



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

Valérie BÉLANGER, MD<sup>1</sup>, Frédérique DUPUIS, BSc<sup>2</sup>, Jean LEBLOND, PhD<sup>2</sup> and Jean-Sébastien ROY, PT, PhD<sup>2,3</sup>

From the <sup>1</sup>Department of Physical Medicine and Rehabilitation (Physiatry), Centre Hospitalier Universitaire de Québec – Université Laval, <sup>2</sup>Center for Interdisciplinary Research in Rehabilitation and Social Integration, and <sup>3</sup>Department 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 sources:** MEDLINE, CINAHL and EMBASE.

**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

Shoulder pain is common in the general population (1), and pathology of the long head of the biceps tendon (LHBT) can be a primary source of shoulder pain, either in isolation or in association with other shoulder pathologies, such as rotator cuff diseases

### LAY ABSTRACT

People with shoulder pain seek medical attention in order to relieve their symptoms and improve their quality of life. However, given the complexity of the shoulder girdle, making the right diagnosis can be challenging. Clinicians and other healthcare practitioners base their approach on the findings of current medical history, as well as physical and ultrasound examinations. Once a structure is identified as a potential pain-generator, a specific therapy can be used. The biceps tendon is one such structure. The aim of this study is to assess the accuracy of physical and ultrasound examinations in diagnosing biceps tendon pathologies. This will help to guide clinical decision-making and may prevent delay in seeking specific treatment approaches.

(2, 3). Most described LHBT pathologies include superior labrum anterior and posterior (SLAP) lesions, tendinosis, dislocation and rupture (4). In the clinical setting, orthopaedic special tests (OSTs) and, more recently, high-resolution ultrasound (HRUS) are used for ruling in or out shoulder disorders, such as LHBT pathologies. While numerous OSTs have been proposed to identify the different LHBT pathologies, HRUS can be used to detect LHBT tendinopathy, dislocation, rupture and intra-articular peritendinous effusion. In rare cases, HRUS can directly diagnose insertional pathology, such as SLAP lesions (5). Eight systematic reviews have been published on the diagnostic accuracy of OSTs for a wide spectrum of shoulder disorders, including LHBT pathologies, most of which were SLAP lesions. The conclusions were that OSTs are neither very specific nor sensitive in diagnosing SLAP lesions (6–13). However, new high-quality diagnostic accuracy studies for OSTs have been conducted in the past few years, and could therefore change the conclusions of these previous systematic reviews. In addition, no systematic review has focused on the accuracy of HRUS in diagnosing LHBT pathologies. To our knowledge, no systematic review has been carried out specifically addressing the diagnosis of LHBT pathologies in clinical practice, including the accuracy of both OSTs and HRUS examinations. A better picture of the current accuracy of clinicians in assessing the LHBT will enable a better selection of diagnostic tools for the clinical evaluation of shoulder pain.

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.

### Search methods for identification of studies

MEDLINE, CINAHL and EMBASE databases were searched for eligible articles from their inception dates to July 2018. Articles had to be written in French or English. The full search strategy is described in Appendix II. The reference lists for every article found in the original electronic search were screened to identify further eligible articles.

### Data collection and analysis

**Selection of studies.** Two review authors independently selected the studies. In case of disagreement, a third author was involved to reach consensus. Articles were selected if they met the selection criteria for population, index test, reference standard, and reported on the diagnostic accuracy of individual index tests for diagnosing a specific LHBT pathology (SLAP lesions, LHBT tendinopathy, dislocation, rupture or effusion). We started with a review of titles, proceeded to abstracts where titles indicated possibly relevant studies, and selected eligible studies after reading their full text.

**Data extraction and management.** Data were extracted by 2 independent authors. If any disagreement occurred during this step, a third reviewer intervened to reach mutual agreement. The extraction decision was based on the possibility of drawing a 2 × 2 table. If the tables were not included in the article, data allowing reconstruction was necessary. If there was any discrepancy between text and tables, articles were removed from analysis unless original authors could be contacted to resolve the issue.

**Quality assessment.** The risk of bias of each study was assessed using Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) by the same 2 independent authors who selected the studies and extracted the data (17). This tool is designed to appraise studies' selection bias and information bias by assessing 4 key domains: patient selection; index test; reference standard; flow of patients through the study and timing of the index test(s) and reference standard. Results are expressed in terms of the methodological quality "high", "low" or "unclear", based on the author's judgement. Authors of reviews are encouraged to tailor QUADAS-2 to their review by developing review-specific guidance on how to assess each signalling question (17). In that respect, after consensus among authors, specific criteria were used for each section (Table I). Gwet's first-order agreement coefficient (Gwet's AC1) was used to calculate interobserver agreement (18).

