Diagnostic accuracy of clinical tests for the diagnosis of hip femoroacetabular impingement/labral tear: a systematic review with meta-analysis

M P Reiman,[1,2] A P Goode,[1] C E Cook,[1] P Hölmich,[3,4] K Thorborg[3,5]

[1]Department of Orthopedic Surgery, Duke University Medical Center, Durham, North Carolina, USA
[2]Duke University Medical Center, Durham, North Carolina, USA
[3]Sports Orthopedic Research Centre—Copenhagen (SORC-C), Arthroscopic Center Amager, Copenhagen University Hospital, Copenhagen, Denmark.
[4]Aspetar Sports Groin Pain Center, Qatar Orthopaedic and Sports Medicine Hospital, Doha, Qatar
[5]Physical Medicine & Rehabilitation Research—Copenhagen (PMR-C), and Departments of Orthopaedic Surgery and Physical Therapy, Copenhagen University Hospital, Copenhagen, Denmark

Correspondence to
Dr Michael P Reiman, Department of Orthopedic Surgery, Duke University School of Medicine, DUMC 104002, Durham, NC 27710, USA; reiman.michael@gmail.com

Accepted 21 November 2014
Published Online First
16 December 2014

ABSTRACT

Background Surgery for hip femoroacetabular impingement/acetabular labral tear (FAI/ALT) is exponentially increasing despite lacking investigation of the accuracy of various diagnostic measures. Useful clinical utility of these measures is necessary to support diagnostic imaging and subsequent surgical decision-making.

Objective Summarise/evaluate the current diagnostic accuracy of various clinical tests germane to hip FAI/ALT pathology.

Methods A computer-assisted literature search of MEDLINE, CINAHL and EMBASE databases using keywords related to diagnostic accuracy of the hip joint, as well as the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) guidelines were used for the search and reporting phases of the study. Quality assessment of bias and applicability was conducted using the Quality of Diagnostic Accuracy Studies-2 (QUADAS-2). Random effects models were used to summarise sensitivities (SN), specificities (SP), diagnostic odds ratio (DOR) and respective confidence intervals (CI).

Results The employed search strategy revealed 21 potential articles, with one demonstrating high quality. Nine articles qualified for meta-analysis. The meta-analysis demonstrated that flexion-adduction-internal rotation (pooled SN ranging from 0.94 (95% CI 0.90 to 0.97) to 0.99 (95% CI 0.98 to 1.00); DOR 5.71 (95% CI 0.84 to 38.86) to 7.82 (95% CI 1.06 to 57.84)) and flexion-internal rotation (pooled SN 0.96 (95% CI 0.81 to 0.99); DOR 8.36 (95% CI 0.41 to 171.3) tests possess only screening accuracy.

Conclusions Few hip physical examination tests for diagnosing FAI/ALT have been investigated in enough studies of substantial quality to direct clinical decision-making. Further high-quality studies across a wider spectrum of hip pathology patients are recommended to discern the confirmed clinical utility of these tests.

Trials registration number PROSPERO Registration # CRD42014010144.

INTRODUCTION

Femoroacetabular impingement (FAI), an abnormal articulation and abutment of the femoral head against the acetabulum, is suggested to contribute to acetabular labral tear (ALT) and chondrolabral injury.[1] The prevalence of ALT in patients with hip or groin pain ranges from 22% to 55%,[2–5] although higher prevalence would be expected in a hip surgeon's consultancy.

Surgery for the correction of FAI has increased significantly over the past decade.[6 7] In fact, an 18-fold increase between 1999 and 2009 was shown in the USA.[8] Given that differential diagnosis for the patient presenting with hip or groin pain is still suggested to be a diagnostic challenge,[9] focus on proper diagnosis would seem warranted. A significant increase in attention has been shown for differential diagnosis of the hip joint and periarticular structures as the primary source for hip-related pain/dysfunction,[10–14] although the principal focus is on radiographic imaging, and, to a lesser extent, clinical examination.[15–17] The diagnostic accuracy of radiographic imaging for FAI/ALT,[13 14] findings of radiographic asymmetries in asymptomatic individuals,[18–44] as well as availability and cost of such imaging merits determination of the clinical utility of the clinical examination for FAI/ALT pathologies.

While the diagnostic accuracy of hip physical examination (HPE) tests for FAI/ALT have previously been described,[45–47] the estimation of test probability and, therefore, true clinical utility has not been synthesised. Therefore, the purpose of this study was to conduct a systematic review and meta-analysis of the literature regarding the diagnostic accuracy of HPE tests for FAI/ALT and describe their clinical utility.

METHODS

Registration
The study was registered on 6 August 2014 with the International Prospective Register of Systematic Reviews (PROSPERO# CRD42014010144). PROSPERO is a database of prospectively registered systematic reviews for health and social topics. The study was registered after pilot search and prior to updated data search and extraction.

Data sources
The *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) guidelines were utilised during the search and reporting phase of this review. The PRISMA statement includes a 27-item checklist that is designed to be used as a basis for reporting systematic reviews of randomised trials,[48] but the checklist can also be applied to multiple forms of research methodologies.[49]

A computer-assisted literature search of MEDLINE, CINAHL and EMBASE databases was performed from inception of each respective database to 10 June 2014. The goal was to optimise the sensitivity our search strategy,[50 51] and to increase the likelihood that all appropriate studies were identified. The search strategy was developed in collaboration with a medical information specialist and used controlled vocabulary and key words related to

diagnostic accuracy of the clinical examination measures relative to hip femoroacetabular impingement and labral tear.

