Most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review

Phillip C Hughes, Nicholas F Taylor and Rod A Green

La Trobe University
Australia

Question: Do clinical tests accurately diagnose rotator cuff pathology? **Design**: A systematic review of investigations into the diagnostic accuracy of clinical tests for rotator cuff pathology. **Participants**: People with shoulder pain who underwent clinical testing in order to diagnose rotator cuff pathology. **Outcome measures**: The diagnostic accuracy of clinical tests was determined using likelihood ratios. **Results**: Thirteen studies met the inclusion criteria. The 13 studies evaluated 14 clinical tests in 89 separate evaluations of diagnostic accuracy. Only one evaluation, palpation for supraspinatus ruptures, resulted in significant positive *and* negative likelihood ratios. Eight of the 89 evaluations resulted in either significant positive *or* negative likelihood ratios. However, none of these eight positive *or* negative likelihood ratios were found in other studies. Of the 89 evaluations of clinical tests 71 (80%) did not result in either significant positive or negative likelihood ratio evaluations across different studies. **Conclusion**: Overall, most tests for rotator cuff pathology were inaccurate and cannot be recommended for clinical use. At best, suspicion of a rotator cuff tear may be heightened by a positive palpation, combined Hawkins/painful arc/infraspinatus test, Napoleon test, lift-off test, belly-press test, or drop-arm test, and it may be reduced by a negative palpation, empty can test or Hawkins-Kennedy test. **[Hughes PC, Taylor NF, Green RA (2008) Most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review. *Australian Journal of Physiotherapy* 54: 159–170]**

Key words: Rotator cuff; Diagnosis, differential, Review

Introduction

Shoulder pain can be a debilitating condition and is estimated to be the third most common cause of musculoskeletal consultation in primary care (Urwin et al 1998). Rotator cuff pathology may be a major cause of shoulder pain. Using tests that are the subject of this review, Ostor et al (2005) found rotator cuff tendinopathy to be present in 85% of patients presenting to a general medical practice with shoulder pain. Murrell and Walton (2001) reported that rotator cuff tears account for up to 50% of major shoulder injuries, but noted that they are sometimes difficult to diagnose.

Two reviews have been completed investigating tests for rotator cuff pathology and both have questioned the diagnostic accuracy of clinical tests of rotator cuff pathology. Dinnes et al (2003) reviewed the diagnostic accuracy of investigations including ultrasound and magnetic resonance imaging without focusing on clinical testing. Hegedus et al (2008) reviewed clinical tests for all shoulder pathology, not just the rotator cuff, but included studies that had not used accepted reference standards such as operation report or magnetic resonance imaging.

A lack of consensus on diagnostic criteria and concordance in clinical assessment may complicate the choice of intervention (Mitchell et al 2005). Accurate clinical testing should facilitate timely and appropriate intervention for patients presenting with shoulder pain and suspected rotator cuff pathology. Therefore, the research question for this review was:

> Do clinical tests accurately diagnose rotator cuff pathology?

Method

Identification and selection of studies

Electronic data bases AMED, CINAHL, Embase, Medline, SportsDISCUS were searched from January, 1966 to April, 2007 (see Appendix 1 on the eAddenda for the search strategy). Two key concepts were used for the search. The two concepts were linked in the search, using the 'and' operator and each concept comprised 'or' operators. The terms in the first concept were rotator cuff or the individual muscles which contribute to the rotator cuff or the names of standard clinical tests for rotator cuff pathology as described by Brukner and Kahn (2006) and Donatelli (2004). The terms in the second concept related to diagnostic accuracy. The search was supplemented by a search of the references of included studies. Three reviewers (PH, RG and NT) independently screened the title and abstract of papers identified in the initial search strategy against the inclusion criteria (Box 1) and potentially relevant studies were retrieved for evaluation of full text. Differences of opinion between reviewers were resolved by consensus.

Box 1. Inclusion criteria

Studies in peer-reviewed journals

English language studies

Human participants

Subjects presenting with shoulder problems

Clinical diagnostic testing for rotator cuff pathology (tear or inflammatory change)

Clinical tests used are primarily for the diagnosis of rotator cuff pathology and may include (but are not restricted to) clinical tests taken from two standard texts (Donatelli 2004, Brukner and Kahn 2006):

- Locking test
- Neer and Welsh impingement test
- Hawkins and Kennedy Impingement test
- Supraspinatus test
- Gilcrest sign
- Gerber's lift-off test
- Patte's test
- Drop-arm test
- External rotation lag sign
- Internal rotation lag sign
- Drop sign (Donatelli 2004)
- Painful arc
- Passive flexion – pain at end of range
- Empty can test
- Impingement test (Brukner & Kahn 2006)

Results of the clinical tests are compared to the findings of a reference standard – MRI or operation report

Sufficient data are presented to allow calculation of specificity and sensitivity for the clinical tests

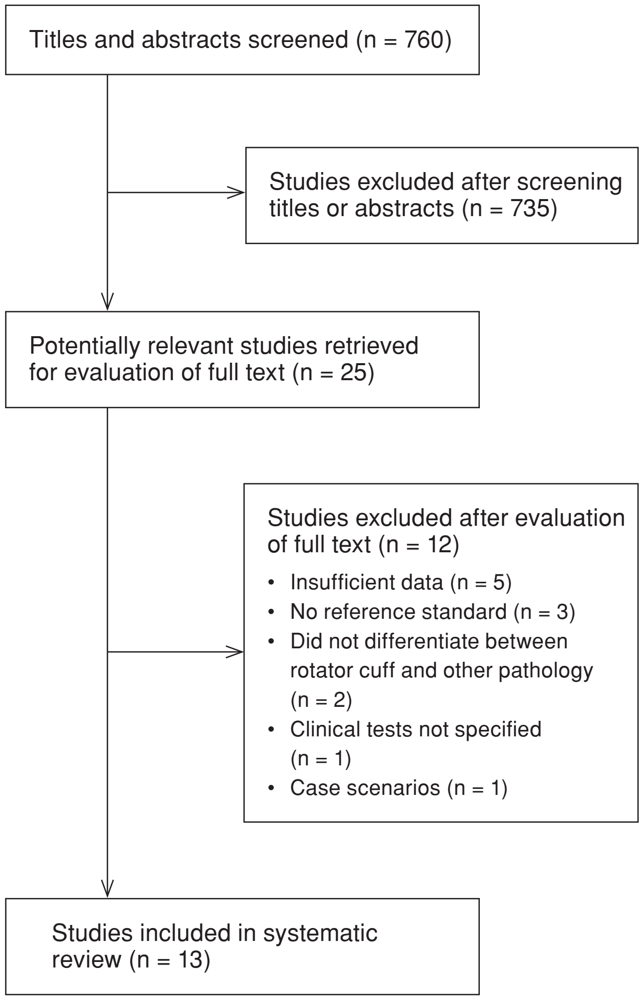

Figure 1. Flow of studies through the review.