### Statistical analysis and data synthesis

A systematic review should not culminate in meta-analysis if there are differences between the studies in terms of the

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

QUADAS-2 item	Specifications
Patient selection	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.
Index tests	Had to be described, or at least, referenced. If neither of those, we attributed a high risk of bias.
Reference standards	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. If the reference standard used in the study was not in this list, it was considered an unclear risk of bias.
Flow and timing	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.

LHBT: long head of the biceps tendon; SLAP: superior labrum anterior and posterior; MR: magnetic resonance; HRUS: high-resolution ultrasound.

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

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

HRUS: high-resolution ultrasound; OSTs: orthopaedic special tests; MR: magnetic resonance; SLAP: superior labrum anterior and posterior; LHBT: long head of the biceps tendon.

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.

Studies with cells containing zero in the 2x2 table lead to statistical model instabilities. A continuity correction, consisting of a small positive number (0.5 as suggested in the literature) was then added to the observed frequency (20).

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.

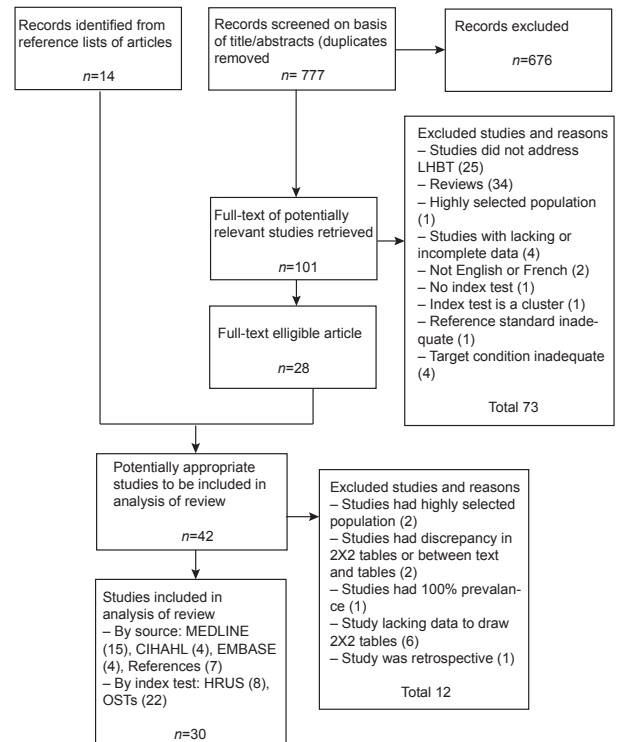
**RESULTS**

*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

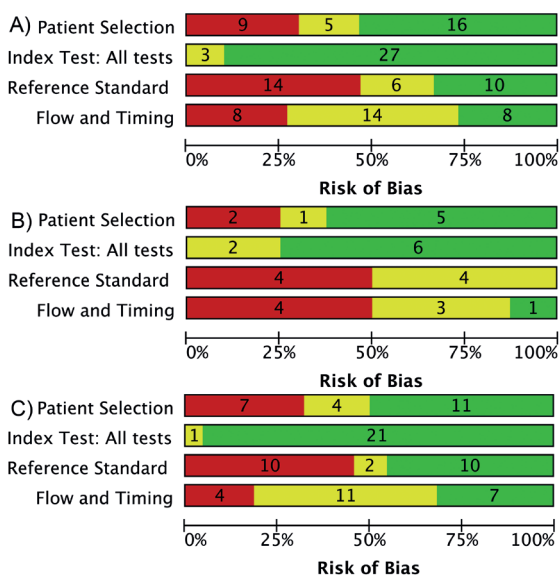


**Fig. 1.** Flow diagram of the bibliographic search. HRUS: high-resolution ultrasound; OSTs: orthopaedic special tests.

**Table III.** Summary of included studies

Study	n of shoulders	Study design	Index test	Reference standard	Target condition
<i>HRUS accuracy</i>					
Armstrong, 2006 (23)	71	Prospective	HRUS	Arthroscopy	Dislocation
Farin, 1995 (44)	266	Prospective	HRUS	Open surgery	Dislocation
Fischer, 2015 (24)	45	Prospective	HRUS	MRI	Dislocation Rupture (complete)
Moosmayer, 2005 (32)	75	Prospective	HRUS	Arthroscopy or open surgery	Dislocation Rupture (partial) Rupture (total)
Naredo, 1999 (43)	36	Prospective	HRUS	MRI	Effusion Dislocation Tendinopathy
Read, 1998 (33)	42	Prospective	HRUS	Arthroscopy	Tendinopathy Dislocation Rupture (total)
Skendzel, 2011 (34)	67	Prospective	HRUS	Arthroscopy	Tendinopathy Rupture (partial) Rupture (total)
Teefey, 2000 (42)	100	Prospective	HRUS	Arthroscopy	Dislocation Rupture (total)
<i>OSTs accuracy</i>					
Arrigoni, 2014 (28)	109	Prospective	ACT Speed	Arthroscopy	Any but SLAP lesion
Chen, 2011 (47)	143	Prospective	Palpation Speed Yergason	HRUS	Tendinopathy
