

# Reliability of Common Provocative Tests for Shoulder Tendinitis

Robert Doxey, MD, MOH, Matt S. Thiese, PhD, MSPH, and Kurt T. Hegmann, MD, MPH

**Objective:** Accurate and reliable shoulder tendinopathy examination maneuvers are essential for diagnosing and treating shoulder pain; however, studies have reported varying results as to the accuracy of common maneuvers. Thus, data from a large, cross-sectional study were used to systematically quantify the reliability and accuracy of clinical diagnostic tests. **Methods:** Baseline data from the WISTAH cohort study were used to evaluate inter-tester reliability and accuracy of common provocative shoulder examination tests compared with a case definition of shoulder tendinitis. **Results:** Inter-tester reliability showed reliable consistency between providers with kappa coefficients between 89.5% and 94.8% for all tests. However, sensitivity was generally poor (3.0% to 60.6%). Specificity was consistently high (96.2% to 99.6%). **Conclusion:** Common shoulder provocative tests show low sensitivity but high specificity, which implies that the primary utility of examination maneuvers is for their negative predictive values.

**Keywords:** physical examination, predictive value, rotator cuff, sensitivity, shoulder

Shoulder disorders are the third most common occupational musculoskeletal complaint<sup>1</sup> and constitute 7% to 10% of all occupational injury and disease.<sup>2</sup> Shoulder pain is also a common complaint involving an estimated 1% to 12% of all primary care encounters.<sup>3</sup> Chronicity is also common, as 40% to 50% of patients report continued symptoms between 6 and 12 months after an initial visit.<sup>3</sup> The associated financial burden varies by case, and, in general, sick leave and other indirect expenses constitute the majority of the overall cost of this disorder.<sup>4</sup> The Occupational Safety and Health Administration (OSHA) has developed a program to estimate these costs through the Safety Pays Program. Estimates for a simple musculoskeletal strain are \$33,140 in direct costs and \$36,454 in indirect costs.<sup>5</sup>

Common provocative examination maneuvers for the rotator cuff attempt to primarily increase impingement on the supraspinatus tendon in the subacromial space (Neer, Hawkins–Kennedy, painful arc) or increase the tension on the tendon by activating the supraspinatus muscle (Jobe or empty can, resisted external rotation of the forearm). The sensitivity of these impingement tests reportedly ranges from 63% to 97% and specificity for these tests ranges from 24% to 96%. The sensitivity of the tension-based provocative test reportedly ranges from 50% to 76% with a specificity range of 87% to 96%. The wide range in these reported data may be due to varying techniques and study weaknesses, thus calling into question the reliability of some of the more common provocative tests.<sup>3,6–8</sup>

Michener et al<sup>9</sup> used surgical confirmation to investigate shoulder examination tests and reported relatively poor kappa

## Learning Objectives

- Become familiar with the occupational importance of shoulder disorders and the common provocative maneuvers used to diagnose shoulder pain.
- Summarize the new findings on diagnostic performance of shoulder tendinopathy examination maneuvers.
- Discuss the implications for the clinical utility of these provocative tests for shoulder pain.

coefficient's ranging between 0.39 and 0.67. They also reported select test sensitivities and specificities ranged widely (50% to 81% and 54% to 87%, respectively), and indicated that the painful arc, external rotation resistance test, and Jobe tests have the best utility and reliability, but have improved accuracy when test results are combined. Litaker et al<sup>10</sup> used arthrography as a reference. However, a surgical or arthrographic standard<sup>11,12</sup> are questionable reference standards for the utility of examination tests among typical outpatient clinical patients, as most patients do not undergo and do not require procedures.

A cross-sectional evaluation of a large population of workers to systematically assess the sensitivity, specificity, and reproducibility of common provocative shoulder maneuvers against a case definition of shoulder tendinitis has not been reported and is the purpose of this report.

