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

Valérie BÉLANGER, MD1, Frédérique DUPUIS, BSc2, Jean LEBLOND, PhD3 and Jean-Sébastien ROY, PT, PhD2,3
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 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 (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.

**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.
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
SLAP II–IV lesions were ascertained by the reference standard. The type II–IV group comprised studies either designed
tomatic, type II–IV and type I–IV lesions studies were isolated
plots. Positive (LR+) and negative (LR–) likelihood ratios were
calculated from the overall Sn and Sp. However, confidence and
interval. Heterogeneity was explored graphically using forest
errors. The credible intervals are narrower than the CI. The
interval the uncertainty caused by sampling and measurement
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 es-
imates 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 opera-
tors 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.

For SLAP lesions, because the degenerative fraying of the
SLAP I lesion is often considered a normal variant and asym-
tomatic, 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.

| Index test | Reference standard | LHBT pathology identified |
|------------|--------------------|---------------------------|
| HRUS       | Surgery (open or arthroscopy) | One of: Tendinopathy, Dislocation, Rupture – partial, Rupture – total, Effusion (bicipital recess) |
| OSTs (for each) | Surgery (open or arthroscopy) | One pathology which each OST is designed to detect (see Appendix I): SLAP lesion, Tendinopathy |
| HRUS       | MR imaging/arthrography | 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.

### 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 III. Summary of included studies

| Study                      | n of shoulders | Study design | Index test | Reference standard | Target condition |
|----------------------------|----------------|--------------|------------|--------------------|------------------|
| HRUS accuracy              |                |              |            |                    |                  |
| 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) |
| 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) |
| Teeffy, 2000 (42)          | 100            | Prospective  | HRUS       | Arthroscopy        | Dislocation Rupture (total) |
| OSTs accuracy              |                |              |            |                    |                  |
| 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 Speed Yergason | Arthroscopy SLAP II-IV |
| Gill, 2007 (29)            | 847            | Prospective  | Palpation Speed Yergason | Arthroscopy Rupture (partial) |
| Guanche, 2003 (36)         | 60             | Prospective  | Crank Palpation Speed Yergason | Arthroscopy SLAP I-IV |
| Kibler, 2009 (37)          | 101            | Prospective  | ACT AST DLST Speed UCT | Arthroscopy SLAP I-IV |
| Kim HA, 2007 (51)          | 176            | Prospective  | ACT Speed UCT Yergason Speed Yergason | HRUS | Any but SLAP lesions |
| Kim SH, 2001 (46)          | 127            | Prospective  | BLII Speed Yergason | Arthroscopy SLAP II-IV |
| Kim YS, 2007 (45)          | 61             | Prospective  | PCT Speed Yergason | Arthroscopy SLAP I-IV |
| Lasbleiz, 2014 (48)        | 39             | Prospective  | Speed Yergason | HRUS | Tendinopathy Any but SLAP lesions |
| McFarland, 2002 (38)       | 409            | Prospective  | ACT AST Speed UCT | Arthroscopy SLAP II-IV |
| Michener, 2011 (22)        | 419            | Prospective  | ACT AST Crank Speed Yergason | Arthroscopy SLAP II-IV |
| Micheroli, 2013 (49)       | 100            | Prospective  | Speed Yergason | HRUS | Tendinopathy Any but SLAP lesion |
| Mimori, 1999 (26)          | 7              | Prospective  | Crank Speed Yergason | Arthroscopy SLAP II-IV |
| Nakawaga, 2005 (27)        | 15             | Prospective  | ACT AST Crank Palpation Speed | Arthroscopy SLAP II-IV |
| Oh, 2008 (31)              | 146            | Prospective  | ACT AST BLII Speed Palpation Speed | Arthroscopy SLAP II-IV |
| Parentis, 2006 (39)        | 132            | Prospective  | ACT AST Crank Speed Yergason | Arthroscopy SLAP I-IV |
| Salaffi, 2010 (50)         | 203            | Prospective  | Speed Yergason | HRUS | Any but SLAP |
| Schlechter, 2009 (40)      | 254            | Prospective  | ACT AST Speed PCT | Arthroscopy SLAP II-IV |
| Sodha, 2017 (41)           | 774            | Prospective  | DLST Speed Yergason | Arthroscopy SLAP I-IV |
| Toprak, 2013 (30)          | 69             | Prospective  | Palpation Speed Yergason | 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.
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. 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 diagnosing 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

| Pathology               | $n$ (studies) | $n$ (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.

