Concurrent Criterion-Related Validity of Physical Examination Tests for Hip Labral Lesions: A Systematic Review

M. Rebecca Leibold, MPT, MTC[1], Peter A. Huijbregts, PT, DPT, OCS, FAAOMPT, FCAMT[2], Richard Jensen, PT, PhD[3]

Hip injuries comprise 5–9% of all injuries sustained by high school athletes[1]. Hip and groin pain are common reasons for people to seek physical therapy treatment. However, differential diagnosis for these symptoms is complex (Table 1). Injuries to the labrum of the hip constitute one of the possible reasons for hip and groin pain. Narvani et al[2] reported that 22% of athletes with groin pain were diagnosed with a labral tear. McCarthy et al[3] found that 55% of their patients with mechanical hip pain had a labral tear. Without premortem information available allowing for correlation with symptom status, these same authors also harvested 54 cadaveric acetabula and found that 52% of these had labral lesions[3]. Santori and Villar[4] reported on 412 arthroscopic surgeries for disabling hip pain of >6 months duration: 76 patients (18%) had acetabular labral tears. With the advent of arthroscopic surgery as an accurate means of diagnosis, hip labral injuries have become of growing interest to the medical profession.

ABSTRACT: Hip injuries are prevalent, especially within the athletic population. Of the hip injuries in this population, some 18–55% are lesions to the labrum of the hip. Clinical diagnosis of hip labral lesions is difficult because data on prevalence are varied. In addition, data on the prevalence of internal and external risk factors are absent as are data on the correlation of these risk factors with labral lesions, making it difficult to gauge the diagnostic utility. The mechanism of injury is often unknown or not specific to labral lesions. Internal risk factors may remain hidden to physical therapists because in most jurisdictions, ordering imaging tests is not within their scope of practice. Anterior inguinal pain seems highly sensitive for the diagnosis of patients with labral lesions but can hardly be considered specific; data on other pain-related and mechanical symptoms clearly have little diagnostic utility, making these data collected during the patient history almost irrelevant to diagnosis. By way of a comprehensive literature review and narrative and systematic analysis of the methodological quality of the retrieved diagnostic utility studies, this paper aimed to determine a diagnostic physical examination test or test cluster based on current best evidence for the diagnosis of hip labral lesions. Current best evidence indicates that a negative finding for the flexion-adduction-internal rotation test, the flexion-internal rotation test, the impingement provocation test, the flexion-adduction-axial compression test, the Fitzgerald test, or a combination of these tests provides the clinician with the greatest evidence-based confidence that a hip labral lesion is absent. Currently, research has produced no tests with sufficient specificity to help confidently rule in a diagnosis of hip labral lesion. Suggestions for future research are provided.

KEYWORDS: Concurrent Criterion-Related Validity, Hip Labral Lesion, Physical Examination, STARD, Systematic Review.

However, clinical diagnosis of patients with hip labral lesions is difficult. For one, demographics for patients with labral injuries are highly variable: Ages reported in the literature ranged from 8 to 72, although most patients were in the fourth decade of life[2,4-22]. Labral lesions may be more common in women. When we combined all studies reviewed in this paper, 60% of patients were women. A higher activity level as found in runners, professional athletes, and those attending the gym 3 times a week has been suggested as a risk factor[2,11]. In addition, the majority of patients with labral pathology do not recall the mechanism of injury that led to their symptoms. Santori and Villar[4] collected data on etiology from 58 of their 76 patients with acetabular labral tears: 29.3% were of unknown etiology, traumatic injury occurred in 25.9%, and in 44.8% the labral lesions were likely degenerative in nature. When patients do recall the mechanism of injury, this may include hyperabduction, twisting, falling, or running, or it may be related to a motor vehicle accident, sports, work, or a direct blow. Other external risk factors noted in the literature include repetitive micro-trauma; sports activities that require frequent hip external rotation such as soccer, golf, hockey, karate, and ballet; running;

[1]Outpatient Physical Therapist, Colorado Athletic Conditioning Clinic, Aurora, CO. [2]Assistant Professor, Online Education Department, University of St. Augustine for Health Sciences, St. Augustine, FL. [3]Dean, Division of Advanced Studies, University of St. Augustine for Health Sciences, St. Augustine, FL.

Address all correspondence and requests for reprints to: Dr. Peter Huijbregts, shelbournephysio@telus

TABLE 1. Differential diagnosis: Causes of hip and groin pain.

Childhood disorders	Infectious conditions
• Congenital dysplasia • Legg-Calve-Perthes disease • Slipped capital femoral epiphysis (SCFE)	• Septic arthritis • Osteomyelitis • Psoas abscess • Hip pyarthrosis • Urinary tract infection
Traumatic conditions	**Inflammatory conditions**
• Subluxation/Dislocation • Fractures of the femoral head • Stress fractures • Loose bodies • Acetabular labral tears • Contusions • Femoral or inguinal hernia • Athletic pubalgia	• Rheumatoid arthritis • Juvenile arthritis • Ankylosing spondylitis • Bursitis • Tendonitis • Pelvic inflammatory disease • Prostatitis • Crohn's disease • Psoriasis • Reiter's syndrome • Systemic Lupus Erythematosus
Degenerative joint disease	**Neurologic conditions**
• Osteoarthritis • Osteolysis	• Radiculopathy • Local nerve entrapment (ilioinguinal, genitofemoral, or lateral femoral cutaneous)
Vascular conditions	**Metabolic conditions**
• Osteonecrosis/avascular necrosis	• Gout • Metabolic bone disease
Neoplasms	**Other causes**
	• Referred pain • Corticosteriod use • Alcoholism • Psychosocial • Gynecologic

hyperextension with or without external rotation; and dislocation[23,24].

There are also internal risk factors; anatomical variations associated with labral lesions are mentioned in the literature (Table 2). Wenger et al[25] noted structural abnormalities in 31 patients with labral tears including acetabular retroversion, coxa valga, abnormal Tonnis[26] angle, small femoral head-neck offset, and incongruent hips. Peelle et al[27] compared radiographs of 78 patients with labral tears confirmed on arthroscopy to those of 22 subjects without hip dysfunction. Of the patients with labral tears, 49% had an osseous abnormality including a lateral center-edge angle <25°, head-neck offset <9 mm, offset ratio <0.17, acetabular retroversion, femoral anteversion, an aspherical femoral head, and a Tonnis osteoarthritis grade (Table 3[26]) of 1 and 2. Patients with labral tears demonstrated significantly smaller lateral center-edge angles ($P=0.008$), larger Tonnis angles ($P=0.02$), and a greater probability of acetabular dysplasia ($P=0.001$) than controls[27]. Ito et al[28] compared 24 patients to 24 control subjects and found that patients had significantly less femoral anteversion ($P<0.001$); they also noted a significant between-group difference for head-neck offset ($P<0.002$). Siebenrock et al[29] also found patients to have a significantly different head-neck offset when compared to a control group ($P=0.01–0.04$). Kassarjian et al[14] studied 42 hips with an antero-superior labral tear: 93% had an abnormal head-neck offset with a mean angle of 69.7°; abnormal was defined as >55°. Acetabular retroversion, femoral anteversion, and abnormal head-neck offset all increase the chance of labral impingement against the acetabular rim, especially with active hip flexion with or without internal rotation. Two different joint morphologies have been proposed as a cause for femoro-acetabular impingement that may lead to labral failure[30]. A larger femoral head may lead to "cam" impingement whereby the head prematurely impacts the antero-superior aspect of the acetabular rim during active hip motions causing acetabular cartilage and labral damage. "Pincer" impingement occurs when a normal femoral head is paired with an abnormal acetabulum (e.g., coxa profunda or acetabular retroversion). This type initially affects only the labrum. Further internal risk factors mentioned in the literature include pelvic instability and degeneration, Legg-Calve Perthes disease, slipped capital femoral epiphysis, and a shallow tapering between the femoral head and neck; one study also reported osteonecrosis as a risk factor[7,31,32].

The most common symptom in patients with labral pathology is anterior inguinal pain, whereas anterior thigh pain, lateral thigh pain, and buttock pain are less prevalent. Burnett et al[7] found that 92% of their patients with labral tears complained of anterior groin pain. Keeney et al[15] reported 97 of 102 patients with groin pain and Fitzgerald[10] also noted anterior groin pain in 48 patients with confirmed labral tears. Pain level has been recorded as moderate to severe and pain has limited the patient's activities. Walking, climbing stairs, running, and twisting motions at the hip have been reported as aggravating factors. Two studies also noted that patients had pain at night. Ito et al[28] found night pain in 14 of 25 patients and Burnett et al[7] reported 71% of their patients had night pain.

