AJSM PreView, published on December 18, 2008 as doi:10.1177/0363546508325153

A Meta-analysis Examining Clinical Test Utility for Assessing Superior Labral Anterior Posterior Lesions

Brent B. Meserve,*† DPT, Joshua A. Cleland,‡§ DPT, PhD, OCS, and Thomas R. Boucher,‖ PhD
From †Department of Rehabilitative Medicine, Dartmouth Hitchcock Medical Center, Lebanon, New Hampshire, ‡Department of Physical Therapy, Franklin Pierce College, Concord, New Hampshire, §Rehabilitation Services of Concord Hospital, Concord, New Hampshire, and ‖Department of Mathematics, Plymouth State University, Plymouth, New Hampshire

Background: The reported accuracy of clinical tests for superior labral anterior posterior lesions is extremely variable. Pooling results from multiple studies of higher quality is necessary to establish the best clinical tests to use.

Hypothesis: Certain clinical tests are superior to others for diagnosing the presence or absence of a superior labral anterior posterior lesion.

Study Design: Meta-analysis.

Methods: A literature search of MEDLINE (1966-2007), CINAHL (1982-2007), and BIOSIS (1995-2007) was performed for (labrum OR labral OR SLAP OR Bankart) AND (shoulder OR shoulder joint OR glenoid) AND (specificity OR sensitivity AND specificity). Identified articles were reviewed for inclusion criteria. Sensitivity and specificity values were recorded from each study and used for meta-analysis.

Results: Six of 198 identified studies satisfied the eligibility criteria. Active compression, anterior slide, crank, and Speed tests were analyzed using receiver operating characteristic curves. The accuracy of the anterior slide test was significantly inferior to that of the active compression, crank, and Speed tests. There was no significant difference in test accuracy found among active compression, crank, and Speed tests. Between studies, methodological scores did not significantly affect sensitivity and specificity values.

Conclusion: The anterior slide test is a poor test for detecting the presence of a labral lesion in the shoulder. Active compression, crank, and Speed tests are more optimal choices. Clinicians should choose the active compression test first, crank second, and Speed test third when a labral lesion is suspected.

Keywords: physical examination; sensitivity; specificity; superior labral anterior posterior (SLAP) lesions

The glenohumeral joint consists of the humeral head, glenoid of the scapula, and the labrum. The purpose of the labrum is to deepen the glenoid and improve joint congruency. The glenohumeral ligaments, joint capsule, rotator cuff muscles, and labrum all provide stability to the glenohumeral joint. The labrum is vulnerable to injury, especially with aggressive overhead activity. Superior labral lesions were first described in 1985 by Andrews et al[1] and identified in 73 overhead-throwing athletes. The prevalence estimates of shoulder pain range from 4% to 8% for primary care physician visits.[46] However, clinically it is difficult to determine what number of these cases is the direct result of labral injury. Magnetic resonance imaging with and without contrast has demonstrated high diagnostic accuracy for detecting labral lesions.[21,25,26,41,48] However, with the cost of this modality for imaging the shoulder averaging US $2400 in 1997,[34] overuse can be problematic and often not result in any alteration in patient treatment.[9] Therefore, a thorough clinical examination using accurate clinical tests is important. Clinical tests designed to detect the presence of a labral tear vary considerably in reported sensitivity and specificity values.[¶] Tests with low accuracy typically do not provide clinicians with useful information for making a diagnosis. When sensitivity and

*Address correspondence to Brent B. Meserve, DPT, Dartmouth-Hitchcock Medical Center, One Medical Drive, Lebanon, NH 03756 (e-mail: Brent.B.Meserve@hitchcock.org).

No potential conflict of interest declared.

The American Journal of Sports Medicine, Vol. X, No. X
DOI: 10.1177/0363546508325153

¶References 2, 8, 11, 15-17, 19, 20, 22, 23, 27, 29-33, 36, 47.

specificity levels are at or below 50%, then the test has a predictive value equal to or lower than chance for detecting the presence or absence of a superior labral anterior posterior (SLAP) lesion. Therefore, meta-analysis of multiple high-quality studies may add additional information.