Screening filters were initially used during assessment of title, abstract and full text documents. The search was further limited to humans and English-only publications. The search strategy for MEDLINE is listed in online supplementary appendix 1.

Selection criteria

Articles examining clinical examination for FAI/ALT pathology were eligible if they met all of the following criteria: (1) included participants with hip pain suspected to be related to hip FAI/ALT, (2) included at least one clinical hip FAI/ALT pathology examination measure, (3) utilised an acceptable reference standard, (4) reported the results in sufficient detail to allow reconstruction of contingency tables to allow calculation of diagnostic accuracy metrics and (5) were written in English.

An article was excluded if: (1) the pathology was associated with a condition located elsewhere (eg, lumbar spine) that referred pain to the hip region, (2) the clinical measures were performed under any form of anaesthesia or on cadavers, (3) specialised instrumentation not readily available to all clinicians were used, (4) physical clinical tests were included as a component of cluster tests, and (5) testing was performed on infants/toddlers.

Data extraction

Two reviewers (MPR, KT) independently performed the search. Since computerised search results for diagnostic accuracy data frequently omit many relevant articles,[52] the reference lists of all selected publications were checked to retrieve relevant publications that were not identified in the computerised search. Grey literature was also manually searched and included publications, posters, abstracts or conference proceedings. To identify relevant articles, titles and abstracts of all identified citations were independently screened. Full-text articles were retrieved if the abstract provided insufficient information to establish eligibility or if the article passed the first eligibility screening.

All criteria were independently applied by two reviewers (MPR, KT) to the full text of the articles that passed the first eligibility screening. Disagreements among the reviewers were discussed and resolved by a third reviewer (CEC). We determined which articles to include (for meta-analysis) using clinical and statistical judgement of study heterogeneity. Clinical judgment criteria involved assessment of similarity of populations, assessment context (eg, test performed a priori), study design (eg, case–control vs case based) and method in which specific tests were applied.[53] In addition, after approval using clinical judgment, studies were statistically pooled when ≥2 studies examined the same index test and diagnosis with the same reference standard.

Risk of bias/quality assessment

Each of the full text articles was independently reviewed by two reviewers (MPR, CEC) and scored with the *Quality Assessment of Diagnostic Accuracy Studies* 2 scores (QUADAS-2) tool.[54] Disagreements among the reviewers were discussed and resolved during a consensus meeting. The QUADAS-2 is a quality assessment tool comprised of four domains: patient selection, index test, reference standard, and flow and timing. The risk of bias is assessed in each of the domains, while the first three domains are also assessed for applicability by indicating a 'low', 'high' or 'unclear' rating. Applicability in the QUADAS-2 refers to whether certain aspects of an individual study are matching or not matching the review question. Unlike the QUADAS-1, the

QUADAS-2 does not utilise a comprehensive quality score, rather an overall judgement of 'low', 'high' or 'unclear' risk. An overall risk rating of 'low risk of bias' or 'low concern regarding applicability' requires the study to be ranked as 'low' on all relevant domains. A 'high' or 'unclear' rating in one or more domains may require that the study be rated as an 'at risk of bias' or having 'concerns regarding applicability.'

Data extraction and analysis

One reviewer (MPR) independently extracted information and data regarding study population, setting, special test performance, pathology, diagnostic reference-standard and number of true positives, false positives, false negatives and true negatives for calculation of sensitivity (SN), specificity (SP), positive likelihood ratios (+LR) and negative likelihood ratios (−LR) when not provided. SN is defined as the percentage of people who test positive for a specific pathology among a group of people who have the pathology. SP is the percentage of people who test negative for a specific pathology among a group of people who do not have the pathology. A positive likelihood ratio (+LR) is the ratio of a positive test result in people with the pathology to a positive test result in people without the pathology. A +LR identifies the strength of a test in determining the presence of a finding, and is calculated by the formula: $SN/(1-SP)$. A negative likelihood ratio (−LR) is the ratio of a negative test result in people with the pathology to a negative test result in people without the pathology, and is calculated by the formula: $(1-SN)/SP$. The higher the +LR and lower the −LR the greater the post-test probability is altered. If analysed independently, tests that demonstrate high SN and low −LR are useful in ruling out a condition (screening). In contrast, tests that demonstrate high SP and high +LR assist in ruling in a condition (confirmation).[55]

It has been suggested that post-test probability can be altered to a minimal degree with +LRs of 1– 2 or −LRs of 0.5–1, to a small degree with +LRs of 2–5 or −LRs of 0.2–0.5, to a moderate degree with +LRs of 5–10 and −LRs of 0.1–0.2) and to a large and almost conclusive degree with +LRs greater than 10 and −LRs less than 0.1.[55] Pretest probability is defined as the probability of the target pathology before a diagnostic test result is known. It represents the probability that a specific patient, with a specific history, presenting to a specific clinical setting, with a specific symptom complex, has a specific pathology.[55] The diagnostic odds ratio (DOR), defined as +LR/−LR, is a single indicator, independent of prevalence, which represents the ratio of the odds of positivity in disease relative to the odds or positivity in the non-diseased. The values for the DOR range from 0, indicating no test discrimination, to infinity with higher scores indicating better discrimination.[56]

