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

Eric J Hegedus

ABSTRACT

Objective To update our previously published systematic review and meta-analysis by subjecting the literature on shoulder physical examination (ShPE) to careful analysis in order to determine each test’s clinical utility.

Methods This review is an update of previous work, therefore the terms in the Medline and CINAHL search strategies remained the same with the exception that the search was confined to the dates November, 2006 through to February, 2012. The previous study dates were 1966 – October, 2006. Further, the original search was expanded, without date restrictions, to include two new databases: EMBASE and the Cochrane Library. The Quality Assessment of Diagnostic Accuracy Studies, version 2 (QUADAS 2) tool was used to critique the quality of each new paper. Where appropriate, data from the prior review and this review were combined to perform meta-analysis using the updated hierarchical summary receiver operating characteristic and bivariate models.

Results Since the publication of the 2008 review, 32 additional studies were identified and critiqued. For subacromial impingement, the meta-analysis revealed that the pooled sensitivity and specificity for the Neer test was 72% and 60%, respectively, for the Hawkins-Kennedy test was 79% and 59%, respectively, and for the painful arc was 53% and 76%, respectively. Also from the meta-analysis, regarding superior labral anterior to posterior (SLAP) tears, the test with the best sensitivity (52%) was the relocation test; the test with the best specificity (95%) was Yergason’s test; and the test with the best positive likelihood ratio (2.81) was the compression-rotation test. Regarding new (to this series of reviews) ShPE tests, where meta-analysis was not possible because of lack of sufficient studies or heterogeneity between studies, there are some individual tests that warrant further investigation. A highly specific test (specificity >80%, LR+ ≥ 5.0) from a low bias study is the passive distraction test for a SLAP lesion. This test may rule in a SLAP lesion when positive. A sensitive test (sensitivity >80%, LR− ≤ 0.20) of note is the shoulder shrug sign, for stiffness-related disorders (osteoarthritis and adhesive capsulitis) as well as rotator cuff tendinopathy. There are six additional tests with higher sensitivities, specificities, or both but caution is urged since all of these tests have been studied only once and more than one ShPE test (ie, active compression, biceps load II) has been introduced with great diagnostic statistics only to have further research fail to replicate the results of the original authors. The belly-off and modified belly press tests for subscapularis tendinopathy, bony apprehension test for bony instability, olecranon-manubrium percussion test for bony abnormality, passive compression for a SLAP lesion, and the lateral Jobe test for rotator cuff tear give reason for optimism since they demonstrated both high sensitivities and specificities reported in low bias studies. Finally, one additional test was studied in two separate papers. The modified dynamic labral shear test, may be diagnostic of labral tears in general, but be sensitive for SLAP lesions specifically.

Conclusion Based on data from the original 2008 review and this update, the use of any single ShPE test to make a pathognomonic diagnosis cannot be unequivocally recommended. There exist some promising tests but their properties must be confirmed in more than one study. Combinations of ShPE tests provide better accuracy, but marginally so. These findings seem to provide support for stressing a comprehensive clinical examination including history and physical examination. However, there is a great need for large, prospective, well-designed studies that examine the diagnostic accuracy of the many aspects of the clinical examination and what combinations of these aspects are useful in differentially diagnosing pathologies of the shoulder.

INTRODUCTION

In 2006, we reviewed shoulder physical examination (ShPE) and in 2008 our work was published in this journal.1 This publication was followed by a series of either similar or otherwise redundant publications, addressing all or dedicated pathognomonic components of shoulder testing.2–7 The majority of those subsequent articles did not meta-analyse the ShPE test’s accuracy, evaluate risk of bias among the studies, or identify studies unique to our 2008 publication.1 The fact that so many review articles analysed the diagnostic accuracy of clinical shoulder tests in a period of three years speaks to the need to clearly address the question. Which physical examination tests provide clinicians with the most value for diagnosis when examining the shoulder?

Since 2006, there have been many changes necessitating an update of the original article. First and foremost, the publication of diagnostic articles on the use of ShPE tests in the clinical examination has continued at a brisk pace resulting in numerous new publications on the accuracy of established
tests and the development of new tests. Next, the methodology by which a systematic review on diagnostic accuracy is conducted has been updated from the Quality of Reporting of Meta-analysis (QUOROM)\(^6\) with the publication of Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA).\(^3\) Third, the criterion standard method of performing a meta-analysis has become a unification\(^10\) of the bivariate model\(^11\) and the hierarchical summary receiver operating characteristic (HSROC) model.\(^12\) Finally, the method by which the quality of individual studies is examined has been updated from the original Quality Assessment of Diagnostic Accuracy Studies (QUADAS)\(^13\) to the newly published QUADAS-2.\(^14\) These changes over the last five years have been extensive but the goal with this systematic review and meta-analysis has remained the same: to analyse the literature on ShPE tests of the shoulder to careful analysis in order to determine their clinical utility in adult (18 or older) patients.

**METHODS**

This systematic review with meta-analysis was conducted and reported according to the protocol outlined by PRISMA\(^9\) using a research question framed by PICOS methodology. PICOS is a pneumonic representing population (eg, adults), intervention (eg, diagnostic test), comparison (eg, control group), outcome (eg, accuracy) and study design (eg, cohort). In order to be eligible for this review, diagnostic accuracy studies, written in English, had to report both the sensitivity and specificity of ShPE tests in adults with shoulder pain due to musculoskeletal pathology. Excluded from this review, were articles using equipment or devices that are not readily available to most clinicians during physical examination and articles in which subjects were tested under anaesthesia or in which subjects were cadavers.

**Study selection**

Since this review is an update of our previous work,\(^1\) the terms and the Medline and CINAHL search strategies remained the same with the exception that the search was confined to the dates November, 2006 through February, 2012. Our previous study dates were 1966 – October, 2006. Further, the original search was expanded, without date restrictions, to include two new databases: EMBASE and the Cochrane Library. A hand search was also conducted which included the authors’ private collections and the searching of previous systematic reviews. Two authors (EH and AW) read titles and abstracts of all database-captured articles applying the a priori inclusion/exclusion criteria and agreement was measured using the \(\kappa\) statistic (figure 1). Disagreement was then resolved by discussion between the two authors and, in the event that agreement could not be reached, a third author (CC) served as the deciding vote. With the remaining articles, the same two authors (EH and AW) read the entire paper and again, a \(\kappa\) value was calculated to measure agreement as to which articles to retain for final analysis (figure 1). Once the final group of 32 articles was determined, 2x2 table data were extracted and saved for meta-analysis. Only data from studies, where the 2x2 data were reported or could be inferred from stated positive likelihood ratios, negative likelihood ratios, positive predictive values, and negative predictive values were retained for meta-analysis. If 2x2 data could not be discerned, the article was excluded from meta-analysis but still retained for systematic review and qualitative analysis.

**Quality assessment**

Once the final group of articles was agreed upon, two authors (EH and AW) independently examined the quality of each article using the QUADAS-2 tool.\(^14\) QUADAS-2 is a 4-phase tool, the last phase of which assists authors of systematic reviews in rating: 1) bias and 2) applicability. The risk of bias is assessed in four key areas: patient selection, index test, reference standard, and flow and timing. Concern for applicability is assessed in three key areas: patient selection, index test, and reference standard. For both categories, risk of bias and concern for applicability, the individual criteria were classified as low risk, high risk, or unclear and the results were presented using tables from the QUADAS web site (www.quadas.org).

**Statistical analysis**

In order to maximise the potential for meta-analysis, we added 2x2 data from our first meta-analysis\(^1\) to data gathered from the 32 additional articles included in this review. Hierarchical summary receiver operating characteristic (HSROC) curve\(^12\) and bivariate\(^11\) models were used to combine estimates of sensitivity (SN), specificity (SP), positive likelihood ratios (+LR), negative likelihood ratios (−LR) and diagnostic OR (DOR) with their 95% CI. Sensitivity measures the proportion of actual positives which are correctly identified as such (eg, the percentage of sick people who are correctly identified as having the condition). Specificity measures the proportion of negatives which are correctly identified (eg, the percentage of healthy people who are correctly identified as not having the condition). Positive likelihood ratio (LR+) dictates how much the odds of the disease increase when a test is positive.\(^15\) The negative likelihood ratio (LR−) dictates how much the odds of the disease decrease when a test is negative.\(^15\) Diagnostic OR express the strength of association between the test result and disease. These models, in the absence of covariates, are different parameterisations of the same model\(^10\) and take into account the correlation between sensitivity and specificity and both the within and the between study variances.\(^16\) The 95% prediction region is graphically provided which is the given probability (ie, 95%) of including the true sensitivity and specificity of a future study.\(^17\) DerSimonian-Laird\(^18\) random-effects models were used where less than four studies were eligible for statistical pooling. Heterogeneity was explored graphically with forest plots and statistically with Cochrane-Q with p<0.10 to indicate significant heterogeneity. When appropriate, meta-regression or subgroup analysis using study level characteristics was used to explore heterogeneity with a p<0.10 to indicate a significant difference in stratified estimates. A p value of <0.10 was decided upon to determine a significance in stratified estimates due to the low power of the test used to detect differences in stratified estimates.\(^19\) A 0.5 was added to all four cells of the 2x2 table when a zero was encountered in any cell as suggested by Cox.\(^20\)