Cook, 2012 (35)	87	Prospective	ACT BLII DLST LTT Speed	Arthroscopy	SLAP I-IV
Fowler, 2010 (25)	101	Prospective	ACT	Arthroscopy	SLAP II-IV
Gill, 2007 (29)	847	Prospective	Palpation Speed	Arthroscopy	Rupture (partial)
Guanche, 2003 (36)	60	Prospective	Crank Palpation Speed	Arthroscopy	SLAP I-IV
Kibler, 2009 (37)	101	Prospective	ACT AST DLST Speed UCT	Arthroscopy	SLAP I-IV
Kim HA, 2007 (51)	176	Prospective	Speed UCT Yergason		Any but SLAP lesions
Kim SH, 2001 (46)	127	Prospective	Speed Yergason	HRUS	Any but SLAP lesions
Kim YS, 2007 (45)	61	Prospective	BLII	Arthroscopy	SLAP II-IV
Lasbleiz, 2014 (48)	39	Prospective	PCT	Arthroscopy	SLAP I-IV SLAP II-IV
McFarland, 2002 (38)	409	Prospective	Speed Yergason	HRUS	Tendinopathy Any but SLAP lesions
Michener, 2011 (22)	419	Prospective	ACT	Arthroscopy	SLAP II-IV
Micheroli, 2013 (49)	55	Prospective	AST Crank	Arthroscopy	SLAP II-IV
Mimori, 1999 (26)	100	Prospective	Speed Yergason	HRUS	Tendinopathy Any but SLAP lesion
Nakawaga, 2005 (27)	7	Prospective	Heuter	Arthroscopy	Any but SLAP lesion
Oh, 2008 (31)	15	Prospective	Crank Mimori	Arthroscopy	SLAP II-IV
Parentis, 2006 (39)	54	Prospective	ACT AST Crank Palpation Speed	Arthroscopy	SLAP II-IV
Salaffi, 2010 (50)	146	Prospective	ACT AST BLII Palpation Speed	Arthroscopy	SLAP I-IV SLAP II-IV
Schlechter, 2009 (40)	132	Prospective	ACT AST Crank Speed	Arthroscopy	SLAP I-IV SLAP II-IV
Sodha, 2017 (41)	203	Prospective	Speed	HRUS	Any but SLAP
Toprak, 2013 (30)	254	Prospective	ACT AST PDT	Arthroscopy	SLAP II-IV
	774	Prospective	DLST	Arthroscopy	SLAP I-IV
	69	Prospective	Palpation	HRUS	Tendinopathy Any but SLAP lesion

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.



**Fig. 2.** Methodological quality graph for accuracy studies: (A) all, (B) high-resolution ultrasound (HRUS), and (C) orthopaedic special tests (OSTs). Graphs show the percentage and number of studies with a high (red), low (green) and unclear (yellow) risk of bias for the 4 items.

studies 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. S1<sup>1</sup>).

**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. S1<sup>1</sup>). 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. S1<sup>1</sup>).

**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. S1<sup>1</sup>).

**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. S1<sup>1</sup>). 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**

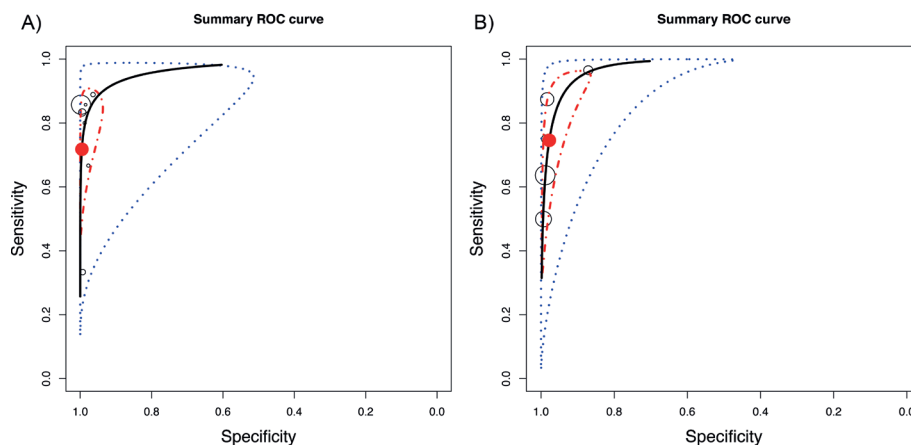
**SLAP I–IV lesions.** Accuracy for diagnosing SLAP I–IV lesions was assessed for 10 OSTs (Fig. S2<sup>1</sup>). The

<sup>1</sup><https://doi.org/10.2340/16501977-2564>

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

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

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.

