

# Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests

E J Hegedus,<sup>1</sup> A Goode,<sup>1</sup> S Campbell,<sup>1</sup> A Morin,<sup>1</sup> M Tamaddoni,<sup>1</sup> C T Moorman III,<sup>1</sup> C Cook<sup>2</sup>

<sup>1</sup> Duke University, Durham, NC, USA; <sup>2</sup> Center for Excellence in Surgical Outcomes, Duke University, Durham, NC, USA

Correspondence to:  
Eric J Hegedus, Assistant  
Professor, Duke University,  
DUMC 3907, Durham, NC  
27710, USA; [eric.hegedus@duke.edu](mailto:eric.hegedus@duke.edu)

Accepted 9 August 2007  
Published Online First  
24 August 2007

## ABSTRACT

**Objective:** To compile and critique research on the diagnostic accuracy of individual orthopaedic physical examination tests in a manner that would allow clinicians to judge whether these tests are valuable to their practice.

**Methods:** A computer-assisted literature search of MEDLINE, CINAHL, and SPORTDiscus databases (1966 to October 2006) using keywords related to diagnostic accuracy of physical examination tests of the shoulder. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was used to critique the quality of each paper. Meta-analysis through meta-regression of the diagnostic odds ratio (DOR) was performed on the Neer test for impingement, the Hawkins—Kennedy test for impingement, and the Speed test for superior labral pathology.

**Results:** Forty-five studies were critiqued with only half demonstrating acceptable high quality and only two having adequate sample size. For impingement, the meta-analysis revealed that the pooled sensitivity and specificity for the Neer test was 79% and 53%, respectively, and for the Hawkins—Kennedy test was 79% and 59%, respectively. For superior labral (SLAP) tears, the summary sensitivity and specificity of the Speed test was 32% and 61%, respectively. Regarding orthopaedic special tests (OSTs) where meta-analysis was not possible either due to lack of sufficient studies or heterogeneity between studies, the list that demonstrates both high sensitivity and high specificity is short: hornblowers's sign and the external rotation lag sign for tears of the rotator cuff, biceps load II for superior labral anterior to posterior (SLAP) lesions, and apprehension, relocation and anterior release for anterior instability. Even these tests have been under-studied or are from lower quality studies or both. No tests for impingement or acromioclavicular (AC) joint pathology demonstrated significant diagnostic accuracy.

**Conclusion:** Based on pooled data, the diagnostic accuracy of the Neer test for impingement, the Hawkins—Kennedy test for impingement and the Speed test for labral pathology is limited. There is a great need for large, prospective, well-designed studies that examine the diagnostic accuracy of the numerous physical examination tests of the shoulder. Currently, almost without exception, there is a lack of clarity with regard to whether common OSTs used in clinical examination are useful in differentially diagnosing pathologies of the shoulder.

physical findings is important to determine a treatment path and because the ability to correctly diagnose the source of shoulder pain can save the patient from further diagnostic tests that are more costly, painful or inconvenient. Physical examination tests or orthopaedic special tests (OSTs) have historically been an integral part of this process. However, despite the fact that studies on the diagnostic accuracy of OSTs in the shoulder have been published at an accelerated rate, the quality of these publications has been reported to be somewhat suspect.<sup>1–3</sup> Further, many studies<sup>4–8</sup> have questioned both the accuracy and reliability of the clinical examination especially as it relates to a pathoanatomical model. Despite the accelerated rate of publication of diagnostic accuracy studies, we are aware of only two previous systematic reviews that address multiple pathologies of the shoulder.<sup>9–11</sup> The two-article review by Tennent *et al*<sup>10, 11</sup> gave thorough descriptions of each test but failed to examine the diagnostic accuracy of the tests and made no comment on the quality of literature supporting the use of individual OSTs. The review by Dinnes *et al*<sup>9</sup> focused primarily on diagnostic imaging but did include 10 articles related to the use of OSTs in the clinical examination process. Our current systematic review includes over four times as many articles as Dinnes *et al*'s study<sup>9</sup>.

The purpose of this article is to subject the literature on OSTs of the shoulder to a systematic review and meta-analysis to provide clinicians with enough information to determine whether these OSTs are appropriate for clinical practice.

## METHODS

### Search strategy

In order to make the retrieval of articles on diagnostic accuracy as comprehensive as possible, a generic search strategy as reported by Haynes *et al*<sup>12</sup> was employed using the MEDLINE, CINAHL and SPORTDiscus databases (1966 to October 2006) through OVID. This generic strategy to find studies on diagnostic accuracy was then combined with a subject-specific strategy addressing the shoulder and pathologies of the shoulder, and physical examination (table 1).<sup>13</sup> In addition to the database searches, personal files were hand searched by one of the authors (EH) for publications, posters or abstracts. The reference lists in review articles were cross-checked and all individual names of each special test were queried using Medline and PubMed.

History and physical examination of patients with shoulder pain has traditionally been a cornerstone of the diagnostic process. Diagnosis based on

**Table 1** MEDLINE search strategy

No.	Search history	Results
1	(shoulder adj3 pain).mp. [mp = title, original title, abstract, name of substance word, subject heading word]	2915
2	exp Shoulder Joint/ or exp Acromion/ or exp Shoulder Impingement Syndrome/ or exp Rotator Cuff/	10 507
3	(labral tear and shoulder).mp. [mp = title, original title, abstract, name of substance word, subject heading word]	60
4	((bursitis or tendonitis) and shoulder).mp. [mp = title, original title, abstract, name of substance word, subject heading word]	550
5	(shoulder adj2 arthritis).mp. [mp = title, original title, abstract, name of substance word, subject heading word]	41
6	(frozen adj2 shoulder).mp. [mp = title, original title, abstract, name of substance word, subject heading word]	343
7	(adhesive adj2 capsulitis).mp. [mp = title, original title, abstract, name of substance word, subject heading word]	219
8	(shoulder adj2 instability).mp. [mp = title, original title, abstract, name of substance word, subject heading word]	647
9	exp Physical Examination/	497 944
10	clinical examination.tw.	16 846
11	sensitiv\$.mp.	741 616
12	diagnos\$.mp.	1 222 753
13	di.fs.	1 369 790
14	(1 or 2 or 3 or 4 or 5 or 6 or 7 or 8) and (9 or 10) and (11 or 12 or 13)	831
15	limit 14 to (humans and English language)	686

### Study selection

All abstracts for 686 articles from Medline, 182 articles from CINAHL, 54 articles from SPORTDiscus and 7 articles from the hand search were reviewed by two of the authors (EH and SC) independently. Agreement between the two authors regarding

which articles to read in full was determined by consensus. Articles were eligible for inclusion if the criterion standard was surgery, magnetic resonance imaging (MRI) or injection (acromioclavicular joint only), at least one physical examination test/special test was studied, if one of the paired statistics of sensitivity and specificity were reported or could be discerned for an individual test, and if the article was in the English language. Studies were excluded if the special test was performed under anaesthesia or in cadavers, if a group of special tests was assigned the status of “composite physical examination” or if the article was a review.

Each selected study was independently assessed by all reviewers. The reviewers were familiar with the literature and, thus, were not blinded to the authors, the date of publication or the journals in which they were published. If there was disagreement as to the final selection, a third author made the conclusive decision. A summary of our search procedure is presented in fig 1 and articles pulled for review based on a consensus of the authors are presented in tables 3–7.

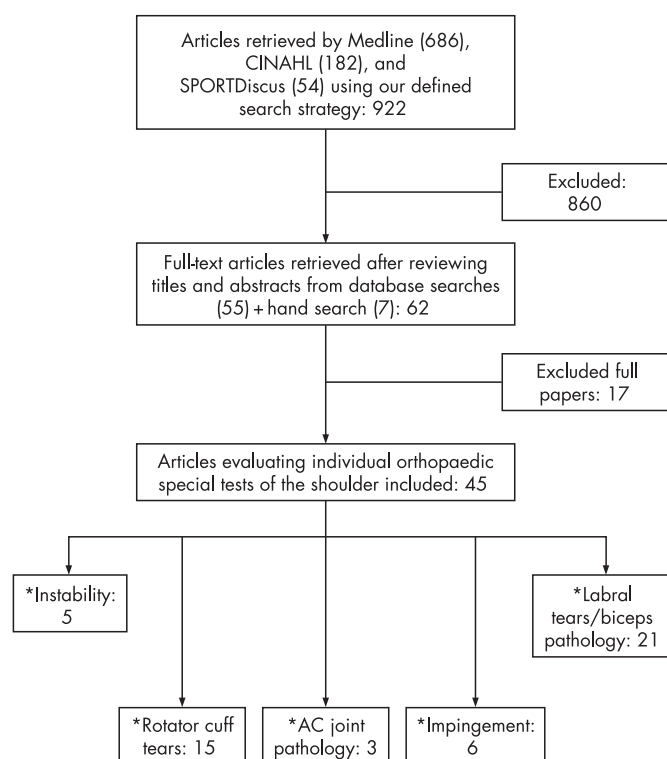
### Quality assessment

After all relevant articles were obtained, the quality of the articles was assessed and data was extracted from each article. The quality of a study was determined by examining that study's internal and external validity. Internal and external validity were evaluated (unmasked) by the primary author (EH) using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool developed by Whiting *et al* (table 2).<sup>14</sup> QUADAS involves individualized scoring of 14 components. Each of the 14 steps is scored as “yes”, “no”, or “unclear”. Individual procedures for scoring each of the 14 items, including operational standards for each question have been published

**Table 2** Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?			
2. Were selection criteria clearly described?			
3. Is the reference standard likely to classify the target condition correctly?			
4. Is the period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?			
5. Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?			
6. Did patients receive the same reference standard regardless of the index test result?			
7. Was the reference standard independent of the index test (ie, the index test did not form part of the reference standard)?			
8. Was the execution of the index test described in sufficient detail to permit replication of the test?			
9. Was the execution of the reference standard described in sufficient detail to permit its replication?			
10. Were the index test results interpreted without knowledge of the results of the reference standard?			
11. Were the reference standard results interpreted without knowledge of the results of the index test?			
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?			
13. Were uninterpretable/intermediate test results reported?			
14. Were withdrawals from the study explained?			

