Diagnostic Accuracy of Clinical Tests for Subscapularis Tears

A Systematic Review and Meta-analysis

Alexandre Lädermann,*†‡§ MD, Philippe Collin,‖ MD, Olivia Zbinden,† MD, Timon Meynard,† MSc, Mo Saffarini,¶ MEng, MBA, FRSM, and Joe Chih-Hao Chiu,# MD, PhD
Investigation performed at La Tour Hospital, Meyrin, Switzerland

Background: Previous systematic reviews and meta-analyses on the diagnostic accuracy of shoulder clinical tests do not reach conclusions regarding subscapularis tears.

Purpose: To compare the diagnostic accuracy of commonly used clinical tests for subscapularis tears.

Study Design: Systematic review; Level of evidence, 3.

Methods: An electronic literature search was conducted using Medline, Embase, and the Cochrane Library/Central. Eligibility criteria were original clinical studies reporting the diagnostic accuracy of clinical tests to diagnose the presence of rotator cuff tears involving the subscapularis.

Results: The electronic literature search returned 2212 records, of which 13 articles were eligible. Among 8 tests included in the systematic review, the lift-off test was most frequently reported (12 studies). Four tests were eligible for meta-analysis: bear-hug test, belly-press test, internal rotation lag sign (IRLS), and lift-off test. The highest pooled sensitivity was 0.55 (95% CI, 0.28-0.79) for the bear-hug test, while the lowest pooled sensitivity was 0.32 (95% CI, 0.13-0.61), for the IRLS. In all tests, pooled specificity was >0.90.

Conclusion: Among the 4 clinical tests eligible for meta-analysis (bear-hug test, belly-press test, IRLS, and lift-off test), all had pooled specificity >0.90 but pooled sensitivity <0.60. No single clinical test is sufficiently reliable to diagnose subscapularis tears.

Registration: PROSPERO (CRD42019137019).

Keywords: clinical tests; subscapularis; rotator cuff tear; diagnostic accuracy; systematic review; meta-analysis

Assessment of history and physical examination are the first steps in diagnosing patients presenting with shoulder pain, which is often the result of degenerative rotator cuff disease.[33] Primary physical examination includes clinical tests that aim to reproduce symptoms to identify which tendons are torn.

More than 180 shoulder clinical tests have been described in the literature.[28] In some instances, the same test is used to diagnose different tendons; in others, the same test may simply have a different name. This heterogeneity of clinical tests, in purpose and terminology, renders the assessment of their diagnostic accuracy difficult and leads clinicians to question their usefulness altogether.[12] Tests commonly used to diagnose subscapularis tears, whether isolated or concomitant with supraspinatus tears, involve active internal shoulder rotation at different

flexion angles.[27] The lift-off test[11] was the first test designed to evaluate the integrity of the subscapularis, followed by the internal rotation lag sign (IRLS),[15] the belly-press test,[10] and a variant of the latter, the Napoleon test.[35] The belly-off sign and bear-hug tests were later described by Scheibel et al[35] and Barth et al,[4] respectively.

Previous systematic reviews[6,14,17] and meta-analyses[12,13] on the diagnostic accuracy of shoulder clinical tests, while providing well-designed analysis of more general shoulder tests, have not yielded conclusions on the reliable detection of subscapularis tears. While a number of recent studies[5,21,39] investigated newer clinical tests used for the diagnosis of subscapularis tears, none compared their diagnostic accuracy across the spectrum of clinical tests available. This systematic review and meta-analysis therefore aims to collect, synthesize, and critically evaluate the literature on the diagnostic accuracy of the clinical tests most commonly utilized for assessing the presence of subscapularis tears and determine any gaps in the literature and directions for future research.

METHODS

This systematic review and meta-analysis adhered to the principles outlined in the handbook of the Cochrane Collaboration[16] and the established guidelines from PRISMA-DTA (Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies).[29] The study protocol, including the search strategy, was registered on PROSPERO (CRD42019137019).

Search Strategy

We conducted an electronic literature search using Medline (1946–), Embase (1980–), and the Cochrane Library/Central on July 7, 2020, using the following search strategy: ("rotator cuff" OR "subscapularis" OR "supraspinatus") AND ("disease" OR "rupture" OR "tear" OR "pathology") AND ("clinical test" OR "clinical examination" OR "physical test" OR "physical examination") (Table 1). The electronic literature search returned 2212 records, of which 710 were duplicates.

The titles and abstracts of the remaining 1502 records were screened by 2 independent reviewers (A.L. and M.S.) to determine relevance according to the following eligibility criteria.

Inclusion Criteria. Each original clinical study had to report at least 1 of the following: (1) true and false positives and true and false negatives; (2) sensitivity and specificity; and/or (3) positive predictive value (PPV) and negative predictive value (NPV) of individual clinical tests (physical examination) against radiographic, arthroscopic, or intraoperative observations. Diagnoses had to focus on the presence of rotator cuff tears involving the subscapularis: either isolated subscapularis tears or anterosuperior tears of the subscapularis and supraspinatus. Patients had to present with shoulder pain, functional impairment, or other evidence of rotator cuff disease.

Exclusion Criteria. Cohorts were excluded if they had patients with shoulder injury <6 weeks, history of shoulder instability, dislocation, rheumatoid arthritis, fracture, fibromyalgia, labral lesion, adhesive capsulitis, tumor, complex regional pain syndrome, or stroke-related disorder. Articles written in languages other than English, French, German, Spanish, or Italian were also excluded.

Study Selection and Data Extraction

Two reviewers (A.L. and M.S.) independently performed the search. The reference lists of all selected publications were checked. Gray literature, systematic reviews, meta-analyses, and guidelines on shoulder clinical tests were searched to retrieve relevant publications not identified in the electronic search. Selection of relevant articles was first performed through titles and then abstracts. Full-text articles were retrieved if the abstract provided insufficient information to establish eligibility or the article passed the first eligibility screening. Disagreements between the reviewers were discussed and resolved by a third independent reviewer (P.C.).

The 2 reviewers independently extracted study characteristics (year of publication, journal, level of evidence, prevalence of subscapularis tears, age, eligibility, reference diagnostic method) and data (true and false positives, true and false negatives, sensitivity, specificity, diagnostic odds ratio, PPV, and NPV). For each finding, the sensitivity, specificity, PPV, NPV, and diagnostic odds ratio with their 95% CIs were recalculated from data in the article, using a continuity correction of 0.5 if applicable.[7]

The 2 reviewers assessed risk of bias on eligible studies using the QUADAS-2 criteria (Quality Assessment of Diagnostic Accuracy Studies).[29] In line with recommendations, the original 14 questions and scoring system were adapted to this study.