Studies were included in the review if they were full reports of English language studies in peer-reviewed journals, involving participants presenting with shoulder pain who underwent clinical diagnostic testing using tests such as, but not restricted to, those proposed for rotator cuff testing in two standard texts, Brukner and Kahn (2006) or Donatelli (2004). Studies were included if they compared the results of clinical testing for rotator cuff pathology with the findings of a reference standard appropriate for rotator cuff injury. Sackett and Haynes (2002) recommend operation report as a reference standard in diagnostic testing, while magnetic resonance imaging has been reported to be to be highly accurate for the detection of rotator cuff lesions (Ardic et al 2006). Studies were only included if they

Table 1. Assessment of methodological quality.*

Question	Rule
Were patients selected consecutively?	Check if consecutive patients with the features of interest were enrolled, or randomly selected from patients presenting with shoulder pain.
Was the decision to perform the reference standard independent of the test results?	Check if all the people who presented with shoulder pain (as opposed to only those with a positive test) received the reference standard.
Was there a valid reference standard?	Check if the all the patients underwent surgery or MRI and were included in the analysis.
Was the test and reference standards measured independently (ie. blind to each other)?	Check if the clinical tests and the reference standard were measured blind to the results of each other. If they were silent on this, accept that they were not blind.
If the reference standard was a later event that the test aimed to predict, was any intervention decision blind to the test result?	Check if there was no treatment between the clinical test and the reference standard. If they were silent on this, accept that there was no treatment between the clinical teat and reference standard.

*adapted from National Health and Medical Research Council 1999

Table 2. Summary of included studies.

Study	Clinical test	Reference standard	Participants
Ardic et al (2006)	Hawkins-Kennedy Neer (impingement)	MRI	n = 58 (59 shoulders) Gender = 13 M, 45 F Age = 55.5 yr
Barth et al (2006)	Bear-Hug Belly-Press Lift-off Napoleon (subscapularis)	Arthroscopy	n = 68 Gender = 49 M, 19 F Age = 45.1 yr
Calis et al (2000)	Hawkins-Kennedy Neer (impingement) Drop-arm Horizontal adduction Painful arc (supraspinatus)	MRI	n = 86 (87 shoulders) Gender = 48 M, 72 F Age = 51.6 yr
Holtby & Razmjou (2004)	Empty can test (supraspinatus)	Operation or arthroscopy	n = 50 Gender = 34 M, 16 F Age = 50 yr
Itoi et al (999)	Empty can test Full can test (supraspinatus)	MRI	n = 136 (143 shoulders) Gender = 105 M, 31 F Age = 43 yr
Itoi et al 2006	Empty can test Full can test (supraspinatus) External Rotation Strength test (infraspinatus) Lift-off (subscapularis)	Arthroscopy	n = 149 (160 shoulders) Gender = not reported Age = 53 yr
Kim et al (2006)	Empty can test Full can test (supraspinatus)	MRI	n = 200 Gender = 84 M, 116 F Age = 59.5 yr
Leroux et al (1995)*	Empty can test (supraspinatus) Patte's test (infraspinatus) Lift-off (subscapularis)	Operation	n = 55 Gender = 33 M, 22 F Age = 51 yr
Lyons & Tomlinson (1992)	Palpation (supraspinatus)	Operation	n = 42 Gender = 25 M, 17 F Age = not reported
MacDonald et al (2000)	Hawkins-Kennedy Neer (impingement)	Arthroscopy	n = 85 Gender = 62 M, 23 F Age = 40 yr
Murrell & Walton (2001)*	Drop-arm sign (supraspinatus)	Operation	n = 400 Gender = not reported Age = not reported
Park et al (2005)	Horizontal adduction Drop-arm sign Hawkins-Kennedy Neer (impingement) External Rotation Strength test (infraspinatus) Painful arc Empty can test (supraspinatus)	Arthroscopy	n = 552 Gender = not reported Age = not reported
Wolf & Agrawal (2001)	Palpation (supraspinatus)	Arthroscopy	n = 109 Gender = 67 M, 42 F Age = 51.2 yr

MRI = magnetic resonance imaging; *Note: Only clinical tests with sensitivity and specificity values were included in the final analysis

reported sensitivity and specificity values (or enough data to calculate sensitivity and specificity values) which allowed the calculation of likelihood values as an indication of the diagnostic accuracy of the clinical tests.

Assessment of methodological quality of studies

To reduce sources of bias, three reviewers independently assessed the included studies for methodological quality using criteria adapted from guidelines for appraising studies concerned with diagnostic tests by the National Health and Medical Research Council (1999). Differences of opinion between reviewers were resolved by consensus. Table 1 outlines the questions and the interpretive rules that were applied to assess quality.

Table 3. Quality of included studies.