Meta-analysis

DerSimonian and Laird[57] random effects models, which incorporate both between and within study heterogeneity, were used to produce summary estimates of SN, SP, +LR, −LR, and diagnostic DOR. We initially attempted to analyse these data with a bivariate/hierarchical receiver operator curve model for those diagnostic tests with at least four studies. However, these models failed to converge and therefore report findings from the random effects models. An I-squared value of >50% and Cochrane's-Q p value of <0.10 were the criteria to indicate significant between-study heterogeneity, of SN and SP and likelihood ratios respectfully. Publication bias was not formally tested due to low power of the tests with limited included studies.[58] No significant threshold effects were found using Spearman correlation coefficients. When a computational problem of empty cells existed, 0.5 was

added to all four cells as suggested by Cox.[59] Summary receiver operating characteristic (SROC) curves was produced when ≥ 4 studies were pooled. These curves summarise the diagnostic performance as a single number; the area under the curve (AUC) with a standard error (SE).[60] AUC $\geq 0.90–1.00$ were considered excellent, $\geq 0.80–90$ considered good, $\geq 0.70–80$ considered fair, $\geq 0.60–0.70$ considered poor and $\geq 0.50–60$ fail. Another measure, the Q* index and accompanied SE, is an additional measured produced which is the point on the SROC curve closest to the ideal left top-left corner (where SN and SP meet).[61] All analyses were conducted by one of the authors (AG), blinded to results of the search, inclusion/exclusion and study quality, in Meta-DiSc V.1.4.[62]

RESULTS
Selection of studies
The systematic search through MEDLINE, CINAHL and EMBASE netted 340 abstracts, and 10 additional abstracts were identified through an extensive manual search. In total, 350

titles were initially retained after duplicates were removed. Abstract and full text review reduced the acceptable papers to 21 (figure 1 and table 1).

This review included 1335 participants/1398 hips across 21 studies, investigating 11 different clinical special tests (tables 2 and 3). The sample size of the studies ranged from 10[63] to 241 participants (table 2).[64] Eight of the studies were retrospective,[65–73] nine were prospective[5 64 74–79] and the study design was unclear in four of the studies (table 1).[3 63 80 81]

Quality scores
Table 1 provides the overall risk of bias score, with only one study[80] demonstrating low risk of bias. It was unclear whether data collection was retrospective or prospective in this study.

RESULTS OF INDIVIDUAL DIAGNOSTIC CLINICAL MEASURES
Thirteen studies investigated the FADDIR test,[65–69 71–73 75–77 79 81] three studies investigated the Flex-IR test,[63 64 74]
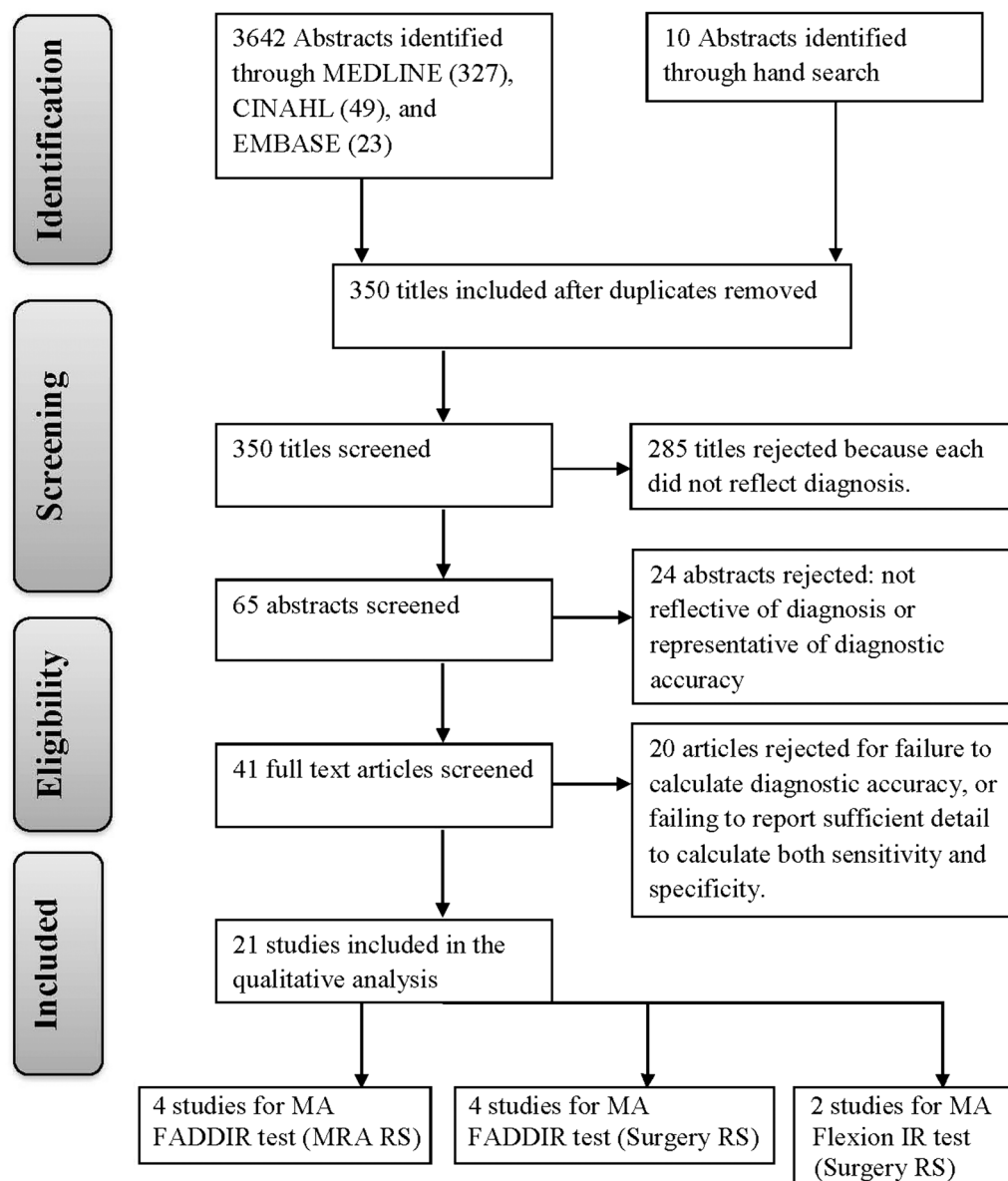