*Address correspondence to Alexandre Lädermann, MD, La Tour Hospital, Av. J.D. Maillard 3, 1217 Meyrin, Switzerland (email: alexandre.laedermann@gmail.com).

[†]Division of Orthopedics and Trauma Surgery, La Tour Hospital, Meyrin, Switzerland.

[‡]Faculty of Medicine, University of Geneva, Geneva, Switzerland.

[§]Orthopedics and Trauma Service, University Hospitals of Geneva, Geneva, Switzerland.

[‖]Centre Hospitalier Privé Saint-Grégoire (Vivalto Santé), Saint-Grégoire, France.

[¶]ReSurg SA, Nyon, Switzerland.

[#]Department of Orthopedic Surgery, Chang Gung Memorial Hospital, Taoyuan.

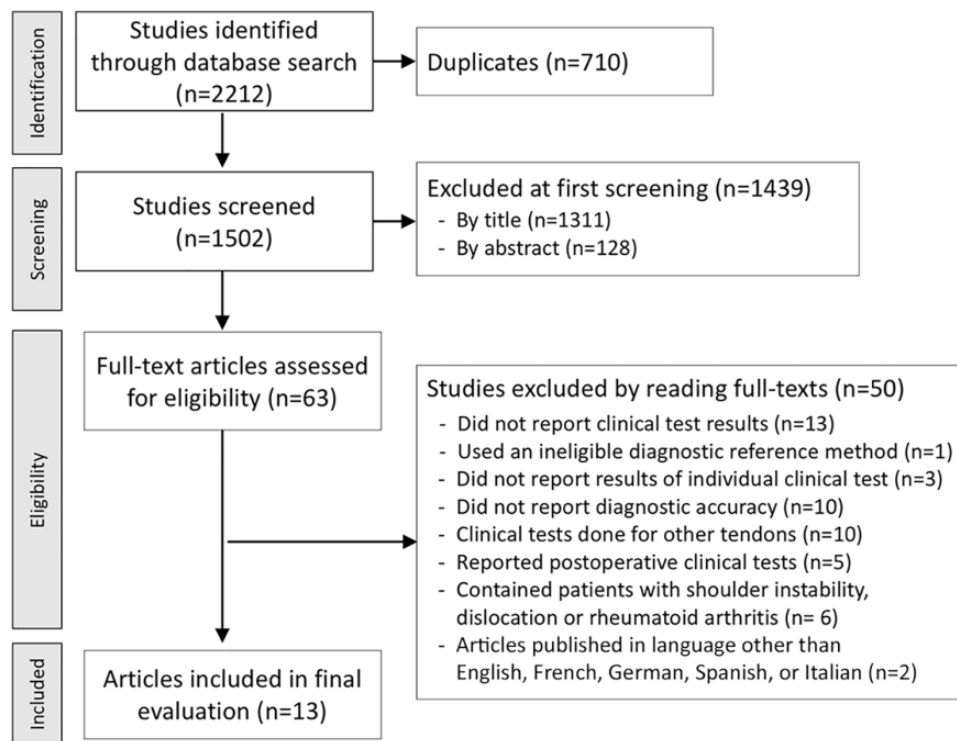Final revision submitted June 4, 2021; accepted June 9, 2021.

Figure 1. PRISMA diagram. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analyses.

Statistical Analysis

Clinical tests were described in summary format when (1) true and false positives and true and false negatives could not be retrieved or (2) tests were reported in only 1 study. A meta-analysis was performed on clinical tests reported in at least 3 studies, for which true and false positives and true and false negatives were described or could be retrieved from corresponding authors. A bivariate random effects approach was taken for the meta-analysis of the pairs of sensitivity and specificity[32] and pairs of PPV and NPV.[24] The main outcomes of interests were the sensitivity/specificity and PPV/NPV for each test, presented with their 95% CIs in forest plots, as well as summary receiver operating characteristic curves, constructed for pairs of sensitivity and specificity. Heterogeneity was investigated visually by examining forest plots. Publication bias could not be evaluated statistically because none of the tests were represented by at least 10 studies.[8] Statistical analyses were performed using R Version 3.5.0 (R Foundation for Statistical Computing) with the *mada* package.

RESULTS

Systematic Review

A total of 1439 articles were excluded by reading their titles or abstracts, and a further 50 were excluded by reading their full text, leaving 13 from which data were extracted for this review (Figure 1). No additional relevant articles were identified from citations in selected studies, gray literature, systematic reviews, meta-analyses, or guidelines.

The 13 eligible studies (Figure 1),[∥] all published between 2006 and 2018, reported diagnostic accuracy for 8 clinical tests: bear-hug test, belly-off sign, belly-press test, IRLS, internal rotation resistance test (IRRT), lift-off test, Napoleon test, and supine Napoleon test. The most frequently cited test was the lift-off test (12 studies), while the least cited tests were the IRRT and supine Napoleon test (1 study each).

The study design was prospective in 11 studies and retrospective in 2 (Table 2). The reference diagnostic method was arthroscopy in 7 studies, ultrasound in 4, and magnetic resonance imaging (MRI) or magnetic resonance arthrography (MRA) in 2. Quality assessment using QUADAS-2 revealed that the risk of bias was low in 3 studies, moderate in 7, and high in 4, owing to flaws regarding patient selection in 5 studies, reference standard in 8, and low and timing in 1 (Table 3).

Meta-analysis

Of the 8 clinical tests, 4 were evaluated in ≥3 studies, which reported true and false positives as well as true and false negatives for any subscapularis tear, and were therefore eligible for meta-analysis: bear-hug test (Figure 2), belly-press test (Figure 3), IRLS (Figure 4), and lift-off test (Figure 5). The most frequently represented test was the

---

∥References 4,5,19,21-23,25,30,34,36-39.