Liu et al[23] reported 90% sensitivity and 85% specificity of physical examination compared to arthroscopy for detecting labral lesions. Physical examination included the apprehension, relocation, load and shift, inferior sulcus sign, and crank tests. Accurate clinical tests are especially important when attempting to identify the presence of labral tears to facilitate the appropriate plan of care, which in this case is likely surgical intervention. Significant improvement in function after arthroscopic debridement or repair of symptomatic labral lesions has been consistently reported.[4,18,37] Furthermore, increased certainty of labral injury will assist clinicians with treatment strategies, that is, immobilization, shoulder strengthening, surgery, and/or activity modification.

Recent systematic reviews have reported considerable variability in sensitivity and specificity values for several clinical tests for shoulder disorders.[8,15,30] However, none of these reviews were performed using meta-analysis. Mirkovic et al[30] used a modified quality assessment scale designed for randomized control trials for their systematic review of shoulder tests. This tool has not been validated for the evaluation of diagnostic tests. Studies should analyze the reported test accuracy data with a reliable and valid instrument and include the study methodological quality into analysis.[6,14]

Bias and data variation exist in diagnostic accuracy studies.[38,39,49,50] The variation in part is owing to threshold differences between studies and inappropriately pooling estimates of sensitivity and specificity across studies. Receiver operating characteristic (ROC) curves are a statistical technique using linear regression to describe the accuracy or performance of a diagnostic test by plotting predicted true-positive rates (y-axis) versus given false-positive rates (x-axis). They are used to control for test threshold differences between studies and provide a means of interpreting the relationship between the 2 measures of test accuracy: sensitivity and specificity.[14,39,40] The purpose of this meta-analysis is to perform a review of the literature, statistically analyze available data using ROC curves, and interpret the findings for clinicians regarding the diagnostic utility of clinical tests for the identification of SLAP lesions. We hypothesize that certain physical tests will be found to be superior to others for diagnosing the presence or absence of a SLAP lesion.

METHODS

Search Strategy

The principal investigator (B.B.M.) performed an extensive computer database literature search to identify all articles written in English that satisfied the eligibility criteria of this meta-analysis. The following search terms were used: (labrum OR labral OR SLAP OR Bankart) AND (shoulder OR shoulder joint OR glenoid) AND (specificity OR sensitivity AND specificity). The electronic databases searched included MEDLINE[28] (1966-June 2007), CINAHL[5] (1982-June 2007), and BIOSIS[3] (1995-June 2007). Articles were first selected according to title relevance. Abstracts and/or full-text articles were then reviewed. Reference tracking was performed on selected articles.

Inclusion/Exclusion Criteria

Studies were eligible for inclusion if they investigated the diagnostic accuracy of at least 1 clinical test for identifying labral lesions of the shoulder and used arthrotomy, arthroscopy, or MRI as reference standards. Studies that did not report clinical test negatives or used "nondiseased" controls were excluded. Studies were also excluded if they were performed on cadavers only. If a clinical test was evaluated in only a single study, then the data were not eligible for meta-analysis and therefore not reported in this study. In addition, studies reporting specificities and sensitivities of 0 or 1 were excluded because in these cases the logit sensitivities or logit specificities used in ROC analysis are undefined. For instance, the logit sensitivity is defined to be log(sensitivity/(1 − sensitivity)), which is undefined when the sensitivity is either 0 or 1. The logit function is the inverse of the logistic function. The logistic function is often used in logistic regression to model how the probability p of an event may be affected by 1 or more explanatory variables: an example would be to have the model $p = P(a + bx)$, where x is the explanatory variable and a and b are model parameters to be fitted. The explanatory variables in this case are the sensitivity and specificity values.

Study Selection

One reviewer (B.M.M.) identified all eligible studies by reviewing each article title (those identified during the database searches) and reading the abstract. If the reviewer was uncertain if the eligibility criterion was satisfied after reading the abstract, the full text was retrieved and examined for the presence of inclusion and exclusion criteria.