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 0.10 to 0.48 and 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 0.86 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. S3<sup>1</sup>). 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).

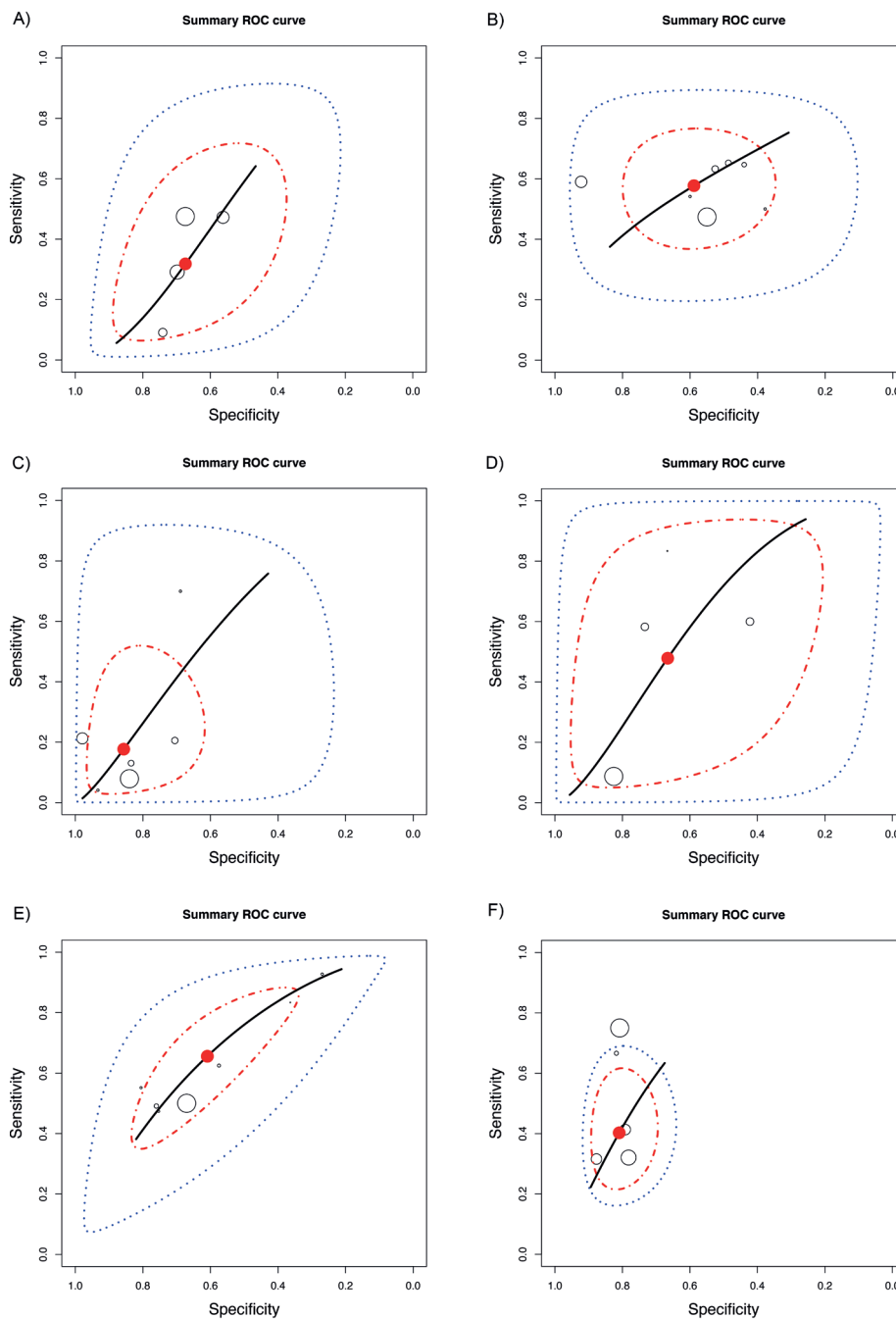
Data were pooled from studies assessing the active compression test, the anterior slide test and the crank test (Table V, Fig. 4). The results indicate a widely variable performance for the 3 tests. The pooled Sn and Sp for the active compression test are 0.59 (95% CI 0.19–0.96) and 0.57 (95% CI 0.18–0.96), 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. S4<sup>1</sup>). 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 V.** Overall orthopaedic special tests' accuracy in characterization of long head of the biceps tendon (LHBT) pathology