## METHODS

This report is from cross-sectional analyses of workers in the WISTAH study. Institutional Review Boards approved the study at the University of Utah and University of Wisconsin-Milwaukee. As detailed WISTAH study methods are previously available, a focused summary of the methods follows.<sup>13</sup> Participants received normal wages while participating in study activities, but no other incentives were paid.

A convenience sample of workers were invited to participate from 15 different companies in Illinois, Utah, and Wisconsin. Workers were recruited regardless of the presence of symptoms. Instead, recruitment goals were to enroll approximately one-third each in low-, medium-, and high-job physical demands positions.

Health data included those from a questionnaire, structured interview, and two physical examinations. The questionnaire included demographics, habits, and psychosocial factors. The structured interview was performed by either an occupational medicine resident or therapist. A symptoms diagram was used with workers to help localize symptoms (see Fig. 1). The research team member who performed the structured interview also performed the first of the two physical examinations.

Two physical examinations were performed on all workers and were conducted with knowledge of the results of the structured interview. Examinations were standardized, with examiners having reviewed videotapes and practiced physical examination maneuvers until reproducibility was documented. The first examination was comprehensive and included all physical examination maneuvers regardless of symptoms. The second examination was performed by an experienced, board-certified occupational medicine physician and assessed both the positive and pertinent negatives from the first examination.

From the The University of Utah, Rocky Mountain Center for Occupational and Environmental Health, Salt Lake City, Utah.

Funding provided, in part, by the Centers for Disease Control and Prevention/ National Institute of Occupational Safety and Health under award numbers 1 U 01 OH007917-01 and 3TC42OH008414.

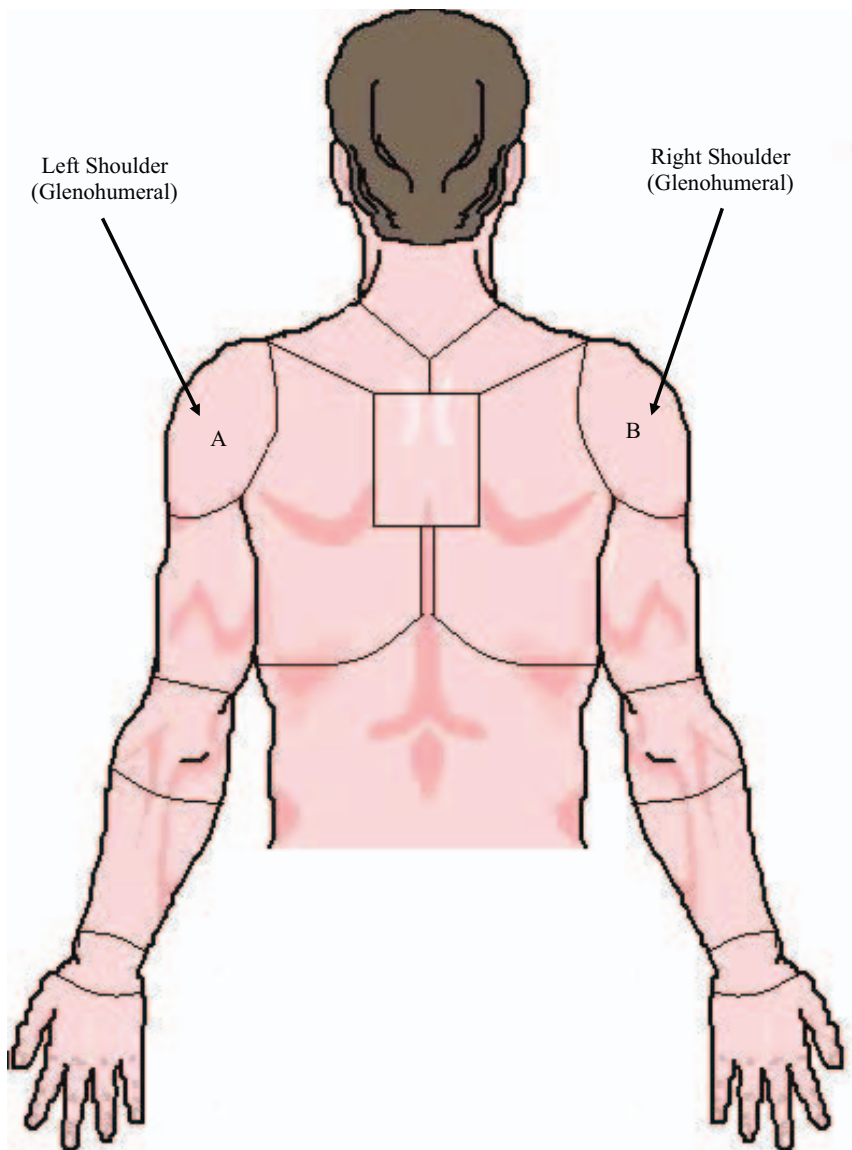
Doxey, Thiese, and Hegmann have no relationships/conditions/circumstances that present potential conflict of interest.