Publication bias was analysed statistically with the Egger\(^21\) test with a p<0.05 to indicate significant publication bias. Threshold effects were tested using Spearman correlation coefficients.\(^22\) Influential studies on summary estimates were assessed with Cooks-d and standardised residuals according to Rabe-Hesketh\(^23\) with sensitivity analyses to determine if influential studies should be removed from the analyses. All statistical analyses were conducted in Stata 11 (Stata, College Station Texas, USA) by one of the authors (AG).
was examined inclusive of any size of tear or classification. Therefore, the rotator cuff concur with this thought but separated these pathologic entities into Rotator Cuff Tears, Tendinopathy, Subacromial impingement, Instability, Labral tears, Biceps pathology, Stiffness-related disorders and Other. The most frequent topics of focus were RCT, Tendinopathy, Subacromial impingement and Labral tears. Many would consider tendinopathy and impingement different labels for the same syndrome and further, that both labels capture a continuum of disease that includes RCTs. We have used the Q2,14 blinded from each other’s assessments. The two authors (EH and AW) independently used the QUADAS-2 (Q2),14 in the areas of risk of bias and concern for applicability (Appendix). Concern for applicability, for this review, was defined as concern for external validity, the degree to which results of a research study can be applied to practice. The two authors (EH and AW) independently used the Q2,14 blinded from each other’s assessments. The number of low risk/concern scores was tallied into a total score for each article and agreement was calculated using a weighted κ statistic. The weighted κ was poor (κ = 0.31 with 95% CI 0.10 to 0.52). Summaries of risk of bias and concern for applicability for each pathological group are presented in figure 2.

Quality assessment – risk of bias and concern for applicability

Each of the 32 papers qualifying for final review was scrutinised, via the QUADAS-2 (Q2),14 in the areas of risk of bias and concern for applicability (Appendix). Concern for applicability, for this review, was defined as concern for external validity, the degree to which results of a research study can be applied to practice. The two authors (EH and AW) independently used the Q2,14 blinded from each other’s assessments. The number of low risk/concern scores was tallied into a total score for each article and agreement was calculated using a weighted κ statistic. The weighted κ was poor (κ = 0.31 with 95% CI 0.10 to 0.52). Summaries of risk of bias and concern for applicability for each pathological group are presented in figure 2.
## Table 1  
Summary of studies

<table>
<thead>
<tr>
<th>Lead author, year</th>
<th>Mean age years (range)</th>
<th>Mean symptom duration</th>
<th>Study design</th>
<th>Criterion standard</th>
<th>ShPE test</th>
<th>LR+</th>
<th>LR−</th>
<th>Author conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michener 24</td>
<td>40.6 (18–83)</td>
<td>33.8 months</td>
<td>Prospective blinded study</td>
<td>Arthroscopy</td>
<td>Hawkins-Kennedy</td>
<td>1.63</td>
<td>0.61</td>
<td>The single tests of painful arc, external rotation resistance, and Neer are useful screening tests to rule out SAIS (subacromial impingement syndrome). The reliability of all tests was acceptable for clinical use. Based on reliability and diagnostic accuracy, the single tests of the painful arc, external rotation resistance, and empty can have the best overall clinical utility. The cut point of 3 or more positive of 5 tests can confirm the diagnosis of SAIS, while less than 3 positive of 5 rules out SAIS.</td>
</tr>
<tr>
<td>Mille 25</td>
<td>55.5 (20–86)</td>
<td>37.5 months</td>
<td>Case-control, same subject, correlation, double-blind</td>
<td>Ultrasound</td>
<td>External rotation lag sign</td>
<td>7.2</td>
<td>0.60</td>
<td>A positive sign would appear to suggest the moderate likelihood of the presence of a full-thickness tear but this conclusion is tenuous based on the small sample size of the study and subsequent wide confidence interval</td>
</tr>
<tr>
<td>Kim YS 26</td>
<td>32.6 (19–54)</td>
<td>NA</td>
<td>Cohort study</td>
<td>Arthroscopy</td>
<td>Passive compression test</td>
<td>5.9</td>
<td>0.21</td>
<td>The passive compression test is a useful and accurate technique for predicting superior labral tears of the shoulder joint.</td>
</tr>
<tr>
<td>Fodor 27</td>
<td>57 (20–84)</td>
<td>NA</td>
<td>Prospective, consecutive subjects</td>
<td>Ultrasound</td>
<td>Neer</td>
<td>10.8</td>
<td>0.48</td>
<td>The Hawkins-Kennedy test is the most sensitive test for identification of subacromial impingement syndrome, while Neer is most specific. With 4 (+) tests, the specificity increases and the sensitivity decreases. No tests were good at distinguishing stages of subacromial impingement.</td>
</tr>
<tr>
<td>Ja 28</td>
<td>NA</td>
<td>NA</td>
<td>Retrospective</td>
<td>Arthroscopy</td>
<td>Shrug sign</td>
<td>3.6</td>
<td>0.12</td>
<td>The shrug sign is a non-specific physical exam sign for shoulder dysfunction and is more commonly associated with glenohumeral OA, adhesive capsulitis, and massive rotator cuff tear.</td>
</tr>
<tr>
<td>Bushnell 29</td>
<td>24 (16–52)</td>
<td>NA</td>
<td>prospective pilot study</td>
<td>Arthroscopy</td>
<td>Bony apprehension test for instability</td>
<td>7.14</td>
<td>0.00</td>
<td>The bony apprehension test can reliably screen for significant osseous lesions.</td>
</tr>
<tr>
<td>Castoldi 30</td>
<td>50.4 (16–89)</td>
<td>NA</td>
<td>Prospective cohort treatment</td>
<td>Arthroscopy</td>
<td>External rotation lag sign (ERLS) – FTT SS</td>
<td>28.00</td>
<td>0.45</td>
<td>The ERLS is highly specific and acceptably sensitive for the diagnosis of full-thickness tears, even in case of an isolated lesion of the supraspinatus tendon.</td>
</tr>
<tr>
<td>Silva 31</td>
<td>55 (24–82)</td>
<td>97.5 days</td>
<td>Prospective</td>
<td>MRI</td>
<td>Neer</td>
<td>0.98</td>
<td>1.10</td>
<td>The Yocum test was the most sensitive for subacromial impingement and the Gerber test for subacromial-subdeltoid bursitis. The Gerber and Patte tests provide the best diagnostic combo. The majority of tests showed low specificity.</td>
</tr>
<tr>
<td>Chew 32</td>
<td>44 (18–75)</td>
<td>9.8 months</td>
<td>prospective cohort</td>
<td>Ultrasound</td>
<td>Neer</td>
<td>1.60</td>
<td>0.60</td>
<td>Diagnosis of supraspinatus pathology may be accomplished with a cluster of three tests: age &gt;39, (+) painful arc, self reported clicking or popping.</td>
</tr>
<tr>
<td>Bak 33</td>
<td>56 FTT, 38 No tear (39–75 FTT; 19–73 control)</td>
<td>13 days</td>
<td>Prospective diagnostic study</td>
<td>Ultrasound</td>
<td>Hawkins-Kennedy</td>
<td>1.04</td>
<td>0.88</td>
<td>BEFORE subacromial lidocaine injection: external rotation lag sign or drop arm test are indicative of a FTT supraspinatus; negative lag sign does not preclude a tear. AFTER subacromial lidocaine injection: specificity improves and sensitivity is reduced for all tests.</td>
</tr>
<tr>
<td>Lead author</td>
<td>Year</td>
<td>Mean age years (range)</td>
<td>Mean symptom duration</td>
<td>Study design</td>
<td>Criterion standard</td>
<td>SHPE test</td>
<td>LR+</td>
<td>LR−</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>------------------------</td>
<td>-----------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-----------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Bartsch</td>
<td>34</td>
<td>58 (SD 11.6)</td>
<td>NA</td>
<td>Prospective, consecutive subjects</td>
<td>Arthroscopy Lift off test</td>
<td>Internal rotation lag sign</td>
<td>Modifi  ed belly press (BPT)</td>
<td>1.90</td>
</tr>
<tr>
<td>Kibler</td>
<td>35</td>
<td>59 (28–64)</td>
<td>NA</td>
<td>Cohort study</td>
<td>Arthroscopy</td>
<td>Belly press</td>
<td>Upper cut</td>
<td>Bear hug</td>
</tr>
<tr>
<td>Chen</td>
<td>36</td>
<td>NA</td>
<td>23 weeks</td>
<td>Prospective, double-blind</td>
<td>Ultrasound</td>
<td>Yergason's</td>
<td>Speed's</td>
<td>Bicipital groove palpation</td>
</tr>
<tr>
<td>Fowler</td>
<td>37</td>
<td>41 35 weeks</td>
<td>Cohort/retrospective</td>
<td>Arthroscopy</td>
<td>Hawkins-Kennedy/RC</td>
<td>Relocation/Bankart lesion</td>
<td>Relocation/Hill-Sachs</td>
<td>O'Brien's/labrum</td>
</tr>
<tr>
<td>Goyal</td>
<td>38</td>
<td>45 (23–62)</td>
<td>2.8 months</td>
<td>Case-control</td>
<td>Ultrasound</td>
<td>Resisted abduction</td>
<td>Adduction stress</td>
<td>4.6</td>
</tr>
</tbody>
</table>