TABLE 2
Description of Included Studies[a]

	Study Design	No. of Patients	Mean Age, y	Reference Test	AS Tears, %	Inclusion Criteria	Exclusion Criteria
Barth (2006)[4]	P	68	45	Arthro	29	Patients scheduled for an arthroscopic procedure between January 2004 and March 2004	Previously operative shoulders and stiff shoulders scheduled for capsular release and lysis of adhesions
Bartsch (2010)[5]	P	50	58[b]	Arthro	30	Patients with subacromial and/or glenohumeral impingement syndrome scheduled for an arthroscopic procedure	Calcifying tendinitis, shoulder stiffness, instability, osteoarthritis, or previous surgery; suspicion or evidence of RC tear and/or stiffness on the contralateral side
Itoi (2006)[19]	R	160	53	Arthro	18	RC tear or cuff tendinitis	—
Kappe (2018)[21]	P	106	57	Arthro	30	Consecutive patients undergoing shoulder arthroscopy at a single institution	Shoulder instability, history of shoulder trauma or surgery, advanced osteoarthritis, or shoulder stiffness
Kim (2007)[22]	P	120	59	US	91	Patients with shoulder pain visiting a rheumatology department	Rheumatoid arthritis, previous trauma
Lasbleiz (2014)[23]	P	39	59	US	5	Ambulatory physiotherapy treatment for degenerative RC disease, age >40 y, shoulder pain >1 mo, degenerative RC disease	Limited range of motion, calcification on radiographs, previous surgery, shoulder instability, humeral fracture, local steroid injections within 30 d, inflammatory joint disease, and neoplastic disorder
Lin (2015)[25]	P	235	51	Arthro	37	Consecutive patients with RC injury	Shoulder stiffness, instability, calcifying tendinitis, and previous surgery; disease on the contralateral shoulder
Miller (2008)[30]	P	37	56	US	33	Shoulder pain, full passive movement, age >18 y	Previous surgery, neurologic symptoms
Salaffi (2010)[34]	P	203	58	US	23	Patients with painful shoulders referred to rheumatology; age, 18 to 70 y	Postoperative pain, diabetes, congenital anomalies, tumor of the shoulder girdle, septic arthritis, inflammatory rheumatic disease
Somerville (2014)[36]	P	139	46	Arthro with MRA	9	Consecutive patients with first-time shoulder complaint at a tertiary care orthopaedic center	Patients who were referred for shoulder replacement surgery
Takeda (2016)[37]	P	130	65	Arthro	40	Patients scheduled to undergo arthroscopic RC repair from February 2013 to February 2015	Shoulder stiffness, osteoarthritis, instability, or a history of shoulder surgery
van Kampen (2014)[38]	P	100	44	MRA	6	Patients with shoulder complaint	Previous diagnosis of shoulder disorders, fractures, frozen shoulder, or arthritis; deficiencies in Dutch; history of shoulder instability
Yoon (2013)[39]	R	312	57	MRI	43	Patients scheduled to undergo arthroscopic RC repair	Severe pain or stiffness or difficulty during clinical or isokinetic muscle performance testing, need of biceps tenotomy or tenodesis, history of shoulder surgery, a symptomatic lesion in the contralateral shoulder, and inflammatory arthritis or disease in the shoulder

[a]Dash indicates the article did not specify the information. Arthro, arthroscopy; AS, anterosuperior; MRA, magnetic resonance arthrography; MRI, magnetic resonance imaging; P, prospective; R, retrospective; RC, rotator cuff; US, ultrasound.
[b]Median.

TABLE 3
Quality Assessment of Studies Using the QUADAS-2[a]

Domain	Barth[4]	Bartsch[5]	Itoi[19]	Kappe[21]	Kim[22]	Lasbleiz[23]	Lin[25]	Miller[30]	Salaffi[34]	Somerville[36]	Takeda[37]	van Kampen[38]	Yoon[39]
Patient selection	–	–	+	–	–	+	–	–	–	–	+	+	+
Index text	–	–	–	–	–	–	–	–	–	–	–	–	–
Reference standard	–	+	+	+	–	+	+	+	+	–	+	–	–
Flow and timing	–	–	+	–	–	–	–	–	–	–	–	–	–
Overall risk of bias[b]	Low	Mod	High	Mod	Low	High	Mod	Mod	Mod	Low	High	Mod	Mod

[a]–, little risk of bias; +, considerable risk of bias; Mod, moderate.
[b]Low, little risk of bias in all 4 domains; moderate, considerable risk of bias in 1 of 4 domains; high, considerable risk of bias in at least 2 of 4 domains.
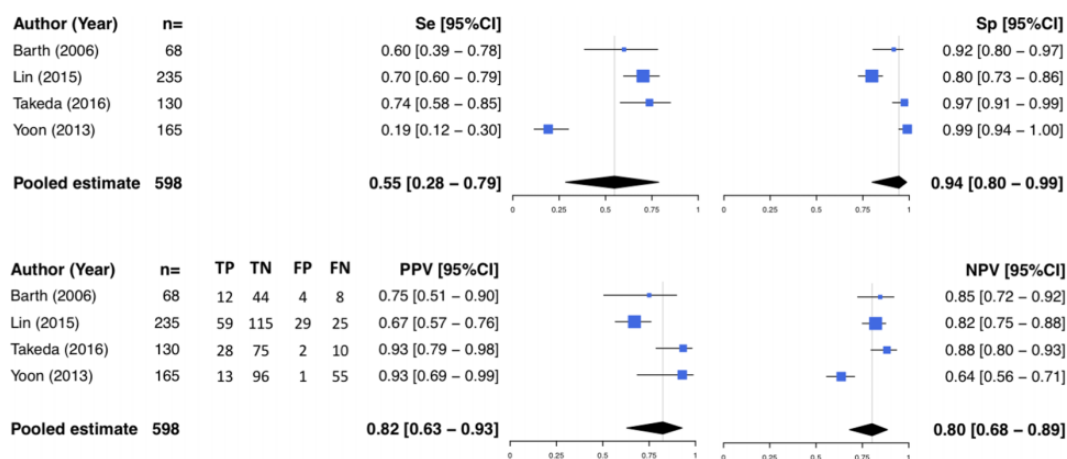