The JOEM editorial board and planners have no financial interest related to this research.

Address correspondence to: Kurt T. Hegmann, MD, MPH, University of Utah, Rocky Mountain Center for Occupational and Environmental Health, 391 Chipeta Way, Suite C, Salt Lake City, UT 84108 (Kurt.Hegmann@hsc.utah.edu)

Copyright © 2018 American College of Occupational and Environmental Medicine

DOI: 10.1097/JOM.0000000000001430



**FIGURE 1.** Symptoms diagram used for workers to localize shoulder pain. Glenohumeral joint pain was required for this study's case definition (areas A and B).

This study's case definition for shoulder tendinitis required a history of glenohumeral joint shoulder pain in the 30 days before enrollment and a positive Jobe test (empty can, or supraspinatus test) on the second examiner's examination.<sup>14</sup> The following examination tests were assessed: Neer, painful arc, and resisted external rotation. Only workers with two completed examinations were included in these analyses.

### Statistical Analysis

STATA statistical analysis software (StataCorp, LLC, College Station, Texas) was used for analyses. Inter-tester reliability for the aforementioned provocative tests was analyzed comparing the two examinations with kappa reliability coefficients. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were determined for each test except Jobe, as it was a component of the case definition.

### RESULTS

A total of 1216 workers were enrolled, but 146 were dropped from these analyses due to incomplete data, leaving a total of

1070 workers for these analyses. Most of the workers were female (66.1%) and the overall mean age was  $42.2 \pm 11.4$  years (see Table 1). The body mass index (BMI) ranged from 16.2 to 58.5 with an average of  $29.4 \pm 6.8$  kg/m<sup>2</sup>. Most (86.8%) were right-handed, while 5.4% considered himself/herself ambidextrous.

Right shoulder pain in the prior month was reported among 256 of 1070 (23.9%) participants and 126 (11.8%) had right shoulder pain at the time of the examination. In the month before enrollment, left shoulder pain was reported among 215 of 1070 (20.1%) participants and 99 (9.3%) had left shoulder pain at the time of the examination.

Table 2 details inter-tester reliability data. The total numbers assessed naturally varied due to the study examination methods described above. The inter-tester reliability ranged from 87.0% to 94.8% for the five tests that were evaluated.

Table 3 details the operant characteristics of the four provocative tests evaluated. For the right shoulder, 66 participants were found to meet the case definition of rotator cuff tendinitis. Almost one-third, 20 of the 66 (30.3%), reported a history of right rotator cuff tendinitis and nine reported a previous right rotator cuff tear.