Reproduced with permission from Whiting P, Rutjes AW, Dinnes J, *et al*. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004;8(25):iii, 1–234.<sup>14</sup>



**Figure 1** Flow diagram of literature screening process. \*The subtotal of articles based on pathology equals 50 because five articles were appropriate for two subcategories based on pathology.

**Table 3** Summary of articles reporting on the diagnostic accuracy of OSTs for pathologies of the shoulder: impingement tests

Test, author [ref] and year	Sample size	Sensitivity/ specificity	LR+/LR–	QUADAS score	Criterion standard
<b>Neer test</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 subjects			10	Surgery
Stage I		86/49	1.69/0.29		
Stage I–III		68/69	2.17/0.47		
Calis <i>et al</i> <sup>21</sup> 2000	125 shoulders			8	Subacromial injection & MRI
Stage I		71/31	1.02/0.94		
Stage I–III		89/31	1.28/0.37		
MacDonald <i>et al</i> <sup>22</sup> 2000	85 subjects			8	Surgery
Stage I		75/48	1.41/0.55		
Stage I–III		77/63	2.05/0.37		
Bak & Fauno <sup>24</sup> 1997	72 shoulders			7	Surgery
Stages I–III		39/98	19.5/0.62		
Leroux <i>et al</i> <sup>26</sup> 1995	55 patients				
Stages I–III		89/NR	NR/NR	10	Surgery
<b>Hawkins–Kennedy test</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 subjects			10	Surgery
Stage I		76/45	1.36/0.55		
Stage I–III		72/66	2.12/0.33		
Calis <i>et al</i> <sup>21</sup> 2000	125 shoulders			8	Subacromial injection & MRI
Stage I		95/31	1.38/0.16		
Stage I–III		92/25	1.23/0.32		
MacDonald <i>et al</i> <sup>22</sup> 2000	85 subjects			8	Surgery
Stage I		92/44	1.65/0.19		
Stages I–III		89/60	2.23/0.18		
Bak & Fauno <sup>24</sup> 1997	72 shoulders			7	Surgery
Stages I–III		80/78	3.62/0.26		
Leroux <i>et al</i> <sup>26</sup> 1995	55 patients			10	Surgery
Stages I–III		87/NR	NR/NR		
<b>Painful arc</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 subjects			10	Surgery
Stage I		71/47	1.33/0.63		
Stage I–III		74/81	3.89/0.33		
Calis <i>et al</i> <sup>21</sup> 2000	125 shoulders			8	Subacromial injection & MRI
Stage I		10/88	44/1.02		
Stage I–III		33/81	1.67/0.84		
<b>Cross-body adduction</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 subjects			10	Surgery
Stage I		25/80	1.25/0.94		
Stage I–III		23/82	1.25/0.95		
Calis <i>et al</i> <sup>21</sup> 2000	125 shoulders			8	Subacromial injection & MRI
Stage I		62/31	0.90/1.23		
Stage I–III		82/28	1.13/0.65		
<b>Speed test</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 subjects			10	Surgery
Stage I		33/70	1.10/0.96		
Stage I–III		38/83	2.29/0.74		
Calis <i>et al</i> <sup>21</sup> 2000	125 shoulders			8	Subacromial injection & MRI
Stage I		52/58	1.24/0.83		
Stage I–III		69/56	1.54/0.57		
<b>Drop arm</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 subjects			10	Surgery
Stage I		14/77	61/1.12		
Stage I–III		27/88	2.25/0.83		
Calis <i>et al</i> <sup>21</sup> 2000	125 shoulders			8	Subacromial injection & MRI
Stage I		4/100	NR/NR		
Stage I–III		8/97	2.79/0.95		
<b>Other tests</b>					
Park <i>et al</i> <sup>25</sup> 2005 — supraspinatus/empty can	552 subjects			10	Surgery
Stage I		25/67	76/1.12		
Stage I–III		44/90	4.4/0.62		

Continued

Table 3 Continued

Test, author [ref] and year	Sample size	Sensitivity/ specificity	LR+/LR–	QUADAS score	Criterion standard
Park <i>et al</i> <sup>25</sup> 2005 – infraspinatus test	552 subjects			10	Surgery
Stage I		25/69	0.81/1.09		
Stage I–III		42/90	4.2/0.64		
Calis <i>et al</i> <sup>21</sup> 2000 – Yergason test	125 shoulders			8	Subacromial injection & MRI
Stage I		10/88	0.45/1.02		
Stage I–III		37/86	2.7/0.73		
Zaslav <sup>23</sup> 2001 – internal rotation resistance strength, non-subacromial and impingement tests	110 subjects	88/96	25/0.12	8	Surgery
Leroux <i>et al</i> <sup>26</sup> 1995 – Yocum test	55 patients	78/NR	NR/NR	10	Surgery
Stages I–III					

NR, not reported or unable to be calculated based on the article examined.

Stage I: impingements I = tendonitis or bursitis.

Stage II: impingement stage II = partial rotator cuff tear.

Stage III: impingement stage III = full or complete rotator cuff tear.

Where possible for tables 3–7, 2×2 tables were reconstructed and 0.5 was added to any cell with 0 subjects. Sensitivity = true positives/(true positives + false negatives).

Specificity = true negatives/(true negatives + false positives). Positive likelihood ratios indicated by LR+ = sensitivity/(1 – specificity). Negative likelihood ratios indicated by LR– = (1 – sensitivity)/specificity.

although a cumulative methodological score is not advocated.<sup>15</sup> Past studies<sup>16–18</sup> have used a score of 7 of 14 or greater “yeses” to indicate a high-quality diagnostic accuracy study whereas scores below 7 were indicative of low quality. Based on the experience of two of the authors (EH and CC) in use of the QUADAS tool in their textbook,<sup>19</sup> we established a higher quality score as 10 of 14 or greater unequivocal “yeses”, whereas below 10 was associated with a lower quality study.

### Statistical analyses

Meta-analysis was performed using dr-ROC software version 2.00 (dr2 Consulting, Glenside, PA, USA). Only studies presenting data representing both sensitivity and specificity were selected for pooling of results. Data were eligible for pooling in three special tests: Neer for impingement, Hawkins–Kennedy for impingement and Speed for labral pathology. Raw data from each individual study for these three tests was placed in a 2×2 table. The dr-ROC software was used to pool sensitivities and specificities using the inverse-variance method, which gives greater weight to individual studies with more subjects. Both the fixed effects and random effects models were used to pool information with similar outcome. The results of the fixed effects analysis are presented here. The diagnostic odds ratio (DOR) and the area under the curve (AUC) of the summary receiver operating characteristic (SROC) curve were both calculated as summary statistics indicating the overall diagnostic power of each of the three tests. Cochran’s Q was used to test for heterogeneity and the I<sup>2</sup> statistic<sup>20</sup> was used to quantify the percentage of variation across the studies that was associated with heterogeneity.

### RESULTS

Forty-five articles met our inclusion criteria including the reporting of one of the paired statistics of sensitivity and specificity or the reporting of raw data that allowed the calculation of one of these paired statistics. Meta-analysis could be performed on three tests that had the requisite statistical homogeneity to draw summary conclusions from the meta-analysis: Neer for impingement, Hawkins–Kennedy for impingement and Speed for a SLAP lesion. Unfortunately, even for these well-known tests, there were only four articles for each test that met the inclusion criteria. The results of the meta-analyses are presented in the following subsections based on pathology.

### Impingement

Four<sup>21–24</sup> of the six articles that addressed the diagnostic accuracy of OSTs for impingement were of lower design/reporting quality by our definition (table 3). Bias in these studies was mostly related to lack of masking on the part of the physician who performed both the OST and the criterion standard to confirm diagnosis. However, in two of the studies addressing impingement,<sup>21, 24</sup> the criterion standard of surgery was not used. Because of consistent bias in articles investigating OSTs designed to detect impingement, we guardedly report either the supraspinatus/empty can or infraspinatus tests may serve as confirmatory tests for impingement due to higher specificity.

In the meta-analysis, only the Neer test and the Hawkins–Kennedy test had homogeneous data from four articles each. The pooled sensitivity and specificity for the Neer test was 0.79 (95% confidence interval (CI) 0.75 to 0.82) and 0.53 (95% CI 0.48 to 0.58), respectively, and for the Hawkins–Kennedy test was 0.79 (95% CI 0.75 to 0.82) and 0.59 (95% CI 0.53 to 0.64), respectively. However, the pooled diagnostic odds ratio (DOR) for both tests is around 1 and the 95% confidence interval crosses 1 indicating that neither test has diagnostic utility for impingement (figs 2 and 3). Figures 4 and 5 show the fitted summary receiver operating characteristic (ROC) curves for the Neer and Hawkins–Kennedy tests, respectively. The area under the curve (AUC) for the Neer test is 0.74 (95% CI 0.70 to 0.78) and for the Hawkins–Kennedy test is 0.76 (95% CI 0.72 to 0.80) confirming the limited usefulness of these tests in diagnosis of impingement.