Labral lesions may also cause mechanical symptoms. Burnett et al[7] reported 89% of patients with labral tears

TABLE 2. Definitions of anatomical variations associated with labral lesions.

Anatomical Variations	Description	Values
Retroverted acetabulum	An acetabulum that is orientated in a more posterior position in reference to the sagittal plane	>15 degrees = abnormal
Tonnis angle (acetabular index of weight-bearing surface)*	A line parallel to the sourcil to a horizontal line through the center of the femoral head	>10 degrees = abnormal
Head-neck offset	Comparison of the radius of the femoral head and the radius of the femoral neck	<7.2mm = abnormal 11.5mm = normal
Coxa profunda and Protrusion acetabuli	An increase in the depth of the acetabulum	15–27mm = normal
Head-neck shaft angle	Angle between one line bisecting the longitudinal axis of the femur and one line bisecting the longitudinal axis of the femoral neck	126–139 degrees = normal Coxa vara<126 Coxa valga>139
Aspherical head	A flattening of the femoral head and increased radius of the anterior portion of the head of the femur	Anterolateral prominence at head-neck junction
Lateral center-edge angle (Wiberg)*	Line through the center of the femoral head to lateral edge of the acetabulum, and vertical line through the center of the femoral head	< 20-25 degrees = abnormal
Anterior center edge angle*	Vertical line extending superiorly from the center of femoral head to a line formed by the tangent to the anterior-most portion of the acetabulum	<20 degrees = abnormal
Femoral anteversion	Angle between the transverse axis of the knee joint and the transverse axis of both femoral condyles	15–20 degrees = normal
Offset ratio	Ratio of head-neck offset distance in relation to the diameter of the femoral head	<0.27 = abnormal

*Used to determine acetabular dysplasia.

TABLE 3. Tonnis osteoarthritis grades.

Grade	Symptoms
Grade 0	No signs of osteoarthritis
Grade 1	Slight narrowing of the joint space. Slight sclerosis of femoral head or acetabulum and slight lipping at the joint margin.
Grade 2	Small cysts, less than 50% joint-space narrowing, moderate loss of femoral head sphericity.
Grade 3	Large cysts, severe narrowing or no joint space, severe deformity of the femoral head and avascular necrosis.

also mentioned a history of a limp. Fitzgerald[10] reported 5 of 55 patients had a limp and Keeney et al[15] noted that 39 of 102 subjects mentioned a limp. Some but not all patients with labral pathology have reported clicking, catching, or locking of the hip with motion. McCarthy et al[20] found that 67% of subjects complained of clicking or locking with hip motion. Of their subjects, 72% had labral tears. However, the authors did not state whether those with clicking indeed had a labral tear. Narvani et al[2] reported that 4 of 4 patients with labral tears noted clicking, but that 2 patients without labral tears also mentioned clicking. Leunig et al[19] reported that 6 of 23 patients with labral tears had locking symptoms. Keeney et al[15] found locking or catching in >50% of their patients, and Farjo[9] reported that 18 of 28 patients who were found to have labral tears upon arthroscopy had mechanical symptoms. Fitzgerald[10] reported that 34 of 64 patients had a click associated with hip pain and were also positive for labral tears.

In summary, data on prevalence of hip labral lesions provided in the literature are highly variable and likely depend strongly on the population studied. Exact data on the prevalence of the internal and external risk factors noted are absent as are hard data for correlation with labral lesions, thereby making it difficult to gauge the diagnostic utility. The mechanism of injury is often unknown or not specific to labral lesions. Internal risk factors may remain hidden to physical therapists, because in most jurisdictions ordering imaging tests is not within the scope of practice. Anterior inguinal pain seems highly sensitive for the diagnosis of patients with labral lesions but can hardly be considered specific. Data on other pain-related and mechanical symptoms clearly have little diagnostic utility, making these data collected during the patient history almost irrelevant to diagnosis. However, inaccurate diagnosis may result in prolonged rehabilitation and associated cost. Clinical tests capable of ruling out labral lesions with sufficient diagnostic confidence would prevent unnecessary arthroscopic surgery that is currently required for accurate diagnosis. Physical therapists, especially in a direct access role, are uniquely positioned in the health care system to clinically rule in or rule out a diagnosis of hip labral lesions and to facilitate appropriate management but due to scope of practice legislation, they are limited to history and physical examination in the diagnostic process. Therefore, the goal of this paper is to determine a diagnostic physical examination test or test cluster based on current best evidence for the diagnosis of hip labral lesions.

Methods and Materials

To establish current best evidence with regard to the physical diagnosis of hip labral lesions, we performed a literature search. Databases used for this search included Medline, CINAHL, the Cochrane Libraries, and LIRN. Search terms included *hip labrum, acetabular labrum, hip labral lesions, hip labral tears, hip limbus, acetabular labral function, hip labrum function, hip labral function, acetabular labrum function, acetabular labral tears, acetabular labral lesions, hip labrum tears, acetabulum labrum tears, acetabulum limbus*, and *acetabular limbus*. All searches were matched against the first search using *hip labrum* as the search phrase.

We restricted our search to the period from January 1990 to March 2007, mainly because a preliminary search did not result in any papers on this topic published prior to 1990. Our literature search was further restricted to articles written in English. Our intent was to find quantitative research on diagnostic utility of physical examination tests for labral lesions of the hip, where these tests were compared to a gold standard or reference test of imaging or arthroscopy. The literature search results are summarized in Table 4.

To provide a quantitative measure of diagnostic accuracy, we have provided the statistical measures of accuracy, sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios. Where the authors did not include these statistical measures, we have calculated these statistics; definitions and calculation of the relevant statistics are provided in Table 5[33].

Flawed studies and the resulting biased diagnostic utility statistics can lead a clinician to misdiagnosis and inappropriate management decisions. This review is a systematic review in that it uses the Standards for Reporting of Diagnostic Accuracy (STARD) tool for methodological quality assessment of the studies retrieved and it takes these STARD scores into account when providing a current best-evidence summary. Designed originally as a prospective tool to improve the methodological quality of diagnostic utility studies, the STARD tool[34] contains 25 items that can also be used retrospectively as a checklist to evaluate methodological quality. However, with no established cut-off values and no research into the reliability of this tool, we acknowledge that the methodological quality assessment it provides is qualitative at best. Therefore, in addition to STARD scores, we provide a narrative discussion of the biases in the diagnostic utility studies retrieved to allow for a current best-evidence synthesis taking into account also potential methodological flaws not addressed or insufficiently addressed in the STARD tool.

Results

Table 4 contains all data on the literature search results for the various search terms used in the five databases selected. The search strategy retrieved 16 articles that met our inclusion criteria. These 16 articles provided research on the diagnostic utility of 9 different physical examination tests for hip labral lesions. Statistical measures related to diagnostic utility for these tests have been summarized in Table 6. Table 7 contains the STARD scores for the 16 studies retrieved.

Flexion-Adduction-Internal Rotation Test (Impingement Test)

Seven studies have reported on the impingement test or sign[5-7,13-15,18]. The premise for this test is that with flexion and adduction of the hip, the femoral head comes in close approximation with the acetabular rim. Internally rotating the hip then places a shearing force on the labrum (Figure 1). Pain in the groin area is considered indicative of labral pathology, including degeneration, fraying, or tearing.

Burnett et al[7] described the impingement test with the patient supine and the examiner passively flexing the hip to 90°, adducting and then internally rotating the hip. Pain in the groin region was considered a positive test. One medical physician examined and performed surgery on all subjects. This study was retrospective: All 66 subjects had arthroscopically confirmed labral tears. The study was performed in an orthopaedic surgery setting, and the gold standard test used in this study was arthroscopic surgery. Of 66 patients, 63 (95.5%) were positive on the impingement test.

Keeney et al[15] did not specifically describe but only referenced the impingement test. The study did not provide operational definitions of a positive or negative test finding. Information on raters was absent. Subjects included in this retrospective study were 101 consecutive patients (102 hips) with persistent inguinal pain, positive impingement test, minimal degeneration on radiographs, and a negative examination for

TABLE 4. Literature search description and results.