The diagnostic accuracy of isolated standard shoulder tests in recreational athletes was overall very poor. A positive response in each of the SLAP tests was observed in almost all of the SLAP positive shoulders. The diagnostic accuracy of isolated standard shoulder tests in recreational athletes was increased when used in combination and was substantially better in this setting. The results of shoulder examinations are variable and statistical analysis may not demonstrate a substantial improvement on the original observations of Codman. Sensitivity was good for the clinical diagnosis of supraspinatus lesions and the modality of diagnostic accuracy was high for lesions of infraspinatus, subscapularis, and the acromioclavicular joint. Specificity was high for SLAP and the modality of diagnostic accuracy was low for the detection of rotator cuff tendinosis.
Table 1

<table>
<thead>
<tr>
<th>Lead author, year</th>
<th>Mean age years (range)</th>
<th>Mean symptom duration</th>
<th>Study design</th>
<th>Criterion standard</th>
<th>SnPE test</th>
<th>LR+</th>
<th>LR−</th>
<th>Author conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawkins-Kennedy for AC joint OA</td>
<td>0.85</td>
<td>1.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawkins-Kennedy for biceps tendinopathy</td>
<td>0.89</td>
<td>1.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lift off for biceps tendinopathy</td>
<td>2.55</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lift off for glenohumeral OA</td>
<td>2.90</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lift off for RC tendinopathy</td>
<td>0.48</td>
<td>1.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neer for AC joint OA</td>
<td>0.97</td>
<td>1.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neer for biceps tendinopathy</td>
<td>1.08</td>
<td>0.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed's for biceps tendinopathy</td>
<td>1.52</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whipple for massive RC tear</td>
<td>1.35</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whipple for RC tendinopathy</td>
<td>1.19</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Hawkins-Kennedy | 0.62 | 0.52 | The Hawkins-Kennedy test was the most accurate test for diagnosing any degree of subacromial impingement syndrome.
| Painful arc of abduction | 0.59 | 1.4 |
| Abduction weakness | 0.76 | 1.24 |
| Abduction pain | 2.21 | 0.34 |
| External rotation weakness | 0.74 | 1.8 |
| External rotation pain | 3.3 | 0.74 |
| Empty can weakness | 1.56 | 0.72 |
| Empty can pain | 0.78 | 1.45 |
| Full can weakness | 1.79 | 0.73 |
| Full can pain | 0.46 | 2.6 |
| Empty can (SS) | 0.64 | 1.34 | Physical examination used for the diagnosis of shoulder pain had low sensitivity and specificity for the detection of rotator cuff tendon tears.
| Lift-off (SB) | 0.081 | 4.67 |
| Yergason's (biceps) | 4.03 | 0.31 |
| Hawkins-Kennedy | 2.15 | 0.51 |
| Empty can (SS) | 0.13 | 0.6 |
| Patte's test | 0.13 | 0.5 |
| Lift-off (SB) | 1.3 | 0.7 |
| Yergason's (biceps) | 1.3 | 0.96 |
| Hawkins-Kennedy | 2.15 | 0.51 |
| Empty can | 1.14 | 0.85 |
| Patte's test | 2.43 | 0.5 |
| Lift-off | 1.45 | 0.85 |
| Speed's | 2.1 | 0.56 |
| SNAPSHOT >3 | 6.1 | 0.3 |
| Walsworth | 0.67 | 2.5 |
| Anterior slide | 2.38 | 0.69 |
| Crank | 1.35 | 0.71 |
| Passive distraction test (PDT) | 8.83 | 0.5 |
| Active compression test (ACT) | 7.38 | 0.45 |
| Active compression test (ACT) | 7.38 | 0.45 |
| Anterior slide test | 10.5 | 0.81 |
| PDT + ACT | 7 | 0.33 |
| The combination of popping or catching with a positive crank or anterior slide result or a positive anterior slide result with a positive active compression or crank test result suggests the presence of a labral tear. The combined absence of popping or catching and a negative anterior slide or crank result suggests the absence of a labral tear.
<p>| The passive distraction test can be used alone or in combo to aid the clinical evaluation and diagnosis of SLAP lesion. |</p>
<table>
<thead>
<tr>
<th>Lead author, year</th>
<th>Mean age years (range)</th>
<th>Mean symptom duration</th>
<th>Study design</th>
<th>Criterion standard</th>
<th>SHPE test</th>
<th>LR+</th>
<th>LR−</th>
<th>Author conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gillooly⁴⁶</td>
<td>53 (17–83)</td>
<td>15</td>
<td>Prospective cohort</td>
<td>Arthroscopy</td>
<td>Lateral Jobe test Combined tests*</td>
<td>7.36</td>
<td>0.21</td>
<td>The lateral Jobe test had a higher sensitivity than the combined tests (Empty can, strength in ER, and subacromial impingement tests). It is a simple, new technique which can improve the clinical diagnosis of rotator cuff tears; *positive result for the combined tests was taken as weakness on supraspinatus testing, weakness in external rotation and pain on subacromial impingement or a combination of two of these and an age greater than 60 years. The presence of a normal OMP (olecranon manubrium percussion test) sign does not negate the need for radiographic studies in patients with shoulder injury. The presence of an abnormal OMP sign suggests the need for appropriate radiographic studies. Coracoid pain test is an easy and reliable test for identifying patients with or without adhesive capsulitis.</td>
</tr>
<tr>
<td>Adams⁴⁷</td>
<td>NA</td>
<td>acute</td>
<td>Prospective</td>
<td>X-rays</td>
<td>Olecranon-Manubrium Percussion Sign</td>
<td>0.15</td>
<td></td>
<td>The presence of a normal OMP (olecranon manubrium percussion test) sign does not negate the need for radiographic studies in patients with shoulder injury. The presence of an abnormal OMP sign suggests the need for appropriate radiographic studies.</td>
</tr>
<tr>
<td>Carbone⁴⁸</td>
<td>40–50</td>
<td>NA</td>
<td>Retrospective</td>
<td>Codman’s criteria, exam &amp;/ or MRI</td>
<td>Coracoid pain test (Adhesive capsulitis)</td>
<td>4.95</td>
<td>0.01</td>
<td>Coracoid pain test is an easy and reliable test for identifying patients with or without adhesive capsulitis.</td>
</tr>
<tr>
<td>Ebinger⁴⁹</td>
<td>49 (14–79)</td>
<td>chronic</td>
<td>Prospective</td>
<td>Arthroscopy</td>
<td>Supine flexion resistance test Speed’s</td>
<td>2.6</td>
<td>0.29</td>
<td>Each of the 5 stand-alone tests and clusters of tests provide minimal to no value in the diagnosis of a SLAP lesion, whether a SLAP-only lesion or a SLAP lesion with or without concomitant findings reference.</td>
</tr>
<tr>
<td>Cook⁵⁰</td>
<td>45</td>
<td>chronic</td>
<td>Prospective, case-based, case-control</td>
<td>Arthroscopy</td>
<td>Active compression Kim II Dynamic labral shear Speed’s</td>
<td>1.3</td>
<td>0.94</td>
<td>Coracoid pain test is an easy and reliable test for identifying patients with or without adhesive capsulitis.</td>
</tr>
<tr>
<td>Gill⁵¹</td>
<td>44 no tear/59 partial tear</td>
<td>NA</td>
<td>Cohort study</td>
<td>Arthroscopy</td>
<td>Palpation Lift off Speed’s</td>
<td>1.13</td>
<td>0.87</td>
<td>Both the empty can test and full can test were considered to be valuable as screening tests to detect a torn rotator cuff, using the positive signs of pain and weakness separately, in spite of their modest overall accuracy.</td>
</tr>
<tr>
<td>Kim, E⁵²</td>
<td>60 (37–83)</td>
<td>&gt;3 months</td>
<td>Prospective</td>
<td>MRI and arthroscopy</td>
<td>Empty can pain or weakness for RC tear (PIT or FTT) Empty can pain for RC tear (PIT or FTT) Empty can pain for RC tear (PIT or FTT)</td>
<td>1.74</td>
<td>0.02</td>
<td>The accuracy of clinical diagnosis of periarticular shoulder conditions is low. Physical exam was unable to differentiate rotator cuff tendinitis from tear, and partial thickness tear from full-thickness tear.</td>
</tr>
<tr>
<td>Naredo⁵³</td>
<td>58 (21–77)</td>
<td>12.5 months</td>
<td>Prospective</td>
<td>Ultrasound</td>
<td>Empty can pain or weakness, SS tear Empty can pain or weakness, SS tendinopathy Lift off, SB tendinopathy Patte’s, IS tendinopathy Patte’s, IS tear</td>
<td>3.28</td>
<td>0.50</td>
<td>The accuracy of clinical diagnosis of periarticular shoulder conditions is low. Physical exam was unable to differentiate rotator cuff tendinitis from tear, and partial thickness tear from full-thickness tear.</td>
</tr>
</tbody>
</table>
Table 1  Continued