### Rotator cuff integrity

Of the 15 studies that examined OSTs to assess rotator cuff integrity, nine<sup>22–25, 27–33</sup> examined the ability of the tests to detect any rotator cuff tear and five<sup>26, 34–37</sup> examined the ability to assess the tear of a specific rotator cuff muscle (table 4). One article<sup>38</sup> reported the sensitivity only (not the specificity) of four OSTs to detect subscapularis lesions alone and also to detect any rotator cuff tear. Eight of the fifteen articles were of high design/reporting quality by our definition. Of the remaining seven articles, issues included a lack of stated inclusion/exclusion criteria, lack of masking of the physician to the results of the OST, and no reporting of intermediate or uninterpretable results. None of the 10 OSTs for rotator cuff pathology that were examined in more than one study proved consistently diagnostic. Two tests, the external rotation lag sign

**Table 4** Summary of articles pulled for review by the authors: rotator cuff integrity

Test, author [ref] and year	Sample size	Sensitivity/specificity	LR+/LR–	QUADAS score	Criterion standard
<b>Supraspinatus/Jobe empty can test</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 patients	53/82	3.0/0.57	10	Surgery
Boileau <i>et al</i> <sup>31</sup> 2004	21 patients	81/NR	NR/NR	10	Surgery
Holtby & Razmjou <sup>27</sup> 2004	50 patients	41/70	1.4/0.84	13	Surgery
Litaker <i>et al</i> <sup>22</sup> 2000	448 patients	64/65	1.8/0.55	10	Double-contrast arthrography
Itoi <i>et al</i> <sup>35</sup> 1999	143 shoulders	89/98	45/0.11	8	Surgery & MRI
Hertel <i>et al</i> <sup>28</sup> 1996	87 patients	84/54	2.0/0.28	7	Surgery
Leroux <i>et al</i> <sup>26</sup> 1995	55 patients	86/50	1.7/0.28	10	Surgery
<b>Lift-off test</b>					
Barth <i>et al</i> <sup>34</sup> 2006	68 shoulders	18/92	2.3/0.89	11	Surgery
Scheibel <i>et al</i> <sup>38</sup> 2005	12 patients	58/NR	NR/NR	6	Surgery & MRI
Hertel <i>et al</i> <sup>28</sup> 1996	53 patients	62/98	31/0.39	7	Surgery
Leroux <i>et al</i> <sup>26</sup> 1995	55 patients	17/60	0.43/1.38	10	Surgery
Gerber & Krushell <sup>36</sup> 1991	13 patients	92/NR	NR/NR	9	Surgery
<b>Speed/Gilcreest palm-up test</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 patients	40/75	1.6/0.80	10	Surgery
Boileau <i>et al</i> <sup>31</sup> 2004	21 patients	48/NR	NR/NR	10	Surgery
Leroux <i>et al</i> <sup>26</sup> 1995	55 patients	63/35	0.97/1.05	10	Surgery
<b>Neer test</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 patients	59/47	1.1/0.87	10	Surgery
MacDonald <i>et al</i> <sup>22</sup> 2000	85 patients	83/51	1.7/0.33	8	Surgery
<b>Hawkins–Kennedy test</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 patients	69/48	1.3/0.65	10	Surgery
MacDonald <i>et al</i> <sup>22</sup> 2000	85 patients	88/43	1.5/0.28	8	Surgery
<b>External rotation lag sign</b>					
Walch <i>et al</i> <sup>37</sup> 1998	54 patients			7	Surgery
Infraspinatus		98/98	49/0.02		
Teres minor		95/72	3.4/0.07		
Hertel <i>et al</i> <sup>28</sup> 1996	87 patients	69/98	35/0.32	7	Surgery
<b>Internal rotation lag sign</b>					
Scheibel <i>et al</i> <sup>38</sup> 2005	12 patients	75/NR	NR/NR	6	Surgery & MRI
Hertel <i>et al</i> <sup>28</sup> 1996	53 patients	97/96	24/0.03	7	Surgery
<b>Rent test</b>					
Wolf & Agrawal <sup>30</sup> 2001	109 patients	96/97	32/0.04	10	Surgery
Lyons & Tomlinson 1992	42 patients	73/66	2.2/0.41	6	Surgery
<b>Napoleon sign</b>					
Barth <i>et al</i> <sup>34</sup> 2006	68 shoulders	25/98	13/0.77	11	Surgery
Scheibel <i>et al</i> <sup>38</sup> 2005	16 patients	69/NR	NR/NR	6	Surgery & MRI
<b>Drop arm</b>					
Park <i>et al</i> <sup>25</sup> 2005	552 patients	35/88	2.9/0.74	10	Surgery
Murrell & Walton <sup>33</sup> 2001	400 patients	10/98	5.0/0.92	5	Surgery
<b>Other tests</b>					
Litaker <i>et al</i> <sup>22</sup> 2000 – supine impingement sign	448 patients	97/9	1.07/0.33	10	Double contrast arthrography
Hertel <i>et al</i> <sup>28</sup> 1996 – drop sign	87 patients	21/92	2.6/0.86	7	Surgery
Park <i>et al</i> <sup>25</sup> 2005 – infraspinatus test	552 patients	51/84	3.2/0.58	10	Surgery
Park <i>et al</i> <sup>25</sup> 2005 – cross-body adduction	552 patients	23/81	1.2/0.95	10	Surgery
Park <i>et al</i> <sup>25</sup> 2005 – painful arc	552 patients	76/62	2.0/0.39	10	Surgery
Barth <i>et al</i> <sup>34</sup> 2006 – bear-hug test	68 shoulders	60/92	7.5/0.32	11	Surgery
Barth <i>et al</i> <sup>34</sup> 2006 – belly press test	68 shoulders	40/98	20/0.61	11	Surgery
Itoi <i>et al</i> <sup>35</sup> 1999 – full-can test	143 shoulders	86/57	2/0.25	8	Surgery & MRI
Walch <i>et al</i> <sup>37</sup> 1998 – hornblower's sign	54 patients	95/92	12/0.05	7	Surgery
Scheibel <i>et al</i> 2005 <sup>38</sup> – belly-off test	16 patients	38/NR	NR/NR	6	Surgery & MRI

Note: the same study may have different numbers of sample size based on how many subjects received the OST.

For the Scheibel *et al* study, two groups were combined to report the sensitivity of the OSTs for any rotator cuff tear, whether partial or complete.

(ERLS) and the drop arm test, demonstrated value as specific tests for a tear of any rotator cuff muscle, and the supine impingement test may have value, when negative, in ruling out a rotator cuff tear. Further, two tests, the bear-hug and belly press tests, appear to be valuable as specific tests for ruling in a

subscapularis muscle tendon tear when positive. The necessary bolus of information to perform a meta-analysis was not available since the studies either examined different OSTs or the studies that examined the same OST focused on the detection of different pathologies.

**Table 5** Summary of articles pulled for review by the authors: glenoid labrum integrity

Test, author [ref] and year	Sample size	Sensitivity/specificity	LR+/LR−	QUADAS score	Criterion standard
<b>Superior labral pathology</b>					
<i>Active compression/O'Brien test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	63/50	1.25/0.75	10	Surgery
Myers <i>et al</i> <sup>46</sup> 2005	37 patients	78/11	0.88/2.0	8	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	54/60	1.35/0.77	9	Surgery
McFarland <i>et al</i> <sup>48</sup> 2002	426 patients	47/55	1.04/0.96	11	Surgery
Morgan <i>et al</i> <sup>49</sup> 1998	102 patients	68/NR	NR/NR	10	Surgery
O'Brien <i>et al</i> <sup>50</sup> 1998	256 patients	99/98	49.5/0.01	5	Combination of radiograph, MRI, surgery was given to some but not all patients
<i>Anterior slide test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	10/82	0.56/1.10	10	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	5/93	0.71/1.02	9	Surgery
McFarland <i>et al</i> <sup>48</sup> 2002	419 patients	8/84	0.50/1.1	11	Surgery
Kibler <sup>51</sup> 1995	226 patients	78/91	8.34/0.24	7	Surgery
<i>Crank test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	13/83	0.85/1.06	10	Surgery
Myers <i>et al</i> <sup>46</sup> 2005	36 patients	35/70	1.15/0.93	8	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	58/72	2.1/0.58	9	Surgery
Guanche <i>et al</i> <sup>52</sup> 2003	60 shoulders	39/67	1.2/0.91	12	Surgery
Mimori <i>et al</i> <sup>53</sup> 2004	32 patients	81/88	6.8/0.22	7	MRI and 15/32 patients received surgery
<i>Relocation/Jobe relocation test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	50/53	1.07/0.94	10	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	75/40	1.25/0.63	9	Surgery
Guanche <i>et al</i> <sup>52</sup> 2003	60 shoulders	36/63	0.97/1.02	12	Surgery
Morgan <i>et al</i> <sup>49</sup> 1998	102 patients	51/NR	NR/NR	10	Surgery
<i>Speed test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	48/67	1.47/0.77	10	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	6/98	3.0/0.96	9	Surgery
Holtby & Razmjou <sup>43</sup> 2004	50 patients	32/75	1.3/0.91	11	Surgery
Guanche <i>et al</i> <sup>52</sup> 2003	60 shoulders	9/74	0.35/1.23	12	Surgery
Morgan <i>et al</i> <sup>49</sup> 1998	102 patients	68/NR	NR/NR	10	Surgery
<i>Biceps tenderness test</i>					
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	25/80	1.3/0.94	9	Surgery
Guanche <i>et al</i> <sup>52</sup> 2003	60 shoulders	48/52	1.0/1.0	12	Surgery
Morgan <i>et al</i> <sup>49</sup> 1998	102 patients	68/NR	NR/NR	10	Surgery
<i>Yergason's test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	13/94	1.9/0.94	10	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	12/98	6.0/0.90	9	Surgery
Holtby & Razmjou <sup>43</sup> 2004	50 patients	43/79	2.1/0.72	11	Surgery
Guanche <i>et al</i> <sup>52</sup> 2003	60 shoulders	12/96	3.0/0.92	12	Surgery
<i>Neer test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	50/52	1.1/0.96	10	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	33/60	0.83/1.12	9	Surgery
<i>Hawkins–Kennedy test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	68/31	0.97/1.07	10	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	50/67	1.5/0.77	9	Surgery
<i>Pain provocation test</i>					
Parentis <i>et al</i> <sup>45</sup> 2006	132 patients	15/90	1.5/0.94	10	Surgery
Mimori <i>et al</i> <sup>53</sup> 2004	32 patients	98/86	7.0/0.02	7	MRI and 15/32 patients received surgery
<i>Compression-rotation test</i>					
Nakagawa <i>et al</i> <sup>47</sup> 2005	54 patients	26/98	13.0/0.76	9	Surgery
McFarland <i>et al</i> <sup>48</sup> 2002	303 patients	24/76	1.0/1.0	11	Surgery
<i>Other tests</i>					
Guanche <i>et al</i> <sup>52</sup> 2003 – anterior apprehension test	60 shoulders	30/63	0.81/1.11	12	Surgery
Myers <i>et al</i> <sup>46</sup> 2005 – RSERT	40 patients	83/82	4.6/0.21	8	Surgery
Kim <i>et al</i> <sup>54</sup> 1999 – biceps load I test	75 patients	91/97	30.3/0.09	9	Surgery
Kim <i>et al</i> <sup>55</sup> 2001 – biceps load II test	127 patients	90/97	30.0/0.10	11	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005 – fulcrum test	54 patients	83/40	1.4/0.43	9	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005 – forced shoulder abduction/Komuro test	54 patients	67/67	2.0/0.49	9	Surgery