Assessment of Methodological Quality and Data Abstraction

Each selected article was scored according to a quality checklist adopted from Irwig and colleagues and the Cochrane Methods Group on Systematic Review of Screening and Diagnostic Tests (see Appendix 1, available in the online version of this article at http://ajs.sagepub.com/supplemental/).[6,14] This checklist consists of 16 items. Each item varied in the total number of possible choices, ranging from 3 to 6. For example, items assigned a 0 to 2 quality range had 3 possible choices, 0 to 3 quality range had 4 choices, 0 to 4 quality range had 5 choices, and 0 to 5 quality range had 6 choices. Therefore, the total possible points obtainable from the 16-item questionnaire equaled 37. Only studies that scored 19 points or more on the quality checklist were included in the meta-analysis. This method is consistent with previous published systematic reviews.[7,44,45] Previous systematic reviews have also used the Cochrane Methods Group scoring system to assess study quality, without assigning a value to checklist items.[42,43]

A second scoring criterion assigned a numeric value of 1 to 4 to each article, indicating the strength of the evidence.[12] Level 1 articles were those that used an independent, blinded comparison to a reference standard of at least 50 consecutive and relevant patients. Level 2 articles were similar in methods to level 1 but contained fewer than 50 patients. Level 3 articles used independent and blind comparison to a reference standard without consecutive patient examinations. Level 4 articles lacked independent comparison and consecutive collection of patients, but the examination was compared with 1 of the above mentioned reference standards.

Statistical Analysis

Sensitivity and specificity values were extracted from each article for analysis. An analysis of covariance (ANCOVA) was conducted with logit sensitivity as dependent variable, logit specificity as covariate, and diagnostic test as the factor. The fitted ANCOVA model was used to predict logit sensitivities at various logit specificities for each of the diagnostic tests. These logit specificity–predicted logit sensitivity pairs were used to calculate specificity and predicted sensitivity values for each diagnostic test and, in turn, to produce an ROC curve for each test. An ROC curve is a plot of sensitivity versus (1 – specificity) for a diagnostic test as the test's discrimination threshold is varied. The ROC can be represented equivalently by plotting the fraction of true positives versus the fraction of false positives for a diagnostic test. Appendix 2 (available in the online version of this article at http://ajs.sagepub.com/supplemental/) provides a glossary of terms. Minitab 14.20 and R 2.1.1 were the computer software programs used for statistical analysis.

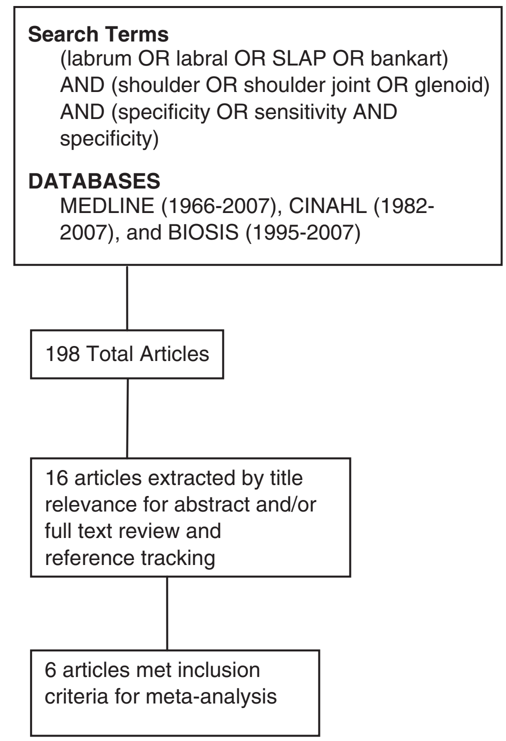

Figure 1. Flowchart of search results.

Study Selection

The literature search and reference tracking revealed 198 potential articles (Figure 1). Many different diagnostic SLAP lesion tests were evaluated over the collection of studies, most with 3 or fewer replications. These tests were

TABLE 1
List of Excluded Studies

Author	N	Reference Standard	Test
Kim et al,[17] 1999	75	Arthroscopy	Biceps load test I
Kim et al,[16] 2001	127	Arthroscopy	Biceps load test II
Nakagawa et al,[32] 2005	54	Arthroscopy	Clunk
Nakagawa et al,[32] 2005	54	Arthroscopy	Forced shoulder abduction and elbow flexion
Parentis et al,[36] 2006	132	Arthroscopy	Jobe relocation
Nakagawa et al,[32] 2005	54	Arthroscopy	Jobe relocation
Guanche and Jones,[11] 2003	60	Arthroscopy	Jobe relocation
Kim et al,[19] 2005	172	Arthroscopy	Kim test
Mimori et al,[29] 1999	32	MRI/arthrography	Pain provocation
Parentis et al,[36] 2006	132	Arthroscopy	Pain provocation
Kim et al,[20] 2007	61	Arthroscopy	Passive compression
Kim et al,[19] 2005	172	Arthroscopy	Posterior jerk
Nakagawa et al,[32] 2005	54	Arthroscopy	Posterior jerk
Holtby and Razmjou,[13] 2004	152	Arthroscopy	Yergason
Guanche and Jones,[11] 2003	60	Arthroscopy	Yergason
Parentis et al,[36] 2006	132	Arthroscopy	Yergason
Nakagawa et al,[32] 2005	54	Arthroscopy	Yergason
Myers et al,[31] 2005	40	Arthroscopy	Supination-external rotation