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

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



**Fig. 4.** Hierarchical summary receiver operating characteristic (ROC) curve examining the diagnostic value of the Speed test for characterization of: (A) superior labrum anterior and posterior (SLAP) I-IV lesions, (B) active compression test for characterization of SLAP II-IV lesions, (C) anterior slide test for characterization of SLAP II-IV lesions, (D) crank test for characterization of SLAP II-IV lesions, (E) Speed test for characterization of any long head of the biceps tendon (LHBT) pathology, but SLAP lesion, and (F) Yergason's manoeuvre in characterization of any pathology but SLAP lesion. 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<sup>1</sup>). 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 Sp. 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 overlay 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 LHBT 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<sup>1</sup>). 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 LHBT; 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 LHBT 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 LHBT tendinopathy, Yergason's manoeuvre proved highly specific. For proximal LHBT 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 decision in practice. Combining the more sensitive clinical information with the more specific data could be quite helpful in improving our ability to diagnose LHBT pathology. Future research on the subject should focus on the development of such clusters.

#### *Conclusion*

In order to rule in LHBT 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

- Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004; 33: 73–81.
- Nho SJ, Strauss EJ, Lenart BA, Provencher MT, Mazzocca AD, Verma NN, et al. Long head of the biceps tendinopathy: diagnosis and management. *J Am Acad Orthop Surg* 2010; 18: 645–656.
- Redondo-Alonso L, Chamorro-Moriana G, Jimenez-Rejano J, Lopez-Tarrida P, Ridao-Fernandez C. Relationship between chronic pathologies of the supraspinatus tendon and the long head of the biceps tendon: systematic review. *BMC Musculoskelet Disord* 2014; 15: 377.
- Sarmento M. Long head of biceps: from anatomy to treatment. *Acta Reumatol Port* 2015; 40: 26–33.
- Brasseur JL. The biceps tendons: From the top and from the bottom. *J Ultrasound* 2012; 15: 29–38.
- Calvert E, Chambers GK, Regan W, Hawkins RH, Leith JM. Special physical examination tests for superior labrum anterior posterior shoulder tears are clinically limited and invalid: a diagnostic systematic review. *J Clin Epidemiol* 2009; 62: 558–563.
- Dessaur WA, Magarey ME. Diagnostic accuracy of clinical tests for superior labral anterior posterior lesions: a systematic review. *J Orthop Sports Phys Ther* 2008; 38: 341–352.
- Gismervik SO, Drogset JO, Granviken F, Ro M, Leivseth G. Physical examination tests of the shoulder: a systematic review and meta-analysis of diagnostic test performance. *BMC Musculoskelet Disord* 2017; 18: 41.
- Hanchard NC, Lenza M, Handoll HH, Takwoingi Y. Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. *Cochrane Database Syst Rev* [cited 2013 Apr 30]. Available from: <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD007427>.
- Hegedus EJ, Goode AP, Cook CE, Michener L, Myer CA, Myer DM, et al. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. *Br J Sports Med* 2012; 46: 964–978.
- Meserve BB, Cleland JA, Boucher TR. A meta-analysis examining clinical test utility for assessing superior labral anterior posterior lesions. *Am J Sports Med* 2009; 37: 2252–2258.