Continued

Table 5 Continued

Test, author [ref] and year	Sample size	Sensitivity/specificity	LR+/LR–	QUADAS score	Criterion standard
Nakagawa <i>et al</i> <sup>47</sup> 2005 – clunk test	54 patients	44/68	1.3/0.84	9	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005 – Ellman's test	54 patients	42/63	1.1/0.92	9	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005 – abduction inferior stability (ABIS)/Feagin test	54 patients	29/90	2.9/0.79	9	Surgery
Kim <i>et al</i> <sup>64</sup> 1999 – biceps tension test	75 patients	33/78	1.5/0.86	9	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005 – jerk test	54 patients	25/80	1.3/0.94	9	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005 – painful arc test	54 patients	21/73	0.78/1.08	9	Surgery
Nakagawa <i>et al</i> <sup>47</sup> 2005 – sulcus sign	54 patients	17/93	2.4/0.89	9	Surgery
Berg & Ciullo <sup>56</sup> 1998 – SLAPrehension	66 patients	88/NR	NR/NR	8	Surgery
<b>Any labral pathology</b>					
<i>Crank test</i>					
Guanche <i>et al</i> <sup>62</sup> 2003	60 shoulders	40/73	1.5/0.82	12	Surgery
Stetson & Templin <sup>40</sup> 2002	65 patients	46/56	1.1/0.96	11	Surgery
Liu <i>et al</i> <sup>69</sup> 1996	62 patients	91/93	13/0.10	11	Surgery
<i>Other tests</i>					
Stetson & Templin <sup>40</sup> – active compression/O'Brien test	65 patients	54/31	0.78/1.5	11	Surgery
Guanche <i>et al</i> <sup>62</sup> 2003 – relocation/Jobe relocation test	60 shoulders	44/87	3.4/0.64	12	Surgery
Hamner <i>et al</i> <sup>41</sup> 2000 – modified relocation/modified Jobe relocation test	13 patients	92/NR	NR/NR	7	Surgery
Guanche <i>et al</i> <sup>62</sup> 2003 – Speed test	60 shoulders	18/87	1.4/0.94	12	Surgery
Guanche <i>et al</i> <sup>62</sup> 2003 – Yergason test	60 shoulders	9/93	1.3/0.9	12	Surgery
Guanche <i>et al</i> <sup>62</sup> 2003 – anterior apprehension test	60 shoulders	40/87	3.1/0.69	12	Surgery
Guanche <i>et al</i> <sup>62</sup> 2003 – biceps tenderness test	60 shoulders	44/40	0.73/1.4	12	Surgery
<b>Posterior labral pathology</b>					
Kim <i>et al</i> <sup>67</sup> 2005 – jerk test	172 shoulders	73/98	36.5/0.26	11	Surgery
Kim <i>et al</i> <sup>67</sup> 2005 – Kim test	172 shoulders	80/94	13.3/0.21	11	Surgery
Meister <i>et al</i> <sup>68</sup> 2004 – posterior impingement sign	69 patients	76/85	5.0/0.29	7	Surgery
<b>Other labral tears and biceps pathology</b>					
Ardic <i>et al</i> <sup>44</sup> 2006 –Speed test	36 patients	69/60	1.7/0.51	12	MRI
Boileau <i>et al</i> <sup>61</sup> 2004 – hourglass test	21 patients	98/NR	NR/NR	10	Surgery
Bennett <sup>42</sup> 1998 – SLAP or biceps Speed test	46 shoulders	90/13.8	1.04/0.72	10	Surgery

### Glenoid labrum integrity/long head of biceps pathology

Fourteen of the 21 studies examining the diagnostic accuracy of OSTs for labral pathology focused on the detection specifically of superior labrum pathology (table 5). Of the remaining seven studies, three<sup>39–41</sup> focused on detection of any labral pathology, two<sup>32–33</sup> examined the integrity of the posterior labrum, two<sup>27–42–43</sup> examined detection of either a long head of the biceps or superior labral anterior to posterior (SLAP) lesion and two<sup>31–44</sup> examined detection of biceps pathology alone. Twelve of the 21 articles were of high design/reporting quality by our definition. Similar to the rotator cuff article group, problems with design of the studies of the glenoid labrum included lack of stated inclusion/exclusion criteria, lack of masking of the physician to the results of the OST and no reporting of intermediate or uninterpretable results. Two OSTs for posterior labral tears, the Kim test and the Jerk test, and one for SLAP lesions, the biceps load II test, appear to be diagnostic but more studies are needed to investigate these tests.

In the meta-analysis, only the Speed test and only as an OST for a SLAP lesion had homogeneous data from four articles describing the diagnostic accuracy of this test. The summary sensitivity and specificity of the Speed test were 0.32 (95% CI 0.24 to 0.42) and 0.61 (95% CI 0.54 to 0.68), respectively. However, the pooled DOR for the Speed test is less than 1 and the 95% confidence interval (CI) crosses 1 indicating that the Speed test has no diagnostic utility for a SLAP lesion (fig 6). Figure 7 shows the fitted summary ROC curve for the Speed test. The AUC for the Speed test is 0.54 (95% CI 0.44 to 0.64)

confirming the use of this test in the diagnosis of a SLAP lesion is no better than chance.

### Instability

Five articles examined OSTs for instability with all articles specifically attempting to identify anterior shoulder instability (table 6). Three<sup>59–61</sup> of the articles were of high design/reporting quality by our definition. There was not sufficient data to perform a meta-analysis in this sub-group of articles but the apprehension, relocation and anterior release tests appear diagnostic of anterior instability, especially when apprehension and not pain is used as the definition for a positive test.

### Acromioclavicular (AC) joint pathology

Three articles<sup>50–64–65</sup> examined the diagnostic accuracy of OSTs for AC joint pathology (table 7). Two<sup>64–65</sup> of the three articles were of high design/reporting quality by our definition. The active compression test may be diagnostic of AC joint pathology but is troublesome in that as the quality of the study improves, the statistics monitoring diagnostic accuracy worsen. There was not sufficient power to perform a meta-analysis in this sub-group of articles.