<table>
<thead>
<tr>
<th>Lead author, year</th>
<th>Mean age years (range)</th>
<th>Mean symptom duration</th>
<th>Study design</th>
<th>Criterion standard</th>
<th>Criterion standard</th>
<th>ShPE test</th>
<th>LR+</th>
<th>LR−</th>
<th>Author conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itoi2,4</td>
<td>53 (16–86)</td>
<td>NA</td>
<td>Retrospective case series</td>
<td>Arthroscopy</td>
<td>SS tear: Full can pain</td>
<td>1.6</td>
<td>0.4</td>
<td></td>
<td>Pain is not useful in locating the sight of a tear. In patients with cuff tendinopathy, the supraspinatus test is most accurate when interpreted with MMT &lt; 5, whereas ERST (infraspinatus) is most accurate with MMT &lt; 4+, and lift-off test (subscapularis) most accurate with MMT &lt; 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS tear: Full can weak</td>
<td>1.8</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS tear: Empty can pain</td>
<td>1.3</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS tear: Empty can weak</td>
<td>1.5</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IS tear: External rotation strength tests weak &lt; 5</td>
<td>1.8</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IS tear: External rotation strength tests weak &lt; 5</td>
<td>1.5</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SB tear: Lift-off test weak &lt; 5</td>
<td>1.9</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SB tear: Lift-off test weak &lt; 5</td>
<td>1.9</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oh5</td>
<td>Mid 40's (17-70's)</td>
<td>NA</td>
<td>Retrospective case control study</td>
<td>Arthroscopy</td>
<td>Biceps groove tenderness</td>
<td>1.7</td>
<td>1.1</td>
<td></td>
<td>No test had a high sensitivity and high specificity; no combination of 2 tests yielded sensitivity/specificity of more than 60%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Speed's</td>
<td>1.06</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yergason</td>
<td>0.92</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Relocation</td>
<td>0.96</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Compression rotation</td>
<td>1.3</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Active compression</td>
<td>1.3</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kibler</td>
<td>0.7</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Biceps lead II</td>
<td>1.4</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anterior apprehension</td>
<td>1.1</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Whipple</td>
<td>1.1</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AC, acromioclavicular; FTT, full-thickness tear; IS, infraspinatus; MMT, manual muscle test; NA, not available; OA, osteoarthritis; PTT, partial-thickness tear; RC, rotator cuff; SB, subscapularis; SLAP, superior labrum anterior posterior; SS, supraspinatus; TM, teres minor.
### Table 2  Alphabeticallist of common shoulder physical examination (ShPE) tests

<table>
<thead>
<tr>
<th>Test name(s)</th>
<th>Pathology</th>
<th>Lead Author</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Risk of Bias* from QUADAS 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC Resisted Extension</td>
<td>AC Joint OA</td>
<td>Jia³⁹</td>
<td>72</td>
<td>85</td>
<td>High</td>
</tr>
<tr>
<td>Active Compression/O'Brien</td>
<td>SLAP</td>
<td>Cook²⁰</td>
<td>91</td>
<td>14</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>SLAP</td>
<td>Schlechter¹⁵</td>
<td>59</td>
<td>92</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>SLAP</td>
<td>Ebinger⁴⁸</td>
<td>94</td>
<td>28</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Type II SLAP</td>
<td>Oh⁵⁵</td>
<td>63</td>
<td>53</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Jia³⁹</td>
<td>53</td>
<td>58</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Kübler¹⁵</td>
<td>61</td>
<td>84</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Fowler²⁷</td>
<td>63</td>
<td>40</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Kübler¹⁵</td>
<td>38</td>
<td>61</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Jia³⁹</td>
<td>68</td>
<td>46</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>AC Joint OA</td>
<td>Walsworth⁴⁴</td>
<td>55</td>
<td>18</td>
<td>Low</td>
</tr>
<tr>
<td>Adduction Stress</td>
<td>AC Joint OA</td>
<td>Goyal³⁸</td>
<td>57</td>
<td>96</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Kübler¹⁵</td>
<td>24</td>
<td>62</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Jia³⁹</td>
<td>50</td>
<td>81</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Kübler¹⁵</td>
<td>48</td>
<td>82</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>SLAP</td>
<td>Schlechter¹⁵</td>
<td>21</td>
<td>98</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Type II SLAP</td>
<td>Oh⁵⁵</td>
<td>21</td>
<td>70</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Walsworth⁴⁴</td>
<td>43</td>
<td>82</td>
<td>Low</td>
</tr>
<tr>
<td>Anterior Slide</td>
<td>Type II SLAP</td>
<td>Oh⁵⁵</td>
<td>62</td>
<td>42</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>SLAP</td>
<td>Fowler²⁷</td>
<td>29</td>
<td>70</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>SLAP</td>
<td>Fowler²⁷</td>
<td>29</td>
<td>70</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Glenohumeral Instability</td>
<td>Jia³⁹</td>
<td>59</td>
<td>96</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Anterior Instability</td>
<td>Jia³⁹</td>
<td>72</td>
<td>96</td>
<td>High</td>
</tr>
<tr>
<td>Apprehension- Anterior</td>
<td>Posterior Instability</td>
<td>Jia³⁹</td>
<td>19</td>
<td>99</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Bear Hug</td>
<td>Kübler¹⁵</td>
<td>79</td>
<td>60</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Kübler¹⁵</td>
<td>37</td>
<td>32</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Kübler¹⁵</td>
<td>15</td>
<td>75</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Shoulder Girdle Tendinopathy</td>
<td>Bartsch¹⁴</td>
<td>86</td>
<td>91</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Kübler¹⁵</td>
<td>31</td>
<td>85</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Kübler¹⁵</td>
<td>15</td>
<td>75</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial Tendinopathy</td>
<td>Goyal³⁸</td>
<td>75</td>
<td>97</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Kübler¹⁵</td>
<td>55</td>
<td>53</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>SLAP</td>
<td>Oh⁵⁵</td>
<td>30</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Apprehension- Posterior</td>
<td>Bony Instability</td>
<td>Bushnell²³</td>
<td>94</td>
<td>84</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Compression-Rotation</td>
<td>Oh⁵⁵</td>
<td>61</td>
<td>54</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Walsworth⁴⁴</td>
<td>61</td>
<td>55</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Cross-body</td>
<td>Cheuv²²</td>
<td>75</td>
<td>61</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tendinopathy</td>
<td>Cheuv²²</td>
<td>22</td>
<td>75</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>AC Joint OA</td>
<td>Jia³⁹</td>
<td>77</td>
<td>79</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>RC Tendinopathy</td>
<td>Jia³⁹</td>
<td>74</td>
<td>66</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinatus</td>
<td>Baks³³</td>
<td>45</td>
<td>70</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Kübler¹⁵</td>
<td>72</td>
<td>98</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Kübler¹⁵</td>
<td>18</td>
<td>53</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Empty Can (pain)</td>
<td>Ito³⁴</td>
<td>78</td>
<td>40</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Kelly⁴⁰</td>
<td>52</td>
<td>33</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>RC Tear</td>
<td>Kim E³²</td>
<td>94</td>
<td>46</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Empty Can (weak)</td>
<td>Ito³⁴</td>
<td>87</td>
<td>43</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Kelly⁴⁰</td>
<td>52</td>
<td>67</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>RC Tear</td>
<td>Kim E³²</td>
<td>76</td>
<td>71</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Michener²⁴</td>
<td>50</td>
<td>87</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Empty Can (pain or weak)</td>
<td>Cheuv²²</td>
<td>83</td>
<td>49</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tendinopathy</td>
<td>Cheuv²²</td>
<td>99</td>
<td>43</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinatus</td>
<td>Baks³³</td>
<td>76</td>
<td>39</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tear</td>
<td>Naredo³³</td>
<td>19</td>
<td>100</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tendinopathy</td>
<td>Kim HA⁴¹</td>
<td>72</td>
<td>45</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tendinopathy</td>
<td>Kim HA⁴¹</td>
<td>31</td>
<td>52</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tendinopathy</td>
<td>Fowler²⁷</td>
<td>50</td>
<td>83</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tendinopathy</td>
<td>Salaffi⁴³</td>
<td>56</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tendinopathy</td>
<td>Naredo³³</td>
<td>79</td>
<td>50</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Supraspinatus Tendinopathy</td>
<td>Goya³⁶</td>
<td>90</td>
<td>37</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Silva³¹</td>
<td>74</td>
<td>30</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>RC Tear</td>
<td>Kim E³²</td>
<td>71</td>
<td>74</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Empty Can (pain and weak)</td>
<td>Ito³⁴</td>
<td>80</td>
<td>50</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Kelly⁴⁰</td>
<td>35</td>
<td>25</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Full Can (pain)</td>
<td>Torn Supraspinatus</td>
<td>Naredo³³</td>
<td>19</td>
<td>100</td>
</tr>
</tbody>
</table>