- Sandrey MA. Special physical examination tests for superior labrum anterior-posterior shoulder tears: an examination of clinical usefulness. *J Athl Train* 2013; 48: 856–858.
- Walton DM, Sadi J. Identifying SLAP lesions: a meta-analysis of clinical tests and exercise in clinical reasoning. *Phys Ther Sport* 2008; 9: 167–176.
- Magee DJ. *Shoulder*. In: Elsevier, editor. *Orthopedic Physical Assessment*. 6th edn. Edmonton, Canada: Saunders; 2014, p. 266–401.
- Beggs I, Bianchi S, Bueno A, Cohen M, Court-Payen M, Grainger A, et al. Musculoskeletal ultrasound: technical guidelines. *Insights Imaging* 2010; 1: 99–141.
- Jacobson JA. Shoulder US: anatomy, technique, and scanning pitfalls. *Radiology* 2011; 260: 6–16.
- Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529–536.
- Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol* 2013; 13: 61.
- Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0.0 2010. Available from: Available from: <http://srdta.cochrane.org/>.
- Sotiriadis A, Papatheodorou SI, Martins WP. Synthesizing Evidence from Diagnostic Accuracy Tests: the SEDATE guideline. *Ultrasound Obstet Gynecol* 2015; 47: 386–395.
- Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; 20: 2865–2884.
- Michener LA, Doukas WC, Murphy KP, Walsworth MK. Diagnostic accuracy of history and physical examination of superior labrum anterior-posterior lesions. *J Athl Train* 2011; 46: 343–348.
- Armstrong A, Teefey SA, Wu T, Clark AM, Middleton WD, Yamaguchi K, et al. The efficacy of ultrasound in the diagnosis of long head of the biceps tendon pathology. *J Shoulder Elbow Surg* 2006; 15: 7–11.
- Fischer CA, Weber MA, Neubecker C, Bruckner T, Tanner M, Zeifang F. Ultrasound vs. MRI in the assessment of rotator cuff structure prior to shoulder arthroplasty. *Journal of Orthopaedics* 2015; 12: 23–30.
- Fowler EM, Horsley IG, Rolf CG. Clinical and arthroscopic findings in recreationally active patients. *Sports Med Arthrosc Rehabil Ther Technol* 2010; 2: 2.
- Mimori K, Muneta T, Nakagawa T, Shinomiya K. A new pain provocation test for superior labral tears of the shoulder. *Am J Sports Med* 1999; 27: 137–142.
- Nakagawa S, Yoneda M, Hayashida K, Obata M, Fukushima S, Miyazaki Y. Forced shoulder abduction and elbow flexion test: a new simple clinical test to detect superior labral injury in the throwing shoulder. *Arthroscopy* 2005; 21: 1290–1295.
- Arrigoni P, Ragone V, D'Ambrosi R, Denard P, Randelli F, Banfi G, et al. Improving the accuracy of the preoperative diagnosis of long head of the biceps pathology: the biceps resisted flexion test. *Joints* 2014; 2: 54–58.
- Gill HS, El Rassi G, Bahk MS, Castillo RC, McFarland EG. Physical examination for partial tears of the biceps tendon. *Am J Sports Med* 2007; 35: 1334–1340.
- Toprak U, Ustuner E, Ozer D, Uyanik S, Baltaci G, Sakizlioglu SS, et al. Palpation tests versus impingement tests in Neer stage I and II subacromial impingement syndrome. *Knee Surg Sports Traumatol Arthrosc* 2013; 21: 424–429.
- Oh JH, Kim JY, Kim WS, Gong HS, Lee JH. The evaluation of various physical examinations for the diagnosis of type II superior labrum anterior and posterior lesion. *Am J Sports Med* 2008; 36: 353–359.
- Moosmayer S, Smith HJ. Diagnostic ultrasound of the shoulder—a method for experts only? Results from an orthopedic surgeon with relative inexpensive compared to operative findings. *Acta Orthop* 2005; 76: 503–508.
- Read JW, Perko M. Shoulder ultrasound: diagnostic accuracy for impingement syndrome, rotator cuff tear, and biceps tendon pathology. *J Shoulder Elbow Surg* 1998; 7: 264–271.
- Skendzel JG, Jacobson JA, Carpenter JE, Miller BS. Long head of biceps brachii tendon evaluation: accuracy of preoperative ultrasound. *AJR Am J Roentgenol* 2011;