### Overall

We examined the quality of forty-five articles reporting on the diagnostic accuracy of OSTs of the shoulder. If our definition

**Table 6** Summary of articles pulled for review by the authors: anterior instability

Test, author [ref] and year	Sample size	Sensitivity/ Specificity	LR+/LR–	QUADAS score	Criterion standard
<b>Relocation/Jobe relocation test</b>					
Farber <i>et al</i> <sup>61</sup> 2006	363 patients	Pain 30/90 App. 81/92	3.0/0.77 10.4/0.20	11	Surgery
Lo <i>et al</i> 2004 <sup>59*</sup>	46 patients	Either pain or app. 46/54	1.0/0.99	10	Combination of clinical exam and radiographs
Speer <i>et al</i> <sup>60</sup> 1994	100 patients	Pain 54/44 App. 67/99	0.96/1.0 67/0.33	10	Surgery
<b>Apprehension</b>					
Farber <i>et al</i> <sup>61</sup> 2006	363 patients	Pain 50/56 App. 72/96	1.1/0.89 20.2/0.29	11	Surgery
Lo <i>et al</i> 2004 <sup>59*</sup>	46 patients	Either pain or app. 53/99	53/0.47	10	Combination of clinical exam and radiographs
<b>Anterior release/surprise test</b>					
Lo <i>et al</i> 2004 <sup>59*</sup>	46 patients	Either pain or app. 64/99	58.6/0.37	10	Combination of clinical exam and radiographs
Gross & Distefano <sup>62</sup> 1997	100 patients	Pain 92/89	8.3/0.09	9	Surgery and examination under anaesthesia
<b>Other tests</b>					
Farber <i>et al</i> <sup>61</sup> 2006 – anterior drawer test	363 patients	Pain 28/71 App. 53/85	0.97/1.0 3.6/0.57	11	Surgery
<b>Anterior-superior instability with an anterior supraspinatus tear (SLAC lesion)</b>					
Savoie <i>et al</i> <sup>63</sup> 2001 – load and shift test	40 patients	98/NR	NR/NR	9	Surgery
Savoie <i>et al</i> <sup>63</sup> 2001 – anterior-superior SLAP test	40 patients	95/NR	NR/NR	9	Surgery
Savoie <i>et al</i> <sup>63</sup> 2001 – active compression/O'Brien test	40 patients	88/NR	NR/NR	9	Surgery
Savoie <i>et al</i> <sup>63</sup> 2001 – Kibler test	40 patients	93/NR	NR/NR	9	Surgery
Savoie <i>et al</i> <sup>63</sup> 2001 – Whipple test	40 patients	100/NR	NR/NR	9	Surgery

App., apprehension.

\*The values from the Lo *et al* study represent the mean of four examiners.

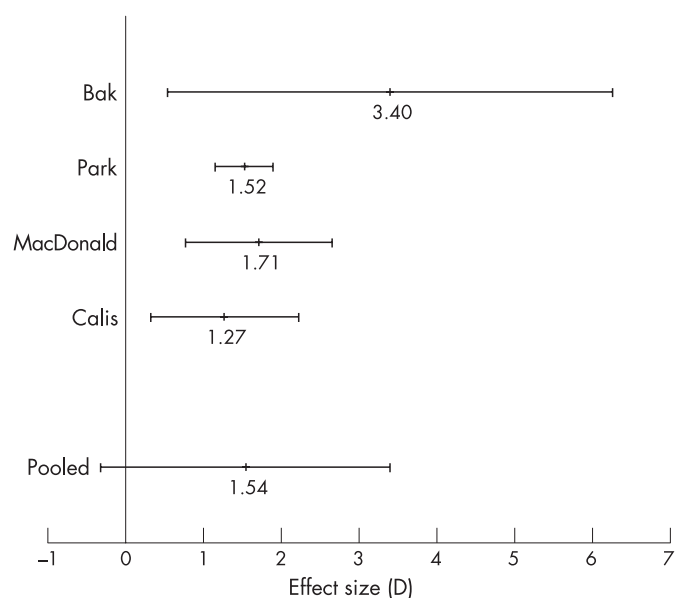
(10/14 or greater on the QUADAS(14) tool) of a higher quality article is accepted, then 22 of 45 articles were of higher quality with 10 of the higher quality articles being published since 2004.

## DISCUSSION

Several studies have called into question the ability of clinicians to diagnose shoulder problems based on pathology.<sup>4–8</sup> Coincidentally, OSTs are designed to do exactly that: to

**Table 7** Summary of articles pulled for review by the authors: acromioclavicular (AC) joint pathology

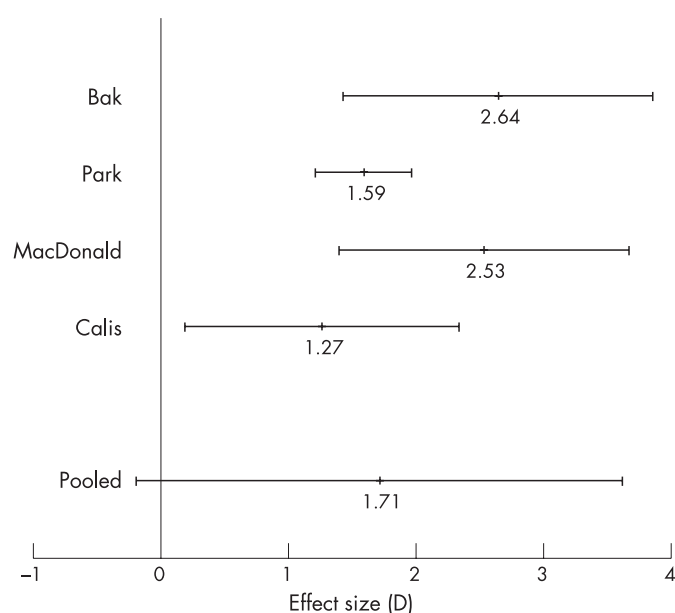
Test, author [ref] and year	Sample size	Sensitivity/specificity	LR+/LR–	QUADAS Score	Criterion standard
<b>Active compression/O'Brien test</b>					
Chronopoulos <i>et al</i> <sup>64</sup> 2004	325 patients	41/95	8.2/0.62	10	AC joint injection
Walton <i>et al</i> <sup>65</sup> 2004	38 patients	16/90	1.6/0.93	13	AC joint injection
O'Brien <i>et al</i> <sup>60</sup> 1998	262 patients	93/96	23.1/0.08	5	Combination of radiograph, MRI, surgery was given to some but not all patients
<b>Other tests</b>					
Chronopoulos <i>et al</i> <sup>64</sup> 2004 – AC resisted extension test	348 patients	72/85	4.8/0.33	10	AC joint injection
Chronopoulos <i>et al</i> <sup>64</sup> 2004 – cross-body adduction test	553 patients	77/79	3.67/0.29	10	AC joint injection
Chronopoulos <i>et al</i> 2004 – Neer test	585 patients	57/41	0.97/1.1	10	AC joint injection
Chronopoulos <i>et al</i> <sup>64</sup> 2004 – Hawkins–Kennedy test	337 patients	47/45	0.85/1.2	10	AC joint injection
Chronopoulos <i>et al</i> <sup>64</sup> 2004 – painful arc test	358 patients	50/47	0.94/1.1	10	AC joint injection
Chronopoulos <i>et al</i> <sup>64</sup> 2004 – drop-arm test	524 patients	35/72	1.25/0.90	10	AC joint injection
Chronopoulos <i>et al</i> <sup>64</sup> 2004 – Speed test	533 patients	24/71	0.83/1.07	10	AC joint injection
Walton <i>et al</i> <sup>65</sup> 2004 – Paxino test	38 patients	79/50	1.58/0.42	13	AC joint injection
Walton <i>et al</i> <sup>65</sup> 2004 – pain with palpation	38 patients	96/10	1.07/0.40	13	AC joint injection



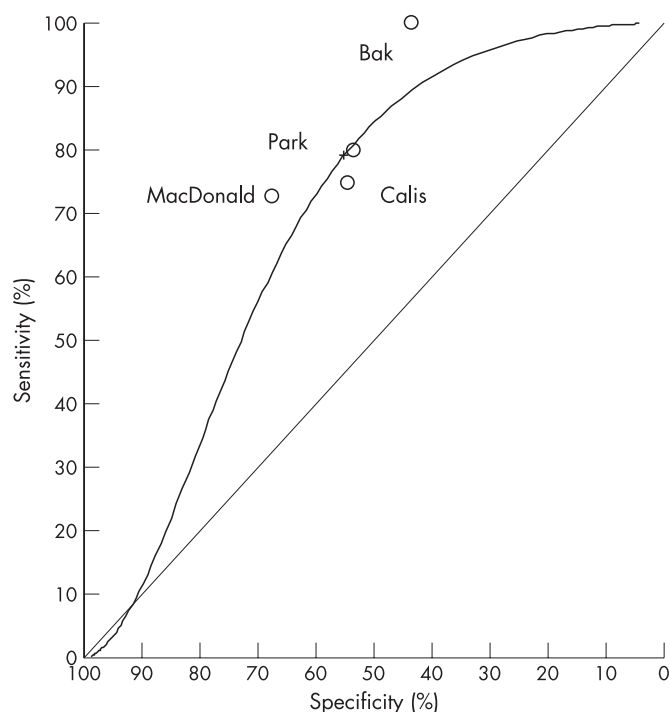
**Figure 2** Diagnostic odds ratio (DOR) of the Neer test for impingement. Individual studies labelled by primary author and combined results labelled as pooled.

differentially diagnose pathologies of the shoulder. Therefore, it is imperative, before we abandon the idea that accurate diagnosis in the shoulder is possible, that we thoroughly examine the body of literature related to OSTs of the shoulder. The most powerful method to accomplish such a goal is meta-analysis. Meta-analysis of three tests (Neer, Hawkins–Kennedy and Speed) examining two pathologies (impingement and SLAP lesion) demonstrated that none of these tests is diagnostic for their stated pathology.