*QUADAS 2 Risk of Bias scores: Low (0-3), Moderate (4-6), High (7-9).*
<table>
<thead>
<tr>
<th>Test name(s)</th>
<th>Pathology</th>
<th>Lead Author</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Risk of Bias* from QUADAS 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Can (weak)</td>
<td>Torn Supraspinatus</td>
<td>Ito24</td>
<td>83</td>
<td>53</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>RC Tear</td>
<td>Kim E52</td>
<td>77</td>
<td>32</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Kelly30</td>
<td>45</td>
<td>75</td>
<td>Low</td>
</tr>
<tr>
<td>Full Can (pain or weak)</td>
<td>Supraspinatus Tendinopathy</td>
<td>Chew32</td>
<td>75</td>
<td>68</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>RC Tear</td>
<td>Kim E52</td>
<td>90</td>
<td>54</td>
<td>Moderate</td>
</tr>
<tr>
<td>Full Can (pain and weak)</td>
<td>RC Tear</td>
<td>Kim E52</td>
<td>59</td>
<td>82</td>
<td>Moderate</td>
</tr>
<tr>
<td>External Rotation Lag Sign</td>
<td>Masssive RC Tear</td>
<td>Jia39</td>
<td>35</td>
<td>89</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinatus</td>
<td>Baks23</td>
<td>45</td>
<td>91</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinatus</td>
<td>Castoldi29</td>
<td>56</td>
<td>98</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>FTT- Infraspinus</td>
<td>Castoldi29</td>
<td>97</td>
<td>93</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>FTT- Teres Minor</td>
<td>Castoldi29</td>
<td>100</td>
<td>93</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>RC Tendinopathy</td>
<td>Jia39</td>
<td>7</td>
<td>84</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinatus/Infraspinus</td>
<td>Miller25</td>
<td>46</td>
<td>94</td>
<td>Moderate</td>
</tr>
<tr>
<td>Hawkins-Kennedy</td>
<td>Supraspinatus Tendinopathy</td>
<td>Chew32</td>
<td>87</td>
<td>32</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinus</td>
<td>Baks23</td>
<td>77</td>
<td>26</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Kelly30</td>
<td>74</td>
<td>50</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Michener24</td>
<td>63</td>
<td>62</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Silva21</td>
<td>74</td>
<td>40</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Fodor27</td>
<td>72</td>
<td>89</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Salaffi43</td>
<td>64</td>
<td>71</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Fowle27</td>
<td>58</td>
<td>72</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>AC Joint OA</td>
<td>47</td>
<td>45</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Jia39</td>
<td>55</td>
<td>38</td>
<td>High</td>
</tr>
<tr>
<td>Internal Rotation Lag Sign</td>
<td>Subscapularis Tendinopathy</td>
<td>Bartsch14</td>
<td>71</td>
<td>60</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinus</td>
<td>Baks23</td>
<td>31</td>
<td>87</td>
<td>High</td>
</tr>
<tr>
<td>Labral Tension</td>
<td>Subscapularis Tear</td>
<td>Miller15</td>
<td>100</td>
<td>84</td>
<td>Moderate</td>
</tr>
<tr>
<td>Lateral Jobe</td>
<td>SLAP</td>
<td>Cook10</td>
<td>28</td>
<td>76</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>RC Tear</td>
<td>Gilli48</td>
<td>81</td>
<td>89</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Partial Biceps Tear</td>
<td>Gilli48</td>
<td>28</td>
<td>89</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Jia39</td>
<td>28</td>
<td>89</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Subscapularis Tendinopathy</td>
<td>Bartsch14</td>
<td>40</td>
<td>79</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subscapularis Tendinopathy</td>
<td>Naredo31</td>
<td>50</td>
<td>84</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subscapularis Tendinopathy</td>
<td>Kim HA42</td>
<td>69</td>
<td>48</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Subscapularis Tendinopathy</td>
<td>Kim HA41</td>
<td>6</td>
<td>23</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subscapularis Tendinopathy</td>
<td>Salaffi43</td>
<td>35</td>
<td>75</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Silva21</td>
<td>68</td>
<td>50</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>RC Tendinopathy</td>
<td>Fowle27</td>
<td>19</td>
<td>90</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Glenohumeral OA</td>
<td>Jia39</td>
<td>29</td>
<td>90</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>RC Tendinopathy</td>
<td>Jia39</td>
<td>10</td>
<td>79</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Neer</td>
<td>Chew32</td>
<td>64</td>
<td>61</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinus</td>
<td>Baks23</td>
<td>60</td>
<td>35</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Kelly30</td>
<td>62</td>
<td>10</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Michener24</td>
<td>81</td>
<td>54</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Silva21</td>
<td>68</td>
<td>30</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Fodor27</td>
<td>54</td>
<td>95</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>AC Joint OA</td>
<td>Jia39</td>
<td>57</td>
<td>41</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Biceps Tendinopathy</td>
<td>Jia39</td>
<td>64</td>
<td>41</td>
<td>High</td>
</tr>
<tr>
<td>Olecranon Manubrium Percussion</td>
<td>Supraspinatus Tendinopathy</td>
<td>Chew32</td>
<td>64</td>
<td>61</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Neer</td>
<td>Chew32</td>
<td>71</td>
<td>81</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinus</td>
<td>Baks23</td>
<td>60</td>
<td>35</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Kelly30</td>
<td>49</td>
<td>33</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Michener24</td>
<td>75</td>
<td>67</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Fodor27</td>
<td>67</td>
<td>80</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>FTT- Supraspinus</td>
<td>Baks23</td>
<td>96</td>
<td>4</td>
<td>High</td>
</tr>
<tr>
<td>Palpation- biceps</td>
<td>Biceps Tendinopathy</td>
<td>Chen</td>
<td>57</td>
<td>74</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Partial T- Biceps</td>
<td>Gilli48</td>
<td>52</td>
<td>54</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Type II SLAP</td>
<td>Oh55</td>
<td>27</td>
<td>66</td>
<td>Low</td>
</tr>
<tr>
<td>Palpation-coracoid</td>
<td>Adhesive Capsulitis</td>
<td>Carbone46</td>
<td>96</td>
<td>87</td>
<td>High</td>
</tr>
<tr>
<td>Passive-Abduction (pain)</td>
<td>Subacromial impingement</td>
<td>Silva21</td>
<td>74</td>
<td>10</td>
<td>Low</td>
</tr>
<tr>
<td>Passive Compression</td>
<td>SLAP</td>
<td>Kim YS28</td>
<td>82</td>
<td>86</td>
<td>Low</td>
</tr>
<tr>
<td>Passive Distraction</td>
<td>SLAP</td>
<td>Schlechter45</td>
<td>53</td>
<td>94</td>
<td>Low</td>
</tr>
<tr>
<td>Patte</td>
<td>Subacromial impingement</td>
<td>Silva21</td>
<td>58</td>
<td>60</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Infraspinatus Tendinopathy</td>
<td>Kim HA42</td>
<td>63</td>
<td>73</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Infraspinatus Tendinopathy</td>
<td>Salaffi43</td>
<td>62</td>
<td>74</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Infraspinatus Tendinopathy</td>
<td>Naredo31</td>
<td>71</td>
<td>90</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Infraspinatus Tendinopathy</td>
<td>Naredo31</td>
<td>36</td>
<td>95</td>
<td>Moderate</td>
</tr>
<tr>
<td>Relocation</td>
<td>Type II SLAP</td>
<td>Oh55</td>
<td>44</td>
<td>54</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Bankart lesion</td>
<td>Fowle27</td>
<td>79</td>
<td>87</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Hill-Sachs Lesion</td>
<td>Fowle27</td>
<td>81</td>
<td>81</td>
<td>High</td>
</tr>
<tr>
<td>Resisted- Abduction (pain)</td>
<td>Subacromial impingement</td>
<td>Kelly30</td>
<td>55</td>
<td>75</td>
<td>Low</td>
</tr>
<tr>
<td>Resisted- Abduction (weak)</td>
<td>Subacromial impingement</td>
<td>Kelly30</td>
<td>38</td>
<td>50</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Subacromial impingement</td>
<td>Silva21</td>
<td>58</td>
<td>20</td>
<td>Low</td>
</tr>
</tbody>
</table>