Figure 1 Flow diagram for study inclusion. IR, internal rotation; MA, meta-analysis; FADDIR, flexion adduction internal rotation test; MRA, MR arthrogram; RS, reference standard.

Review

Table 1 Primary purpose(s) and risk of bias in included studies

Study	Data collection	Primary purpose of study	Risk of bias
Aprato et al[65]	Retrospective	Diagnostic accuracy of MRA for FAI	High
Ayeni et al[80]	Unclear	Diagnostic accuracy of symptomatic maximal squat (clinical test) and cam FAI	Low
Barton et al[66]	Prospective	Diagnostic accuracy of α angle for cam FAI	High
Beaule et al[81]	Unclear	Diagnostic accuracy of CT for FAI	High
Chan et al[74]	Prospective	Diagnostic accuracy of MRA for ALT	High
Domayer et al[67]	Retrospective	Diagnostic accuracy of radiography for FAI	High
Hananouchi et al[75]	Prospective	Diagnostic accuracy of FADDIR test across four patient cohorts	High
Keeney et al[68]	Retrospective	Diagnostic accuracy of MRA for ALT and concurrent hip pathology	High
Laude et al[69]	Retrospective	Description of surgical treatment and outcomes for FAI	High
Leunig et al[76]	Prospective	Diagnostic accuracy of MRA for ALT	High
Martin et al[77]	Prospective	Diagnostic accuracy of clinical tests	High
Maslowski et al[78]	Prospective	Diagnostic accuracy several clinical tests	High
McCarthy and Busconi[3]	Unclear	Determine relationship between clinical findings, radiological tests and surgical findings	High
Myrick and Nissen[70]	Retrospective	Diagnostic accuracy of THIRD test	High
Narvani et al[5]	Prospective	Determine prevalence of ALT in sports patients presenting with groin pain	High
Nogier et al[64]	Prospective, descriptive study	Description epidemiology of patients with mechanical hip symptoms	High
Peters et al[71]	Retrospective	Description of clinical outcomes post-surgical treatment for FAI	High
Petersilge et al[63]	Unclear	Diagnostic accuracy of MRA for ALT	High
Sink et al[72]	Retrospective	Clinical presentation and diagnosis of FAI	High
Troelsen et al[79]	Prospective	Diagnostic accuracy of clinical tests and US compared with MRA	High
Wang et al[73]	Retrospective	Diagnostic accuracy of clinical examination, along with arthroscopic treatment for ALT	High

dx, diagnosis; FAI, femoroactabular impingement; MRA, MR angiogram.

one study investigated the Bilateral lower extremity squat test,[80] three studies investigated the FABER test,[77–79] one study investigated the Scour test,[78] one study investigated the IR with overpressure test,[78] one study investigated the Resisted SLR test,[78] one study investigated the Thomas test,[3] one study investigated the THIRD test,[70] one study investigated the IR-flexion-axial compression test,[5] and one study investigated the Trochanteric tenderness test.[77] Table 2 reports which tests each study investigated, as well as the characteristics of the study participants and the reference standard utilised. Table 3 reports the tests investigated, their diagnostic accuracy, reference standard utilised and the pretest to post-test probability change after use of the tests in each study for each test investigated.

META-ANALYSIS
Table 4 provides diagnostic properties and total sample sizes of the nine studies included in the meta-analysis. Four studies (n=188 participants)[65 68 79 81] investigated the use of the FADDIR test with MR angiogram (MRA) reference standard, four studies (n=319 participants)[68 69 71 76] investigated the use of the FADDIR test with surgery reference standard, and two studies (n=27 participants)[63 74] investigated the use of the Flex IR test with surgery reference standard.

Figures 2 and 3 illustrate the SROC curves for FADDIR test with MRA reference standard and surgery reference standard, respectively. This figure illustrates the relationship between SN and 1-SP (false-positive rate) for the included studies in pooled analyses. The AUC was considered fair (AUC=0.76 ((SE=0.19)) for the FADDIR with MRA reference and poor (AUC=0.65 ((SE=1.0)) for FADDIR with surgery reference. The point at which SN and SP were equal (Q*) was 0.70 (SE=0.19) the MRA reference standard and 0.61 (SE=0.81) for surgery reference standard.