Figure 2. Forest plot representing the diagnostic accuracy of the bear-hug test. FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; Se, sensitivity; Sp, specificity; TN, true negative; TP, true positive.
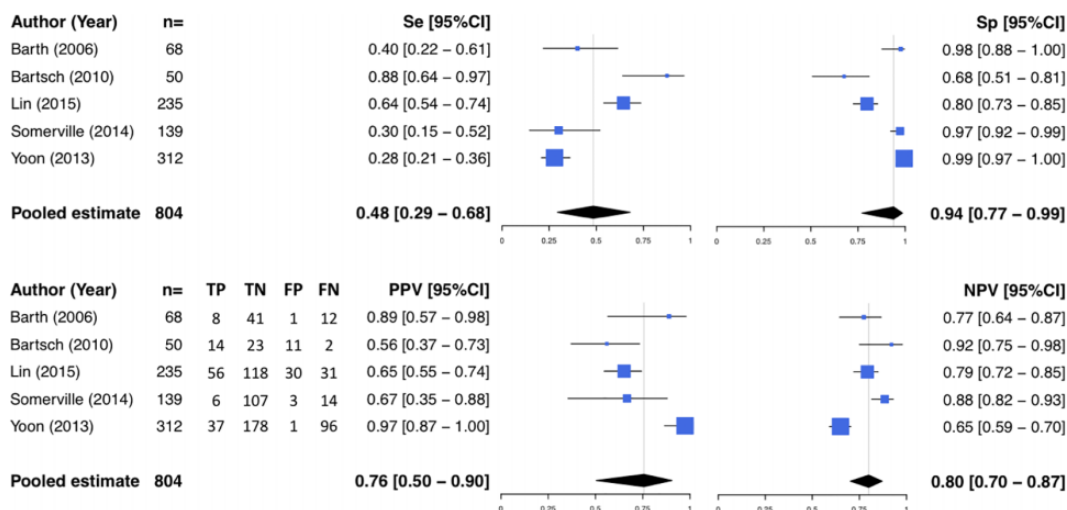

Figure 3. Forest plot representing the diagnostic accuracy of the belly-press test. See Figure 2 for abbreviations.
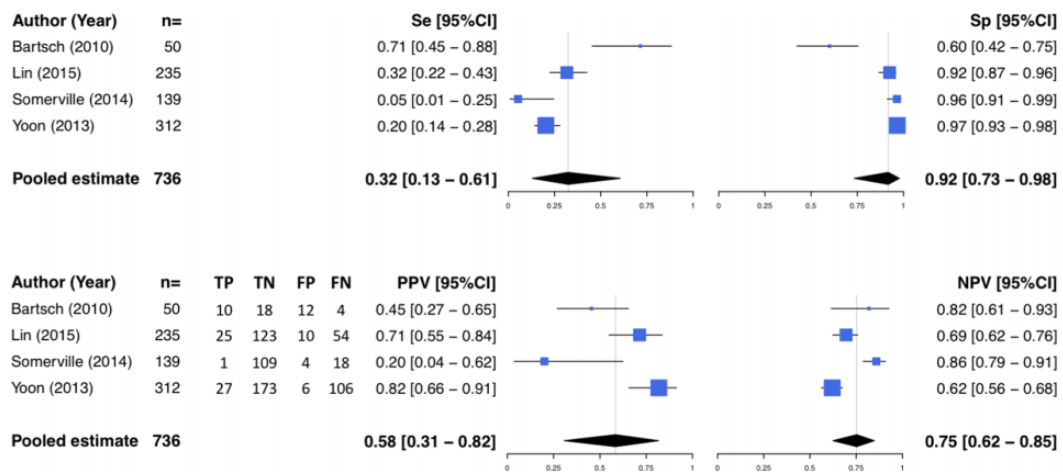
Figure 4. Forest plot representing the diagnostic accuracy of the internal rotation lag sign. See Figure 2 for abbreviations.
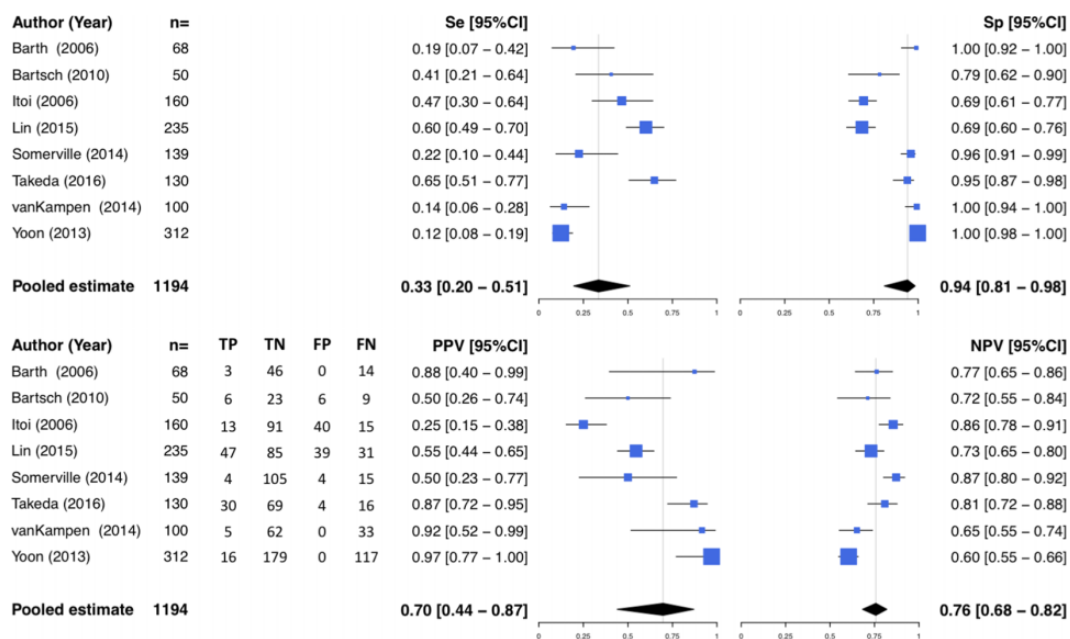

Figure 5. Forest plot representing the diagnostic accuracy of the lift-off test. See Figure 2 for abbreviations.

lift-off test (8 studies), while the least represented were the bear-hug and IRLS tests (4 studies each). The level of evidence was 1 or 2 in 5 studies and 3 or 4 in 3 studies (Table 2). The reference diagnostic method was arthroscopy in 6 studies and MRI or MRA in 2 studies. According to QUADAS-2 criteria, the risk of bias was low in 2 studies, moderate in 4, and high in 2 (Table 3).