In addition to the meta-analyses, we elected to perform a comprehensive systematic review using the QUADAS<sup>14</sup> document to assist with the quality assessment of 44 articles that

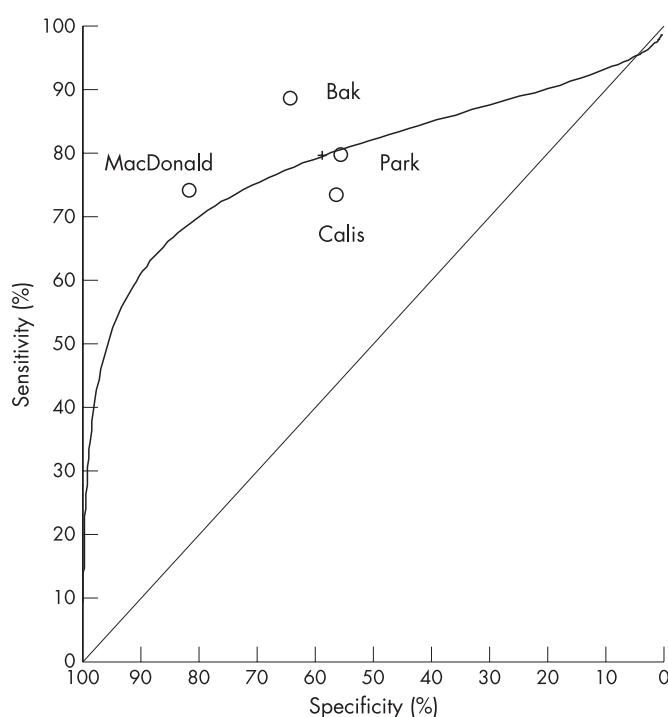


**Figure 3** Diagnostic odds ratio (DOR) of the Hawkins–Kennedy test for impingement. Individual studies labelled by primary author and combined results labelled as pooled.

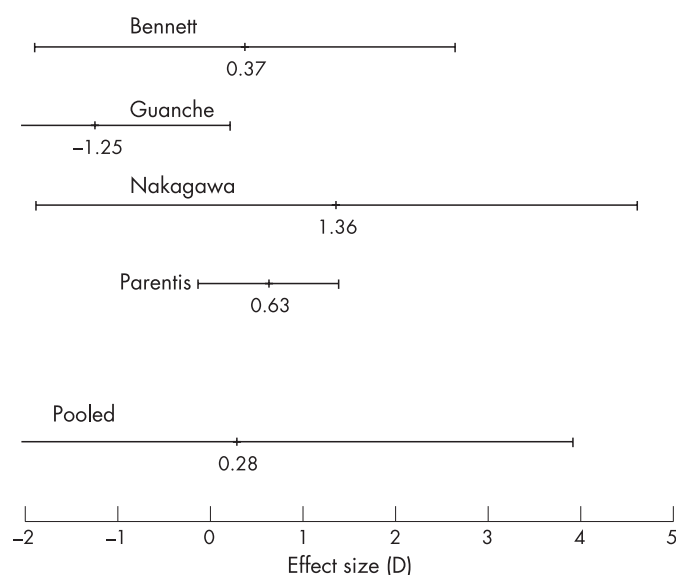


**Figure 4** Summary receiver operating characteristic (ROC) curve for the Neer test with studies represented by an open circle and labelled by primary author.

reported on the diagnostic accuracy of almost 50 OSTs. For ease of analysis and for the convenience of the reader, we divided these OSTs into five categories based on pathology: impingement, rotator cuff pathology, labral/biceps pathology, instability and AC joint pathology. Impingement, the final common



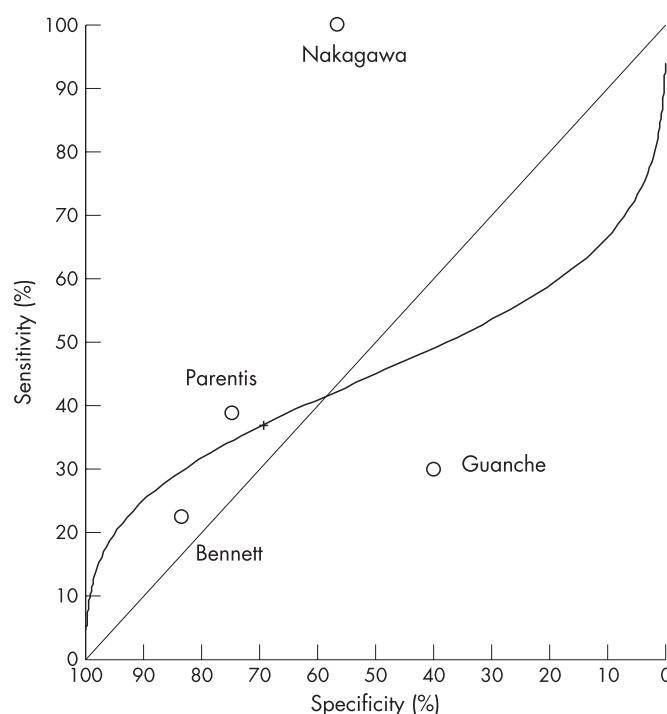
**Figure 5** Summary receiver operating characteristic (ROC) curve for the Hawkins–Kennedy test with studies represented by an open circle and labelled by primary author.



**Figure 6** Diagnostic odds ratio (DOR) of the Speed test for superior labral (SLAP) pathology. Individual studies labelled by primary author and combined results labelled as pooled.

pathway for many pathologies of the shoulder, was subdivided in most of the studies into stages I–III based on Neer's original classification.<sup>66</sup> Stage I was defined as subacromial bursitis or tendonitis. Stage II was defined as a partial rotator cuff tear and Stage III was defined as a full-thickness or complete tear of the rotator cuff. Each successive stage is considered a worsening progression along a continuum of pathology. We elected to report the values from articles that specifically reported diagnostic values for Stage III impingement in the pathological category of "rotator cuff pathology" but reported all other individual stages and combined stage data in the "impingement" category. When examining likelihood ratios of the OSTs that attempt to detect impingement, there is one non-subacromial impingement test (internal rotation resistance strength test) and no subacromial impingement tests that improve the post-test probability of detecting subacromial impingement by a moderate or large amount.<sup>67</sup> The internal rotation resistance strength test was examined in only one article<sup>23</sup> judged to be of lower quality (QUADAS 8/14) so the reported values should be viewed with caution. As for the subacromial impingement tests, value can sometimes be found in an OST with either high sensitivity or high specificity.<sup>68–69</sup> OSTs with high sensitivity are valuable as a screen where a negative test can rule out a pathology while OSTs with a high specificity can be used as a confirmatory test where a positive finding rules in the pathology.<sup>69</sup> When viewed in this context as either a screen or a confirmatory test, no impingement test seems to serve as a screen and either the supraspinatus/empty can or infraspinatus tests may serve as confirmatory tests for impingement. We urge caution with this conclusion since there is only one study that examined the diagnostic accuracy of each of the supraspinatus/empty can and infraspinatus tests.

As previously mentioned, studies that reported on the diagnostic accuracy of Stage III impingement were grouped with studies reporting on rotator cuff pathology. One study<sup>4</sup> estimated that rotator cuff lesions account for 70% of painful shoulder episodes. Of the nine OSTs for rotator cuff pathology that were examined in more than one study, none consistently exhibited likelihood ratios that would modify post-test probability



**Figure 7** Summary receiver operating characteristic (ROC) curve for the Speed test with studies represented by an open circle and labelled by primary author.

of detecting a tear of the rotator cuff by a moderate or large amount.<sup>67</sup> However, in one study<sup>37</sup> with small sample size and numerous design faults, the hornblower's sign was diagnostic of severe degeneration or absence of the teres minor muscle. In this same study,<sup>37</sup> the external rotation lag sign (ERLS) was found to be diagnostic of an infraspinatus muscle tear. A second study<sup>28</sup> demonstrated value in the ERLS as a specific test for any rotator cuff tear. Further, two tests, the bear-hug and belly press tests, were shown in one well-conducted study<sup>34</sup> to be valuable as specific tests for ruling in a subscapularis muscle tendon tear when positive, and the supine impingement test<sup>32</sup> was found to be a sensitive test to screen for any rotator cuff tear. Unfortunately, despite the high quality of the Barth *et al*<sup>34</sup> study and the Litaker *et al*<sup>32</sup> study, both were underpowered according to Flahault *et al*.<sup>70</sup>

The glenoid labrum works in conjunction with the biceps and glenohumeral ligaments to provide shoulder stability.<sup>71</sup> As with overlapping, pathology-based diagnoses such as impingement and Stage III rotator cuff tear, we made the decision to separate studies that examined the detection of instability from those that examined pathology of the glenoid labrum. Of all possible labral pathologies, the superior labral anterior-to-posterior or SLAP lesion was the most researched, being the focus of 12 of the 21 studies analysed in the category of labral and biceps pathology. Eleven OSTs were examined in more than one study and of those 11, the active compression, anterior slide, crank and compression–rotation tests had likelihood ratios indicating a moderate or large effect on the post-test probability of diagnosing a SLAP lesion. Unfortunately, each test failed to perform consistently well when examined in more stringent studies. Of the remaining single-study OSTs designed to diagnose a SLAP lesion, the biceps load I<sup>54</sup> and the biceps load II<sup>55</sup> modified the post-test probability by a large amount<sup>67</sup> and appear to be useful in diagnosing a SLAP lesion. However, both

## What is already known on this topic

- ▶ Orthopaedic special tests (OSTs) are used extensively in clinical practice to detect shoulder pathology.
- ▶ OSTs are reported extensively upon in peer-reviewed articles and textbooks.
- ▶ Varying levels of diagnostic accuracy have been reported for individual OSTs.
- ▶ The literature examining the diagnostic accuracy of OSTs is generally of poor quality.