Figure 4 illustrates pretest to post-test probability changes utilising the FADDIR test with a surgery reference standard. Pretest to post-test probability for SN remained unchanged. Pretest to post-test probability demonstrated a notable shift, although the CIs were quite large.

DISCUSSION
Our study examined the current literature investigating clinical examination tests for FAI/ALT. Owing to study heterogeneity and variable reference standards only two of the 11 different tests qualified for meta-analysis (FADDIR and Flexion IR tests). These, as well as most of the other tests, were predominantly only SN and not SP. Only two tests (in 9 studies) qualified for meta-analysis, and none of those tests significantly shifted post-test probability of the diagnosis of FAI/ALT. In fact, none of the tests investigated in our review is capable of significantly altering post-test probability. Additionally, the studies investigating all of the tests in our review, whether performed for the purpose of determining test diagnostic accuracy or not, were at risk for bias and low quality. A significant need for improved quality studies investigating these clinical tests exists.

The purpose of performing clinical testing is to aid with determination of whether a particular pathology exists or not. When the clinician is appraising evidence about diagnostic tests they should consider several key concepts:[55 82] did the participating individuals present a diagnostic dilemma? how much will different levels of the diagnostic test raise or lower the pretest probability of disease? will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting? are the results applicable to the patients in my practice? will the test results change my management strategy? and will the patients be better off as a result of the test?

Participants presenting to a clinicians practice with a diagnostic dilemma require utilisation of tests with properties capable of differentiating those with versus those without the disease. In order to properly assess these participants it is necessary to determine if they were drawn from a common group in which it

Table 2 Demographics and characteristics of included studies

Study	Participants/mean duration of symptoms	Pathology	Reference standard	Reference standard SN/SP (95% CI)	Clinical test investigated
Aprato et al[65]	41 participants (average age range 23–25 between surgical dislocation and arthroscopy groups; 17 females) NR	FAI	MRA 1.5 T; *Cam*: α angle >50°; *local pincer*: (+) crossover sign and (−) posterior wall sign and α angle <50°; *pincer*: (+) crossover and posterior wall signs or coxa profunda and α angle of <50°; *mixed*: local or global pincer associated with a cam	100/100 (α angle) 91 (NR)/86 (NR) (ALT)	▶ FADDIR test
Ayeni et al[80]	76 consecutive prospectively enrolled participants/78 hips (mean age 38.3 years, 39 females) presenting to adult outpatient clinic over 3 month period. NR	FAI	MRI or MRA 1.5 T; α angle >55° and/or loss of femoral head–neck offset <9.0 mm	NR in study MRI (ALT): 70 (62 to 77)/82 (69 to 94) MRA (ALT): 83 (79 to 88)/57 (44 to 70)[86]	▶ Bilateral LE squat (maximum depth)
Barton et al[66]	68 participants (age range 17 to 60 years, 37 females) from prospective database of all patients undergoing joint-preserving surgery and all having (+) MRA for ALT. NR	FAI	MRA 1.5 T; α angle>50.5°	NR in study MRA (ALT): 83 (79 to 88)/57 (44 to 70)[86]	▶ FADDIR test
Beaule et al[81]	30 participants/36 hips (mean age 40.7 years, 36 hips, 13 females) with painful, non-dysplastic hips. >3 months	ALT	MRA 1.5 T	NR in study MRA (ALT): 83 (79 to 88)/57 (44 to 70)[86]	▶ FADDIR test
Chan et al[74]	30 participants (mean age 41 years, 13 females) with unclear intra-articular hip pain. NR	ALT, AVN	MRA 1.5 T; Arthroscopy (n=17)	97 (77 to 100)/25 (3 to 80)*	▶ Flexion-IR test
Domayer et al[67]	49 participants/60 hips (mean age 28±10.2 years, 29 females) with clinical signs and symptoms suggestive of FAI NR	FAI	MRI 1.5 T; cam FAI=α angle >55°	NR in study MRI (ALT): 70 (62 to 77)/82 (69 to 94)(86)	▶ FADDIR test
Hananouchi et al[75]	69 participants/107 hips (83 painful hips, 24 non-painful hips) prospectively examined (mean age 57.2 years, 54 females) with various pathology NR	FAI, dysplasia	MRI 3 T; FAI: crossover sign, α angle >50°, aspherical femoral head, or CEA >40° Dysplasia: CEA <25°	NR in study MRI (0.5 to 3 T) (ALT): 66 (59 to 73)/79 (67 to 91)[86]	▶ FADDIR test
Keeney et al[68]	101 consecutive participants/102 hips (mean age 37.6 years, 71 females) with clinical signs and symptoms suggestive of ALT. 21.6 months	Variable: ALT, chondral defect, synovitis	MRA 1.5 T	NR in study MRA (ALT): 83 (79 to 88)/57 (44 to 70)[86] Arthroscopy NA	▶ FADDIR test
Laude et al[69]	97 subjects/100 hips (mean age 33.4 years, range 16 to 56 years, 47 females) with persistent hip pain NR	ALT	Arthroscopy	NA	▶ FADDIR test
Leunig et al[76]	23 participants (mean age 40±2 years, 14 females) with ARS 3.5 years	Variable: ALT, dysplasia, arthritic changes	Surgery	NA	▶ FADDIR test
Martin et al[77]	105 participants (mean age 42±15 years, 24 females) potential surgical candidates due to high suspicion of intra-articular pathology (from subjective history) 1.9 years	Variable: ALT, FAI, arthritic changes	Intra-articular Injection with < or >50% pain relief	NR in study	▶ FABER test ▶ FADDIR test ▶ Trochanteric tenderness test
Maslowski et al[78]	50 participants (mean age 60.2 years, 30 females) with signs/symptoms suggestive of intra-articular pathology NR	Variable: ALT, FAI, arthritic changes, AVN	Intra-articular injection with ≥80% pain relief	NR in study	▶ FABER test ▶ Scour test ▶ IR with overpressure ▶ Resisted SLR test