	Hip Labrum	Ace Labrum	Hip Labral Lesion	Hip Labral Tears	Hip Limbus	Ace Labral Function	Hip Labrum Function	Hip Labral Function	Ace Labrum Function	Ace Labral Tears	Ace Labral Lesions	Hip Labrum Tears	Acu Labrum Tears	Acu Limbus	Ace Limbus
• Medline	150	143	59	105	16	29	41	40	41	79	44	49	48	9	8
• Cinahl	0	12	0	0	0	0	0	0	0	15	0	0	0	0	0
• Cochrane	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
• LIRN-EBSCO Host	0	6	0	0	0	0	0	0	0	3	0	0	0	0	0
• LIRN Proquest	0	3	1	3	0	0	1	2	1	3	1	2	0	0	0
Total Selected	10	0	5	1	0	0	0	0	0	0	0	0	0	0	0

Ace=acetabular. Although a number were selected originally, many were duplicates of earlier findings and were not retained. CINAHL-Cumulative Index to Nursing and Allied Health Literature; LIRN-Library and Information Resources Network; Acu-Acetabulum.

TABLE 5. Definition and calculation of statistical measures used to express diagnostic test utility.

Statistical measure	Definition	Calculation
Accuracy	The proportion of people who were correctly identified as either having or not having the disease or dysfunction	(TP + TN) / (TP + FP + FN + TN)
Sensitivity	The proportion of people who have the disease or dysfunction who test positive	TP / (TP + FN)
Specificity	The proportion of people who do not have the disease or dysfunction who test negative	TN / (FP + TN)
Positive predictive value	The proportion of people who test positive and who have the disease or dysfunction	TP / (TP + FP)
Negative predictive value	The proportion of people who test negative and who do not have the disease or dysfunction	TN / (FN + TN)
Positive likelihood ratio	How likely a positive test result is in people who have the disease or dysfunction as compared to how likely it is in those who do not have the disease or dysfunction	Sensitivity/(1-specificity)
Negative likelihood ratio	How likely a negative test result is in people who have the disease or dysfunction as compared to how likely it is in those who do not have the disease or dysfunction	(1-sensitivity)/specificity

TP = true positive; TN = true negative; FP = false positive; FN = false negative

TABLE 6. Diagnostic utility data hip labral lesion tests studied.

	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Positive Likelihood Ratio	Negative Likelihood Ratio
Flexion-Adduction Internal Rotation Test							
Burnett et al[7]	0.95 (63/66)	0.95 (63/66)	NC	1.00 (63/63)	0 (0/3)	NC	NC
Keeney et al[15]	0.91 (93/102)	1.00 (93/93)	0 (0/9)	0.91 (93/102)	NC	1.00	NC
Beck et al[6]	1.00 (19/19)	1.00 (19/19)	NC	1.00 (19/19)	NC	NC	NC
Ito et al[13]	0.96 (24/25)	0.96 (24/25)	NC	1.00 (24/24)	0 (0/1)	NC	NC
Kassarjian et al[14]	1.00 (42/42)	1.00 (42/42)	NC	1.00 (42/42)	NC	NC	NC
Beaule et al[5]	0.97 (35/36)	1.00 (35/35)	0 (0/1)	0.97 (35/36)	NC	1.00	NC
Leunig et al[18]	0.64 (18/28)	1.00 (18/18)	0 (0/10)	0.64 (18/28)	NC	1.00	NC
Impingement Provocation Test							
Leunig et al[19]	0.82 (18/22)	1.00 (18/18)	0 (0/4)	0.82 (18/22)	NC	1.00	NC
Flexion-Internal Rotation Test							
Santori & Villar[4]	1.00 (76/76)	1.00 (76/76)	NC	1.00 (76/76)	NC	NC	NC
Chan et al[8] 1. Gold standard test MRA	0.83 (25/30)	1.00 (25/25)	0 (0/5)	0.83 (25/30)	NC	1.00	NC

(continued)

TABLE 6. (Continued)

	Accuracy	Sensitivity	Specificity	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Positive Likelihood Ratio	Negative Likelihood Ratio
2. Gold standard test arthroscopy	0.94 (16/17)	1.00 (16/16)	0 (0/1)	0.94 (16/17)	NC	1.00	NC
Petersilge et al[21]	0.9 (9/10)	1.00 (9/9)	0 (0/1)	0.9 (9/10)	NC	1.00	NC
Hase & Ueo[12]	0.70 (7/10)	1.00 (7/7)	0 (0/3)	0.70 (7/10)	NC	1.00	NC
Flexion, Adduction, Axial Compression							
Hase & Ueo[12]	1.00 (10/10)	1.00 (10/10)	NC	1.00 (10/10)	NC	NC	NC
Palpation-Posterior to Greater Trochanter							
Hase & Ueo[12]	0.8 (8/10)	0.8 (8/10)	NC	1.00 (8/8)	0 (0/2)	NC	NC
Flexion, Internal Rotation, Axial Compression							
Narvani et al[2]	0.5 (9/18)	0.75 (3/4)	0.43 (6/14)	0.27 (3/11)	0.86 (6/7)	1.32	0.58
Thomas Test							
Narvani et al[2]	ID	0.25	ID	ID	ID	ID	ID
MFIR							
Suenaga et al[22]	0.38 (23/60)	0.38 (23/60)	NC	1.00 (23/23)	0 (0/37)	NC	NC
Guanche &Sikka[11]	1.00 (8/8)	1.00 (8/8)	NC	1.00 (8/8)	NC	NC	NC
MFER							
Suenaga et al[22]	0.27 (16/60)	0.38 (23/60)	NC	1.00 (16/16)	0 (0/44)	NC	NC
Fitzgerald Test							
Fitzgerald[10]	0.96 (54/56)	1.00 (54/54)	NC	1.00 (54/54)	0 (0/2)	NC	NC

NC=not calculated; MFIR=maximal flexion-internal rotation; MFER=maximal flexion-external rotation; ID=insufficient data.

tendon pathology. The study was performed in a surgical orthopaedic setting with a gold standard test of arthroscopy. Of 102 hips positive with impingement testing, 93 had labral tears on arthroscopy.

Beck et al[6] studied the impingement sign but again only referenced the test description. They did not provide an operational definition of a positive or negative test and also did not report who performed the clinical testing. Subjects involved in this retrospective study were those with a diagnosis of femoro-acetabular impingement confirmed by MRA (described as an abnormality of the acetabulum or femur and their biomechanical relationship), who underwent surgical dislocation of the hip. The study was performed in an orthopaedic surgical setting. The gold standard test was surgical findings. Of 19 subjects, all had positive impingement testing and all 19 also had intra-operatively confirmed labral lesions.

Ito et al[13] described the impingement test as hip internal rotation followed by passive flexion to 90° and adduction. A positive test was indicated by sharp groin pain, rated by a medical physician. Subjects included in this retrospective study

TABLE 7. STARD scores retrieved studies.

ITEM #	Beaule et al[5]	Leunig et al1[9]	Beck et al[6]	Burnett et al[7]	Chan et al[8]	Santori & Villar[4]	Petersilge et al[21]	Ito et al1[3]	Suenaga et al[22]	Narvani et al[2]	Leunig et al[18]	Hase & Ueo[12]	Fitzgerald[10]	Kassarjian et al[14]	Guanche & Sikka[11]	Keeney et al[15]
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1
3	1	0	1	0	0	0	0	0	0	0	1	0	1	1	1	0
4	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	0	1	1	0	0	0	0	0	0	1	0	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
8	1	0	0	1	0	0	1	1	0	1	1	0	0	1	0	1
9	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1
10	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1
11	0	0	0	1	1	0	1	0	0	0	1	0	0	1	0	1
12	1	1	0	1	1	0	0	1	1	1	1	0	0	0	0	0
13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
14	0	0	1	1	0	0	1	1	1	0	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
17	0	1	0	0	1	0	0	0	0	0	0	1	0	1	1	0
18	1	1	0	0	1	0	0	1	1	1	0	0	0	0	0	1
19	0	1	0	0	1	1	0	1	1	1	0	0	0	0	0	1
20	0	1	1	1	0	1	0	0	0	1	NA	1	1	NA	0	0
21	1	1	0	1	1	0	0	0	1	1	1	0	0	0	0	0
22	NA	NA	0	1	0	0	0	NA	NA	0	NA	NA	1	NA	NA	1
23	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA
24	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
25	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1
TOTAL	14/21	10/21	10/22	13/22	12/22	7/22	10/22	11/21	11/21	12/22	16/21	8/21	10/22	15/20	9/21	16/22

NA=not appropriate

were surgical patients with femoro-acetabular impingement, limited range of motion, and, in most cases, a positive impingement test. This study was performed in an orthopaedic surgical setting and the gold standard test used was surgical findings. Of 25 subjects, 24 had positive impingement testing; 25 had labral damage confirmed intra-operatively, 17 of which were specifically noted to be labral tears.