Study	Were the patients selected consecutively?	Was the decision to perform the reference standard independent of the test results?	Was there a valid reference standard?	Were the test and reference standards measured independently?	If the reference standard was a later event that the test aimed to predict, was any intervention decision blind to the test result?
Ardic et al (2006)	Y	Y	Y	Y	Y
Barth et al (2006)	Y	Y	Y	N	Y
Calis et al (2000)	Y	Y	Y	N	Y
Holtby & Razmjou (2004)	Y	N	Y	Y	Y
Itoi et al (1999)	Y	Y	Y	N	Y
Itoi et al (2006)	N	Y	Y	N	Y
Kim et al (2006)	Y	N	Y	N	Y
Leroux et al (1995)	Y	Y	Y	N	Y
Lyons & Tomlinson (1992)	N	N	Y	N	Y
MacDonald et al (2000)	Y	N	Y	N	Y
Murrell & Walton (2001)	Y	Y	Y	N	Y
Park et al (2005)	N	Y	Y	N	Y
Wolf & Agrawal (2001)	Y	N	Y	N	Y

Table 4. Distribution of likelihood ratios for 89 evaluations of diagnostic accuracy for clinical tests of rotator cuff pathology.

	+LR		
	< 5	5–10	> 10
–LR			
> 0.2	71	4	6
0.1–0.2	5	0	0
< 0.1	2	0	1

Pale blue area = +LR > 10 or –LR < 0.1; Dark blue area = +LR >10 *and* –LR <0.1

Data analysis

Data were extracted from each included study using a standard form developed for the review. One reviewer extracted data, which were then checked by a second reviewer. Data extracted included research designs of the included trials, the clinical test and reference standards used, participant demographics, diagnostic criteria of the clinical tests, the degree of tear, and sensitivity and specificity values.

Diagnostic accuracy was determined using likelihood ratios. Where they were not reported, positive likelihood ratios (+LR) and negative likelihood ratios (–LR) were calculated from sensitivity and specificity values. Likelihood ratios are clinically useful statistics for summarising diagnostic accuracy (Deeks and Altman 2004) and are considered to be the best indices of diagnostic validity (Riddle and Stratford 1999). They assess the accuracy of a diagnostic test in terms of shifting the pre-test probability of the patient truly having the condition of interest (Einstein et al 1997). The following guidelines have been suggested for the interpretation of likelihood ratios (Jaeschke et al 1994):

Significant shift = +LR greater than 10, and –LR less than 0.1
Small shift = +LR between 5 and 10, and –LR between 0.1 and 0.2
Smaller shift = +LR between 2 and 5, and –LR between 0.2 and 0.5
Rarely important shifts = +LR between 1 and 2, and –LR between 0.5 and 1
Irrelevant shifts = LR close to 1

Based on these guidelines, a clinical test was considered to be diagnostically accurate if it had a positive likelihood ratio greater than 10 and/or a negative likelihood ratio less than 0.1. Where 2 × 2 tables were provided in the studies or case, and control numbers could be confidently calculated to integer numbers from reported sensitivity and specificity values, 95% confidence intervals were calculated and reported. Meta-analysis was performed if there was homogeneity of methods used and tests investigated across studies.

Results

Identification and selection of studies

The search strategy yielded 760 studies. Initial screening reduced this to 25 studies by application of the inclusion and exclusion criteria to title and abstract. Full copies of the 25 studies were obtained. Twelve of these 25 studies were omitted because they did not meet the inclusion criteria, leaving a final yield of 13 studies (Figure 1). No studies were obtained from a secondary search of the references of these studies.

Five of the 25 studies were omitted because insufficient data were presented to allow calculation of sensitivity and

Table 5. Sensitivity, specificity, and likelihood ratios for impingement tests.

Clinical test	Study	n (shoulders)	Diagnostic criteria	Degree of tear	Sensitivity (%)	Specificity (%)	+LR (95% CI)	–LR (95% CI)
Hawkins-Kennedy	Calis et al (2000)	87	Pain	Zlatkin Stage 1	95.2	30.7	1.37	0.16
		87	Pain	Zlatkin Stage 2	87.5	23.0	1.14	0.54
		87	Pain	Zlatkin Stage 3	100	35.7	1.56	0.00
	MacDonald et al (2000)	85	Pain	Severity not stated	87.5	42.6	1.53 (1.17 to 1.99)	0.29 (0.10 to 0.88)
	Park et al (2005)	552	Pain	Any severity	71.5	66.3	2.12	0.43
		552	Pain	PTT	75.4	44.4	1.36	0.55
		552	Pain	FTT	68.7	48.3	1.33	0.65
Horizontal adduction	Calis et al (2000)	87	Pain	Zlatkin Stage 1	61.9	30.7	0.89	1.24
		87	Pain	Zlatkin Stage 2	83.3	23.0	1.08	0.73
		87	Pain	Zlatkin Stage 3	90.0	28.5	1.26	0.35
	Park et al (2005)	552	Pain	Any severity	22.5	82.0	1.25	0.95
		552	Pain	PTT	16.7	78.5	0.78	1.06
		552	Pain	FTT	23.4	80.8	1.22	0.95
Neer	Calis et al (2000)	87	Pain	Zlatkin Stage 1	71.4	30.7	1.03	0.93
		87	Pain	Zlatkin Stage 2	91.6	26.9	1.25	0.31
		87	Pain	Zlatkin Stage 3	90.0	28.5	1.26	0.35
	MacDonald et al (2000)	85	Pain	Severity not stated	83.3	50.8	1.69 (1.24 to 2.31)	0.33 (0.13 to 0.83)
	Park et al (2005)	552	Pain	Any severity	68	68.7	2.19	0.47
		552	Pain	PTT	75.4	47.5	1.44	0.52
		552	Pain	FTT	59.3	47.2	1.12	0.86

+LR = positive likelihood ratio, –LR = negative likelihood ratio; Zlatkin Stage 1 = increased signal intensity in the tendon without any thinning or irregularity, Zlatkin Stage 2 = increased MRI signal intensity in the tendon with thinning or irregularity, Zlatkin Stage 3 = complete disruption of the supraspinatus tendon; PTT = partial thickness tear; FTT = full thickness tear

specificity values for the clinical tests (Bryant et al 2002, Frost et al 1999, Hertel et al 1996, Norregaard et al 2002, Scheibel et al 2005). Arithmetical errors in data presentation by Hertel et al (1996) meant that sensitivity and specificity could not be calculated with confidence. Three studies used a reference standard that did not meet the inclusion criteria (Bak & Faunl 1997, Litaker et al 2000, Walch et al 1998). One study did not specify the clinical tests used (Malhi and Khan 2005). One study was not primarily testing for rotator cuff pathology and did not differentiate between rotator cuff and other shoulder pathology (Zaslav 2001). One study used case scenarios rather than human subjects (Razmjou et al 2006), and one study did not discriminate between labral and rotator cuff pathology (Meister et al 2004).