TABLE 2
Patient Characteristics and Reported Data From Included Studies

Study	Level of Evidence[42]	Mean Symptom Duration	Gender Distribution, Male/Female	Patient Population	Mean Age, y	Quality Score	N	Reference Standard	Test	Sensitivity	Specificity
McFarland et al,[27] 2002	3	Not specified	252/174	Shoulder pain	45	30	426	Arthroscopy	Active compression	0.47	0.55
									Anterior Slide	0.08	0.84
Parentis et al,[36] 2006	1	Not specified	98/34	Shoulder pain excluding stiffness	42	28	132	Arthroscopy	Active compression	0.625	0.5
									Speed	0.478	0.674
									Crank	0.125	0.826
									Anterior slide	0.1	0.815
Stetson et al,[47] 2002	3	12 months	45/20	Shoulder pain	46	25	65	Arthroscopy	Active compression	0.54	0.31
									Crank	0.46	0.56
Myers et al,[31] 2005	3	Not specified	39/1	Shoulder in pain for athletes; excluded individuals older than 50 years or those with athletic injury	24	30	40	Arthroscopy	Active compression	0.778	0.111
									Crank	0.346	0.7
Guanche and Jones,[11] 2003	3	Not specified	48/11	Shoulder pain excluding stiffness	38	27	60	Arthroscopy	Active compression	0.63	0.73
									Crank	0.4	0.73
									Speed	0.18	0.87
Nakagawa et al,[32] 2005	3	Not specified	52/2	Shoulder pain in throwing athletes, excluded major shoulder trauma	23	32	54	Arthroscopy	Active compression	0.54	0.6
									Anterior slide	0.05	0.93
									Crank	0.58	0.72
									Speed	0.04	0.99

not included in the analysis because the small number of replicated studies for each of these tests would make it difficult to detect any statistically significant differences among them (Table 1). These tests included active compression, anterior apprehension, anterior slide, biceps load test I and II, Speed's, clunk, compression-rotation, crank, forced shoulder abduction and elbow flexion, jerk, Jobe relocation, Kim, pain provocation, passive compression, posterior jerk, Yergason's, and supination-external rotation. Seven of these tests were evaluated individually in separate studies (biceps load test I and II, clunk, forced shoulder abduction and elbow flexion, Kim, passive compression, and supination-external rotation).[16,17,19,20,31,32] In other words, the diagnostic accuracy of these clinical tests was not evaluated in more than 1 study. Pain provocation[29,36] and posterior jerk[19,32] tests were evaluated across 4 studies, Jobe relocation across 3 studies,[11,32,36] and Yergason's across 4 studies.[11,13,32,36] Six studies[11,27,31,32,36,47] reported sensitivity and specificity data for 4 tests: Speed, active compression, crank, and anterior slide. All satisfied the eligibility and quality criteria for the meta-analysis with methodological quality scores ≥19. Magee[24] has described the execution of these clinical tests in detail (see Appendix 3, available in the online version of this article at http://ajs.sagepub.com/supplemental/). Patient and study characteristics for these 6 studies are presented in Table 2.

TABLE 3
Frequency Table of Study Methodological Score for Quality Categories

	Quality Score Category			
	25-27	28	30	32
Frequency	5	4	4	4

There is large variability from study to study in the reported sensitivities and specificities for the same diagnostic test. This makes it necessary to control for the effect of the study on the sensitivities and specificities of each diagnostic test. This was done by selecting tests with a small number of studies distributed as uniformly as possible across the diagnostic tests. Also, this meta-analysis is concerned with the effect of methodological quality on the reported study results. Studies were grouped by methodological score quality into 4 categories (25-27, 28, 30, and 32). This produces the best fit to a balanced distribution of replication and methodological score categories. The summary of these groupings is in Table 3. Studies distributed as uniformly as possible across the methodological score categories, when cross-tabulated with diagnostic test, were selected. Therefore, 17 test comparisons across 6 individual studies were chosen for ROC analysis. This collection of data achieves a good balance over the 2 confounders (methodological quality and study) and against the factor (diagnostic test). In particular, all of the selected studies evaluated the active compression diagnostic test. This aided in testing the effect of individual studies on the diagnostic test accuracy.