Kassarjian et al[14] studied a flexion, adduction, and internal rotation test. Operational definition of a positive test finding was pain unspecified as to location and limited range of motion. Information on raters was absent. Subjects (42 hips) in this retrospective study were patients with MRA-confirmed femoroacetabular impingement**.** The study was performed in a hospital setting and the gold standard test used was MRA. The impingement test was positive in all hips studied. All hips had antero-superior labral tears on MRA. Although not used in our calculation of diagnostic utility statistics, of interest is that 11 of the subjects had surgery and that all 11 had confirmed labral tears.

Beaule et al[5] reported on the impingement sign referencing but not providing a description of this test. Although pain was considered a positive response, a clear operational definition of a positive test or the location of pain was not provided. Information on raters was absent. Subjects included in this prospective study were 30 consecutive patients. In these 30 patients, 36 painful non-dysplastic hips were evaluated. The study setting was not mentioned. The gold standard tests were MRA and, for some patients, surgical findings. Of 36 hips with a positive impingement test, 35 had labral tears confirmed with MRA; 21 patients had surgery, which confirmed all labral tears[5]. We calculated diagnostic utility statistics using MRA as the reference test.

Leunig et al[18] reported on the impingement test. The test was described as first internally rotating the hip with the patient supine, next passively flexing the hip to 90°, and then adding adduction. A sharp groin pain indicated a positive test. Information on raters was absent. Subjects included in this prospective study were patients with developmental dysplasia (n=14) of the hip and femoroacetabular impingement (n=14). The study was performed in a clinical orthopaedic setting. Reference testing used was MRA. All subjects in both groups had positive impingement testing, and 9 patients in each group presented with labral tears confirmed by MRA.

FIGURE 1. (LEFT) Flexion-adduction-internal rotation test.

FIGURE 2. (RIGHT) Impingement provocation test. for the postero-inferior labrum.

Impingement Provocation Test

Whereas the impingement test would seem tailored to provocation of the antero-superior labrum, Leunig et al[19] described and studied a test with proposed specific effects on the anterior or posterior parts of the hip labrum. These authors reported on the impingement provocation test described as flexion, adduction, and internal rotation for the antero-superior acetabular rim and hyperextension, abduction, and external rotation for the postero-inferior rim (Figure 2). Discomfort and apprehension were mentioned as a positive test, but the authors did not specifically define a positive test or location of discomfort. Information on raters was absent. Subjects included 22 patients with acetabular rim syndrome, characterized by acetabular impingement and groin pain; all subjects had intermittent groin pain, full range of motion, and positive impingement testing. The setting for this prospective study was not mentioned. Gold standard testing was by way of surgical findings. Of 22 subjects, 18 had labral tears or degeneration or both as confirmed intra-operatively.

Flexion-Internal Rotation Test

Some authors have studied a modified form of the impingement test performed with the patient supine, hips flexed to 90°, and then internally rotated. Unlike the impingement test, this modified test has no adduction component (Figure 3). As with the impingement test discussed above, pain in the groin with this test has generally been considered indicative of labral degeneration, fraying, or tearing.

Santori and Villar[4] studied a test consisting of 90° flexion and internal rotation. Pain and normal range of motion were considered a positive test indicative of a labral tear. Information on raters was absent. Subjects in this retrospective study were patients with disabling hip symptoms for greater than six months and surgically confirmed labral lesions; hence, gold standard testing used was arthroscopy. All 76 patients in their study with surgically detected labral tears had a positive clinical test.

Chan et al[8] prospectively studied the same test. A positive test was defined as pain reproduction with normal range of motion. The authors did not describe location of pain, and information on raters was absent. Subjects were patients suspected of labral tears and the study was performed in a hospital setting. The clinical test was positive in all 30 subjects; MRA used as gold standard testing revealed 25 labral tears. Of the 25 patients with MRA evidence of labral tears, 17 had arthroscopic surgery and in 16 of those, arthroscopy as a gold standard test revealed labral tears. Using two gold standard tests yielded two different sets of diagnostic utility statistics, both provided in Table 6.

Petersilge et al[21] also studied the same test. Pain reproduction was considered a positive test, but again location of the pain was not mentioned. Information on raters was absent. Subjects included in this retrospective study were those with hip pain and with MRA and surgical confirmation of labral pathology. Gold standard testing was surgical findings, consisting of arthroscopy in 7 patients and arthrotomy in 3 patients. The authors reported positive findings in 9 subjects that included 7 labral tears, 1 avulsion, and l degenerated labrum; l labrum was without pathological findings.

Hase and Ueo[12] studied three different clinical tests. One was a modified form of the impingement test. Pain was considered a positive test and information on raters was absent. Subjects (n=10) in this retrospective study were patients arthroscopically diagnosed with labral tear. This study was performed in a surgical setting. Gold standard testing was arthroscopy. This clinical test was positive in 7 patients.

Flexion-Adduction-Axial Compression Test

In the above retrospective study, Hase and Ueo[12] also examined a test consisting of axial compression of the hip joint in 90° of flexion and slight adduction; pain was considered a positive test (Figure 4). All patients in this study had pain

FIGURE 3. Flexion-internal rotation test.

FIGURE 4. Flexion-adduction-axial compression test.

with this maneuver, and as noted above all were confirmed via arthroscopy to have labral tears.

Palpation Posterior to Greater Trochanter

In the same retrospective study, Hase and Ueo[12] also studied tenderness to palpation just posterior to the greater trochanter (Figure 5). Tenderness was positive in 8 patients of the 10 with confirmed labral lesions who served as subjects for this study.

Flexion-Internal Rotation-Axial Compression Test

Narvani et al[2] prospectively studied a test consisting of internal rotation, flexion, and axial compression; reproduction of pain or discomfort indicated a positive test (Figure 6). An orthopaedic surgeon performed the clinical tests. Subjects (n=18) were all active patients who went to the gym 3 times a week or who were professional athletes and had presented to a sports clinic with groin pain. Arthroscopy was used as the gold standard in this study. The test resulted in positive findings in 3 of 4 patients with confirmed labral tears via arthroscopy, but this clinical test was also positive in 8 of 14 patients who were without labral tears.

Thomas Test

In the prospective study discussed above, Narvani et al[2] also studied the Thomas test. The authors described this test as extending the hip from a flexed position. Operational definitions for a positive or negative test were not provided. The authors reported no correlation between the Thomas test and the presence of a labral tear, with sensitivity only of this test reported by the authors as 25%.

Maximum Flexion-Internal Rotation Test

Suenaga et al[22] performed two clinical tests in their retrospective study: one involved maximum flexion and internal rotation (MFIR). Data on test performance and interpretation were not provided, and data on raters were also absent. Subjects (n=60) were patients with dysplastic osteoarthritis who underwent an acetabular transposition osteotomy. This study setting was a hospital, and the gold standard testing was arthroscopy. The test was positive in 23 patients, even though all 60 patients demonstrated complete or incomplete labral tears upon arthroscopic examination.

Guanche and Sikka[11] studied a test described as forced flexion of the hip with internal rotation (Figure 7). A positive test was indicated by pain exacerbation and reproduction of the patient's pain. Data on raters were absent. Subjects included in this retrospective study were 8 high-level runners, described as either Olympic-level or having run 5 marathons, and the gold standard testing was arthroscopy. All subjects reported pain with the test, and arthroscopy confirmed labral tears in all 8 subjects, located antero- and postero-superiorly.

Maximum Flexion-External Rotation Test

In the retrospective study discussed above, Suenaga et al[22] also studied a test consisting of maximum flexion and external rotation (Figure 8). Data on test performance and interpretation were again not provided; the test was positive in 16 patients.

Fitzgerald Test

Fitzgerald[10] reported two different tests to determine if patients had an anterior

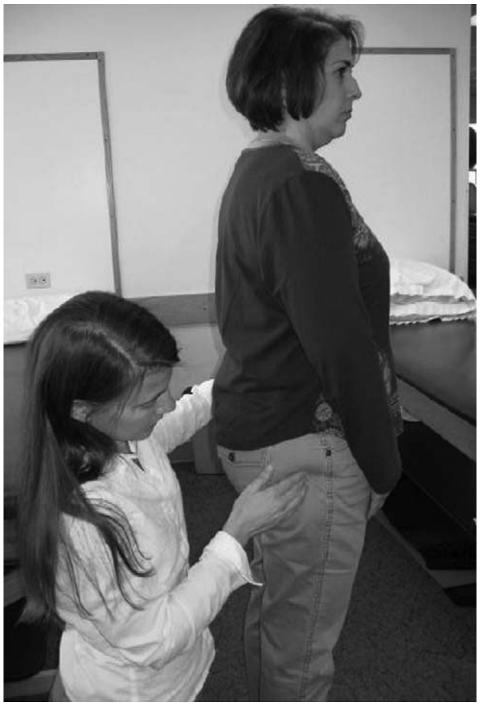

FIGURE 5. (ABOVE LEFT) Palpation posterior to greater trochanter.