Continued

Review

Table 2 Continued

Study	Participants/mean duration of symptoms	Pathology	Reference standard	Reference standard SN/SP (95% CI)	Clinical test investigated
McCarthy and Busconi[3]	59 participants (mean age 37 years, 32 females) with refractory hip pain NR	ALT, synovitis, loose bodies, arthritic changes	Arthroscopy	NA	▶ Thomas test
Myrick and Nissen[70]	100 participants (age range 10–40 years, gender NR) with signs/symptoms suggestive of ALT NR	ALT	Arthroscopy	NA	▶ THIRD test
Narvani et al[5]	18 participants (mean age 30.5±8.5 years, 5 females) with groin pain NR	ALT	MRA 1.0 T	NR in study MRA (0.5 to 3 T) (ALT): 87 (84 to 90)/ 64 (54 to 74)[86]	▶ IR-flexion-axial compression test
Nogier et al[64]	241 participants (mean age 35±10 years, 111 females) with mechanical hip pain All >4 months	Dysplasia, FAI	AP pelvic, false profile and lateral axial radiographs; arthroscan or arthro-MRI prescribed at investigator's discretion	NR in study	▶ Flexion-IR test
Peters et al[71]	94 participants/96 hips (mean age 28 years, 40 females) with signs/symptoms suggestive of FAI NR	FAI, articular cartilage lesions/ arthritic changes	Open surgical dislocation	NA	▶ FADDIR test
Petersilge et al[63]	10 participants (mean age 38.4 years, 5 females) with unexplained intra-articular hip pain NR	ALT, dysplasia, ganglion cyst	Arthroscopy	NA	▶ Flexion-IR test
Sink et al[72]	35 participants (mean age 16 years, 30 females) with signs/symptoms suggestive of FAI 3 months to 3 years	Variable: ALT, FAI, cartilage damage	MRI/MRA, AP pelvic radiograph	NR in study MRI (0.5 to 3 T) (ALT) : 66 (59 to 73)/79 (67 to 91)[86]	▶ FADDIR test
Troelsen et al[79]	18 participants (age range 32 to 56 years, 16 females) with previous PAO NR	Dysplasia, ALT	MRA	MRA (0.5 to 3 T) (ALT): 87 (84 to 90)/64 (54 to 74)[86]	▶ FABER test ▶ FADDIR test
Wang et al[73]	21 participants (mean age 37.1 years, range 17 to 65 years, 12 females) with definite pain around hip joint, predominantly in the groin 12.4 months	ALT, FAI	Cross table and frog leg radiographs: cam FAI= α angle >50°; pincer=coxa profunda, cross-over sign or CEA >40°	NR in study	▶ FADDIR test

*Values calculated by authors of this study.
α, alpha angle; ARS, acetabular rim syndrome; AVN, avascular necrosis; CEA, center edge angle; Dx, diagnosis; FABER, flexion, abduction, external rotation test; FADDIR, flexion-adduction-internal rotation impingement test; IR, internal rotation; LE, lower extremity; MRA, MR angiogram; NA, not applicable; NR, not reported; PAO, periacetabular osteotomy surgery.

Table 3 Summary of articles reporting on the diagnostic accuracy of OSTs for hip impingement/labral tear tests