This collection of studies is also fairly well balanced in terms of methodological score category versus diagnostic test. The individual studies were consistent across diagnostic tests in their methodological quality. This suggests a consistency in methodological quality across diagnostic tests, which was verified when the effect of methodological quality on study results was not found to be significant.

RESULTS

Analysis of variance of the effect of study methodological quality on each of the logits found no significant effect (logit sensitivity, P = .658; logit specificity, P = .2529). Residual diagnostics were consistent with model assumptions. Separate analyses of variance found no significant difference between studies for either logit (logit sensitivity, P = .7163; logit specificity, P = .2310). Residual diagnostics were again consistent with model assumptions. The effects due to study methodological quality and study were not found to be significant. Therefore, the logits were not corrected for an effect due to either.

The logits were negatively correlated (–0.85), suggesting differing thresholds among the studies and that an ROC analysis is appropriate. The ANCOVA with logit sensitivity as the response, logit specificity as the covariate, and diagnostic test as the factor found a highly significant slope for

TABLE 4
Analysis of Covariance Results for Individual Effects on Logit Sensitivity

Coefficient	Estimate	SE	*t*	*P*
Logit specificity	–0.6018	0.1389	–4.333	<.001
Active compression	0.2879	0.2240	1.285	.223
Anterior slide	–1.667	0.4846	–3.440	<.005
Crank	–0.3012	0.3649	–0.826	.425
Speed	–0.4333	0.5295	–0.818	.429

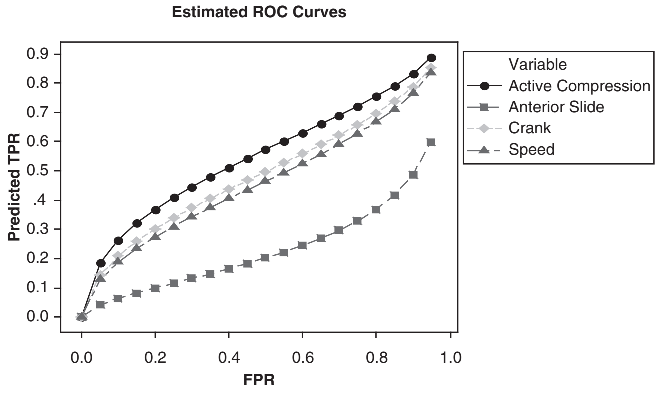

Figure 2. Estimated receiver operating characteristic (ROC) curves for the active compression, anterior slide, crank, and Speed tests.

logit specificity and a highly significant effect of the anterior slide test on logit sensitivity. The overall P value of the ANCOVA was <.0001. Fit of the model was good, with an adjusted R^2 of 0.8371. The ANCOVA results for each of the explanatory variables are given in Table 4. The Shapiro-Wilks test for normality indicates the residuals are consistent with a normal distribution (P = .61). Other residual diagnostics were also consistent with model assumptions.

The parameter estimates in Table 4 were used to predict, for each diagnostic test, true-positive rates at each sequence of false-positive rates. These predicted true-positive/false-positive pairs for each diagnostic test define the estimated ROCs for that diagnostic test. The plot of these true-positive/false-positive pairs is the estimated ROC curve for that test. The estimated ROC curves of the 4 tests are given in Figure 2. The anterior slide test was found to be statistically inferior to the active compression, crank, and Speed with predicted true-positive rates and consistently lower for all false-positive rates. No significant differences were found for ROC curves among the active compression, crank, or Speed tests. The active compression test was found to have the best ROC curve.

DISCUSSION/CONCLUSION

The studies selected for this ROC analysis were more recent than were those not selected. The selected studies examined multiple tests and were consistent in their methodological quality, ensuring a consistency in both methodological quality and effect on results due to the study. This was reflected in the results of the statistical analysis. It is an encouraging sign that higher quality studies of diagnostic tests for tears of the labrum are now being performed. The ROC analysis compared 4 diagnostic tests: active compression, anterior slide, crank, and Speed. The diagnostic utility of the anterior slide was found to be significantly inferior to the other 3; among the latter, no significant differences were detected.