FIGURE 6. (ABOVE RIGHT) Flexion-internal rotation-axial compression test.

FIGURE 7. (BOTTOM LEFT) Maximum flexion-internal rotation test.

FIGURE 8. (BOTTOM RIGHT) Maximum flexion-external rotation test.

or posterior labral tear. One of the tests was described as flexion, external rotation, and full abduction of the hip, followed by the hip being extended, internally rotated, and adducted (Figures 9A and 9B). If this maneuver was painful and presented with or without an audible click, it was considered indicative of an anterior labral tear. The second test was described as extension, abduction, and external rotation from a fully flexed, adducted, and internally rotated position. Pain reproduction with or without an audible click was considered indicative of a posterior labral tear. No clear data on rater(s) were provided. Subjects in this retrospective study were 56 patients with a diagnosis of a labral tear. Gold standard test was findings on arthrotomy and arthroscopy. Of the 56 subjects, 54 had a positive labral maneuver; however, the authors did not correlate the location of the labral tears with findings on the two proposed location-specific tests.

DISCUSSION

As we noted in the introduction, the goal of this review is to determine a current best-evidence diagnostic physical examination test, or perhaps test cluster, for the diagnosis of hip labral lesions. Crucial to this stated goal is an evalua-

FIGURE 9. Fitzgerald test for anterior labrum: **A.** Start position: Flexion-external rotation-abduction. **B.** End position: Extension-internal rotation-adduction (Posterior labral test, extension-abduction-external rotation from flexion-adduction-internal rotation not depicted).

tion of the research validity of the retrieved research papers. Domholdt[35] defined research validity as the extent to which conclusions of a study are believable and useful. We will discuss three areas specific to these diagnostic utility studies on physical examination tests for labral lesions where research validity can be threatened: construct validity, external validity, and statistical conclusion validity. In addition to this narrative review, we will use the implications of the systematic review using the STARD criterion list to provide a current best-evidence summary.

Construct Validity

A construct is an artificial framework that is not directly observable. The main threat to construct validity in diagnostic utility research is the discrepancy between the construct as *labeled* and the construct as *implemented*[35].

Labral lesion as a pathology requiring surgery or conservative management?

In the 16 studies retrieved, arthroscopic findings were the exclusive gold standard test in 12; three studies used MRA[5,14,18], and one study used a mixed gold standard of MRA and arthroscopic findings[8]. As therapists, we are not only interested in identifying those patients who will benefit from surgical management but we would also like to know which patients might benefit from conservative management as could be provided by a therapist. However, it is clear that most studies retrieved here discussed the hip labral lesion as a pathology requiring surgery. With no studies available on the natural history or the effect of conservative management of hip labral lesions, due in part to the absence of an evidence-based, generally agreed-upon clinical diagnostic test or test cluster, we have to rely on basic science data to gauge the validity of the construct of a hip labral lesion that may in some cases be amenable to conservative management.

When reviewing relevant basic science literature, we have to note that there is minimal information addressing the function of the labrum. Some studies[36,37] have demonstrated the labrum's ability to seal the hip joint against fluid expression from the joint space, thus preserving the joint's cartilaginous layers from high stresses, more evenly distributing loads across the joint, and maintaining joint lubrication.

Ferguson et al[38] used a finite element analysis to determine the significance of the labrum's function with respect to joint loads. They assessed the rate at which the acetabulum and femur approached one another under specific loads and found that joints in models without a labrum approximated 40% faster than joints with a labrum. In models without a labrum, the layers of cartilage also were compressed 35% more than in those with a labrum[38]. The contact peak pressure was 18% higher in models without a labrum[38]. With joint pressures increased, fluid was expressed from the cartilage causing tissue consolidation, which was proposed to lead to cartilage delamination as a precursor to degenerative changes. In contrast, models with an intact labrum showed resistance to the fluid being expressed from the cartilage layers, which slowed the deformation rate and limited joint stresses[38].

In another study, Ferguson et al[39] used bovine acetabular labrums and again reported the labrum's ability to maintain interstitial fluid and resist the flow of fluid out of the cartilage with compression testing. Adeeb et al[40] noted that the function of the labrum is to reduce the pressure gradient and fluid flow out of the joint tissues thereby decreasing the tensile stresses across the joint's contact areas. Adeeb et al[40] also suggested decreased circumferential stresses at the joint's perimeter are due to the labrum's function of confining the joint. Jones[41] noted that both the labrum and joint capsule have important functions of stabilizing the joint by creating a vacuum seal.

So the available basic science research, mathematical models, and expert opinion seem to agree that with the absence of the labrum, fluid is expressed out of the joint at a greater rate predisposing the cartilage to damage and possible premature arthritic changes. This evidence would seem to support the construct of the labral lesion as a pathology requiring surgical intervention.

In contrast, there is the construct of a labral lesion potentially amenable to conservative management. Again we have to rely on basic science evidence. Blood supply is essential to establishing healing potential. The acetabular labrum receives its blood supply via the obturator artery and the superior and inferior gluteal arteries[42], and there is also a possible role for the medial femoral circumflex artery[43]. McCarthy et al[42] reported a vast supply of vessels reaching the acetabulum, the labrum, and capsular sulcus. There were, however, no vessels visualized that penetrated the labrum. Kelly et al[44] also reported that the acetabular labrum is essentially avascular, although they did report some peripheral branches on the capsular edge and distal aspect of the labrum. Peterson et al[45] performed immunostaining of laminin (a component of blood vessels) and found it to be positive in the outer third of the labrum and negative in the inner two-thirds of the labrum, thus confirming the inner two-thirds of the labrum as avascular, and suggesting that only tears in the outer third have the potential to heal.

Also relevant to these two constructs of the labral lesion with opposing clinical implications is the great variety in the retrieved studies with regard to a positive finding on the gold standard test. This variety comes from the fact that the studies provide no operational definitions for these positive findings; labral lesions have included labral tears (complete or incomplete), "lesions," damage, degeneration, and avulsion. Considering this variety of mostly poorly defined lesions and the possible rehabilitation potential of at least some labral lesions, the unspoken assumption that a positive clinical test would automatically indicate the need for surgical referral and management should be questioned.

Labral lesion as a painful lesion

Although the operational definition of a positive test finding generally leaves something to be desired in the retrieved studies, most often, pain is indicated or inferred as a positive test finding. For pain to occur due to a labral lesion, the labrum needs to in fact be innervated. The obturator nerve and a branch of the nerve to the quadratus femoris muscle are believed to innervate the acetabular labrum[44]. Kim and Azuma[46] found Vater-Pacini and Golgi-Mazzoni corpuscles (pressure receptors), Ruffini corpuscles (temperature and deep sensory receptors), and Krause corpuscles (temperature receptors) throughout the labrum, but more prominently in the anterior and superior regions; they suggested that these nerve endings within the acetabular labrum might provide pain and proprioceptive sensations. Based on this basic science evidence, it is conceivable that not all clinically relevant labral lesions, especially those not located in the antero-superior labrum, would present with pain even on provocative testing as described above.

Reference test: Arthroscopy or MRA?

Cook et al[33] reported that using an inappropriate reference test is one of the common methodological mistakes made in diagnostic utility studies. The gold standard or reference test used should be the best test available to identify all those patients with the specific disorder. Studies retrieved have mainly used arthroscopic findings as the reference test, but some studies have used MRA[5,14,18]; one study[8] used both MRA and arthroscopy as the reference test.

Burnett et al[7] noted positive MRA findings in 48 of their 66 subjects. However, upon arthroscopy, all 66 patients were found to have labral tears, yielding a sensitivity of 79% for MRA compared to the reference test of arthroscopy. Chan et al[8] found labral tears upon arthroscopy in 16 of 17 patients with a positive MRA, yielding 100% sensitivity and 94% specificity for MRA. Petersilge et al[21] noted positive MRA findings in 10 subjects and labral lesions upon surgery in 9 patients, yielding a sensitivity of 100% but a specificity of 0% for MRA. Leunig et al[19] reported MRA specificity to be greater than 70% for detecting labral pathology and sensitivity to be 63% for labral tears and 92% for labral degeneration. Keeney et al[15] reported 71% sensitivity and a specificity of 44% to detect labral pathology with MRA as compared to a reference test of surgical findings.

None of the studies examined here provided a clear operational definition of the MRA testing procedures used or the level of expertise of the radiologist reviewing the results. Petersilge et al[21] noted that improper distention of the joint, not injecting gadolinium intra-articularly to provide distinction between the labrum and the joint capsule, not observing the joint in three planes, and the level of experience of the radiologist and orthopedic surgeon might all affect the diagnostic utility of MRA. Whatever the reason, it is evident that MRA is an inferior reference test when compared to arthroscopy and that the results from diagnostic utility studies using MRA as a reference test carry less value.