Table 2 summarises the clinical tests investigated, the reference standard used, and the participants investigated. Four studies evaluated participants with subacromial syndrome, and seven studies investigated supraspinatus testing. Four studies used magnetic resonance imaging as a reference standard, while nine studies relied on an operation report. The number of participants in the studies ranged from 42 to 552, with a mean sample size of 156. The age of the participants ranged from 24 to over 77 years. In the ten studies that provided figures, there were 403 female participants and 520 males.

Quality of studies

Table 3 presents the quality of the included studies. In eight of the 13 studies, the decision to perform the reference standard was independent of the results of the clinical tests, so that all the people who presented with shoulder pain (as opposed to only those with a positive test) received the reference standard. In five studies, only those who had surgery were included. Only two studies described blinded measurement of the clinical tests and reference standards (Ardic et al 2006, Holtby & Razmjou 2004).

Diagnostic accuracy of the clinical tests

The 13 studies included in this review yielded 89 evaluations of diagnostic accuracy for 14 clinical tests. This reflected the evaluation of multiple tests (singly, in combination, or both) in individual studies; and often the evaluation of particular tests or combinations of tests in relation to more than one

Table 6. Sensitivity, specificity, and likelihood ratios for supraspinatus tests.

Clinical test	Study	n (shoulders)	Diagnostic criteria	Degree of tear	Sensitivity %	Specificity %	+LR (95% CI)	–LR (95% CI)
Drop-arm	Calis et al (2000)	87	Pain and weakness	Zlatkin Stage 1	4.4	100	∞	0.96
		87	Pain and weakness	Zlatkin Stage 2	6.2	96.1	1.59	0.98
		87	Pain and weakness	Zlatkin Stage 3	15.0	100	∞	0.85
	Murrell & Walton (2001)	400	Not specified	Not specified	10.0	98.0	5.00	0.92
	Park et al (2005)	552	Sudden drop or severe pain	Any severity	26.9	88.4	2.32	0.83
		552	Sudden drop or severe pain	PTT	14.3	77.5	0.64	1.11
		552	Sudden drop or severe pain	FTT	34.9	87.5	2.79	0.74
Full can test	Itoi et al (1999)	143	Pain	FTT	66	64	1.82 (1.29 to 2.57)	0.54 (0.33 to 0.87)
		143	Weakness (< grade 5)	FTT	77	74	2.98 (2.06 to 4.29)	0.31 (0.17 to 0.57)
		143	Pain/weak/both	FTT	86	57	2.01 (1.56 to 2.60)	0.25 (0.11 to 0.57)
	Itoi et al (2006)	160	Pain	Not stated	80	50	1.60 (1.11 to 2.31)	0.40 (0.24 to 0.66)
		160	Weakness (< grade 5)	Not stated	83	53	1.78 (1.21 to 2.63)	0.32 (0.19 to 0.53)
	Kim et al (2005)	200	Pain	FTT	55.5	77.8	2.50	0.57
		200	Pain	PTT or FTT	71.2	67.9	2.22	0.42
		200	Weakness	FTT	59.9	81	3.15	0.50
		200	Weakness	PTT or FTT	77.3	67.9	2.41	0.33
		200	Pain and weakness	FTT	41.6	90.5	4.38	0.65
		200	Pain and weakness	PTT or FTT	59.1	82.1	3.30	0.50
		200	Pain or weakness	FTT	73.7	68.3	2.32	0.39
		200	Pain or weakness	PTT or FTT	89.4	53.7	1.93	0.20
Painful arc	Calis et al (2000)	87	Pain	Zlatkin Stage 1	9.5	88.4	0.82	1.02
		87	Pain	Zlatkin Stage 2	37.5	73.0	1.39	0.86
		87	Pain	Zlatkin Stage 3	45	78.5	2.09	0.70
	Park et al (2005)	552	Pain	Any severity	73.5	81.1	3.89	0.33
		552	Pain	PTT	67.4	47.0	1.27	0.69
		552	Pain	FTT	75.8	61.8	1.98	0.39

Clinical test	Study	n (shoulders)	Diagnostic criteria	Degree of tear	Sensitivity %	Specificity %	+LR (95% CI)	–LR (95% CI)
Empty can test (supraspinatus or Jobe test)	Holtby & Razmjou (2004)	50	Pain and weakness	PTT	62	54	1.35 (0.79 to 2.28)	0.70 (0.39 to 1.31)
		50	Pain and weakness	FTT	41	70	1.37 (0.63 to 2.93)	0.84 (0.53 to 1.33)
		50	Pain and weakness	Massive FTT	88	70	2.93 (1.64 to 5.11)	0.17 (0.01 to 0.17)
	Itoi et al (1999)	143	Pain	FTT	63	46	1.17 (0.86 to 1.58)	0.81 (0.50 to 1.29)
		143	Weakness	FTT	77	68	2.38 (1.72 to 3.30)	0.34 (0.18 to 0.63)
		143	Pain/weak/both	FTT	89	50	1.77 (1.42 to 2.21)	0.23 (0.09 to 0.59)
	Itoi et al (2006)	160	Pain	Not stated	78	40	1.30 (0.95 to 1.76)	0.56 (0.32 to 0.96)
		160	Weakness	Not stated	87	43	1.53 (1.11 to 2.11)	0.30 (0.17 to 0.55)
	Kim et al (2005)	200	Pain	FTT	79.6	60.3	2.01	0.34
		200	Pain	PTT or FTT	93.9	46.3	1.75	0.13
		200	Weakness	FTT	59.9	88.9	5.40	0.45
		200	Weakness	PTT or FTT	75.8	70.9	2.60	0.34
		200	Pain or weakness	FTT	83.9	58.7	2.03	0.27
		200	Pain or weakness	PTT or FTT	98.5	43.3	1.74	0.03
		200	Pain and weakness	FTT	55.5	90.5	5.84	0.49
		200	Pain and weakness	PTT or FTT	71.2	73.9	2.73	0.39
	Leroux et al (1995)	55	Pain	Tendinitis	86	50	1.72	0.28
		55	Functional impairment	Not stated	79	67	2.39	0.31
	Park et al (2005)	552	Weakness	Any severity	44.1	89.9	4.37	0.62
		552	Weakness	PTT	32.1	67.8	1.00	1.00
		552	Weakness	FTT	52.6	82.4	2.99	0.58
Palpation	Lyons & Tomlinson (1992)	42	Palpation of tendon defect	From none to massive	91	75	3.64 (1.09 to 12.17)	0.12 (0.04 to 0.37)
	Wolf & Agrawal (2001)	109	Palpation of tendon defect	FTT	95.7	96.8	29.91 (7.69 to 117.99)	0.04 (0.01 to 0.17)