The highest pooled sensitivity was 0.55 (95% CI, 0.28-0.79) for the bear-hug test, while the lowest pooled sensitivity was 0.32 (95% CI, 0.13-0.61) for the IRLS. There was considerable variation in reported sensitivity; for each clinical test, there was no overlap in the 95% CIs cited by ≥ 2 studies. The highest pooled specificity was 0.94, achieved

by the bear-hug (95% CI, 0.80-0.99), belly-press (95% CI, 0.77-0.99), and lift-off (95% CI, 0.81-0.98) tests, while the lowest pooled specificity was 0.92 (95% CI, 0.73-0.98) for the IRLS. In all tests, pooled specificity was >0.90. By setting the threshold for sensitivity and specificity at >0.80, none of the tests met both criteria.

The highest pooled PPV was 0.82 (95% CI, 0.63-0.93) for the bear-hug test, while the lowest pooled PPV was 0.58 (95% CI, 0.31-0.82) for the IRLS. The highest pooled NPV was 0.80 (95% CI, 0.70-0.87) for the belly-press, while the lowest pooled NPV was 0.75 (95% CI, 0.62-0.85) for the IRLS. When the threshold for PPV and NPV was set at >0.80, only the bear-hug test met both criteria (Table 4).

TABLE 4
Diagnostic Accuracy of Clinical Tests for Subscapularis Tears vs Reference Observations[a]

Clinical Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	DOR (95% CI)
Bear-hug test					
Kappe (2018)[21]	0.52 (—)	0.85 (—)	0.59 (—)	0.81 (—)	2.0 (1.2-3.2)
Takeda (2016)[37]	0.74 (0.58-0.85)	0.97 (0.91-0.99)	0.93 (0.79-0.98)	0.88 (0.80-0.93)	105.0 (21.6-509.3)
Lin (2015)[25]	0.70 (0.60-0.79)	0.80 (0.73-0.86)	0.67 (0.57-0.76)	0.82 (0.75-0.88)	9.4 (5.0-17.4)
Yoon (2013)[39]	0.19 (0.12-0.30)	0.99 (0.94-1.00)	0.93 (0.69-0.99)	0.64 (0.56-0.71)	22.7 (2.9-178.2)
Barth (2006)[4]	0.60 (0.39-0.78)	0.92 (0.80-0.97)	0.75 (0.51-0.90)	0.85 (0.72-0.92)	16.5 (4.2-64.2)
Belly-off sign					
Kappe (2018)[21]	0.31 (—)	0.97 (—)	0.83 (—)	0.77 (—)	4.6 (1.3-16.5)
Bartsch (2010)[5]	0.87 (0.62-0.96)	0.91 (0.78-0.97)	0.81 (0.57-0.93)	0.94 (0.81-0.98)	69.3 (10.4-464.4)
Belly-press test					
Kappe (2018)[21]	0.34 (—)	0.96 (—)	0.79 (—)	0.77 (—)	3.7 (1.3-10.1)
Lin (2015)[25]	0.64 (0.54-0.74)	0.80 (0.73-0.85)	0.65 (0.55-0.74)	0.79 (0.72-0.85)	7.1 (3.9-12.9)
Somerville (2014)[36]	0.30 (0.15-0.52)	0.97 (0.92-0.99)	0.67 (0.35-0.88)	0.88 (0.82-0.93)	15.3 (3.4-68.1)
Yoon (2013)[39]	0.28 (0.21-0.36)	0.99 (0.97-1.00)	0.97 (0.87-1.00)	0.65 (0.59-0.70)	68.6 (9.3-507.8)
Bartsch(2010)[5]	0.88 (0.64-0.97)	0.68 (0.51-0.81)	0.56 (0.37-0.73)	0.92 (0.75-0.98)	14.6 (2.8-76.0)
Barth (2006)[4]	0.40 (0.22-0.61)	0.98 (0.88-1.00)	0.89 (0.57-0.98)	0.77 (0.64-0.87)	27.3 (3.1-240.9)
Somerville (2014)[36,b]	0.50 (0.22-0.79)	0.96 (0.91-0.98)	0.44 (0.19-0.73)	0.97 (0.92-0.99)	23.4 (4.5-121.8)
Lasbleiz (2014)[23,b,c]	0.40 (0.05-0.85)	0.74 (0.57-0.87)	0.18 (0.02-0.52)	0.89 (0.72-0.98)	—
Lasbleiz (2014)[23,b]	0.60 (0.15-0.95)	1.00 (0.90-1.00)	1.00 (0.29-0.71)	0.94 (0.81-0.99)	—
IRLS					
Kappe (2018)[21]	0.41 (—)	0.91 (—)	0.65 (—)	0.78 (—)	2.4 (1.3-4.4)
Lin (2015)[25]	0.32 (0.22-0.43)	0.92 (0.87-0.96)	0.71 (0.55-0.84)	0.69 (0.62-0.76)	5.5 (2.5-12.1)
Somerville (2014)[36]	0.05 (0.01-0.25)	0.96 (0.91-0.99)	0.20 (0.04-0.62)	0.86 (0.79-0.91)	2.0 (0.3-13.4)
Yoon (2013)[39]	0.20 (0.14-0.28)	0.97 (0.93-0.99)	0.82 (0.66-0.91)	0.62 (0.56-0.68)	6.9 (2.8-16.8)
Bartsch (2010)[5]	0.71 (0.45-0.88)	0.60 (0.42-0.75)	0.45 (0.27-0.65)	0.82 (0.62-0.93)	3.5 (0.9-12.9)
Somerville (2014)[36,b]	0.00 (0.00-0.37)	0.96 (0.90-0.98)	0.08 (0.00-0.48)	0.93 (0.88-0.97)	1.3 (0.1-25.1)
Miller (2008)[30,b]	1.00 (—)	0.84 (—)	0.28 (—)	1.00 (—)	—
IRRT[d]					
Lin (2015)[25]					
0°	0.61 (0.51-0.71)	0.76 (0.69-0.83)	0.61 (0.50-0.70)	0.77 (0.69-0.83)	5.1 (2.9-9.1)
90°	0.77 (0.66-0.84)	0.80 (0.73-0.86)	0.69 (0.59-0.78)	0.86 (0.79-0.91)	13.4 (6.9-25.9)
Lift-off test					
Kappe (2018)[21]	0.35 (—)	0.98 (—)	0.90 (—)	0.76 (—)	8.7 (1.3-56.7)
Takeda (2016)[37]	0.65 (0.51-0.77)	0.95 (0.87-0.98)	0.87 (0.72-0.95)	0.81 (0.72-0.88)	28.5 (9.3-88.0)
Lin (2015)[25]	0.60 (0.49-0.70)	0.69 (0.60-0.76)	0.55 (0.44-0.65)	0.73 (0.65-0.80)	3.3 (1.8-5.9)
Lasbleiz (2014)[23,e]	0.75 (0.19-0.99)	0.91 (0.76-0.98)	0.50 (0.12-0.88)	0.97 (0.84-1.00)	—
Somerville (2014)[36]	0.22 (0.10-0.44)	0.96 (0.91-0.99)	0.50 (0.23-0.77)	0.87 (0.80-0.92)	6.8 (1.7-27.9)
van Kampen (2014)[38]	0.14 (0.06-0.28)	1.00 (0.94-1.00)	0.92 (0.52-0.99)	0.65 (0.55-0.74)	20.5 (1.1-382.5)
Yoon (2013)[39]	0.12 (0.08-0.19)	1.00 (0.98-1.00)	0.97 (0.77-1.00)	0.60 (0.55-0.66)	50.4 (3.0-848.4)
Bartsch (2010)[5]	0.41 (0.21-0.64)	0.79 (0.62-0.90)	0.50 (0.25-0.74)	0.72 (0.55-0.84)	2.5 (0.7-9.3)
Salaffi (2010)[34]	0.35 (0.25-0.48)	0.75 (0.67-0.82)	0.85 (0.70-0.90)	0.21 (0.16-0.20)	—
Kim (2007)[22,e]	0.06 (—)	0.23 (—)			
Barth (2006)[4]	0.19 (0.07-0.42)	1.00 (0.92-1.00)	0.88 (0.40-0.99)	0.77 (0.65-0.86)	22.4 (1.1-460.6)
Itoi (2006)[19,c]	0.47 (0.30-0.64)	0.69 (0.61-0.77)	0.25 (0.15-0.38)	0.86 (0.78-0.91)	2.0 (0.9-4.5)
Itoi (2006)[19,f]	0.78 (0.60-0.90)	0.59 (0.50-0.67)	0.29 (0.20-0.40)	0.93 (0.85-0.97)	24.8 (1.2-531.8)
Itoi (2006)[19,g]	0.09 (0.02-0.23)	1.00 (0.97-1.00)	1.00 (0.34-1.00)	0.83 (0.77-0.88)	4.9 (1.9-12.6)
Lasbleiz (2014)[23,b]	0.50 (0.07-0.93)	1.00 (0.90-1.00)	1.00 (0.16-1.00)	0.94 (0.81-0.99)	—
Somerville (2014)[36,b]	0.28 (0.09-0.59)	0.95 (0.90-0.98)	0.25 (0.07-0.59)	0.95 (0.90-0.98)	6.8 (1.3-35.6)
Napoleon test					
Takeda (2016)[37]	0.63 (0.49-0.75)	0.90 (0.81-0.95)	0.80 (0.65-0.90)	0.79 (0.69-0.86)	14.7 (5.8-37.2)
Barth (2006)[4]	0.25 (0.11-0.47)	0.98 (0.89-1.00)	0.83 (0.44-0.97)	0.76 (0.64-0.85)	15.7 (1.7-144.9)