Reference test: Arthroscopy

Clear operational definition of what is considered a positive finding on arthroscopy is lacking in the studies retrieved. In addition, the normal anatomy of the labrum has been a topic of debate in the literature. There has been discussion on the normal shape of the labrum, the presence of partial separation or sulci between the labrum and the acetabulum, and even absence of the labrum and its significance with regard to pathology[36,47-52]. In the absence of consensus, we can also question the validity of arthroscopic findings as the gold standard test.

Lack of rater blinding

Cook et al[33] also reported absence of rater blinding as a common methodological mistake in diagnostic utility studies. Lack of rater blinding in the studies retrieved occurred in several. Of the studies retrieved only five were prospective[2,5,8,18,19]; all the other studies were retrospective. Retrospective studies allow the raters access to not only the findings of the clinical test studied but also to all other physical examination and surgical data. However, even in the prospective studies, it is not clear whether the raters had access to patient data in addition to the clinical test results. The construct as labeled might be diagnostic utility of a single physical examination test but access to other data makes the construct as implemented diagnostic utility of the whole examination process including said clinical test. Even though this construct obviously is more consistent with the clinical situation, it may also place an overestimated diagnostic utility value on the physical examination test studied.

External Validity

External validity deals with the degree to which study results can be generalized to different subjects, settings, and times[35]. However, in the studies retrieved, there are also issues related to operational definition of the test and test findings.

Although not specifically mentioned in multiple studies, we can safely assume that in most if not all cases the setting for the study was a secondary-care level center such as an orthopaedic or sports medicine surgical setting. Often subjects in these studies had chronic complaints for which they might already have received an unsuccessful course of conservative management perhaps including physical therapy; however, data on this are lacking. Suenaga et al[22] studied patients scheduled for an acetabular transposition osteotomy. Other studies included patients with co-morbidities, such as dysplastic osteoarthritis, acetabular dysplasia, acetabular rim syndrome, femoro-acetabular impingement, and developmental dysplasia of the hip[6,13,14,18-20,22]. Accurate diagnosis of these pathologies is clearly outside the realm of the physical therapy practice. Therefore, we have to question our ability to generalize findings from such secondary settings to the primary level, direct-access or even the referral-based environment in which most physical therapists currently practice.

Although again with no specific information provided in multiple studies, we can also safely assume that the rater was mostly the surgeon performing the surgical intervention or ordering the MRA. Although less relevant, we have to question external validity of these studies to physical therapists performing the same clinical tests.

Only some of the studies provided an operational definition of test performance. None provided an indication of the force involved during testing. Clear operational definition of what was considered a positive or negative test finding was also frequently lacking as was the possibility of a third category of test finding, the indeterminate test. As again indicated by Cook et al[33], absence of this third category can overestimate diagnostic utility findings. However, unclear test findings are a common clinical reality that is not acknowledged in studies using dichotomous test results only.

Statistical Conclusion Validity

Using inappropriate statistical tools for data analysis is a threat to statistical conclusion validity[35]. In the studies retrieved, this threat to research validity would seem the least relevant of the three. During the process of retrieval of potentially relevant studies, we had to drop some studies from our selection, although each initially seemed promising, they all provided insufficient data to calculate diagnostic utility statistics. All studies included allowed for some diagnostic utility calculation, but due to the study design, in most cases not all diagnostic utility statistics are included in Table 6. It is interesting to note that for most studies the relevant statistical values were not provided; rather, we had to calculate each value.

Systematic Review: STARD Criterion List Scores

Supplementing the above narrative review, the STARD criterion list allows us to systematically assess methodological quality of the studies retrieved. Table 7 provides the STARD scores for all studies. The studies retrieved yielded STARD scores ranging from 7 to 16; >50% of the studies had scores in the midrange of 10 to 13.

Although the STARD scoring system does not provide consensus-based cut-off values indicating acceptable methodological quality, the studies scoring below this midrange would seem to contribute little to a current best-evidence synthesis. Three studies[4,11,12] fell below this midrange level: Guanche and Sikka, Santori and Villar, and Hase and Ueo. On the other hand, those studies scoring above midrange could be assumed to provide greater value due to the higher methodological quality. Four studies[5,14,15,18] exceeded this midrange: Keeney et al, Leunig et al, Kassarijian et al, and Beaule et al. The remaining 9 studies scored in the midrange and may be of some benefit, but we need to be aware of the modest methodological quality.

Current Best-Evidence Summary

Based on the narrative discussion of research validity above, it is evident that all studies retrieved suffer from methodological flaws to some extent. Some of these flaws mainly affect our interpretation of study findings, whereas others are sufficiently significant that we have to discard a study from consideration altogether.

We discussed how all studies seemed to assume that a labral lesion as indicated by a positive clinical test poses an automatic surgical indication. That would mean that a physical therapist upon finding a hip labral test positive, if of course this test was sufficiently diagnostic, would have to refer the patient for a surgical consult. But perhaps there are some labral lesions that might respond to non-surgical management. Basic science indicates a possible healing potential for peripheral labral lesions. Clinically, this means that despite a positive test finding, especially on a test with established moderate specificity, a trial of conservative management may yet be indicated.

We also noted that not all labral lesions are necessarily painful on testing. Mainly the antero-superior labrum is nociceptively innervated. Although some authors have studied tests with purported specific effects on different portions of the labrum, these authors have not linked clinical test to reference test findings that would allow us to calculate diagnostic utility statistics. Clinically, this means that a negative test finding, even on a highly sensitive test, does not exclude labral lesions, especially those not located antero-superiorly.

We have discussed the lack of rater blinding. Only five of the studies retrieved[2,5,8,18,19] were prospective studies: Chan et al, Beaule et al, Leunig et al, Leunig et al, and Narvani et al. Although we would prefer to use data from prospective rather than retrospective studies in our current best-evidence synthesis, as discussed above this is an issue of limited importance. Clinically, a therapist would not use the finding on one clinical test in isolation to determine a diagnosis but rather would use data collected in the whole history and physical examination process as the researchers likely did in all studies retrieved. We should be aware that this does lead to inflated diagnostic utility data for the tests studied and we should adjust the importance we place upon an isolated test finding accordingly.

In clinical practice, a test is not always positive or negative. Not using a third category of indeterminate findings in all studies retrieved artificially inflates diagnostic utility data[33]. Clinically this means that our confidence in using sensitivity and specificity data from this research to rule out or in a diagnosis should be decreased.

Using an orthopaedic surgeon and not a therapist as the rater in all studies may affect research validity. More important is that patients included were not representative of the patients seen in normal physical therapy settings. Cook et al[33] noted how this spectrum or selection bias would lead to an overestimation of sensitivity data. Clinically, this means that even a negative finding on a highly sensitive test may not provide the therapist with sufficient diagnostic confidence to rule out a labral lesion.

Some threats are sufficiently significant that they cause us to exclude studies from our current best-evidence synthesis. We discussed how MRA is an inferior gold standard test as compared to arthroscopic findings but how even arthroscopy as a gold standard test needs to be questioned. All studies retrieved used arthroscopic surgery as the reference test except for three[5,14,18]: Beaule et al, Kassarjian et al, and Leunig et al. This leads us to exclude these three studies from our current best-evidence synthesis.

Another critical issue is insufficient operational definition of test performance and interpretation of test findings. In the studies by Keeney et al, Beck et al, and Suenaga et al[6,15,22], operational definitions were insufficient for replication and we decided to also exclude these studies.

With regard to the systematic assessment of methodological quality of the studies retrieved, the studies by Guanche and Sikka, Santori and Villar, and Hase and Ueo all scored below the midrange and we, therefore, discarded these from our best-evidence synthesis[4,11,12]. Only four studies[5,14,15,18] scored high in the STARD scoring system: Beaule et al, Kassarjian et al, Keeney et al, and Leunig et al. Illustrating the qualitative nature of the STARD tool is the fact that above we have discarded all four of these studies, due to critical methodological flaws revealed in the narrative review of research validity. In all, this process has left us with 7[2,7,8,10,13,19,21] of the original 16 studies upon which to base our best-evidence synthesis: Burnett et al, Chan et al, Fitzgerald, Leunig et al, Narvani et al, Ito et al, and Petersilge et al. All studies had only midrange STARD scores indicating the need for improving methodological quality in future studies.

