SHOULDER"[TIAB] OR "BICEPS"[TIAB]) OR
  - ("SHOULDER"[TIAB] OR "BICEPS"[TIAB] OR TENDON* OR TENDINO*) OR
  - (BICEPS OR LHB OR LHBT) AND
  - ("Diagnosis"[Mesh] OR "diagnosis" [Subheading]) OR
  - ("Diagnostic Test Approval"[Mesh]) OR
  - ("Dimensional Measurement Accuracy"[Mesh]) OR
  - (VALUE* OR ACCURACY OR RELIABILITY OR FEASIBILITY OR SENSITIVITY OR SPECIFICITY OR REPRODUCIBILITY OR EFFICIENCY OR RATIO OR ROC) OR
  - (MEASURE* OR TOOL* OR TEST OR TESTS) OR
  - (DIAGNOSTIC* OR DIAGNOSIS) AND
  - (English[lang] OR French[lang]) Language

**Search strategy for CINAHL**

- (MH "Physical Examination") OR
  - (MH "Ultrasonography") OR
  - ("PHYSICAL EXAMINATION" OR "CLINICAL TEST*" OR "CLINICAL EXAMINATION" OR "CLINICAL MEASURE*") OR
  - (ULTRASOUND* OR ULTRASONIC OR ULTRASONOGRAPHY) AND
  - (MH "Biceps Brachii Muscles") OR (MH "Biceps Tendonitis") OR
  - (BICEPS AND (LONG OR LHB)) AND
  - (MH "Shoulder") OR (MH "Shoulder Dislocation") OR (MH "Shoulder Pain") OR (MH "Shoulder Injuries") OR (MH "Shoulder Joint") AND
  - (MH "Diagnosis") OR (MH "Diagnosis, Musculoskeletal") OR
  - (MH "Reliability and Validity") OR (MH "Interrater Reliability") OR (MH "Test-Retest Reliability") OR (MH "Reproducibility of Results") OR (MH "Sensitivity and Specificity") OR (MH "Predictive Validity") OR (MH "Productivity") OR (MH "Odds Ratio") OR (MH "Validity") OR
  - (RELIABILITY OR VALIDITY OR ACCURACY OR EFFICIENCY OR REPRODUCIBILITY OR RATIO OR VALUE* OR SENSITIVITY OR SPECIFICITY) OR
  - (MH DIAGNOS*) AND
  - (English[lang] OR French[lang]) Language