* QUADAS 2: Quality Assessment of Diagnostic Accuracy Studies 2
2. The greatest risk of bias was most often associated with the Q2 items Patient Flow and Reference Standard. The greatest concern in the category of applicability was also the reference standard with the addition of the index test. Patient flow concerns become apparent when there was an indeterminate or excessive time between the issuing of the index test and the reference standard was most often due to a failure to use the data reported in the original article. The concern for bias in patient flow concerns became apparent when there was an indeterminate or excessive time between the issuing of the index test and the reference standard was most often due to a failure to use the data reported in the original article. The concern for bias in patient flow concerns was an inability to reconstruct the 2x2 tables accurately from the index test and reference standard, or did not make clear whether all patients were included in the analysis. Often, there was an inability to reconstruct the 2x2 tables accurately from the data reported in the original article. The concern for bias in the reference standard was most often due to a failure to use a double blind design (issuer of the criterion standard was not blinded to index test result) or the failure to use the criterion standard to confirm diagnosis. The obvious gain in popularity of diagnostic ultrasound (n=12 studies in this review) had the deleterious effect of increasing concern for bias since ultrasound is not the criterion standard for shoulder diagnosis. Lastly, the concern for applicability as it relates to the index test is because the authors failed to describe the index test.

### Statistical analysis

**Overall**

Publication bias was not found to be evident with graphical or in statistical analysis. However, publication bias cannot be completely ruled out since these tests have decreased statistical power when analysing less than 10 studies. No significant negative correlations were found to indicate the influence of threshold effects. Table 3 presents the results of meta-analysis for the individual ShPE tests by diagnosis, number of studies and sample size for the analyses.

**Subacromial impingement**

The Neer, Hawkins-Kennedy and painful arc tests for subacromial impingement were summarised for their diagnostic properties and associations. The strongest summary sensitivity was for the Hawkins-Kennedy test (0.80; 0.72, 0.86). However, the value was merely on the sensitivity threshold (80%) for assisting in ruling out subacromial impingement but because of poor specificity, the LR- value shows this test to have little effect on post-test probability to rule out subacromial impingement when negative. In fact, none of the three diagnostic tests demonstrated the likelihood ratios that would be unlikely to result in important changes in post-test probability. The pooled DOR

---

**Table 2 Continued**

<table>
<thead>
<tr>
<th>Test name(s)</th>
<th>Pathology</th>
<th>Lead Author</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Risk of Bias* from QUADAS 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resist-External Rotation/Infraspinus test (pain)</td>
<td>Subacromial impingement</td>
<td>Kelly</td>
<td>33</td>
<td>90</td>
<td>Low</td>
</tr>
<tr>
<td>resist-external rotation</td>
<td>SLAP</td>
<td>Cook</td>
<td>28</td>
<td>76</td>
<td>Moderate</td>
</tr>
<tr>
<td>Resist-ER/Infraspinus test (weak)</td>
<td>Subacromial impingement</td>
<td>Kelly</td>
<td>55</td>
<td>25</td>
<td>Low</td>
</tr>
<tr>
<td>Resist-Lift-off (pain)</td>
<td>Torn Infraspinus</td>
<td>Itoi</td>
<td>54</td>
<td>53</td>
<td>Moderate</td>
</tr>
<tr>
<td>Resist-Lift-off (weak)</td>
<td>Torn Subacromialis</td>
<td>Itoi</td>
<td>54</td>
<td>53</td>
<td>Moderate</td>
</tr>
<tr>
<td>shoulder Shrug</td>
<td>Glenohumeral OA</td>
<td>Jia</td>
<td>34</td>
<td>57</td>
<td>Low</td>
</tr>
<tr>
<td>Torn Infraspinus</td>
<td>Itoi</td>
<td>54</td>
<td>53</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Adhesive Capsulitis</td>
<td>RC Tendinopathy</td>
<td>Jia</td>
<td>34</td>
<td>53</td>
<td>Moderate</td>
</tr>
<tr>
<td>Resist-Internal Rotation</td>
<td>Massive RC Tear</td>
<td>Jia</td>
<td>34</td>
<td>53</td>
<td>Low</td>
</tr>
<tr>
<td>Speed</td>
<td>SLAP</td>
<td>Cook</td>
<td>28</td>
<td>76</td>
<td>Moderate</td>
</tr>
<tr>
<td>type II SLAP</td>
<td>Oh</td>
<td>35</td>
<td>32</td>
<td>66</td>
<td>Low</td>
</tr>
<tr>
<td>Labral Tear</td>
<td>Kibler</td>
<td>15</td>
<td>29</td>
<td>69</td>
<td>Low</td>
</tr>
<tr>
<td>SLAP</td>
<td>Ebing</td>
<td>43</td>
<td>38</td>
<td>37</td>
<td>Low</td>
</tr>
<tr>
<td>Biceps Tendinopathy</td>
<td>Gill</td>
<td>34</td>
<td>61</td>
<td>37</td>
<td>Low</td>
</tr>
<tr>
<td>Biceps Tendinopathy</td>
<td>Kibler</td>
<td>15</td>
<td>61</td>
<td>37</td>
<td>Low</td>
</tr>
<tr>
<td>Biceps Tendinopathy</td>
<td>Jia</td>
<td>34</td>
<td>61</td>
<td>37</td>
<td>Low</td>
</tr>
<tr>
<td>Biceps Tendinopathy</td>
<td>Goyal</td>
<td>38</td>
<td>61</td>
<td>37</td>
<td>High</td>
</tr>
<tr>
<td>Biceps Tendinopathy</td>
<td>Salaffi</td>
<td>41</td>
<td>61</td>
<td>37</td>
<td>Moderate</td>
</tr>
<tr>
<td>Supine Flexion Resistance</td>
<td>SLAP</td>
<td>Ebing</td>
<td>43</td>
<td>61</td>
<td>Low</td>
</tr>
<tr>
<td>Upper Cut</td>
<td>Biceps Tendinopathy</td>
<td>Kibler</td>
<td>15</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Labral Tear</td>
<td>Kibler</td>
<td>15</td>
<td>56</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Yergason</td>
<td>Biceps Tendinopathy</td>
<td>Chen</td>
<td>76</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Biceps Tendinopathy</td>
<td>Kibler</td>
<td>15</td>
<td>78</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Biceps Tendinopathy</td>
<td>Kim HA</td>
<td>42</td>
<td>78</td>
<td>78</td>
<td>High</td>
</tr>
<tr>
<td>Biceps Tendinopathy</td>
<td>Kim HA</td>
<td>42</td>
<td>78</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Labral Tear</td>
<td>Kibler</td>
<td>15</td>
<td>70</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Type II SLAP</td>
<td>Oh</td>
<td>35</td>
<td>87</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Yacum</td>
<td>Subacromial impingement</td>
<td>Silva</td>
<td>14</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>Subacromial impingement</td>
<td>Fedor</td>
<td>47</td>
<td>92</td>
<td>78</td>
<td>Moderate</td>
</tr>
<tr>
<td>Whipple</td>
<td>Type II SLAP</td>
<td>Oh</td>
<td>55</td>
<td>42</td>
<td>Low</td>
</tr>
<tr>
<td>RC Tendinopathy</td>
<td>Jia</td>
<td>34</td>
<td>33</td>
<td>78</td>
<td>High</td>
</tr>
<tr>
<td>Massive RC Tear</td>
<td>Jia</td>
<td>34</td>
<td>33</td>
<td>78</td>
<td>High</td>
</tr>
</tbody>
</table>

*Bias: High = score of high risk of bias in 3 or 4 of total 4 categories; Moderate = score of high risk of bias in 2 of total 4 categories; Low = score of high risk of bias in 0 or 1 of total 4 categories. The 4 categories are: 1. Patient selection 2. Index test 3. Reference standard 4. Flow and timing.