tests have been examined in only one study each with small sample size and the biceps load I test<sup>54</sup> was performed only on patients who had dislocated their shoulder. The biceps load II test<sup>55</sup> may be the most promising but other OSTs like the anterior slide and active compression tests have performed far worse when used by other than the originator of the test. Beyond superior labral pathology, two studies<sup>57–58</sup> examined OSTs for posterior labral tears. The Kim test,<sup>57</sup> the posterior impingement sign<sup>58</sup> and the Jerk test all modified post-test probability by a moderate to large amount.<sup>67</sup> Again, however, we urge caution since each OST has been studied only once and in two of the three tests, the originator of the OST was also the author of the paper. As for OSTs that were studied as diagnostic for non-specific labral tears, the Crank test was again promising in one study<sup>39</sup> but not in two others.<sup>40–52</sup>

Instability may come from a labral tear, trauma or a connective disease like Ehlers–Danlos syndrome. Instability can be multidirectional or unidirectional. Unidirectional instability is most often in the anterior direction.<sup>72</sup> Not coincidentally, all of the instability studies in our review attempted to assess the diagnostic accuracy of OSTs for anterior instability. Three of the OSTs, apprehension, relocation and anterior release, could be viewed as merely a progression of the preceding test, respectively. With both the apprehension test and the relocation test, the use of “apprehension” as a positive test for anterior instability improves both the sensitivity and specificity over the use of “pain” as a positive sign. With the use of “apprehension” as a positive test for anterior instability, both tests modify the post-test probability a moderate to large amount.<sup>67</sup> The anterior release test appears to be a strong diagnostic test regardless of whether “pain” or “apprehension” is used as the definition of a positive test. The remaining OSTs designed to detect anterior instability with an anterior supraspinatus tear, designated a superior labral anterior cuff (SLAC) lesion, appear to all be sensitive and have value as a screening test when negative but all of the data come from one underpowered study and no specificity values were reported.

Finally, AC joint pathology is a common source<sup>73</sup> of shoulder pain and can be a contributor to outlet or subacromial impingement or can be an entity in itself, often confounding shoulder diagnosis. None of the OSTs appear valuable as a diagnostic test based on the likelihood ratios. However, in three studies,<sup>50–64–65</sup> the active compression test was shown to be a specific test that would rule in the AC joint as a source of shoulder pain if positive. Pain with palpation of the AC joint may serve as a screening test for AC joint pathology when negative, but surprisingly, only one study<sup>65</sup> with a small sample size exists to confirm this clinically common use of palpation.

## What this study adds

- ▶ This is the most comprehensive systematic review with meta-analysis of the diagnostic value of individual orthopaedic special tests to date.
- ▶ Meta-analysis for the Neer test of impingement, the Hawkins–Kennedy test of impingement and the Speed test for a SLAP lesion shows these tests to have no discriminatory ability for shoulder diagnosis.
- ▶ Meta-analysis for other OSTs was not possible either because there is not enough diagnostic accuracy research about the test or because statistical heterogeneity between studies did not allow for summary results.

Recommendations are as follows:

- ▶ The Hawkins–Kennedy test may serve as a screen and either the supraspinatus/empty can or infraspinatus test may serve as a confirmatory test for impingement.
- ▶ The supine impingement test may be valuable, when negative, as a screen for any rotator cuff tear.
- ▶ The ERLS test may have value as a specific test for any rotator cuff tear.
- ▶ The hornblower’s sign may be diagnostic of severe degeneration or absence of the teres minor muscle.
- ▶ The external rotation lag sign (ERLS) may be diagnostic of an infraspinatus muscle tear.
- ▶ The bear-hug and belly press tests may be valuable as specific tests for ruling in a subscapularis muscle tear.
- ▶ The biceps load II test appears diagnostic for SLAP lesions.
- ▶ The apprehension, relocation and anterior release tests all appear to be diagnostic of anterior instability, especially when apprehension is used as a “positive” test instead of pain.
- ▶ For AC joint pathology, pain with palpation may be valuable as a screen when negative due to high sensitivity and the active compression test may have value as a confirmatory test when positive due to its high specificity.

## CONCLUSIONS

After an extensive qualitative review and meta-analysis of OSTs of the shoulder, there are very few that appear to be diagnostically discriminatory and, therefore, useful in the clinic. Either the supraspinatus/empty can or infraspinatus test may serve as a confirmatory test for impingement. For rotator cuff tears, the hornblower’s sign may be diagnostic of severe degeneration or absence of the teres minor muscle, the external rotation lag sign (ERLS) may be diagnostic of an infraspinatus muscle tear, and the bear-hug and belly press tests may be valuable for ruling in a subscapularis muscle tear. Further, two tests may have value as tests for any rotator cuff tear, the ERLS as a specific confirmatory test and the supine impingement test as a screening test. Of all the pathologies of the shoulder, glenoid labrum pathology and more specifically SLAP lesions have generated the most enthusiasm in researchers. Many of the OSTs have shown great promise in studies conducted by the originator of the article only to prove far less diagnostic in future studies. With caution, we say that the biceps load II test is diagnostic for SLAP lesions. With regard to anterior instability, the apprehension, relocation and anterior release tests all appear to be diagnostic especially when apprehension is used as a “positive” test instead of pain. For AC joint pathology, pain with palpation may be valuable as a screen when negative due to high sensitivity and the active compression test may have value as a confirmatory test when positive due to its high

specificity. Overall, these recommendations should be viewed as a guide and not an absolute since only two studies<sup>25, 64</sup> in our entire review are adequately powered to detect an OST that has high sensitivity or specificity<sup>70</sup>, one of which<sup>64</sup> is a case-control design, which has been shown to overestimate diagnostic accuracy.<sup>74, 75</sup> We repeat the words of McAlister *et al*<sup>76</sup> from 1999, "Clearly we need large methodologically robust studies on history and physical examination" (p1723).