Although calculated where possible, the statistics of accuracy, predictive values, and likelihood ratios are less relevant to our best-evidence synthesis. The accuracy of a diagnostic test provides a quantitative measure of its overall value, but because it does not differentiate between the diagnostic value of positive and negative test results, its value with regard to the diagnostic decision-making process is minimal[33]. The prevalence in the clinical population being examined with a specific test has to be identical to the prevalence in the study population from which the predictive values were derived before we can justifiably use predictive values as a basis for diagnostic decisions. Considering the issue of spectrum bias discussed above, the usefulness is limited to those situations where we can justifiably make assumptions on similarity of prevalence, allowing us to virtually disregard these statistical data in our diagnostic decision-making process[33].

Likelihood ratios (LR) can be either positive or negative. A positive likelihood ratio indicates a shift in probability favoring the existence of a disorder if the test is found to be positive. Conversely, a negative likelihood ratio indicates a shift in probability favoring the absence of a disorder if the test is found to be negative. Table 8 provides the shifts in probability that a patient does or does not have a particular disorder given a positive or negative test associated with a specific range of positive and negative likelihood ratios[33]. A review of Table 6 yields LR varying from 0.58–1.32, values that provide only minimal shifts in the likelihood of a labral lesion being present or absent.

Relevant to our current best-evidence synthesis based on these remaining seven studies is the fact that physical examination tests that demonstrate high sensitivity are clinically useful screening tools in that they can be used for ruling out selected a diagnosis. With a highly sensitive test, there are few false negatives. On the other hand, highly specific tests are appropriate for "ruling in" a finding, because the likelihood of a false positive finding is low. Tests intended to diagnose a labral lesion of the hip may be false positive because the test detected pathology other than a labral lesion, e.g., capsuloligamentous impingement, labral degeneration without a clear tear, or unspecified osteochondral lesions of the hip. The presence of such lesions would be expected to decrease specificity of a test. This ability of highly sensitive and highly specific tests to rule out a condition or rule in a condition is captured in a mnemonic:

- **SnNOUT**: With highly Sensitive tests, a Negative result will rule a dis order **OUT**.
- **SpPIN**: With highly Specific tests, a Positive result will rule a disorder **IN**.

With regard to a best-evidence synthesis, we therefore seek to identify highly

TABLE 8. Diagnostic value guidelines (reprinted with permission from: Cook C, Cleland J, Huijbregts P. Creation and critique of studies of diagnostic accuracy: Use of the STARD and QUADAS methodological quality assessment tools. *J Manual Manipulative Ther* 2007;15:93–102).

LR+	Interpretation
> 10	Large and often conclusive increase in the likelihood of disease
5 - 10	Moderate increase in the likelihood of disease
2 - 5	Small increase in the likelihood of disease
1 - 2	Minimal increase in the likelihood of disease
1	No change in the likelihood of disease
LR–	**Interpretation**
1	No change in the likelihood of disease
0.5 - 1.0	Minimal decrease in the likelihood of disease
0.2 - 0.5	Small decrease in the likelihood of disease
0.1 - 0.2	Moderate decrease in the likelihood of disease
< 0.1	Large and often conclusive decrease in the likelihood of disease

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

TABLE 9. Current best-evidence synthesis: Sensitivity data.

Test	Study	Sensitivity
Flexion-adduction-internal rotation test	Burnett et al[7]	0.95
	Ito et al[13]	0.96
Impingement provocation test	Leunig et al[19]	1.00
Flexion-internal rotation test	Chan et al[8]	1.00
	Petersilge et al[21]	1.00
Flexion-adduction-axial compression test	Narvani et al[2]	1.00
Fitzgerald test	Fitzgerald[10]	1.00

sensitive and specific tests to rule out or rule in a diagnosis, respectively[33].

Five clinical tests studied in these seven studies yielded a high value for sensitivity (Table 9) with values ranging from 0.95–1.00. Although as discussed above, we have to consider that these sensitivity data are inflated due a variety of methodological flaws, current best evidence still indicates that a negative result with one of these tests will likely rule out a disorder. Although not studied as a test cluster, logically speaking, our confidence in ruling out a diagnosis of hip labral lesion should increase if more of these five tests are found to be negative.

With regard to specificity, only a few studies allowed for calculation of this statistic and only one of these studies yielded a value for specificity greater than zero: Narvani et al[2] produced a specificity of 0.43 for the flexion-internal rotation-axial compression test. We have to conclude that this midrange value provides little, if any, input in the diagnostic decision-making process, especially in light of the fact discussed above that various methodological flaws may have led to inflated specificity values. Currently, our best-evidence synthesis shows that there are no tests that, when positive, can be used to confidently clinically diagnose a hip labral lesion.

Limitations

Limitations of this systematic literature review and the subsequently provided best-evidence summary involve the literature search strategy presented. Articles may have been missed based on the omission of certain search phrases and key words. Limiting the search to English-language articles only may have omitted relevant articles written in another language. Limitations also relate to the studies retrieved: many articles did not include relevant information needed for interpretation of findings, most notably information on test performance and operational definition of test findings. The absence of consensus-based cut-off values on the methodological quality assessment tool used means that the interpretation of the studies retrieved remains based largely on the narrative qualitative assessment of their methodological quality provided in the discussion section of this paper. Finally, the STARD tool used for this review is a prospective tool designed to outline the features required for an unbiased diagnostic accuracy study. In contrast, the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) list is a retrospective tool used to critique the methodological rigor of a diagnostic accuracy study[33]. As a result of this difference in purpose, one could argue that the QUADAS tool would have been a more appropriate tool to use for the methodological assessment portion of this review paper.

Conclusion

Current best evidence indicates that a negative finding for the flexion-adduction-internal rotation test, the flexion-internal rotation test, the impingement provocation test, the flexion-adduction-axial compression test, the Fitzgerald test, or a combination of these tests provides the clinician with the greatest evidence-based confidence that a hip labral lesion is absent. Currently, research has produced no tests with sufficient specificity to help confidently rule in a diagnosis of hip labral lesion.

Our review of the literature and critical analysis of research validity provide directions for future research. Most importantly, future research needs to provide clear operational definitions of test performance and interpretation of test findings. Also, we strongly suggest the consistent use of arthroscopic findings as a reference test with the caveat that description and interpretation of arthroscopic findings needs to be standardized to a greater degree. A third category of indeterminate findings on the clinical tests studied would provide for more realistic diagnostic utility data, and using therapists as raters would increase external validity for use of these tests in the physical therapy setting. Also important is that future studies be done prospectively in physical therapy settings thereby decreasing spectrum bias but also allowing the research to produce information on specificity not provided by retrospective studies in surgical settings using only subjects with arthroscopically confirmed labral lesions. Solely studying physical examination tests or test clusters also would provide for data on diagnostic utility not influenced by other findings in history and physical examination.

The above line of research would allow for evidence-based diagnosis of hip labral lesions, which is required for controlled clinical trials that should be done to determine if and which patients with hip labral lesions would benefit from either conservative or surgical management. Diagnostic utility research into studies that might be able to more specifically diagnose labral lesions with regard to location might provide another avenue to treatment-based classification of these patients.

Finally, the fact that our systematic analysis using the STARD methodological quality assessment tool and our narrative analysis provided contradictory findings indicates that use of the STARD tool as the sole method for methodological quality assessment in systematic reviews of diagnostic utility studies needs to be carefully considered. Currently, all items are scored equally on this tool but we suggest weighing the items higher that represent what we in this review considered fatal flaws to get a more valid representation of study quality. This review clearly indicates that additional development and reliability and validity testing of this tool are required.

Acknowledgments

This paper was written in partial fulfillment for the primary author's Master of Health Science degree at the University of St. Augustine for Health Sciences in St. Augustine, Florida.