AC, acromioclavicular; ER, external rotation; OA, osteoarthritis; RC, rotator cuff; SLAP, superior labrum anterior to posterior.
Reviews for any of these three tests indicates the discriminative diagnostic ability to determine a positive test result among those with subacromial impingement when compared with those without subacromial impingement is unlikely to occur. Figure 3 (Neer), figure 4 (Hawkins-Kennedy) and figure 5 (painful arc) illustrate the included studies with both the 95% confidence and prediction regions indicating the probable wide variability of the true sensitivity and specificity in future studies.

Meta-regression was conducted for both the Neer and Hawkins-Kennedy tests in order to determine if the summary DOR was biased as a result of differing reference standards. For the Neer test, there was a substantially greater DOR among the studies which used the gold standard of surgery for index test confirmation (4.85 (95% CI 3.46 to 6.79)) than other reference standards (1.28 ((95% CI 0.31 to 5.19)). The ratio of DORs was strong (8.79 (95% CI 0.87 to 16.14)) and the stratified estimates were statistically significant (p=0.07). Similarly, the DOR for the Hawkins-Kennedy test was stronger among those studies with the gold standard of surgery (6.41 (95% 3.33 to 12.35) than for studies using other than the gold standard (3.14 (95% 1.37 to 7.22)). However, the stratified estimates were not significantly (p=0.18) different from one another.

SLAP lesions
None of the 8 ShPE tests for which meta-analysis was possible (table 3) demonstrated sensitivity values that would likely rule out a SLAP lesion with a negative test. Yergason’s test had the strongest summary specificity (95.3; 90.6,98.1), but again, the sensitivity was so poor that the LR+ demonstrates insignificant ability of this test to rule in a SLAP lesion when positive. All eight diagnostic tests for a SLAP lesion had likelihood ratios and DORs that were weak and their CI contained the null value (table 3).

The active compression test analysis found the O’Brien et al study to have a large Cooks-D and standardised residuals influencing the summary estimates. Cooks-D is a measure of the influence that a particular study may have on the model parameters and can be used to check for particularly influential studies. Sensitivity analysis, with removal of the O’Brien et al study, resulted in substantial attenuation of the DOR from 3.14 (95% CI 0.42 to 23.40) to 1.19 (95% CI 0.76 to 1.86). As such, this study was not included in summary estimates for the Active Compression test. Figure 6 illustrates the HSROC curves of the Active Compression test both with and without the outlier study.60

Anterior instability
Statistical pooling was done individually for three tests for the diagnosis of anterior instability: the apprehension, relocation and surprise tests. The surprise test demonstrated the strongest sensitivity (81.8; 69.1, 90.9), and therefore, negative likelihood ratio (0.25; 0.08–0.78) that would likely rule out anterior instability when negative. All three tests demonstrated the ability to rule in anterior instability due to strong specificity.
The apprehension test had the strongest positive likelihood ratio (17.2; 10.02, 29.55) and was the only one of the three in which the CI did not contain the null value. The apprehension test had the strongest DOR (53.6; 24.3, 118.3), indicating some evidence for this test’s overall diagnostic discriminative performance.

Significant heterogeneity was found in the properties and associations for the relocation test. Subgroup analysis, accomplished by removing the study by Lo et al61 based upon the non-criterion reference standard used, did not improve the overall heterogeneity.

**Labral tear**

In pooled analyses, the crank test for labral tear demonstrated limited ability to rule in a labral tear with a +LR of 2.4 and specificity of 76%, indicating a likely small change in post-test probability.

**Tendinopathy**

In pooled analyses, the Hawkins-Kennedy test for tendinopathy demonstrated no evidence for the ability to rule in or out, change post-test probability or have overall diagnostic discriminative performance.

**Table 3** Summary estimates from meta-analysis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Test</th>
<th>No. Studies</th>
<th>Sample Size (n)</th>
<th>SN(95% CI)</th>
<th>SP(95% CI)</th>
<th>+LR(95% CI)</th>
<th>-LR(95% CI)</th>
<th>DOR(95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impingement</td>
<td>Neer*</td>
<td>7(n=946)</td>
<td>0.72(0.60, 0.81)</td>
<td>0.60(0.40, 0.77)</td>
<td>1.79(1.24, 2.58)</td>
<td>0.47(0.39, 0.56)</td>
<td>3.83(2.51, 5.94)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H-K*</td>
<td>7(n=944)</td>
<td>0.80(0.72, 0.86)</td>
<td>0.56(0.45, 0.67)</td>
<td>1.84(1.49, 2.26)</td>
<td>0.35(0.27, 0.46)</td>
<td>5.18(3.64, 7.35)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Painful Arc*</td>
<td>4(n=756)</td>
<td>0.53(0.31, 0.74)</td>
<td>0.76(0.60, 0.84)</td>
<td>2.25(1.24, 4.08)</td>
<td>0.62(0.37, 1.03)</td>
<td>3.66(1.24, 10.81)</td>
<td></td>
</tr>
<tr>
<td>SLAP</td>
<td>Active Compression*</td>
<td>6(n=782)</td>
<td>0.67(0.51, 0.80)</td>
<td>0.37(0.22, 0.54)</td>
<td>1.06(0.90, 1.25)</td>
<td>0.89(0.67, 1.20)</td>
<td>1.19(0.76, 1.86)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Speeds†</td>
<td>4(n=327)</td>
<td>0.20(0.05, 0.53)</td>
<td>0.78(0.58, 0.90)</td>
<td>0.90(0.43, 1.90)</td>
<td>1.03(0.86, 1.23)</td>
<td>0.87(0.35, 2.55)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anterior Slide*</td>
<td>4(n=831)</td>
<td>0.17 (0.03, 0.55)</td>
<td>0.86(0.81, 0.89)</td>
<td>1.20(0.22, 6.51)</td>
<td>0.97(0.96, 1.36)</td>
<td>1.24(0.16, 9.47)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crank†</td>
<td>4(n=282)</td>
<td>0.34(0.19, 0.53)</td>
<td>0.75(0.65, 0.83)</td>
<td>1.36(0.84, 2.21)</td>
<td>0.88(0.69, 1.12)</td>
<td>1.54(0.75, 3.18)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yergason’s</td>
<td>3(n=246)</td>
<td>12.4(6.60, 20.6)</td>
<td>95.3(90.6, 98.1)</td>
<td>2.49(0.97, 6.40)</td>
<td>0.91(0.84, 0.99)</td>
<td>2.67(0.99, 7.73)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relocation</td>
<td>3(n=246)</td>
<td>51.6(41.2, 61.8)</td>
<td>52.4(44.0, 60.6)</td>
<td>1.13(0.88, 1.45)</td>
<td>0.93(0.72, 1.20)</td>
<td>1.23(0.72, 2.11)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biceps Palpation</td>
<td>2(n=114)</td>
<td>38.6(26.0, 52.4)</td>
<td>66.7(52.9, 78.6)</td>
<td>1.06(0.66, 1.68)</td>
<td>0.95(0.74, 1.22)</td>
<td>1.13(0.51, 2.50)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Compression Rotation†</td>
<td>2(n=355)</td>
<td>24.5(13.8, 38.3)</td>
<td>78.0(72.9, 82.5)</td>
<td>2.81(0.20, 39.70)</td>
<td>0.87(0.66, 1.16)</td>
<td>3.39(0.15, 74.78)</td>
<td></td>
</tr>
<tr>
<td>Anterior Instability</td>
<td>Relocation†</td>
<td>3(n=509)</td>
<td>64.6(54.9, 73.4)</td>
<td>92.0(86.8, 93.0)</td>
<td>5.48(0.56, 53.8)</td>
<td>0.55(0.24, 1.27)</td>
<td>10.64(0.32, 354.10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Apprehension</td>
<td>2(n=409)</td>
<td>65.6(52.7, 77.1)</td>
<td>95.4(93.3, 97.8)</td>
<td>17.21(10.02, 29.55)</td>
<td>0.39(0.22, 0.68)†</td>
<td>53.60(24.29, 118.30)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surprise</td>
<td>2(n=128)</td>
<td>81.8(69.1, 90.9)</td>
<td>86.1(72.1, 94.7)</td>
<td>5.42(0.96, 30.52)†</td>
<td>0.25(0.08, 0.78)†</td>
<td>28.10(7.71, 102.45)</td>
<td></td>
</tr>
<tr>
<td>Tendinopathy</td>
<td>H-K</td>
<td>3(n=738)</td>
<td>65.6(50.3, 70.5)</td>
<td>62.8(57.3, 68.1)</td>
<td>1.86(1.47, 2.34)</td>
<td>0.46(0.36, 0.60)</td>
<td>4.68(3.35, 6.53)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Labral Tear</td>
<td>Crank</td>
<td>3(n=187)</td>
<td>57.3(47.2, 67.0)</td>
<td>72.6(61.8, 81.8)</td>
<td>2.44(0.69, 8.59)</td>
<td>0.51(0.21, 1.22)</td>
<td>5.81(0.47, 71.50)</td>
</tr>
</tbody>
</table>