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 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). The Sn and Sp for each test were, respectively: 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). 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). 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, n | Shoulders, n | 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. 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. S5¹). 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...
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 0.32, indicating a large increase in the post-test 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 overplay 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 0.32 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).

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
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. 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 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

1. 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.

2. 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.

3. 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.

4. Sarmiento M. Long head of biceps: from anatomy to treatment. Acta Reumatol Port 2015; 40: 26–33.

5. Brasseur JL. The biceps tendons: From the top and from the bottom. J Ultrasound 2012; 15: 29–38.

6. 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.

7. 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.

8. Gismervik SO, Drosgot 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.

9. Hanchard NC, Lenza M, Handoll HH, Takwoingi Y. Physical examination for partial tears of the biceps tendon. Cochrane Database Syst Rev [cited 2013 Apr 30]. Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD007427.

10. 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.

11. Reserve 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.

12. 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.

13. 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.

14. Magee DJ. Shoulder. In: Elsevier, editor. Orthopedic Physical Assessment. 6th edn. Edmonton, Canada: Saunders; 2014, p. 266–401.

15. 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.

16. Jacobson JA. Shoulder US: anatomy, technique, and scanning pitfalls. Radiology 2011; 260: 6–16.

17. 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.

18. 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.

19. 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://srdata.cochrane.org/.

20. Sotiriadis A, Papaioannou SI, Martins WP. Synthesizing Evidence from Diagnostic Accuracy TEsts: the SEDATE guideline. Ultrasound Obstet Gynecol 2015; 47: 386–395.

21. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001; 20: 2865–2884.

22. 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.

23. 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.

24. 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.

25. Fowler EM, Horsley IG, Rolf CG. Clinical and arthroscopic findings in recreationally active patients. Sports Med Arthrosc Rehabil Ther Technol 2010; 2: 2.

26. Mimori K, Muneta T, Nakagawa T, Shionomiya K. A new pain provocation test for superior labral tears of the shoulder. Am J Sports Med 1999; 27: 137–142.

27. Nakagawa S, Yongeda 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.

28. Arrigon P, Tagone V, D’Ambrosi R, Devidi 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.

29. 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.

30. Toprak U, Ustuner E, Cizer 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.

31. Oh JH, Kim JY, Kim WS, Gom 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.

32. 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.

33. 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.
35. Cook C, Beaty 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. Guanche 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 labrum 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. Teevey 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 ultrasonound in the diagnosis of biceps long head tendinitis. Ultrasound Med Biol 2011; 37: 1392–1396.

48. Lasbleiz S, Quintero N, Eak A, 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. Micheli 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 Radioly 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. |
| 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 rotatess 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 is greater with elbow extended than flexed. |
| 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 increases during supination. |
| 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 is more severe in pronated position or provoked only in that position. |
| 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 considered positive if pain is felt deep in the shoulder. |
| 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 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 considered positive if it creates or exacerbates the patient’s usual pain. |
| 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 if 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. |
| Speed test                               | SLAP (Type 2), rupture, tendinopathy | Test is considered positive if no supination occurs. 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 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 ULTRASONOG*) AND |
| ("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) AND |
| ("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 |
| ("Diagnosis 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) AND |
| (English[lang] OR French[lang]) |

| 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) 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 |
| (MW DIAGNO*) AND |
| (English[lang] OR French[lang]) |