Sensitivity values of the active compression test ranged from 47% to 78%. Crank test sensitivity values ranged from 13% to 58%. Speed test sensitivity values were the lowest, ranging from 4% to 48%. Specificity values were higher. Active compression specificity values ranged from 11% to 73%. Crank test specificity values ranged from 56% to 83%. Speed test specificity values were the highest, ranging from 67% to 99%. Positive crank and Speed test results strongly suggest the presence of a labral tear. No test in the analysis strongly excluded the presence of labral tear.

Studies not included in this analysis have reported higher sensitivity and specificity values for these tests. O'Brien et al[33] examined 268 patients with shoulder pain and 55 patients with no shoulder pain with the active compression test. The test was reported to have 100% sensitivity and 96% specificity. This study was excluded from our analysis because 100% sensitivity values are undefined in ROC analysis. Liu et al[22] used the crank test to examine 62 patients with refractory shoulder pain over 3 months; 91% sensitivity and 93% specificity values were reported. This study was also excluded from our analysis because it examined only 1 test.

Some studies not included in this analysis have reported high sensitivity and/or specificity values for other clinical tests for labral tears. Berg and Ciullo[2] described a SLAPprehension sign in which pain is reproduced with active horizontal shoulder adduction and internal rotation with a pronated forearm. This study reported a sensitivity value of 87.5% in a retrospective review of 66 patients with SLAP lesions. Kim et al[16,17,19,20] examined the effectiveness of 4 clinical tests for labral tears: biceps load, biceps load II, passive compression, and the Kim tests. Sensitivity and specificity values ranged from 80% to 89.7% and 85.7% to 96.9%, respectively. Therefore, these tests may be good alternatives to the active compression, Speed, anterior slide, or crank tests. These clinical tests were developed by the respective authors of the studies and had not been further evaluated at the time of this systematic review. Replication of these studies is necessary to determine the diagnostic utility of these tests.

When clinicians suspect the presence of a labral tear, the active compression test should be chosen first, crank test second, and Speed test third. The anterior slide test was significantly inferior to the others and should be avoided. The active compression is more sensitive and thus should be used initially to rule out a labral tear. The Speed test is the most specific and should be used to rule in a labral tear if the active compression test or crank test has positive results. A positive sign with all 3 tests increases the probability that a labral tear is present. Biceps load, passive compression, and Kim tests may be good alternatives; however, the accuracy of these tests has not been replicated.

It is difficult to determine what makes one SLAP lesion test better than another. Mechanical studies have examined the content validity of the active compression test and found it to be poor.[10] However, the active compression test has been found to be one of the best tests for diagnosing SLAP lesions.[35]

STUDY LIMITATIONS

Articles not written in English were excluded. It is possible that articles supporting or refuting the diagnostic utility of these tests could be published in languages other than English. All studies were selected and assessed for methodological quality by 1 reviewer (B.M.M.). This creates the possibility of reviewer bias. Varying "positive" test criteria, that is, pain, click, or both, among the primary studies is another potential source of error and bias among original studies. It is also possible that tests may have been executed in ways that are inconsistent with the description of the original authors. The variability between authors in choosing different reference standards, that is, arthrotomy, arthroscopy, and MRI, may create a confounding variable in these findings. In addition, the standard way SLAP lesions are diagnosed has changed over time with technological advancement. These factors can potentially affect the results of this systematic review. The decision to omit tests that were replicated in 3 or fewer studies potentially eliminates tests with good accuracy. Inclusion criteria may have been too rigid. Only studies with a balanced distribution of methodological quality were included.

CLINICAL MESSAGES

- Active compression, crank, and Speed tests are more accurate for detecting labral tears than is the anterior slide test.
- Reported sensitivity and specificity values ranged from low to high.
- Active compression test is the most sensitive and Speed test the most specific.
- Bicep load, passive compression, and Kim tests may be good alternatives, but more research is warranted.

REFERENCES

1. Andrews JR, Carson WG, Jr, McLeod WD. Glenoid labrum tears related to the long head of the biceps. *Am J Sports Med*. 1985;13(5):337-341.
2. Berg EE, Ciullo JV. A clinical test for superior glenoid labral or "SLAP" lesions. *Clin J Sport Med*. 1998;8(2):121-123.
3. BIOSIS [database online]. Philadelphia, Pa: Thomson Reuters; 2008.
4. Carreira DS, Mazzocca AD, Oryhon J, Brown FM, Hayden JK, Romeo AA. A prospective outcome evaluation of arthroscopic Bankart repairs: minimum 2-year follow-up. *Am J Sports Med*. 2006; 34(5):771-777.