**TABLE 1.** Baseline Characteristics of Subjects (1,070 Total)

Demographic	Total
Age: Range: 18–68 years, Mean ± SD: 42.2 ± 11.4	
Male	363 (33.9%)
Female	707 (66.1%)
BMI, kg/m <sup>2</sup> : Range: 16.2–58.5 Mean ± SD: 29.4 ± 6.8	
BMI <18.5	10 (0.9%)
BMI 18.5–25	290 (27.1%)
BMI 25–30	348 (32.5%)
BMI >30	422 (39.4%)
Right handed	921 (86.8%)
Left handed	89 (7.9%)
Ambidextrous	60 (5.4%)
Left shoulder	
Pain last 30 days	215 (20.1%)
Current pain	99 (9.3%)
Current pain intensity, Mean ± SD: 4.9 ± 2.7	
Hx of tendinitis	32 (3.0%)
Hx of rotator cuff tear	18 (1.7%)
Numbness and tingling current	22 (2.1%)
Numbness and tingling past 30 days	48 (4.5%)
History of clavicle fracture	17 (1.6%)
History of humerus fracture	5 (0.5%)
History of shoulder dislocation	18 (1.7%)
Right shoulder	
Pain in last 30 days	256 (23.9%)
Current pain	126 (11.8%)
Current pain intensity, Mean ± SD: 5.0 ± 2.7	
Hx of tendinitis	53 (5.0%)
Hx of rotator cuff tear	31 (2.9%)
Numbness and tingling current	22 (2.1%)
Numbness and tingling past 30 days	49 (4.6%)
History of clavicle fracture	20 (1.9%)
History of humerus fracture	7 (0.7%)
History of shoulder dislocation	28 (2.6%)

BMI, body mass index.

Two workers reported a previous right clavicle fracture. Six reported a previous right shoulder dislocation. Of the 66 workers, none reported a history of a humeral fracture.

A total of 58 workers met the case definition of left rotator cuff tendinitis. Of those 58 workers, 11 (19.0%) also met the case definition bilaterally and were also included in the analyses of the 66 workers who met the case definition on the right side. A total of 10 of the 58 (17.2%) reported a previous left rotator cuff tear. None of these workers reported a previous left clavicle fracture. Two

reported a previous left shoulder dislocation and one reported a history of a humeral fracture.

Sensitivity results for shoulder tendinitis were poor, especially with the external rotation test at 3.0% to 6.9%. The painful arc test had the highest sensitivity at 44.1% to 60.6%, but revealed a 16.5% difference with higher sensitivity on the right than left. Abnormal range of motion had a similar incongruity with a 15.5% difference from right to left.

Specificity results were high, ranging from 96.2% to 99.6% with good consistency between right and left shoulders. Both PPV and NPV followed the sensitivity and specificity results with lower values for PPV and good results for NPV.

### DISCUSSION

Our analysis of common provocative shoulder examination measures revealed low sensitivities and high specificities. Compared with a case definition of shoulder tendinitis, the painful arc test has the highest sensitivity at 44.1% to 60.6%, followed by Neer impingement (39.7% to 43.9%). External rotation has the lowest sensitivity (3.0% to 6.9%). However, specificities are all greater than 95%. The supraspinatus (Jobe or empty can test) was not assessed, as it was part of the case definition. These data suggest the primary utilities of the examination maneuvers are specificity, NPV, and negative likelihood ratios, which are not well taught in clinical training. These results also suggest these examination maneuvers are not readily usable for purposes of screening workers for prevalent disease, and raise concerns about reproducible case definitions used for both clinical trials and epidemiological investigations.

These results add to a growing body of evidence regarding the inconsistency or unreliability of many of these examination maneuvers used as a diagnostic method for shoulder pathology.<sup>3,6–8</sup> This analysis has contributed to the inconsistent body of data surrounding these diagnostic methods for diagnosis of shoulder problems.