(continued)

<div align="center">Table 4 (continued)</div>

Clinical Test	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	DOR (95% CI)
Supine Napoleon test					
Takeda (2016)[37]	0.84 (0.72-0.92)	0.96 (0.89-0.99)	0.94 (0.83-0.98)	0.90 (0.82-0.95)	134.4 (33.8-533.5)

[a]Unless specified otherwise, all authors considered lack of strength/weakness a positive test result. Dashes indicate data not reported. DOR, diagnostic odds ratio; IRLS, internal rotation lag sign; IRRT, internal rotation resistance test; NPV, negative predictive value; PPV, positive predictive value.

[b]Full-thickness tears.

[c]Pain was used as a criterion for a positive test result.

[d]IRRT at 0° of abduction and 0° of external rotation is performed with the arm at the side and the elbow flexed to 90°. IRRT at maximal 90° of abduction and maximal external rotation is performed with the shoulder at maximal 90° of abduction and maximal external rotation and the elbow flexed to 90°.

[e]We followed the authors' categorization as lift-off tests; however, passive lift-off tests correspond to IRLS.

[f]Authors graded manual muscle strength from normal amount of resistance to applied force (grade 5) to no muscle contraction (grade 0). This cohort had weakness grade <5.

[g]Authors graded manual muscle strength from normal amount of resistance to applied force (grade 5) to no muscle contraction (grade 0). This cohort had weakness grade <2.

Clinical and methodological heterogeneities were considerable for all tests (Figures 2-5).

Unpooled Data

There were insufficient data on the belly-off sign, IRRT at 0° and 90°, the Napoleon test, and the supine Napoleon test to be included in the meta-analysis. For the belly-off sign, Bartsch et al[5] reported sensitivity and specificity to be >0.80, while Kappe et al[21] noted a sensitivity of 0.31 and a sensitivity of 0.97. For the IRRT at 0° and 90°, Lin et al[25] cited sensitivity of 0.62 and 0.77, respectively, and specificity of 0.76 and 0.81, respectively. For the Napoleon test, Barth et al[4] indicated sensitivity to be 0.25 and specificity to be 0.98, while Takeda et al[37] reported sensitivity to be 0.63 and specificity to be 0.90. For the supine Napoleon test, Takeda et al cited sensitivity and specificity as >0.80.

DISCUSSION

The most important finding of this study was that no single clinical test is sufficiently reliable to diagnose subscapularis tears. It is possible that using several in combination could reduce reliance on costly or lengthy radiologic assessments,[26] but this would need well-evidenced studies to establish. The present systematic search yielded 13 articles reporting the diagnostic accuracy of 8 clinical tests for subscapularis tears, of which 4 tests were eligible for meta-analysis: bear-hug test, belly-press test, IRLS, and lift-off test. All 4 tests had pooled specificity >0.90 but pooled sensitivity <0.60, suggesting that none are individually reliable to diagnose subscapularis tears. These tests are commonly used to diagnose subscapularis tears by inducing active internal rotation of the shoulder at different flexion angles.[27] The lift-off test[11] was the first test designed to evaluate the integrity of the subscapularis, followed by the IRLS[15] and the belly-press test,[10] the latter of which the Napoleon test[35] is a modified version. The belly-off sign and bear-hug test were later described by Scheibel et al[35] and Barth et al,[4] respectively.