5. CINAHL [database online]. Ipswich, Mass: EBSCO Publishing; 2008.
6. Cochrane Methods Group on Systematic Review of Screening and Diagnostic Tests. Recommended methods. Available at: som.flinders .edu.au/fusa/cochrane/. Retrieved June 1996.
7. de Graaf I, Prak A, Bierma-Zeinstra S, Thomas S, Peul W, Koes B. Diagnosis of lumbar spinal stenosis: a systematic review of the accuracy of diagnostic tests. *Spine*. 2006;31(10):1168-1176.
8. Dinnes J, Loveman E, McIntyre L, Waugh N. The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. *Health Technol Assess (Winchester, England)*. 2003;7(29):iii, 1-166.
9. Gillan MG, Gilbert FJ, Andrew JE, et al. Influence of imaging on clinical decision making in the treatment of lower back pain. *Radiology*. 2001;220(2):393-399.
10. Green RA, Taylor NF, Mirkovic M, Perrott M. An evaluation of the anatomic basis of the O'Brien active compression test for superior labral anterior and posterior (SLAP) lesions. *J Shoulder Elbow Surg*. 2008;17(1):165-171.
11. Guanche CA, Jones DC. Clinical testing for tears of the glenoid labrum. *Arthroscopy*. 2003;19(5):517-523.
12. Holleman DR Jr, Simel DL. Does the clinical examination predict airflow limitation? *JAMA*. 1995;273(4):313-319.
13. 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(3):231-236.
14. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol*. 1995;48(1):119-130; discussion 131-132.
15. Jones GL, Galluch DB. Clinical assessment of superior glenoid labral lesions: a systematic review. *Clin Orthop Relat Res*. 2007;455:45-51.
16. 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(2):160-164.
17. Kim SH, Ha KI, Han KY. Biceps load test: a clinical test for superior labrum anterior and posterior lesions in shoulders with recurrent anterior dislocations. *Am J Sports Med*. 1999;27(3):300-303.
18. Kim SH, Ha KI, Kim SH, Choi HJ. Results of arthroscopic treatment of superior labral lesions. *J Bone Joint Surg Am*. 2002;84(6):981-985.
19. Kim SH, Park JS, Jeong WK, Shin SK. The Kim test: a novel test for posteroinferior labral lesion of the shoulder. A comparison to the jerk test. *Am J Sports Med*. 2005;33(8):1188-1192.
20. Kim YS, Kim JM, Ha KY, Choy S, Joo MW, Chung YG. The passive compression test: a new clinical test for superior labral tears of the shoulder. *Am J Sports Med*. 2007;35(9):1489-1494.
21. Kirkley A, Litchfield R, Thain L, Spouge A. Agreement between magnetic resonance imaging and arthroscopic evaluation of the shoulder joint in primary anterior dislocation of the shoulder. *Clin J Sport Med*. 2003;13(3):148-151.
22. Liu SH, Henry MH, Nuccion S, Shapiro MS, Dorey F. Diagnosis of glenoid labral tears: a comparison between magnetic resonance imaging and clinical examinations. *Am J Sports Med*. 1996;24(2):149-154.
23. Liu SH, Henry MH, Nuccion SL. A prospective evaluation of a new physical examination in predicting glenoid labral tears. *Am J Sports Med*. 1996;24(6):721-725.
24. Magee DJ. *Orthopedic Physical Assessment*. 4th ed. Philadelphia, Pa: WB Saunders; 2002.
25. Magee T, Williams D, Mani N. Shoulder MR arthrography: which patient group benefits most? *AJR Am J Roentgenol*. 2004;183(4):969-974.
26. Magee TH, Williams D. Sensitivity and specificity in detection of labral tears with 3.0-T MRI of the shoulder. *AJR Am J Roentgenol*. 2006;187(6):1448-1452.
27. McFarland EG, Kim TK, Savino RM. Clinical assessment of three common tests for superior labral anterior-posterior lesions. *Am J Sports Med*. 2002;30(6):810-815.
28. MEDLINE [database online]. Baltimore, Md: Community of Science; 2008.
29. Mimori K, Muneta T, Nakagawa T, Shinomiya K. A new pain provocation test for superior labral tears of the shoulder. *Am J Sports Med*. 1999;27(2):137-142.