A Cochrane review on the subject assessed 33 different studies dated up to 2010 in order to determine if a simple shoulder maneuver could reliably identify a common shoulder diagnosis. Out of these 33 studies, they found 170 different combinations of tests and target conditions. Only 6 of the 170 test and target condition combinations were found to be duplicated in all of the 33 different studies assessed. This indicates a lack of agreement with the purpose and application of these shoulder maneuvers.<sup>8</sup>

In examining a related study data for each test, we found no comparable studies that included this maneuver in the protocol. Two comparable studies reported a sensitivity for Yergason of 37% to 41% and a specificity of 79% to 87%,<sup>15,16</sup> which somewhat followed the

**TABLE 2.** Inter-Tester Reliability for Five Provocative Shoulder Examination Tests

Results	Total Assessed*	Examiner 1 + Results	Examiner 2 + Results	Inter-tester Reliability %
Right shoulder				
Abnormal range of motion	317	17	31	93.6
+ Yergason	518	27	22	94.8
+ External rotation Weakness	245	19	4	93.9
+ Neer	587	77	54	89.6
+ Painful arc	589	67	72	92.7
+ Jobe	591	90	81	87.0
Left shoulder				
Abnormal range of motion	291	7	21	93.8
+ Yergason	513	27	16	94.7
+ External rotation Weakness	253	6	21	92.8
+ Neer	564	61	38	89.5
+ Painful arc	566	45	48	94.2
+ Jobe	570	59	71	89.5

\*Only patients with test results from both providers were used in the analysis. Patients with a normal history and examination were fully examined by one physician or physical therapist.

**TABLE 3.** Operant Characteristics of Provocative Tests for Rotator Cuff Tendinitis\*

Results	Positive Examination (n)	Sensitivity	Specificity	PPV	NPV	Rotator Cuff Tendinitis
Right shoulder						
Abnormal range of motion	34	25.8%	98.3%	50.0%	95.3%	66
+ Yergason	24	16.7%	98.7%	45.8%	94.7%	66
+ External rotation weakness	6	3.0%	99.6%	33.3%	94.0%	66
+ Neer	52	43.9%	97.7%	45.8%	96.4%	66
+ Painful arc	67	60.6%	97.3%	59.7%	97.4%	66
Left shoulder						
Abnormal range of motion	23	10.3%	98.3%	26.1%	95.0%	58
+ Yergason	17	19.0%	99.4%	64.7%	95.5%	58
+ External rotation weakness	10	6.9%	99.4%	40.0%	94.9%	58
+ Neer	39	39.7%	98.4%	59.0%	96.6%	58
+ Painful arc	43	44.1%	96.2%	56.5%	96.6%	58

NPV, negative predictive value; PPV, positive predictive value.

\*Jobe test not included in calculation, as it is part of the diagnostic criteria.

same pattern of our values for sensitivity, 16.7% to 19.0%, and 98.7% to 99.4% for specificity. Two comparable studies were found for external rotator weakness with values for sensitivity and specificity at 56% to 94% and 87% to 95%, respectively,<sup>9,17</sup> yet our values for external rotator weakness differed significantly with a sensitivity range of 3.0% to 6.9% and a specificity range of 99.4% to 99.6%. Two comparable studies of Neer impingement syndrome reported a sensitivity of 28% to 89% and a specificity of 32% to 96%,<sup>11,15</sup> whereas our study found a sensitivity range of 39.7% to 43.9% and specificity range of 97.7% to 98.4%, and inconsistencies from the right shoulder to the left shoulder. Lastly, two comparable studies reported a sensitivity of 32% to 75% and a specificity of 67% to 82% for the painful arc,<sup>9,15</sup> with which our results were comparable with a sensitivity of 44.1% to 60.6% and a specificity of 96.2% to 97.3%.

Inter-tester reliability in this study showed better examination reproducibility than other like studies.<sup>9</sup> This is comparable to the study by Michener et al,<sup>9</sup> which also performed inter-tester reliability on four of the five tests used in our study, abnormal ROM being excluded. The results for the kappa coefficient in that study ranged from 40% to 67%, which is significantly different than our range of 87% to 94.8% for these same tests. The improved results in our study may likely be due to the implementation of standardized examinations, which likely enhanced reproducibility. However, as the examinations were not blinded, the potential for some influence of prior positive results to have carried forward to the second examination could have occurred despite requiring the recording of results regardless of positive or negative results.