Test, authors	Participants	SN/SP (95% CI)	+LR/−LR	PPV/NPV	Reference standard	Pre to post-test probability
Anterior impingement—labral tear (FADDIR) test						
Aprato et al[65]	41	99 (88 to 99)/6 (1 to 40)*	1.1/0.22*	82.1 (68 to91)/50 (6 to 95)	MRA (FAI)	Pretest=80% Post-test=82.1%
Barton et al[66]	68	99 (90 to 100)/2 (0 to 16)*	1.0/0.59*	63 (51 to 74)/50 (6 to 95)	MRA (α angle>50.5°) for FAI	Pretest=62.1% Post-test=63%
Beaule et al[81]	30	100/all (+) (FAI) 99 (88 to 100)/25 (3 to 80)* (labral tear)	NA/NA (FAI) 1.3/0.06* (labral tear)	96 (84 to 99)/50 (5 to 95)	MRA (labral tear)	Pretest=93.4% Post-test=95.9%
Domayer et al[67]	49/60 hips	99 (90 to 100)/3 (0 to 24)*	1.0/0.35*	75 (62 to 84)/50 (5 to 95)	MRI (α angle >55°) for FAI	Pretest=73.4% Post-test=74.6%
Hananouchi et al[75]	69/107 hips	56 (37 to 73)/83 (31 to 98)*	3.3/0.53*	97 (75 to 100)	MRI (FAI)	Pretest=92.6% Post-test=96.7%
Keeney et al[68]	101	99 (94 to 100)/2 (0 to 18)*	1.0/0.3*	76 (66 to 84)/50 (5 to 95)	MRA (labral tear)	Pretest=75.3% Post-test=76.1%
Laude et al[69]	97/100 hips	99 (95 to 100)/6 (0 to 40)*	1.1/0.09*	93 (86 to 96)/50 (5 to 95)	Arthroscopy (labral tear)	Pretest=91.7% Post-test=92.6%
Leunig et al[76]	23	97 (79 to 100)/13 (1 to 60)*	1.1/0.23*	84 (64 to 94)/50 (5 to 95)	Surgery (labral tear)	Pretest=80.4% Post-test=84.1%
Martin et al[77]	105	78 (59 to 89)/10 (3 to 29)	0.86/2.3	53 (38 to 67)/25 (7 to 59)	Intra-articular injection (FAI and labral tear)	Pretest=43.8% Post-test=52.5%
Peters et al[71]	94/96 hips	99 (95 to 100)/3 (0 to 25)*	1.0/0.18*	85 (77 to 91)/50 (5 to 95)	Surgery (FAI)	Pretest=84.2% Post-test=85.1%
Sink et al[72]	35	99 (87 to 100)/17 (2 to 69)* (FAI) 97 (77 to 100)/4 (0 to 28)* (labral tear)	1.2/0.09* (FAI) 1.0/0.76* (labral tear)	57 (39 to 73)/50 (5 to 95)	AP radiograph (pincer FAI) MRI (labral tear)	Pretest=55.0% Post-test=56.9%
Troelsen et al[79]	18	58 (36 to 78)/75 (20 to 97) *	2.3/0.56*	96 (68 to 100)/17 (4 to 49)	MRA (labral tear)	Pretest=94.4% Post-test=95.5%
Wang et al[73]	21	96 (73 to 99)/6 (1 to 37)* FAI	1.0/0.7*	61 (41 to 79)/50 (5 to 95)	Radiographs (labral tear)	Pretest=58.7% Post-test=61.4%
Bilateral LE squat (maximum depth)						
Ayeni et al[80]	76	75 (58 to 87)/41 (28 to 56)*	1.3/0.61*	47.1 (34.1, 60.5)/70.4 (51.5, 84.1)	MRI or MRA (FAI)	Pretest=41% Post-test=47.1%
FABER test						
Maslowski et al[78]	50	52.5 (37.5 to 67.1)/54.5 (28 to 78.7)	1.1/0.72	80.8 (62.1, 91.5)/24 (11.5, 43.4)	Intra-articular injection (FAI and labral tear)	Pretest=78.0% Post-test=80.8%
Martin et al[77]	105	60 (41 to 77)/18 (7 to 39)	0.73/2.2	45.5 (29.8, 62.0)/28.6 (11.7, 54.6)	Intra-articular injection (FAI and labral tear)	Pretest=31.9% Post-test=45.5%
Troelsen et al[79]	18	42 (22 to 64)/75 (20 to 97)*	1.7/0.78*	93.8 (59.8, 99.3)/12.5 (2.9, 40.2)	MRA (labral tear)	Pretest=94.4% Post-test=93.8%
Flexion-internal rotation test						
Chan et al[74]	30	98 (84 to 100)/8 (1 to 48)*	1.1/0.23*	82.3 (65.6, 91.9)	MRA (labral tear)	Pretest=79.7% Post-test=82.3%
Chan et al[74]	17	97 (77 to 100)/25 (3 to 80)*	1.3/0.12*	91.7 (70.6, 98.1)/50 (5.5, 95.5)	Arthroscopy (labral tear)	Pretest=86.8% Post-test=91.7%
Nogier et al[64]	241	70 (62 to 77)/44 (34 to 54)	1.3/0.69	67.9 (60.3, 74.8)/46.4 (36.2, 57.0)	Radiographs (FAI)	Pretest=44.2% Post-test=67.9%
Petersilge et al[63]	10	95 (66 to 100)/25 (3 to 80)	1.3/0.20	86.4 (57.1, 96.8)/50 (5.5, 94.5)	Arthroscopy (labral tear)	Pretest=79.2 Post-test=86.4%