REFERENCES

1. Anderson K, Strickland SM, Warren R. Hip and groin injuries in athletes. *Am J Sports Med* 2001;29:521–533.
2. Narvani A, Tsiridis E, Kendall S, Chaudhuri R, Thomas P. A preliminary report on prevalence of acetabular labrum tears in sports patients with groin pain. *Knee Surg Sports Traumatol Arthroscop* 2003;11:403–408.
3. McCarthy JC, Noble PC, Schuck MR, Wright J, Lee J. The role of labral lesions in development of early degenerative hip disease. *Clin Orthop* 2001;393:25–37.
4. Santori N, Villar R. Acetabular labral tears: Result of arthroscopic partial limbectomy. *Arthroscopy* 2000;16:11–15.
5. Beaule P, Zaragoza E, Motamedi K, Copelan N, Dorey FJ. Three-dimensional computed tomography of the hip in the assessment of femoracetabular impingement. *J Orthop Res* 2005;23:1286–1292.
6. Beck M, Leunig M, Parvizi J, Boutier V, Wyss D, Ganz R. Anterior femoroacetabular impingement. Part II. Midterm results of surgical treatment. *Clin Orthop* 2004;418:67–73.
7. Burnett RSJ, Della Rocca GJ, Prather H, Curry M, Maloney WJ, Clohisy JC. Clinical presentation of patients with tears of the acetabular labrum. *J Bone Joint Surg* 2006;88A:1448–1457.
8. Chan YS, Lien LC, Hsu HL, et al. Evaluating hip labral tears using magnetic resonance arthrography: A prospective study comparing hip arthroscopy and magnetic resonance arthrography diagnosis. *Arthroscopy* 2005;21:1250.e1–1250.e8.
9. Farjo LA, Glick JM, Sampson TG. Hip arthroscopy for acetabular labral tears. *Arthroscopy* 1999;15:132–137.
10. Fitzgerald RH. Acetabular labrum tears. *Clin Orthop* 1995;311:60–68.
11. Guanche CA, Sikka RS. Acetabular labral tears with underlying chondromalacia: A possible association with high-level running. *Arthroscopy* 2005;21:580–585.
12. Hase T, Ueo T. Acetabular labral tear: Arthroscopic diagnosis and treatment. *Arthroscopy* 1999;15:138–141.
13. Ito K, Leunig M, Ganz R. Histopathologic features of the acetabular labrum in femoroacetabular impingement. *Clin Orthop* 2004;429:262–271.
14. Kassarjian A, Yoon LS, Belzile E, Connolly SA, Millis MB, Palmer WE. Triad of MR arthrographic findings in patients with camtype femoroacetabular impingement. *Radiology* 2005;236:588–592.
15. Keeney JA, Peelle MW, Jackson J, Rubin D, Maloney WJ, Clohisy JC. Magnetic resonance arthrography versus arthroscopy in the evaluation of articular hip pathology. *Clin Orthop* 2004;429:163–169.
16. Klaue K, Durnin C, Ganz R. The acetabular rim syndrome. *J Bone Joint Surg* 1991;73B:423–429.
17. Kloen P, Leunig M, Ganz R. Early lesions of the labrum and acetabular cartilage in osteonecrosis of the femoral head. *J Bone Joint Surg* 2002;84B:66–69.
18. Leunig M, Podeszwa D, Beck M, Werlen S, Ganz R. Magnetic resonance arthrography of labral disorders in hips with dysplasia and impingement. *Clin Orthop* 2004;418:74–80.
19. Leunig M, Werlen S, Ungersböck A, Ito K, Ganz R. Evaluation of the acetabular labrum by MR arthrography. *J Bone Joint Surg* 1997;79B:230–234.
20. McCarthy J, Lee J. Acetabular dysplasia: A paradigm of arthroscopic examination of chondral injuries. *Clin Orthop* 2002;405:122–128.
21. Petersilge CA, Haque MA, Petersilge WJ, Lewin JS, Lieberman JM, Buly R. Acetabular labral tears: Evaluation with MR arthrography. *Radiology* 1996;200:231–235.
22. Suenaga E, Noguchi Y, Jingushi S, et al. Relationship between the maximum flexion-internal rotation test and the torn acetabular labrum of a dysplastic hip. *J Orthop Sci* 2002;7:26–32.
23. Lewis C, Sahrmann S. Acetabular labral tears. *Phys Ther* 2006;86:110–121.
24. Mason JB. Acetabular labral tears in the athlete. *Clin Sports Med* 2001;20:779–790.
25. Wenger DE, Kendell KR, Miner MR, Trousdale RT. Acetabular labral tears rarely occur in the absence of bony abnormalities. *Clin Orthop* 2004;426:145–150.
26. Tonnis D, Dortmund D, Heinecke A. Acetabular and femoral anteversion: Relation-

ship with osteoarthritis of the hip. *J Bone Joint Surg* 1999;81A:1747–1770.

27. Peelle MW, Della Rocca GJ, Maloney WJ, Curry MC, Clohisy JC. Acetabular and femoral radiographic abnormalities associated with labral tears. *Clin Orthop* 2005;441:327–333.
28. Ito K, Minka MA, Leunig M, Werlen S, Ganz R. Femoroacetabular impingement and the cam-effect: An MRI-based quantitative anatomical study of the femoral head-neck offset. *J Bone Joint Surg* 2001;83B:171–176.
29. Siebenrock K, Wahab KH, Werlen S, Kalhor M, Leunig M, Ganz R. Abnormal extension of the femoral head epiphysis as a cause of cam impingement. *Clin Orthop* 2004;418:54–60.
30. Lavigne M, Parvizi J, Beck M, Siebenrock KA, Ganz R, Leunig M. Anterior femoroacetabular impingement. Part I. Techniques of joint preserving surgery. *Clin Orthop* 2004;418:61–66.
31. Paluska SA. An overview of hip injuries in running. *Sports Med* 2005;35:991–1014.
32. Bharam S. Labral tears, extra-articular injuries, and hip arthroscopy in the athlete. *Clin Sports Med* 2006;25:279–292.
33. Cook C, Cleland J, Huijbregts P. Creation and critique of studies of diagnostic accuracy: Use of the STARD and QUADAS methodological quality assessment tools. *J Manual Manipulative Ther* 2007;15:93–102.
34. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Radiology* 2003;226:24–28.
35. Domholdt E. *Physical Therapy Research: Principles and Applications*. Philadelphia, PA: WB Saunders, 1993.
36. Ferguson S, Bryant JT, Ganz R, Ito K. The acetabular labrum seal: A poroelastic finite element model. *Clin Biomech* 2000;15:463–468.
37. Ferguson S, Bryant JT, Ganz R, Ito K. An *in vitro* investigation of the acetabular labral seal in hip joint mechanics. *J Biomech* 2003;36:171–178.
38. Ferguson S, Bryant JT, Ganz R, Ito K. The influence of the acetabular labrum on hip joint cartilage consolidation: A poroelastic finite element model. *J Biomech* 2000;33:953–960.
39. Ferguson S, Bryant J, Ito K. The material properties of the bovine acetabular labrum. *J Orthop Res* 2001;19:887–896.
40. Adeeb SM, Ahmed EYS, Matyas J, Hart DA, Frank CB, Shrive NG. Congruency effects on load-bearing in diarthrodial joints. *Comput Methods Biomech Biomed Engin* 2004;7:147–157.
41. Jones D. Neonatal hip stability and the Barlow test. *J Bone Joint Surg* 1991;73B:216–218.
42. McCarthy J, Noble P, Aluisio FV, Schuck M, Wright J, Lee JA. Anatomy, pathologic features, and treatment of acetabular labral tears. *Clin Orthop* 2003;406:38–47.
43. Stranding S, ed. *Gray's Anatomy: The Anatomical Basis of Clinical Practice*. 39th ed. London, UK: Elsevier Churchill Livingstone, 2005.
44. Kelly BT, Shapiro GS, Digiovanni CW, Buly RL, Potter HG, Hannafin JA. Vascularity of the hip labrum: A cadaveric investigation. *Arthroscopy* 2005;21:3–11.
45. Petersen W, Petersen F, Tillmann B. Structure and vascularization of the acetabular labrum with regard to the pathogenesis and healing of labral lesions. *Arch Orthop Trauma Surg* 2003;123:283–288.
46. Kim YT, Azuma H. The nerve endings of the acetabular labrum. *Clin Orthop* 1995;320:176–181.
47. Czerny C, Hofmann S, Urban M, et al. MR arthrography of the adult acetabular capsular-labral complex: Correlation with surgery and anatomy. *Am J Roentgenol* 1999;173:345–349.
48. Aydingoz U, Ozturk M. MR imaging of the acetabular labrum: A comparative study of both hips in 180 asymptomatic volunteers. *Eur Radiol* 2001;11:567–574.
49. Ghebontni L, Roger B, El-khoury J, Brasseur JL, Grenier PA. MR arthrography of the hip: Normal intra-articular structures and common disorders. *Eur Radiol* 2000;10:83–88.
50. Lecouvet FE, Vande Berg BC, Malghem J. MR imaging of the acetabular labrum: Variations in 200 asymptomatic hips. *Am J Roentgenol* 1996;167:1025–1028.
51. Mitchell B, McCrory P, Brukner P, O'Donnell J, Colson E, Howells R. Hip joint pathology: Clinical presentation and correlation between magnetic resonance arthrography, ultrasound, and arthroscopic findings in 25 consecutive cases. *Clin J Sport Med* 2003;13:152–156.
52. Byrd JT. Labral lesions: An elusive source of hip pain. Case reports and literature review. *Arthroscopy* 1996;12:603–612.