Study strengths include a large sample with a population-based, multistate approach rather than a clinic-based assessment. The systematic capture of data, including use of structured interviews, is another significant study strength. Another strength is the multiple standardized examiners performing the examinations among workers in three states. A limitation is that the second examiner was not blinded to the first examiner's results.

### CONCLUSION

This study found good specificities, but poor sensitivities for common provocative shoulder examination tests. The primary use of these examination maneuvers is their NPVs, which is a concept that is not well taught in medical training.

### REFERENCES

1. Somerville L, Bryant D, Willits K, Johnson A. Protocol for determining the diagnostic validity of physical examination maneuvers for shoulder pathology. *BMC Musculoskelet Disord*. 2013;14:60.

2. Walker-Bone K, Cooper C. Hard work never hurt anyone—or did it? A review of occupational associations with soft tissue musculoskeletal disorders of the neck and upper limb. *Ann Rheum Dis*. 2005;64:1112–1117.
3. May S, Chance-Larse K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. *Physiotherapy*. 2010;96:179–190.
4. Kuijpers T, van Tulder MW, van der Heijden GJMG, Bouter LM, van der Windt DAWM. Cost of shoulder pain in primary care consultants: a prospective cohort study in The Netherlands. *BMC Musculoskelet Disord*. 2006;7:83.
5. Department of Labor. The Occupational Safety and Health Administration (OSHA) Safety Pays Program. Available at: <https://www.osha.gov/dcsps/smallbusiness/safetypays/>. Accessed May 6, 2017.
6. Dinnes J, Loveman E, McIntyre L, Waugh N. The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. *Health Technol Assess*. 2003;7:1–166. iii.
7. Hegedus EJ, Goode AP, Cook CE, et al. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. *Br J Sports Med*. 2012;46:964–978.
8. Hanchard NCA, Lenza M, Handoll HHG, Takwoingi Y. Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. *Cochrane Database Syst Rev*. 2013;(4):CD007427.
9. Michener L, Walsworth MK, Doukas WC, Murphy KP. Reliability and diagnostic accuracy of physical examination tests and combination of tests for subacromial impingement. *Arch Phys Med Rehabil*. 2009;90:1898–18903.
10. Litaker D, Pioro M, El Bilbeisi H, Brems J. Returning to the bedside: using the history and physical examination to identify rotator cuff tears. *J Am Geriatr Soc*. 2000;48:1633–1637.
11. Villafane JH, Valdes K, Anselmi F, Piralì C, Negrini S. The diagnostic accuracy of five tests for diagnosing partial-thickness tears of the supraspinatus tendon: a cohort study. *J Hand Ther*. 2015;28:247–252.
12. Kim HA, Kim SH, Seo Y-I. Ultrasonographic findings of painful shoulders and correlation between physical examination and ultrasonographic rotator cuff tear. *Mod Rheumatol*. 2007;17:213–219.
13. Garg A, Hegmann KT, Wertsch JJ, et al. The WISTAH hand study: a prospective cohort study of distal upper extremity musculoskeletal disorders. *BMC Musculoskelet Disord*. 2012;13:1471–2474.
14. Applegate K, Thiese MS, Merryweather AS, et al. Association between cardiovascular disease risk factors and rotator cuff tendinopathy. *J Occup Environ Med*. 2017;59:154–160.
15. Calis M, Akgun K, Birtane M, Karacan I, Calis H, Tuzun F. Diagnostic values of clinical diagnostic tests in subacromial impingement syndrome. *Ann Rheum Dis*. 2000;59:44–47.
16. Kibler WB, Sciasca A, Hester P, Dome D, Jacobs C. Clinical utility of traditional and new tests in the diagnosis of biceps tendon injuries and superior labrum anterior and posterior lesions in the shoulder. *Am J Sports Med*. 2009;37:1840–1847.
17. Iagnocco A, Coari G, Leone A, Valesini G. Sonographic study of painful shoulder. *Clin Exp Rheumatol*. 2003;21:355–358.