The bear-hug test, designed by Barth et al,[4] is the newest of all tests in the meta-analysis. The belly-off sign, Napoleon test, and supine Napoleon test are more recent but lacked sufficient data to be included in the meta-analysis. The bear-hug test appears to be the most promising, based on pooled results from 4 series (598 patients), with best sensitivity (0.55), specificity (0.94), PPV (0.82), and NPV (0.80). The Napoleon test also had promising accuracy. As for the 4 other tests, sensitivity is the diagnostic weakness of the bear-hug test, so the test cannot be used alone to diagnose the presence of subscapularis tears.

Existing studies reporting the diagnostic accuracy of clinical tests for combined IRTT and belly-press test[1] and combined belly-press, bear-hug, and lift-off tests[9] yielded mixed results, with sensitivity of 0.46 and 0.81, respectively. An electromyographic study[31] found that the belly-press, bear-hug, and lift-off tests all activate the integrity of the subscapularis and concluded that these 3 tests can be used interchangeably. A comprehensive meta-analysis on shoulder clinical tests published in 2012 concluded that a combination of clinical tests marginally improves test accuracy.[13] Although medical history and physical examination have limited diagnostic accuracy, they can give useful indications in interpreting clinical tests.[2,14,18,20]

The IRLS constitutes the passive version of the lift-off test (also known as the Gerber test). The 2 tests had equivalent pooled sensitivity (0.32 vs 0.33), specificity (0.92 vs 0.94), and NPV (0.75 vs 0.76), but the IRLS had lower PPV (0.70) than the lift-off test (0.58). This could be explained by a greater familiarity with the lift-off test, which was the most frequently reported. Unlike the lift-off test, the belly-press test and its modified versions (also known as the Napoleon test and the supine Napoleon test) can be performed in the presence of pain or stiffness.[3] Data on the supine Napoleon test from a single study are very promising, with a diagnostic accuracy of 0.84 for sensitivity, 0.96 for specificity, 0.94 for PPV, and 0.90 for NPV, although the risk of bias for this study[37] was high.

Publication bias could not be evaluated statistically; however, studies on clinical tests do not involve medical

devices or treatments, which make them less prone to publication bias. In fact, the wide range of sensitivity (0.0-100), with the rather symmetrical distribution of data, suggests that publication bias was low.

Clinical heterogeneity was low for mean patient age, ranging from 45 to 65 years, but considerable for patient selection, as some series comprised patients who were diagnosed with rotator cuff disease or scheduled to undergo surgery,[4,5,19,23,25,37,39] while others included patients consulting for shoulder pain.[34,36,38] The prevalence of subscapularis tears was higher in series on patients who were diagnosed with rotator cuff disease or scheduled to undergo surgery (mean, 34%; range, 5%-43%) than on patients presenting with shoulder pain (mean, 15%; range, 6%-23%).

Methodological heterogeneity was considerable, given the use of 4 reference diagnostic methods (arthroscopy, MRI, MRA, ultrasound) (Table 1), missing information regarding blinding and/or timing of surgery relative to clinical testing (5 of 10 studies) (Table 4), and subjective thresholds in assessing muscle weakness in clinical tests, which could explain the high variability in sensitivity for all 5 clinical tests. Itoi et al[19] drew attention to the issue of intraobserver repeatability, which the authors assessed in a previous work (correlation coefficient, 0.71). Given that sensitivity was the diagnostic weakness of all pooled tests, combining tests may not improve diagnostic accuracy. Comparing the performance of the painful shoulder with the contralateral shoulder could, however, help circumvent subjectivity in clinical testing.[31]

The quality of any meta-analysis relies on the quality of available studies. Of the 8 studies in the meta-analysis, 5 had a level of evidence of 1 or 2. Furthermore, quality assessment using QUADAS-2 revealed that most studies presented flaws regarding patient selection or diagnostic reference standard or failed to specify blinding and time to surgery, rendering the risk of bias moderate to high in 8 of the 10 studies. Given the small number of primary studies available for pooling, heterogeneity could not be evaluated by hierarchical or bivariate random effects modeling. Other limitations include the high prevalence of rotator cuff disease and comorbidities, as well as the lack of intra- and interobserver repeatability. We therefore recommend that future studies on diagnostic accuracy of clinical tests evaluate repeatability and take into account surgeon experience. Despite these limitations, this study adhered to the standard methodology for systematic reviews and diagnostic meta-analysis outlined in the handbooks of the Cochrane Collaboration[16] and the established guidelines from the PRISMA-DTA.[29]

All tests displayed poor sensitivity, demonstrating that the diagnostic accuracy of clinical tests in evaluating the presence of subscapularis tears is limited, and radiographic assessment remains necessary. Four of the 8 tests—belly-off sign, IRRT, Napoleon test, and supine Napoleon test—could not be pooled for statistical analysis, as too few studies were identified. These tests, which show early promise in the identification of subscapularis tears, would be better understood through future well-designed research.

CONCLUSION

Only 4 tests were eligible for meta-analysis: bear-hug test, belly-press test, IRLS, and lift-off test. All 4 tests had pooled specificity >0.90 but pooled sensitivity <0.60, suggesting that none are individually reliable in diagnosing subscapularis tears. Well-designed studies assessing combinations of tests and less expensive imaging solutions could lead to more reliable clinical diagnosis of subscapularis tears and reduce the reliance on costly or lengthy radiologic assessments.