- 197: 942–948.
35. Cook C, Beatty S, Kissenberth MJ, Siffri P, Pill SG, Hawkins RJ. Diagnostic accuracy of five orthopedic clinical tests for diagnosis of superior labrum anterior posterior (SLAP) lesions. *J Shoulder Elbow Surg* 2012; 21: 13–22.
  36. Guanache CA, Jones DC. Clinical testing for tears of the glenoid labrum. *Arthroscopy* 2003; 19: 517–523.
  37. Ben Kibler W, Sciascia AD, Hester P, Dome D, Jacobs C. Clinical utility of traditional and new tests in the diagnosis of biceps tendon injuries and superior labrum anterior and posterior lesions in the shoulder. *Am J Sports Med* 2009; 37: 1840–1847.
  38. McFarland EG, Kim TK, Savino RM. Clinical assessment of three common tests for superior labral anterior-posterior lesions. *Am J Sports Med* 2002; 30: 810–815.
  39. Parentis MA, Glousman RE, Mohr KS. An evaluation of the provocative tests for superior labral anterior posterior lesions. *Am J Sports Med* 2006; 34: 265–268.
  40. Schlechter JA, Summa S, Rubin BD. The passive distraction test: a new diagnostic aid for clinically significant superior labral pathology. *Arthroscopy* 2009; 25: 1374–1379.
  41. Sodha S, Srikumaran U, Choi K, Borade AU, McFarland EG. Clinical Assessment of the Dynamic Labral Shear Test for Superior Labrum Anterior and Posterior Lesions. *Am J Sports Med* 2017; 45: 775–781.
  42. Teefey SA, Hasan SA, Middleton WD, Patel M, Wright RW, Yamaguchi K. Ultrasonography of the rotator cuff: a comparison of ultrasonographic and arthroscopic findings in one hundred consecutive cases. *J Bone Joint Surg Am* 2000; 82: 498–504.
  43. Naredo AE, Aguado P, Padrón M, Bernad M, Uson J, Mayordomo L, et al. A comparative study of ultrasonography with magnetic resonance imaging in patients with painful shoulder. *J Clin Rheumatol* 1999; 5: 184–192.
  44. Farin PU, Jaroma H, Harju A, Soimakallio S. Medial displacement of the biceps brachii tendon: Evaluation with dynamic sonography during maximal external shoulder rotation. *Radiology* 1995; 195: 845–848.
  45. Kim Y, Kim J, Ha K, Choy S, Joo M, Chung Y. The passive compression test: a new clinical test for superior labral tears of the shoulder. *Am J Sports Med* 2007; 35: 1489–1494.
  46. Kim SH, Ha KI, Ahn JH, Kim SH, Choi HJ. Biceps load test II: A clinical test for SLAP lesions of the shoulder. *Arthroscopy* 2001; 17: 160–164.
  47. Chen HS, Lin SH, Hsu YH, Chen SC, Kang JH. A comparison of physical examinations with musculoskeletal ultrasound in the diagnosis of biceps long head tendinitis. *Ultrasound Med Biol* 2011; 37: 1392–1398.
  48. Lasbleiz S, Quintero N, Ea K, Petrover D, Aout M, Laredo JD, et al. Diagnostic value of clinical tests for degenerative rotator cuff disease in medical practice. *Ann Phys Rehabil Med* 2014; 57: 228–243.
  49. Micheroli R, Kyburz D, Ciurea A, Dubs B, Toniolo M, Bisig S, et al. Correlation of findings in clinical and high resolution ultrasonography examinations of the painful shoulder. *Arthritis Rheum* 2013; 65: S50–S51.
  50. Salaffi F, Ciapetti A, Carotti M, Gasparini S, Filippucci E, Grassi W. Clinical value of single versus composite provocative clinical tests in the assessment of painful shoulder. *J Clin Rheumatol* 2010; 16: 105–108.
  51. Kim HA, Kim SH, Seo YI. Ultrasonographic findings of painful shoulders and correlation between physical examination and ultrasonographic rotator cuff tear. *Mod Rheumatol* 2007; 17: 213–219.
  52. McGee S. Simplifying likelihood ratios. *J Gen Intern Med* 2002; 17: 647–650.
  53. Holtby R, Razmjou H. Accuracy of the Speed's and Yergason's tests in detecting biceps pathology and SLAP lesions: comparison with arthroscopic findings. *Arthroscopy* 2004; 20: 231–236.
  54. Bennett WF. Specificity of the Speed's test: arthroscopic technique for evaluating the biceps tendon at the level of the bicipital groove. *Arthroscopy* 1998; 14: 789–796.
  55. Bureau NJ, Ziegler D. Economics of Musculoskeletal Ultrasound. *Current Radiol Reports* 2016; 4: 44.
  56. Drolet P, Martineau A, Lacroix R, Roy JS. Reliability of ultrasound evaluation of the long head of the biceps tendon. *J Rehabil Med* 2016; 48: 554–558.
  57. Hegedus EJ, Cook C, Lewis J, Wright A, Park JY. Combining orthopedic special tests to improve diagnosis of shoulder pathology. *Phys Ther Sport* 2015; 16: 87–92.
  58. Somerville L, Bryant D, Willits K, Johnson A. Protocol for determining the diagnostic validity of physical examination maneuvers for shoulder pathology. *BMC Musculoskelet Disord* 2013; 14: 60.