+LR = positive likelihood ratio, –LR = negative likelihood ratio; Zlatkin Stage 1 = increased signal intensity in the tendon without any thinning or irregularity, Zlatkin Stage 2 = increased MRI signal intensity in the tendon with thinning or irregularity, Zlatkin Stage 3 = complete disruption of the supraspinatus tendon; PTT = partial thickness tear; FTT = full thickness tear pathology (eg, partial thickness tear and complete rupture, respectively). Meta-analysis was not performed due to the variety of methods and tests used across the studies.

Only one evaluation of diagnostic accuracy produced a positive likelihood ratio above 10 *and* a negative likelihood ratio less than 0.1 (Table 4). The test involved palpation for diagnosing rupture of the supraspinatus tendon (Wolf and Agrawal 2001). However, this result was not found in the only other study involving palpation (Lyons and Tomlinson 1992).

Eight other evaluations of diagnostic accuracy produced a positive likelihood ratio above 10 *or* a negative likelihood ratio less than 0.1. Six of these evaluations produced a positive likelihood ratio above 10: combined Hawkins/painful arc/infraspinatus test, Napoleon, lift-off, belly-press, and drop-arm test (evaluated twice in the one study). Two other evaluations produced a negative likelihood ratio less than 0.1: empty can test and the Hawkins-Kennedy test. However, in none of the tests was this diagnostic accuracy found in another study. Of the 89 evaluations 71 (80%) resulted in a positive likelihood ratio less than five and a negative likelihood ratio greater than 0.2 suggesting that they were inaccurate.

Impingement tests: The sensitivity, specificity, and likelihood ratios for impingement tests are presented in Table 5. The only impingement test to produce a positive likelihood ratio above 10 *or* a negative likelihood ratio below 0.1 was the Hawkins-Kennedy test (the shoulder is passively flexed to 90 degrees and passively internally rotated – pain indicates a positive test). However, this result was not found in six other evaluations across three studies (Calis et al 2000, MacDonald et al 2000, Park et al 2005). The Neer (passive over-pressure of shoulder flexion) and horizontal adduction tests were shown to be inaccurate for the diagnosis of rotator cuff impingement in 13 evaluations across three studies (Calis et al 2000, MacDonald et al 2000, Park et al 2005).

Supraspinatus tests: The sensitivity, specificity, and likelihood ratios for supraspinatus tests are presented in Table 6. Two evaluations of the drop-arm test for supraspinatus pathology (active shoulder abduction to 90 degrees, then return – dropping the arm down with pain indicates a positive test) produced a positive likelihood ratio above 10 *or* a negative likelihood ratio below 0.1 (Calis et al 2000). These results were not found in five other evaluations across three studies (Calis et al 2000, Murrell and Walton 2001, Park et al 2005).

The empty can test, also known as the supraspinatus test or Jobe test (resisted shoulder abduction in internal rotation), demonstrated diagnostic accuracy only once in 21 evaluations across six studies (Holtby and Razmjou 2004, Itoi et al 1999, Itoi et al 2006, Kim et al 2006, Leroux et al 1995, Park et al 2005). Kim et al (2005) reported a negative likelihood ratio of 0.03, using pain or weakness as a criterion, with full or partial thickness tears.

The full can test (resisted shoulder abduction in external rotation) demonstrated a lack of diagnostic accuracy in 13 evaluations of diagnostic accuracy, using pain and/or weakness as criteria, across three studies (Itoi et al 1999, Itoi et al 2006, Kim et al 2006).

The painful arc test (a painful segment in the range of active shoulder abduction) demonstrated a lack of diagnostic

Table 7. Sensitivity, specificity, and likelihood ratios for infraspinatus tests.

Clinical test	Study	n (shoulders)	Diagnostic criteria	Degree of tear	Sensitivity %	Specificity %	+LR (95% CI)	–LR (95% CI)
infraspinatus test (External Rotation Strength test)	Itoi et al (2006)	149	Pain		54	54	1.00 (0.72 to 1.40)	1.00 (0.75 to 1.33)
		149	Weakness		84	53	1.76 (1.37 to 2.26)	0.30 (o.18 to 0.53)
	Park et al (2005)	552	Pain, weakness or positive ERLS	Any severity	41.6	90.1	4.20	0.65
		552	Pain, weakness or positive ERLS	PTT	19.4	69.1	0.63	1.17
		552	Pain, weakness or positive ERLS	FTT	50.5	84	3.16	0.59
Patte's test	Leroux et al (1995)	55	Pain	Tendonitis	92	30	1.31	0.27
		55	Functional impairment	Not reported	83	61	2.13	0.38

+LR = positive likelihood ratio, –LR = negative likelihood ratio; PTT = partial thickness tear; FTT = full thickness tear; ERLS = external rotation lag sign

Table 8. Sensitivity, specificity, and likelihood ratios for subscapularis tests.