SN= sensitivity, SP=specificity, +LR=positive likelihood ratio, -LR=negative likelihood ratio, DOR=diagnostic odds ratio, CI=confidence interval, SLAP=……., *HSROC/Bivariate models and all others use DerSimonian-Laird random-effects models. †indicates those studies and properties demonstrating significant heterogeneity (p>0.10).
DISCUSSION
This is the first study on diagnostic accuracy of which we know that has incorporated HSROC/bivariate models as the criterion standard during performance of a meta-analysis of ShPE tests. We feel that the use of this criterion standard promotes increased attention on and isolation of studies that demonstrate results dramatically outside others of similar context. Of particular interest, is the dramatic change in both the 95% CI and 95% prediction region of the active compression test for a SLAP lesion when the original study60 is eliminated (figure 6). Further, this study60 is a good example of the effect of bias on estimates of diagnostic accuracy given that the publication possesses examples of at least seven kinds of bias. Most notable of these biases, is partial verification bias which has been shown to overestimate the diagnostic accuracy of a test.62

For each diagnostic category, the overall results of this systematic review and meta-analysis indicate that a few tests are helpful to confirm or screen for a given diagnosis. There is a preponderance of evidence about individual physical examination tests that could not be combined for the meta-analysis. For those tests, we have used the diagnostic values and risk of bias from the Q2 to determine which tests are recommended for use as a screen or those recommended as a confirmatory test using the benchmarks of specificity >80%, sensitivity >80%, LR+ ≥ 5.0 and LR− ≤ 0.20. The list is short, and confidence in the diagnostic accuracy estimates is tenuous.

From the meta-analysis portion of this review, the Hawkins-Kennedy initially appears to be of value in ruling out subacromial impingement when negative. However, the LR− is poor and further, a strong argument can be made that subacromial impingement is not a valuable diagnosis but rather a cluster of diagnoses.63 The diagnosis of subacromial impingement encompasses such a broad range of pathologies, from bursitis to a complete rotator cuff tear,64 that a label of subacromial impingement may not help guide treatment.65 Yergason’s test, used for detection of a SLAP lesion, has high (95%) pooled specificity. However, the sensitivity is so low, that a positive test modifies the post-test probability of detecting a SLAP lesion only a small amount. In a similar perspective to subacromial impingement, authors have argued that tests results for SLAP may be effected by the percentage of different forms of Snyder classifications present within the sample.50 Therefore, the only tests that appear to have good clinical utility are the apprehension, relocation, and surprise tests to diagnose anterior instability and these tests are primarily a continuum of the apprehension test. When a patient registers apprehension with this test, the relocation manoeuvre should then decrease apprehension, whereupon, the relocation force is reversed causing a surprised reaction (surprise test) by the patient as the apprehension reappears.

While the results of the meta-analysis were, perhaps, not inspiring to the clinician searching for diagnostic answers, there are some individual tests that warrant further investigation.

Table 4  Best* Test Combinations and Reported Value for Various Pathologies

<table>
<thead>
<tr>
<th>Test Combination</th>
<th>Pathology</th>
<th>Lead Author</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive LR</th>
<th>Negative LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive Distraction and Active Compression</td>
<td>SLAP</td>
<td>Schlecter65</td>
<td>70</td>
<td>90</td>
<td>7.00</td>
<td>.11</td>
</tr>
<tr>
<td>Compression-rotation AND Apprehension AND Speed</td>
<td>Type II SLAP</td>
<td>Ok55</td>
<td>25</td>
<td>92</td>
<td>3.13</td>
<td>0.82</td>
</tr>
<tr>
<td>Anterior Slide AND Cranck</td>
<td>Labral Tear</td>
<td>Walsworth44</td>
<td>34</td>
<td>91</td>
<td>3.75</td>
<td>0.73</td>
</tr>
<tr>
<td>Apprehension AND Relocation</td>
<td>Labral Tear</td>
<td>Guanche66</td>
<td>38</td>
<td>93</td>
<td>5.43</td>
<td>0.67</td>
</tr>
<tr>
<td>Age≥30, Painful Arc, Self-report of Popping or Clicking</td>
<td>Supraspinatus Tendinopathy</td>
<td>Chew32 ≥ 2 positive tests; 3 positive tests</td>
<td>75, 38</td>
<td>81, 99</td>
<td>3.82, 32.20</td>
<td>0.32, 0.63</td>
</tr>
<tr>
<td>Age≥65 AND Weakness in ER (Infraspinatus Test) AND Night Pain</td>
<td>RC Tear</td>
<td>Litaker47</td>
<td>49</td>
<td>95</td>
<td>9.84</td>
<td>0.54</td>
</tr>
<tr>
<td>Hawkins-Kennedy, Neer, Painful Arc, Empty Can, Resisted ER</td>
<td>Subacromial impingement</td>
<td>Michener64 ≥ 3 positive tests</td>
<td>75</td>
<td>74</td>
<td>2.93</td>
<td>0.34</td>
</tr>
<tr>
<td>Lift-off and/or Resisted IR</td>
<td>Subscapularis Tendinopathy; Subscapularis Tear</td>
<td>Naredo51; Naredo53</td>
<td>50, 50</td>
<td>84, 95</td>
<td>3.13, 10.0</td>
<td>0.60, 0.53</td>
</tr>
<tr>
<td>Apprehension AND Relocation</td>
<td>Anterior Instability</td>
<td>Farber66</td>
<td>81</td>
<td>98</td>
<td>39.68</td>
<td>0.19</td>
</tr>
</tbody>
</table>

*Best is defined as the highest sensitivity, specificity, or both from the studies with the least bias.
The posterior apprehension test for posterior instability demonstrated a higher specificity and positive likelihood ratio but these values came from a high bias study.39 Another highly specific test, but from a low bias study,40 is the passive distraction test for a SLAP lesion. This test may rule in a SLAP lesion when positive. Sensitive tests of note are the shoulder shrug sign, for stiffness-related disorders (osteoarthritis and adhesive capsulitis) as well as rotator cuff tendinopathy and the Whipple test for massive rotator cuff tears. However, the diagnostic properties of the Whipple test come from a high bias study.39 Other tests of possible value from high bias studies included the AC resisted extension,39 the resisted belly press,38 and coracoid palpation.48 There are six additional tests of interest since minor muscle, and the active compression test may have value as a confirmary test for AC joint pathology when positive due to its high specificity.

Despite some cause for optimism when looking at some of the individual studies and tests, the more prudent method may be to look at clusters or combinations of tests, since that resembles more closely, the way in which most ShPE tests are used in the clinic. Table 4, while not all-inclusive, shows the best test combinations to date for detecting various pathologies.

Unfortunately, even many of these test clusters modify the post-test probability by a small to minimal amount. Of note in this group of clustered tests is the combination of age>39, painful arc, and self-report of popping and clicking and the combination of the apprehension and relocation tests, both of which produce a large post-test shift toward the diagnoses of supraspinatus tendinopathy, and anterior instability, respectively.

LIMITATIONS
Any review is limited by the quality of studies contained therein. Many of the studies in this review had issues with the reference standard and subject flow and timing. There was clearly a rise in the use of diagnostic ultrasound as a criterion standard, and evidence to supports its use is currently poor.56–58 Further, we limited our articles to those in the English language which may make this review more prone to dissemination bias. However, publication bias was not found to be evident with graphical or in statistical analysis. Finally, this is the first meta-analysis on diagnostic accuracy of ShPE tests that was performed using the Q2 document. The original authors piloted the Q2 on five studies and found that reliability varied considerably.44 Our weighted $\kappa$ ($\kappa=0.31; 0.10, 0.52$) was likewise only fair.

CONCLUSIONS
Based on data from our original review and this update, the use of any single ShPE test to make a pathognomonic diagnosis cannot be unequivocally endorsed due to continued quality issues in publications. Some ShPE tests are beginning to stand up to scrutiny and time but there are far more tests that need to be validated in more than one study. Combinations of ShPE tests provide better accuracy, but marginally so. These findings seem to provide support for stressing a comprehensive clinical examination including history and clinical examination. However, there is a great need for large, prospective, well-designed studies that examine the diagnostic accuracy of the many aspects of the clinical examination and what combinations of these aspects are useful in differentially diagnosing pathologies of the shoulder.

Acknowledgements The authors would like to acknowledge Ms Connie Schardt for her invaluable assistance in the search process and the authors from the original paper whose initial work was foundational: S Campbell, A Morin, M Tamaddoni, C T Moorman III.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

References to this paper are available online at http://bjsm.bmjgroup.com

References to this paper are available online at http://bjsm.bmjgroup.com


978
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

Eric J Hegedus, Adam P Goode, Chad E Cook, et al.

doi: 10.1136/bjsports-2012-091066