REFERENCES

1. Aagaard KE, Hanninen J, Abu-Zidan FM, Lunsjo K. Physical therapists as first-line diagnosticians for traumatic acute rotator cuff tears: a prospective study. *Eur J Trauma Emerg Surg*. 2018;44(5):735-745.
2. Bakhsh W, Nicandri G. Anatomy and physical examination of the shoulder. *Sports Med Arthrosc Rev*. 2018;26(3):e10-e22.
3. Barth J, Audebert S, Toussaint B, et al. Diagnosis of subscapularis tendon tears: are available diagnostic tests pertinent for a positive diagnosis? *Orthop Traumatol Surg Res*. 2012;98(8):S178-S185.
4. Barth JR, Burkhart SS, De Beer JF. The bear-hug test: a new and sensitive test for diagnosing a subscapularis tear. *Arthroscopy*. 2006;22(10):1076-1084.
5. Bartsch M, Greiner S, Haas NP, Scheibel M. Diagnostic values of clinical tests for subscapularis lesions. *Knee Surg Sports Traumatol Arthrosc*. 2010;18(12):1712-1717.
6. Beaudreuil J, Nizard R, Thomas T, et al. Contribution of clinical tests to the diagnosis of rotator cuff disease: a systematic literature review. *Joint Bone Spine*. 2009;76(1):15-19.
7. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ*. 2004; 329(7458):168-169.
8. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882-893.
9. Faruqui S, Wijdicks C, Foad A. Sensitivity of physical examination versus arthroscopy in diagnosing subscapularis tendon injury. *Orthopedics*. 2014;37(1):e29-e33.
10. Gerber C, Hersche O, Farron A. Isolated rupture of the subscapularis tendon. *J Bone Joint Surg Am*. 1996;78(7):1015-1023.
11. 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(3):389-394.
12. Gismervik SO, Drogset JO, Granviken F, Ro M, Leivseth G. Physical examination tests of the shoulder: a systematic review and meta-analysis of diagnostic test performance. *BMC Musculoskelet Disord*. 2017;18(1):41.
13. 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(14):964-978.
14. Hermans J, Luime JJ, Meuffels DE, Reijman M, Simel DL, Bierma-Zeinstra SM. Does this patient with shoulder pain have rotator cuff disease? The Rational Clinical Examination systematic review. *JAMA*. 2013;310(8):837-847.
15. Hertel R, Ballmer FT, Lombert SM, Gerber C. Lag signs in the diagnosis of rotator cuff rupture. *J Shoulder Elbow Surg*. 1996;5(4):307-313.
16. Higgins JPT, Green S; Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Updated March 2011. http://handbook.cochrane.org
17. Hughes PC, Taylor NF, Green RA. Most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review. *Aust J Physiother*. 2008;54(3):159-170.
18. Itoi E. Rotator cuff tear: physical examination and conservative treatment. *J Orthop Sci*. 2013;18(2):197-204.

19. Itoi E, Minagawa H, Yamamoto N, Seki N, Abe H. Are pain location and physical examinations useful in locating a tear site of the rotator cuff? *Am J Sports Med*. 2006;34(2):256-264.

20. Jain NB, Wilcox RB 3rd, Katz JN, Higgins LD. Clinical examination of the rotator cuff. *PM R*. 2013;5(1):45-56.

21. Kappe T, Sgroi M, Reichel H, Daexle M. Diagnostic performance of clinical tests for subscapularis tendon tears. *Knee Surg Sports Traumatol Arthrosc*. 2018;26(1):176-181.

22. Kim HA, Kim SH, Seo YI. Ultrasonographic findings of painful shoulders and correlation between physical examination and ultrasonographic rotator cuff tear. *Mod Rheumatol*. 2007;17(3):213-219.

23. Lasbleiz S, Quintero N, Ea K, et al. Diagnostic value of clinical tests for degenerative rotator cuff disease in medical practice. *Ann Phys Rehabil Med*. 2014;57(4):228-243.

24. Leeflang MM, Deeks JJ, Rutjes AW, Reitsma JB, Bossuyt PM. Bivariate meta-analysis of predictive values of diagnostic tests can be an alternative to bivariate meta-analysis of sensitivity and specificity. *J Clin Epidemiol*. 2012;65(10):1088-1097.

25. Lin L, Yan H, Xiao J, Ao Y, Cui G. Internal rotation resistance test at abduction and external rotation: a new clinical test for diagnosing subscapularis lesions. *Knee Surg Sports Traumatol Arthrosc*. 2015; 23(4):1247-1252.

26. Liu F, Dong J, Shen WJ, Kang Q, Zhou D, Xiong F. Detecting rotator cuff tears: a network meta-analysis of 144 diagnostic studies. *Orthop J Sports Med*. 2020;8(2):2325967119900356.

27. Longo UG, Berton A, Ahrens PM, Maffulli N, Denaro V. Clinical tests for the diagnosis of rotator cuff disease. *Sports Med Arthrosc Rev*. 2011;19(3):266-278.

28. McFarland EG, Selhi HS, Keyurapan E. Clinical evaluation of impingement: what to do and what works. *J Bone Joint Surg Am*. 2006;88(2): 432-441.

29. McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA statement. *JAMA*. 2018;319(4):388-396.

30. Miller CA, Forrester GA, Lewis JS. The validity of the lag signs in diagnosing full-thickness tears of the rotator cuff: a preliminary investigation. *Arch Phys Med Rehabil*. 2008;89(6):1162-1168.

31. Pennock AT, Pennington WW, Torry MR, et al. The influence of arm and shoulder position on the bear-hug, belly-press, and lift-off tests: an electromyographic study. *Am J Sports Med*. 2011;39(11): 2338-2346.

32. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.

33. Roquelaure Y, Ha C, Rouillon C, et al. Risk factors for upper-extremity musculoskeletal disorders in the working population. *Arthritis Rheum*. 2009;61(10):1425-1434.

34. Salaffi F, Ciapetti A, Carotti M, Gasparini S, Filippucci E, Grassi W. Clinical value of single versus composite provocative clinical tests in the assessment of painful shoulder. *J Clin Rheumatol*. 2010;16(3): 105-108.

35. Scheibel M, Magosch P, Pritsch M, Lichtenberg S, Habermeyer P. The belly-off sign: a new clinical diagnostic sign for subscapularis lesions. *Arthroscopy*. 2005;21(10):1229-1235.

36. Somerville LE, Willits K, Johnson AM, et al. Clinical assessment of physical examination maneuvers for rotator cuff lesions. *Am J Sports Med*. 2014;42(8):1911-1919.

37. Takeda Y, Fujii K, Miyatake K, Kawasaki Y, Nakayama T, Sugiura K. Diagnostic value of the supine Napoleon test for subscapularis tendon lesions. *Arthroscopy*. 2016;32(12):2459-2465.

38. van Kampen DA, van den Berg T, van der Woude HJ, et al. The diagnostic value of the combination of patient characteristics, history, and clinical shoulder tests for the diagnosis of rotator cuff tear. *J Orthop Surg Res*. 2014;9:70.

39. Yoon JP, Chung SW, Kim SH, Oh JH. Diagnostic value of four clinical tests for the evaluation of subscapularis integrity. *J Shoulder Elbow Surg*. 2013;22(9):1186-1192.