Clinical test	Study	n (shoulders)	Diagnostic criteria	Degree of tear	Sensitivity %	Specificity %	+LR (95% CI)	–LR (95% CI)
Bear-hug test	Barth et al (2006)	68	Weakness	Not reported	60	91.7	7.23 (2.64 to 19.65)	0.44 (0.25 to 0.75)
Belly-press test	Barth et al (2006)	68	Weakness	Not reported	40	97.9	19.05 (2.57 to 143.63)	0.61 (0.43 to 0.88)
Lift-off test	Barth et al (2006)	68	Weakness	Not reported	17.6	100	∞	0.82 (0.66 to 1.03)
	Itoi et al (2006)	149	Pain		46	69	1.52 (0.95 to 2.44)	0.77 (0.54 to 1.11)
		149	Weakness		79	59	1.91 (1.44 to 2.53)	0.37 (0.18 to 0.75)
	Leroux et al (1995)	55	Weakness	Not reported	0	61	0	1.64
Napoleon test	Barth et al (2006)	68	Weakness	Not reported	25	97.9	11.90 (1.49 to 96.31)	0.77 (0.59 to 0.99)

+LR = positive likelihood ratio, –LR = negative likelihood ratio

Table 9. Sensitivity, specificity, and likelihood ratios for combination tests.

Clinical test	Study	n (shoulders)	Diagnostic criteria	Degree of tear	Sensitivity %	Specificity %	+LR (95% CI)	–LR (95% CI)
Hawkins-Kennedy or Neer	MacDonald et al (2000)	85	Pain	Severity not stated	87.5	37.7	1.35 (1.04 to 1.75)	0.42 (0.16 to 1.10)
	Ardic et al (2006)	59	Pain	Severity not stated	78.3	50.0	1.57 (0.87 to 2.81)	0.43 (0.19 to 0.95)
Hawkins-Kennedy and Neer	MacDonald et al (2000)	85	Pain	Severity not stated	83.3	55.7	1.88 (1.35 to 2.63)	0.30 (0.12 to 0.75)
Hawkins-Kennedy or Painful arc or infraspinatus test	Park et al (2005)	352	All 3 tests positive	FTT	32.7	98	16.35	0.69
			2 out of 3 positive		34.6	90.3	3.57	0.72
			1 out of 3 positive		23.5	70.3	0.79	1.09

+LR = positive likelihood ratio, –LR = negative likelihood ratio; FTT = full thickness tear, PTT = partial thickness tear accuracy for supraspinatus pathology in six evaluations across two studies (Calis et al 2000, Park et al 2005).

The two studies investigating palpation of the supraspinatus tendon for a tendon rupture both reported high sensitivity values (Lyons and Tomlinson 1992, Wolf and Agrawal 2001). Wolf and Agrawal (2001) also found high specificity, thus producing the most accurate result reported in the review; a +LR of 29.91 and a –LR of 0.04 .

Infraspinatus tests: The sensitivity, specificity, and likelihood ratios for infraspinatus tests are presented in Table 7. The infraspinatus test (resisted external rotation with the arm at the side and elbow flexed to 90 degrees) was inaccurate in five evaluations across two studies (Itoi et al 2006, Park et al 2005). Patte's test (resisted external rotation in 90 degrees shoulder flexion, with the elbow supported by the examiner) also was inaccurate (Leroux et al 1995).

Subscapularis tests: The sensitivity, specificity, and likelihood ratios for subscapularis tests are presented in Table 8. The bear-hug, belly-press, Napoleon, and lift-off tests are variants of subscapularis testing, involving active internal rotation of the shoulder in different positions of shoulder flexion. The evaluation of diagnostic accuracy for the lift-off test produced mixed results. The findings of Barth et al (2006) indicated the lift-off test to be an accurate test for diagnosing subscapularis pathology, using weakness as a criterion, with a sensitivity of 100%, resulting in a positive likelihood ratio of infinity. However, these results were not found by Itoi et al (2006) or Leroux et al (1995). The belly-press (positive likelihood ratio of 19.05) and Napoleon (positive likelihood ratio of 11.9) tests produced positive likelihood ratios greater than 10, while the bear-hug did not (Barth et al 2006).

Combination tests: The sensitivity, specificity, and likelihood ratios for combination tests are presented in Table 9. Combinations of clinical tests produced one accurate result in six evaluations of diagnostic accuracy. When the Hawkins-Kennedy and/or Neer tests were used together, they were diagnostically inaccurate (Ardic et al 2006, MacDonald et al 2000). Park et al (2005) investigated combinations of the Hawkins-Kennedy, painful arc, and infraspinatus tests. When all tests were positive, the positive likelihood ratio of 16.35 demonstrated diagnostic accuracy for full thickness tears.

Discussion

The results indicate that although one evaluation showed a number of the clinical tests to be diagnostically accurate, these findings were not found by other evaluations. Furthermore, the methodological quality of the studies reported in this review was only fair which may have tended to overestimate diagnostic accuracy due to various forms of bias. Despite these methodological shortcomings, the reported accuracy of the clinical tests was still generally poor. Overall, the majority of clinical tests used to diagnose rotator cuff pathology were inaccurate.

The positive likelihood ratio above 10 found in some evaluations, suggests that a positive test for combined Hawkins/painful arc/infraspinatus tests, Napoleon, lift-off, belly-press, or drop-arm tests may increase the likelihood that rotator cuff pathology is present, ie, the clinician has greater confidence than before doing the test that the condition is ruled in. The negative likelihood ratio below 0.1 shown in evaluations of the Hawkins-Kennedy and empty can tests suggests that a negative test may reduce the likelihood that rotator cuff pathology is present, ie, the clinician has greater confidence than before doing the test that the condition is ruled out. However, none of the clinical tests demonstrating diagnostic accuracy with a positive likelihood ratio above 10 or a negative likelihood ratio below 0.1 was found in a second study. For example, the evidence of diagnostic accuracy for the empty can test was not supported in 20 other evaluations across six studies. Similarly, other evaluations of the drop-arm test, the Hawkins-Kennedy test, and the lift-off test did not support the isolated findings of diagnostic accuracy. It is important to consider the possibility that they may represent a Type 1 error (ie, accepting that the clinical test accurately diagnoses rotator cuff pathology when it does not).