30. Mirkovic M, Green R, Taylor N, Perrott M. Accuracy of clinical tests to diagnose superior labral anterior and posterior (SLAP) lesions. *Phys Ther Rev*. 2005;10(1):5-14.
31. Myers TH, Zemanovic JR, Andrews JR. The resisted supination external rotation test: a new test for the diagnosis of superior labral anterior posterior lesions. *Am J Sports Med*. 2005;33(9):1315-1320.
32. Nakagawa S, Yoneda M, Hayashida K, Obata M, Fukushima S, Miyazaki Y. Forced shoulder abduction and elbow flexion test: a new simple clinical test to detect superior labral injury in the throwing shoulder. *Arthroscopy*. 2005;21(11):1290-1295.
33. O'Brien SJ, Pagnani MJ, Fealy S, McGlynn SR, Wilson JB. The active compression test: a new and effective test for diagnosing labral tears and acromioclavicular joint abnormality. *Am J Sports Med*. 1998;26(5):610-613.
34. Oh CH, Schweitzer ME, Spettell CM. Internal derangements of the shoulder: decision tree and cost-effectiveness analysis of conventional arthrography, conventional MRI, and MR arthrography. *Skeletal Radiol*. 1999;28(12):670-678.
35. Pandya NK, Colton A, Webner D, Sennett B, Huffman GR. Physical examination and magnetic resonance imaging in the diagnosis of superior labrum anterior-posterior lesions of the shoulder: a sensitivity analysis. *Arthroscopy*. 2008;24(3):311-317.
36. Parentis MA, Glousman RE, Mohr KS, Yocum LA. An evaluation of the provocative tests for superior labral anterior posterior lesions. *Am J Sports Med*. 2006;34(2):265-268.
37. Paxinos A, Walton J, Rutten S, Muller M, Murrell GA. Arthroscopic stabilization of superior labral (SLAP) tears with biodegradable tack: outcomes to 2 years. *Arthroscopy*. 2006;22(6):627-634.
38. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.
39. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174(4):469-476.
40. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine*. Philadelphia, Pa: Churchill Livingstone; 2000.
41. Sasaki T, Saito Y, Yodono H, et al. Labral-ligamentous complex of the shoulder: evaluation with double oblique axial MR arthrography. *Acta Radiol*. 2003;44(4):435-439.
42. Scholten RJ, Deville WL, Opstelten W, Bijl D, van der Plas CG, Bouter LM. The accuracy of physical diagnostic tests for assessing meniscal lesions of the knee: a meta-analysis. *J Fam Pract*. 2001;50(11):938-944.
43. Scholten RJ, Opstelten W, van der Plas CG, Bijl D, Deville WL, Bouter LM. Accuracy of physical diagnostic tests for assessing ruptures of the anterior cruciate ligament: a meta-analysis. *J Fam Pract*. 2003;52(9):689-694.
44. Sehgal N, Shah RV, McKenzie-Brown AM, Everett CR. Diagnostic utility of facet (zygapophysial) joint injections in chronic spinal pain: a systematic review of evidence. *Pain Physician*. 2005;8(2):211-224.
45. Shah RV, Everett CR, McKenzie-Brown AM, Sehgal N. Discography as a diagnostic test for spinal pain: a systematic and narrative review. *Pain Physician*. 2005;8(2):187-209.
46. Speed C. Shoulder pain. *Clin Evid*. 2005;14:1543-1560.
47. Stetson WB, Templin K. The crank test, the O'Brien test, and routine magnetic resonance imaging scans in the diagnosis of labral tears. *Am J Sports Med*. 2002;30(6):806-809.
48. Waldt S, Burkart A, Lange P, Imhoff AB, Rummeny EJ, Woertler K. Diagnostic performance of MR arthrography in the assessment of superior labral anteroposterior lesions of the shoulder. *AJR Am J Roentgenol*. 2004;182(5):1271-1278.
49. Westwood ME, Whiting PF, Kleijnen J. How does study quality affect the results of a diagnostic meta-analysis? *BMC Med Res Methodol*. 2005;5(1):20.
50. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Int Med*. 2004;140(3):189-202.