**Competing interests:** None.

## REFERENCES

- Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;**323**:157–62.
- Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. *BMJ* 2002;**324**:477–80.
- Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;**274**:645–51.
- Ostor AJ, Richards CA, Prevost AT, *et al*. Interrater reproducibility of clinical tests for rotator cuff lesions. *Ann Rheum Dis* 2004;**63**:1288–92.
- Groenier KH, de Winter AF, Winters JC, *et al*. Complaint-severity and cervical spine problems successfully classified patients with shoulder complaints. *J Clin Epidemiol* 2004;**57**:730–6.
- de Winter AF, Jans MP, Scholten RJ, *et al*. Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement. *Ann Rheum Dis* 1999;**58**:272–7.
- Winters JC, Groenier KH, Sobel JS, *et al*. Classification of shoulder complaints in general practice by means of cluster analysis. *Arch Phys Med Rehabil* 1997;**78**:1369–74.
- Norregaard J, Krogsgaard MR, Lorenzen T, *et al*. Diagnosing patients with longstanding shoulder joint pain. *Ann Rheum Dis* 2002;**61**:646–9.
- Dinnes J, Loveman E, McIntyre L, *et al*. The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. *Health Technol Assess* 2003;**7**(29):iii, 1–166.
- Tennent TD, Beach WR, Meyers JF. A review of the special tests associated with shoulder examination. Part II: laxity, instability, and superior labral anterior and posterior (SLAP) lesions. *Am J Sports Med* 2003;**31**:301–7.
- Tennent TD, Beach WR, Meyers JF. A review of the special tests associated with shoulder examination. Part I: the rotator cuff tests. *Am J Sports Med* 2003;**31**:154–60.
- Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. *BMJ* 2004;**328**:1040.
- Jackson JL, O'Malley PG, Kroenke K. Evaluation of acute knee pain in primary care. *Ann Intern Med* 2003;**139**:575–88.
- Whiting P, Rutjes AW, Dinnes J, *et al*. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004;**8**(25):iii, 1–234.
- Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol* 2005;**5**:19.
- de Graaf I, Prak A, Bierma-Zeinstra S, *et al*. Diagnosis of lumbar spinal stenosis: a systematic review of the accuracy of diagnostic tests. *Spine* 2006;**31**:1168–76.
- Sehgal N, Shah RV, McKenzie-Brown AM, *et al*. Diagnostic utility of facet (zygapophysial) joint injections in chronic spinal pain: a systematic review of evidence. *Pain Physician* 2005;**8**:211–24.
- Shah RV, Everett CR, McKenzie-Brown AM, *et al*. Discography as a diagnostic test for spinal pain: a systematic and narrative review. *Pain Physician* 2005;**8**:187–209.
- Cook C, Hegedus E. *Orthopedic physical examination tests: an evidence-based approach*. Upper Saddle River, NJ: Prentice Hall, 2007.
- Higgins JP, Thompson SG, Deeks JJ, *et al*. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60.
- Calis M, Akgun K, Birtane M, *et al*. Diagnostic values of clinical diagnostic tests in subacromial impingement syndrome. *Ann Rheum Dis* 2000;**59**:44–7.
- MacDonald PB, Clark P, Sutherland K. An analysis of the diagnostic accuracy of the Hawkins and Neer subacromial impingement signs. *J Shoulder Elbow Surg* 2000;**9**:299–301.
- Zaslav KR. Internal rotation resistance strength test: a new diagnostic test to differentiate intra-articular pathology from outlet (Neer) impingement syndrome in the shoulder. *J Shoulder Elbow Surg* 2001;**10**:23–7.
- Bak K, Faunl P. Clinical findings in competitive swimmers with shoulder pain. *Am J Sports Med* 1997;**25**:254–60.
- Park HB, Yokota A, Gill HS, *et al*. Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *J Bone Joint Surg Am* 2005;**87**:1446–55.
- Leroux JL, Thomas E, Bonnel F, *et al*. Diagnostic value of clinical tests for shoulder impingement syndrome. *Rev Rhum Engl Ed* 1995;**62**:423–8.
- Holtby R, Razmjou H. Validity of the supraspinatus test as a single clinical test in diagnosing patients with rotator cuff pathology. *J Orthop Sports Phys Ther* 2004;**34**:194–200.
- Hertel R, Ballmer FT, Lombert SM, *et al*. Lag signs in the diagnosis of rotator cuff rupture. *J Shoulder Elbow Surg* 1996;**5**:307–13.
- Lyons AR, Tomlinson JE. Clinical diagnosis of tears of the rotator cuff. *J Bone Joint Surg Br* 1992;**74**:414–5.
- Wolf EM, Agrawal V. Transdeltoid palpation (the rent test) in the diagnosis of rotator cuff tears. *J Shoulder Elbow Surg* 2001;**10**:470–3.
- Boileau P, Ahrens PM, Hatzidakis AM. Entrapment of the long head of the biceps tendon: the hourglass biceps: a cause of pain and locking of the shoulder. *J Shoulder Elbow Surg* 2004;**13**:249–57.
- Litaker D, Piro M, El Bilbeisi H, *et al*. Returning to the bedside: using the history and physical examination to identify rotator cuff tears. *J Am Geriatr Soc* 2000;**48**:1633–7.
- Murrell GA, Walton JR. Diagnosis of rotator cuff tears. *Lancet* 2001;**357**:769–70.
- Barth JR, Burkhart SS, De Beer JF. The bear-hug test: a new and sensitive test for diagnosing a subscapularis tear. *Arthroscopy* 2006;**22**:1076–84.
- Itoi E, Kido T, Sano A, *et al*. Which is more useful, the "full can test" or the "empty can test," in detecting the torn supraspinatus tendon? *Am J Sports Med* 1999;**27**:65–8.
- Gerber C, Krushell RJ. Isolated rupture of the tendon of the subscapularis muscle. Clinical features in 16 cases. *J Bone Joint Surg Br* 1991;**73**:389–94.
- Walch G, Boulahia A, Calderone S, *et al*. The 'dropping' and 'hornblower's' signs in evaluation of rotator-cuff tears. *J Bone Joint Surg Br* 1998;**80**:624–8.
- Scheibel M, Magosch P, Pritsch M, *et al*. The belly-off sign: a new clinical diagnostic sign for subscapularis lesions. *Arthroscopy* 2005;**21**:1229–35.
- Liu SH, Henry MH, Nuccione SL. A prospective evaluation of a new physical examination in predicting glenoid labral tears. *Am J Sports Med* 1996;**24**:721–5.
- 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**:806–9.
- Hammer DL, Pink MM, Jobe FW. A modification of the relocation test: arthroscopic findings associated with a positive test. *J Shoulder Elbow Surg* 2000;**9**:263–7.
- Bennett WF. Specificity of the Speed's test: arthroscopic technique for evaluating the biceps tendon at the level of the bicipital groove. *Arthroscopy* 1998;**14**:789–96.
- Holtby R, Razmjou H. Accuracy of the Speed's and Yergason's tests in detecting biceps pathology and SLAP lesions: comparison with arthroscopic findings. *Arthroscopy* 2004;**20**:231–6.
- Ardic F, Kahraman Y, Kacar M, *et al*. Shoulder impingement syndrome: relationships between clinical, functional, and radiologic findings. *Am J Phys Med Rehabil* 2006;**85**:53–60.
- Parentis MA, Glousman RE, Mohr KS, *et al*. An evaluation of the provocative tests for superior labral anterior posterior lesions. *Am J Sports Med* 2006;**34**:265–8.
- 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**:1315–20.
- Nakagawa S, Yoneda M, Hayashida K, *et al*. Forced shoulder abduction and elbow flexion test: a new simple clinical test to detect superior labral injury in the throwing shoulder. *Arthroscopy* 2005;**21**:1290–5.
- McFarland EG, Kim TK, Savino RM. Clinical assessment of three common tests for superior labral anterior-posterior lesions. *Am J Sports Med* 2002;**30**:810–5.
- Morgan CD, Burkhart SS, Palmeri M, *et al*. Type II SLAP lesions: three subtypes and their relationships to superior instability and rotator cuff tears. *Arthroscopy* 1998;**14**:553–65.
- O'Brien SJ, Pagnani MJ, Fealy S, *et al*. The active compression test: a new and effective test for diagnosing labral tears and acromioclavicular joint abnormality. *Am J Sports Med* 1998;**26**:610–3.
- Kibler WB. Specificity and sensitivity of the anterior slide test in throwing athletes with superior glenoid labral tears. *Arthroscopy* 1995;**11**:296–300.
- Guanche CA, Jones DC. Clinical testing for tears of the glenoid labrum. *Arthroscopy* 2003;**19**:517–23.
- Mimori K, Muneta T, Nakagawa T, *et al*. A new pain provocation test for superior labral tears of the shoulder. *Am J Sports Med* 1999;**27**:137–42.
- 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**:300–3.
- Kim SH, Ha KI, Ahn JH, *et al*. Biceps load test II: A clinical test for SLAP lesions of the shoulder. *Arthroscopy* 2001;**17**:160–4.
- Berg EE, Ciullo JV. A clinical test for superior glenoid labral or 'SLAP' lesions. *Clin J Sport Med* 1998;**8**:121–3.
- Kim SH, Park JS, Jeong WK, *et al*. 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**:1188–92.
- Meister K, Buckley B, Batts J. The posterior impingement sign: diagnosis of rotator cuff and posterior labral tears secondary to internal impingement in overhand athletes. *Am J Orthop* 2004;**33**:412–5.
- Lo IK, Nonweiler B, Woolfrey M, *et al*. An evaluation of the apprehension, relocation, and surprise tests for anterior shoulder instability. *Am J Sports Med* 2004;**32**:301–7.
- Speer KP, Hannafin JA, Altchek DW, *et al*. An evaluation of the shoulder relocation test. *Am J Sports Med* 1994;**22**:177–83.
- Farber AJ, Castillo R, Clough M, *et al*. Clinical assessment of three common tests for traumatic anterior shoulder instability. *J Bone Joint Surg Am* 2006;**88**:1467–74.
- Gross ML, Distefano MC. Anterior release test. A new test for occult shoulder instability. *Clin Orthop Relat Res* 1997:105–8.
- Savoie FH3rd, Field LD, Atchinson S. Anterior superior instability with rotator cuff tearing: SLAC lesion. *Orthop Clin North Am* 2001;**32**:457–61, ix.

64. **Chronopoulos E**, Kim TK, Park HB, *et al.* Diagnostic value of physical tests for isolated chronic acromioclavicular lesions. *Am J Sports Med* 2004;**32**:655–61.
65. **Walton J**, Mahajan S, Paxinos A, *et al.* Diagnostic values of tests for acromioclavicular joint pain. *J Bone Joint Surg Am* 2004;**86A**:807–12.
66. **Neer CS2nd**. Impingement lesions. *Clin Orthop Relat Res* 1983;(173):70–7.
67. **Jaeschke R**, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;**271**:703–7.
68. **Obuchowski NA**, Graham RJ, Baker ME, *et al.* Ten criteria for effective screening: their application to multislice CT screening for pulmonary and colorectal cancers. *AJR Am J Roentgenol* 2001;**176**:1357–62.
69. **Davidson M**. The interpretation of diagnostic test: a primer for physiotherapists. *Aust J Physiother* 2002;**48**:227–32.
70. **Flahault A**, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005;**58**:859–62.
71. **Parentis MA**, Mohr KJ, El Attrache NS. Disorders of the superior labrum: review and treatment guidelines. *Clin Orthop Relat Res* 2002;(400):77–87.
72. **Tzannes A**, Murrell GA. Clinical examination of the unstable shoulder. *Sports Med* 2002;**32**:447–57.
73. **Bigliani LU**, Nicholson GP, Flatow EL. Arthroscopic resection of the distal clavicle. *Orthop Clin North Am* 1993;**24**:133–41.
74. **Lijmer JG**, Mol BW, Heisterkamp S, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.
75. **Rutjes AW**, Reitsma JB, Di Nisio M, *et al.* Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;**174**:469–76.
76. **McAlister FA**, Straus SE, Sackett DL. Why we need large, simple studies of the clinical examination: the problem and a proposed solution. CARE-COAD1 group. Clinical Assessment of the Reliability of the Examination-Chronic Obstructive Airways Disease Group. *Lancet* 1999;**354**:1721–4.

## Commentary

This paper attempts to distinguish the good scientific “wood” from the “trees” with respect to examining the shoulder. There are not many publications of good scientific rigour in this area. The authors are to be commended on their efforts.

**Professor George Murrell**, St George Hospital, University of New South Wales, Department of Orthopaedic Surgery, Kogarah, NSW, Australia; murrell.g@ori.org.au

### Date for your diary

#### International Forum on Quality & Safety in Health Care

22–25 April 2008

Le Palais de Congrès de Paris

Paris, France

<http://www.internationalforum.bmj.com>

#### Why attend?

- ▶ Join over 1000 healthcare professionals from over 40 countries worldwide
- ▶ Learn from experienced leaders and improvement experts
- ▶ Find out what is current in quality and safety
- ▶ Gain new skills and tools for creating change in your organisation
- ▶ Take home practical solutions for improvement methods
- ▶ Network with like-minded colleagues

This is a premier event for those dedicated to improving quality and safety in healthcare. Our focus is on practical, take-home ideas that will leave you inspired to implement valuable, lasting change on your return to the workplace.

#### Register online from January 2008 onwards

For more information on the forum visit - <http://www.internationalforum.bmj.com>