The positive likelihood ratio above 10 and the negative likelihood ratio below 0.1 for palpation demonstrated in Wolf and Agrawal (2001) suggests that a positive test indicates that rotator cuff rupture is more likely to be present, while a negative test indicates that it is less likely to be present. However, the other study investigating palpation did not produce the same level of diagnostic accuracy. In other areas of clinical practice, such as the physical examination of people with spinal disorders, palpation has exhibited generally low levels of reliability (May et al 2006). Palpation is a technique that is dependent on the skill of the assessor (Downey et al 1999). The success reported for palpation in Wolf and Agrawal (2001) may not be reproduced to the same degree in other clinical situations. A lack of reproducibility of clinical tests may also have contributed to the poor diagnostic accuracy demonstrated by many of the other clinical tests.

A recent systematic review examined studies concerning the accuracy of clinical tests for the shoulder, including rotator cuff and impingement tests (Hegedus et al 2008). The current review differs from the review by Hegedus et al (2008) in a number of aspects. The current review is concerned solely with rotator cuff pathology, whereas Hegedus et al (2008) included all shoulder pathology. Hegedus et al (2008) also reported the results of studies that used computed tomography results (Walch et al 1998) or double contrast arthrography (Litaker et al 2000) as reference standards. We required both sensitivity and specificity to be provided for a study to be included, whereas Hegedus et al (2008) included studies with only one of these provided. Finally, there was concern that one of the key conclusions of Hegedus et al (2008) – that the empty can test could serve as a confirmatory test for impingement – may have been based on a typographical error, with the specificity of 98% not appearing to reflect the value of 50% reported in the original study (Itoi et al 1999). Despite these differences, Hegedus et al (2008) examined 10 of the 13 papers included in the current review and, overall, the poor accuracy of clinical tests for rotator cuff pathology demonstrated in the current review was found in the results of Hegedus et al (2008).

A possible explanation for the poor diagnostic accuracy found in this review could be that the tests are not anatomically valid. A recent systematic review on the anatomical basis of clinical tests of the shoulder found that there was a lack of evidence for anatomical validity for supraspinatus testing, and likely none for impingement (Green et al 2008). Further enquiry into the anatomical basis of clinical tests for rotator cuff pathology may be a worthwhile direction for the development of accurate clinical tests. However, it may be unrealistic to expect a test to be able to isolate a single structure in order to implicate it in pathology. In their proposed model of impingement, Brukner and Khan (2006) detail the intricate anatomical and functional relationships between structures in the shoulder. Other information, such as mechanism of injury, pain behaviour, and location of pain when combined with clinical tests might provide a more accurate indication of clinical patterns. A suite of criteria, not just clinical tests, may prove to be of greater use in the clinic.

It may be that our present emphasis on pathological diagnosis at the shoulder is misguided. Perhaps clinicians should be describing signs and symptoms and speculating on pathology rather than trying to localise a specific pathologic structure. This is the approach recommended by Maitland and Banks (2001) for the treatment of spinal conditions, where a pathological diagnosis can only be made in about 15% of patients (Waddell 2004). A diagnostic triage with patients categorised as having backache or serious pathology has been proposed (Wadell 2004). A similar approach might be worthy of consideration for patients presenting with shoulder pain.

In conclusion, overall, most tests for rotator cuff pathology were inaccurate and cannot be recommended for clinical use. At most, suspicion of a rotator cuff tear may be heightened by a positive palpation, combined Hawkins/painful arc/infraspinatus test, Napoleon test, Lift-off test, belly-press test, or drop-arm test and it may be reduced by a negative palpation, empty can test or Hawkins-Kennedy test. The poor accuracy of clinical tests for rotator cuff pathology could be related to a lack of anatomical validity of the tests or it may be that the close relationships of structures in the shoulder may make it difficult to identify specific pathologies with clinical tests.

eAddenda: Appendix 1 available at www.physiotherapy.asn.au/AJP

Correspondence: Phillip Hughes, School of Physiotherapy, La Trobe University, Australia 3086. Email: phughes@fastmail.fm

References

Ardic F, Kahraman Y, Kacar M, Kahraman MC, Findikoglu G, Yorganciogl ZR (2006). Shoulder impingement syndrome: relationships between clinical, functional, and radiologic findings. *American Journal of Physical Medicine and Rehabilitation* 85: 53–60.

Bak K, Faunl P (1997) Clinical findings in competitive swimmers with shoulder pain. *American Journal of Sports Medicine* 25: 254–260.

Barth JR, Burkhart SS, De Beer JF (2006) The Bear-Hug Test: A new and sensitive test for diagnosing a subscapularis tear. *Arthroscopy* 22: 1076–1084.

Brukner P, Kahn K (Eds) (2006) Clinical Sports Medicine (3rd edn). Sydney: McGraw-Hill.

Bryant L, Shnier R, Bryant C, Murrell GA (2002) A comparison of clinical estimation, ultrasonography, magnetic resonance imaging, and arthroscopy in determining the size of rotator cuff tears. *Journal of Shoulder and Elbow Surgery* 11: 219–224.

Calis M, Akgun K, Birtane M, Karacan I, Cali H, Tuzun F (2000) Diagnostic values of clinical diagnostic tests in subacromial impingement syndrome. *Annals of the Rheumatic Diseases* 59: 44–47.

Deeks JJ, Altman DG (2004) Diagnostic tests 4: likelihood ratios. *BMJ* 329: 168–169.

Dinnes J, Loverman E, McIntyre L, Waugh N (2003) The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. *Health Technology Assessment* 7: iii), 1–166.

Donatelli RA (Ed) (2004) Physical Therapy of the Shoulder (4th edn). St Louis: Churchill Livingstone.

Downey BJ, Taylor NF, Niere KR (1999) Manipulative physiotherapists can reliably palpate nominated lumbar spinal levels. *Manual Therapy* 4: 151–156.