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

OST	LHBT injury	Description of OST
Active compression test (O'Brien)	SLAP (Type 2)	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.
Anterior slide test (Kibler)	SLAP	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.
Biceps load II (Kim II)	SLAP	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.
Crank test	SLAP	Patient in supine position, arm elevated to 160° in scapular plane, elbow flexed. Examiner applies an axial load to humerus with one hand while the other hand rotates the humerus medially and laterally. Test considered positive with pain on rotation.
Dynamic labral shear test (O'Driscoll)	SLAP	Patient in supine or sitting position, arm at the side and unsupported, elbow flexed to 90°. Examiner fully laterally rotates the arm with 90° abduction in the scapular plane. Next, examiner abducts the arm to 120°, takes the arm into maximum horizontal abduction. Using his/her other hand, he/she applies an anterior shear load to the joint. Test considered positive if pain or click is induced between 90° and 120° abduction.
Forced shoulder abduction test	SLAP	Patient seated. Examiner passively abducts the patient's shoulder fully with elbow in full extension. Next, keeping the same shoulder position, elbow is flexed. Test considered positive if pain is greater with elbow extended than flexed.
Labral tension test	SLAP	Patient in supine position, 120° arm abduction, neutral forearm. Examiner holds the patient's hand and resists patient's supination. Test considered positive if pain increases during supination.
Mimori test	SLAP	Patient positioned with arm abducted 90–100°. Examiner laterally rotates the arm by holding the wrist, taking the forearm in full supination and then full pronation. Test is considered positive if pain is more severe in pronated position or provoked only in that position.
Passive compression test	SLAP	Patient in side-lying position, the tested arm uppermost. Examiner stabilizes the shoulder over the scapula and clavicle with 1 hand, while the other hand holds the arm in 30° abduction at the elbow. Passively, patient's shoulder is laterally rotated and arm pushed proximally and extended by the examiner's hand on the elbow. Test is considered positive if pain or click is felt in the glenohumeral joint.
Passive distraction test	SLAP	Patient lies supine, arm abducted to 150°, elbow extended and forearm supinated. Examiner stabilizes the humerus while pronating the forearm. Test is considered positive if pain is felt deep in the shoulder.
Gilchrest's sign	Tendinopathy	Patient standing with a 2–3 kg weight over the head, shoulder in full external rotation and arm straight. Patient lowers the arm to the side in the coronal plane. Test is considered positive when it elicits the usual patient discomfort or pain in the bicipital groove.
Groove palpation	SLAP, rupture, tendinopathy	Patient positioned with 20–30° shoulder abduction, 90° elbow flexion and full forearm supination. Examiner holds the forearm with 1 hand while palpating the shoulder with the other. Acromion is first palpated and then the greater tuberosity of the humerus (38). Test is considered positive if it creates or exacerbates the patient's usual pain.
Heuter's sign	Rupture	Patient positioned with neutral shoulder, 90° elbow flexion and forearm pronation. Examiner resists elbow flexion. Test is considered positive if no supination occurs.
Speed test	SLAP (Type 2), rupture, tendinopathy	Patient positioned with arm forward flexed 90°, first forearm supinated, then pronated. Examiner resists an eccentric movement sequentially in the 2 positions. Test is considered positive if it increases or creates the patient's usual tenderness in the bicipital groove, especially with the arm supinated.
Upper cut test	Any lesion	Patient positioned with shoulder in neutral by the side, elbow flexed to 90°, forearm supinated and hand in a fist. Examiner puts a hand over the fist to resist the patient while the patients tries to bring quickly the hand up and toward the chin. Test is considered positive if pain or a painful pop is elicited over the anterior shoulder.
Yergason's manoeuvre	Dislocation, tendinopathy	Patient positioned with elbow flexed to 90° and stabilized against the thorax, forearm pronated. With 1 hand, the examiner resists supination while the patient also laterally rotates the arm against resistance. The other hand palpates the bicipital groove. Test is considered positive if the tendon is felt to "pop out" of the groove (dislocation) or induces patient's usual pain (tendinopathy).

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 ULTRASONOGR*) <b>AND</b>	
("Shoulder"[Mesh] OR "Shoulder Joint"[Mesh] OR "Shoulder Dislocation"[Mesh] OR "Tendinopathy"[Mesh] OR "Tendon Injuries"[Mesh] OR BICEPS) OR ("SHOULDER"[TIAB] OR "BICEPS"[TIAB]) OR ("SHOULDER"[TIAB] OR "BICEPS"[TIAB] OR TENDON* OR TENDINO*) OR (BICEPS OR LHB OR LHBT) <b>AND</b>	Target condition set
("Diagnosis"[Mesh] OR "diagnosis" [Subheading]) OR ("Reproducibility of Results"[Mesh] OR "Sensitivity and Specificity"[Mesh] OR "Predictive Value of Tests"[Mesh] OR "Efficiency"[Mesh]) OR ("Diagnostic Test Approval"[Mesh]) OR ("Dimensional Measurement Accuracy"[Mesh]) OR (PREDICTIVE OR RELIABILITY OR FEASIBILITY OR SENSITIVITY OR SPECIFICITY OR REPRODUCIBILITY OR EFFICIENCY OR RATIO OR ROC) OR (VALUE* OR ACCURACY OR RELIABILITY OR SENSITIVITY OR SPECIFICITY OR VALID* OR VALUE* OR RATIO) OR (MEASURE* OR TOOL* OR TEST OR TESTS) OR (DIAGNOSTIC* OR DIAGNOSIS) <b>AND</b>	Accuracy description set
(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) <b>AND</b>	Index tests set
(MH "Biceps Brachii Muscles") OR (MH "Biceps Tendonitis") OR (BICEPS) OR (BICEPS AND (LONG OR LHB)) <b>AND</b>	Target condition set
(MH "Shoulder") OR (MH "Shoulder Dislocation") OR (MH "Shoulder Pain") OR (MH "Shoulder Injuries") OR (MH "Shoulder Joint") <b>AND</b>	Patient description set
(MH "Diagnosis") OR (MH "Diagnosis, Musculoskeletal")	Accuracy description set
(MH "Reliability") 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 (MW DIAGNO*) <b>AND</b>	
(English[lang] OR French[lang])	Language