Einstein AJ, Bodian CA, Gill J (1997) The relationship amongst performance measures in the selection of diagnostic tests. *Archives of Pathology Laboratory Medicine* 121: 64–74.

Frost P, Andersen J, Lundorf E (1999) Is supraspinatus pathology as defined by magnetic resonance imaging associated with clinical sign of shoulder impingement? *Journal of Shoulder and Elbow Surgery* 8: 565–568.

Green R, Shanley K, Taylor N, Perrot M (2008) The anatomical basis for clinical tests assessing musculoskeletal function of the shoulder. *Physical Therapy Reviews* 13: 17–24.

Hegedus E, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman C, Cook C (2008) Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. *British Journal of Sports Medicine,* 42: 80–92.

Hertel R, Ballmer FT, Lombert SM, Gerber C (1996) Lag signs in the diagnosis of rotator cuff rupture. *Journal of Shoulder Elbow Surgery* 5: 307–313.

Holtby R, Razmjou H (2004) Validity of the supraspinatus test as a single clinical test in diagnosing patients with rotator cuff pathology. *Journal of Orthopaedic Sports Physical Therapy* 34: 194–200.

Itoi E, Kido T, Sano A, Urayama M, Sato K (1999) Which is more useful the 'full can test' or the 'empty can test,' in detecting the torn supraspinatus tendon? *American Journal of Sports Medicine* 27: 65–68.

Itoi E, Minagawa H, Yamamoto N, Seki N, Abe H (2006) Are pain location and physical examinations useful in locating a tear site of the rotator cuff? *American Journal of Sports Medicine* 34: 256–264.

Jaeschke R, Guyatt GH, Sacket DL (1994) User's guide 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 working group. *JAMA* 271: 703–707.

Kim E, Jeong HJ, Lee KW, Song JS (2006) Interpreting positive signs of the supraspinatus test in screening for torn rotator cuff. *Acta Medica Okayama* 60: 223–228.

Leroux JL, Thomas E, Bonnel F, Blotman F (1995) Diagnostic value of clinical tests for shoulder impingement syndrome. *Revue du Rhumatisme (English Edition)* 62: 423–428.

Litaker D, Pioro M, El Bilbeisi H, Brems J (2000) Returning to the bedside: using the history and physical examination to identify rotator cuff tears. *Journal of the American Geriatrics Society* 48: 1633–1637.

Lyons AR, Tomlinson JE (1992) Clinical diagnosis of tears of the rotator cuff. *Journal of Bone Joint Surgery–British Volume* 74: 414–415.

May S, Littlewood C, Bishop A (2006) Reliability of procedure in the examination of non-specific low back pain: a systematic review. *Australian Journal of Physiotherapy* 52: 91–102.

MacDonald PB, Clark P, Sutherland K (2000) An analysis of the diagnostic accuracy of the Hawkins and Neer subacromial impingement signs. *Journal of Shoulder Elbow Surgery* 9: 299–301.

Maitland GD, Banks K (2000) Maitland's Vertebral Manipulation (6th edn). Oxford: Harcourt.

Malhi AM, Khan R (2005) Correlation between clinical diagnosis and arthroscopic findings of the shoulder. *Postgraduate Medical Journal* 81: 657–659.

Meister K, Buckley B, Batts J (2004) The posterior impingement sign: diagnosis of rotator cuff and posterior labral tears secondary to internal impingement in overhand athletes. *American Journal of Orthopedics* 33: 412–415.

Mitchell C, Adebajo A, Hay E, Carr A (2005) Shoulder pain: diagnosis and management in primary care. *BMJ* 331: 1124–1128.

Murrell GA, Walton JR (2001) Diagnosis of rotator cuff tears. *Lancet* 357: 769–770.

National Health and Medical Research Council (1999) How to review the evidence: systematic identification and review of the scientific literature. *Handbook series on preparing clinical practice guidelines*. Available at: http://www.nhmrc.gov.au/publications/subjects/clinical.htm

Norregaard J, Krogsgaard MR, Lorenzen T, Jensen EM (2002) Diagnosing patients with longstanding shoulder joint pain. *Annals of the Rheumatic Diseases* 61: 646–649.

Ostor AJ, Richards CA, Prevost AT, Speed CA, Hazleman BL (2005) Diagnosis and relation to general health of shoulder disorders presenting to primary care. *Rheumatology* 44: 800–805.

Park HB, Yokota A, Gill HS, El Rassi G, McFarland EG (2005) Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *Journal of Bone Joint Surgery–American Volume* 87: 1446–1455.

Razmjou H, Haines T, Holtby R (2006) Diagnostic and therapeutic decision-making: exploring the role of pretest probability in patients with rotator cuff pathology. *Physiotherapy Canada* 58: 196–204.

Riddle D, Stratford PW (1999) Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. *Physical Therapy* 79: 939–948.

Sacket DL, Haynes RB (2002) The architecture of diagnostic research. *BMJ* 324: 539–541.

Scheibel M, Magosch P, Pritsch M, Lichtenberg S, Habermeyer P (2005) The belly-off sign: A new clinical diagnostic sign for subscapularis lesions. *Arthroscopy* 21: 1229–1235.

Urwin M, Synons D, Allison T, Brammah TH, Roxby M (1998) Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Annals of Rheumatic Disease* 57: 649–655.

Wadell G (2004) The Back Pain Revolution (2nd edn). Edinburgh: Churchill Livingstone.

Walch G, Boulahia A, Calderone S, Robinson A (1998) The ‘dropping’ and ‘hornblower’s’ signs in evaluation of rotator-cuff tears. *Journal of Bone and Joint Surgery (British)* 80: 624–628.

Wolf EM, Agrawal V (2001) Transdeltoid palpation (the rent test) in the diagnosis of rotator cuff tears. *Journal of Shoulder Elbow Surgery* 10: 470–473.

Zaslav KR (2001) Internal rotation resistance strength test: a new diagnostic test to differentiate intra-articular pathology from outlet (Neer) impingement syndrome in the shoulder. *Journal of Shoulder and Elbow Surgery* 10: 23–27.