Continued

Review

Table 3 Continued

Test, authors	Participants	SN/SP (95% CI)	+LR/−LR	PPV/NPV	Reference standard	Pre to post-test probability
Internal rotation-flexion-axial compression test						
Narvani et al[5]	18	75 (19 to 99)/43 (18 to 72)*	1.3/0.58*	27.3 (9.7, 56.6)/85.7 (48.7, 97.4)	MRA (labral tear)	Pretest=16.7% Post-test=27.3%
Internal rotation with overpressure						
Maslowski et al[78]	50	91 (68 to 99)/18 (5 to 40)	1.1/0.5	45.5 (31.7, 59.9)/71.4 (35.9, 91.8)	Intra-articular injection (FAI and labral tear)	Pretest=39.2% Post-test=45.5%
Resisted straight leg raise test						
Maslowski et al[78]	50	58 (34 to 79)/29 (11 to 54)	0.87/1.28	40.6 (25.5, 57.7)/50 (29, 71)	Intra-articular injection (FAI and labral tear)	Pretest=44% Post-test=40.6%
Scour test						
Maslowski et al[78]	50	50 (26 to 74)/29 (12 to 51)	0.70/1.72	35.5 (21.1, 53.1)/42.1 (23.1, 63.7)	Intra-articular injection (FAI and labral tear)	Pretest=22% Post-test=35.5%
THIRD test						
Myrick and Nissen[70]	100	98 (93 to 100)/75 (19 to 99)	3.9/0.03*	98.9 (94.3, 99.8)/60 (23.1, 88.2)	Arthroscopy (labral tear)	Pretest=94% Post-test=98.9%
Thomas test						
McCarthy and Busconi[3]	59	89 (73 to 97)/92 (75 to 99)*	11.1/0.12*	94 (80 to 99)/86 (67 to 96)*	Arthroscopy (labral tear)	
Trochanteric tenderness (not tender)						
Martin et al[77]	105	57 (39 to 74)/45 (27 to 65)	1.1/0.93	55.6 (37.7, 72.4)/47.6 (28.3, 67.6)	Intra-articular injection (FAI and labral tear)	Pretest=54.2% Post-test=55.6%

*Calculations performed by authors of this study.
ARS, acetabular rim syndrome; F, female; FABER, flexion, abduction, external rotation test; FADDIR, flexion-adduction-internal rotation impingement test, LE, lower extremity; +LR, positive likelihood ratio; −LR, negative likelihood ratio; MRA, MR angiogram; NA, not applicable; NPV, negative predictive value; NR, not reported; OST, orthopaedic special tests; PPV, positive predictive value; SN, sensitivity (%); SP, specificity (%);VAS, visual analogue scale.

Table 4 Pooled diagnostic properties for the diagnosis of FAI/ALT

Measure	# studies (sample size)	SN (95% CI)	SP (95% CI)	+LR (95% CI)	ue−LR (95% CI)	DOR (95% CI)	PPV (95% CI)	Post-test change
FADDIR (MRA)	4 (65, 68, 79, 81) (n=188)	0.94 (0.90 to 0.97)	0.09 (0.02 to 0.23)	1.02 (0.96 to 1.08)	0.45 (0.19 to 1.09)	5.71 (0.84 to 38.86)	0.833 (0.771 to 0.885)	Pretest=84% Post-test=83%
FADDIR (surgery)	4 (68, 69, 71, 76) (n=319)	0.99 (0.98 to 1.00)	0.05 (0.01 to 0.18)	1.04 (0.97 to 1.12)	0.14 (0.02 to 0.93)	7.82 (1.06 to 57.84)	0.90 (0.89 to 0.90)	Pretest=90% Post-test= 90%
Flexion IR	2 (63, 74) (n=27)	0.96 (0.81 to 0.99)	0.25 (0.01 to 0.81)	1.28 (0.72 to 2.27)	0.15 (0.01 to 1.99)	8.36 (0.41 to 171.3)	0.90 (0.73 to 0.98)	Pretest=87% Post-test=90%

DerSimoninian-Laird random-effects models used throughout.
DOR, diagnostic OR; FADDIR, flexion, adduction, internal rotation test; +LR, positive likelihood ratio; −LR, negative likelihood ratio; SN, sensitivity; SP, specificity.
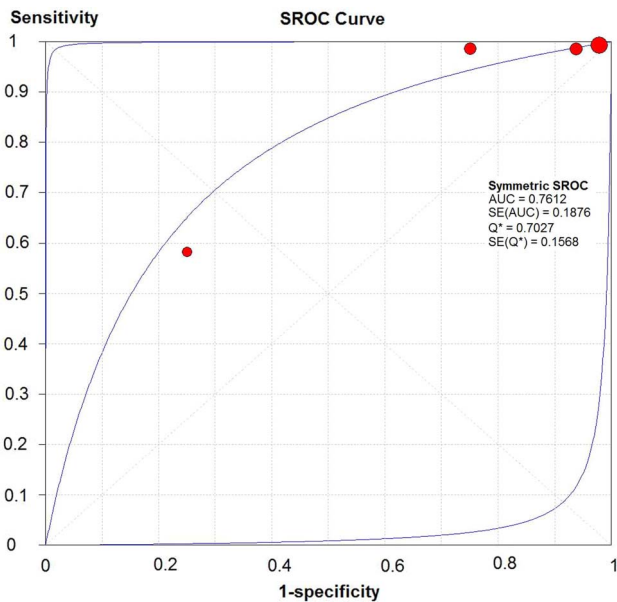
Figure 2 Summary receiver operating characteristic curve for FADDIR test (MRA reference standard).

is not known whether the condition of interest is present or absent.[55 82] Participants in all 21 studies of this review were of high suspicion for various types of intra-articular joint pathology, most that were of high suspicion for FAI/ALT due to groin pain and/or other subjective symptoms (eg, clicking, catching) that are highly suggestive of these pathologies.[3 5 68] The pretest probability in the investigated tests ranged from 17%[5] to 94%,[79] but was higher than 55% in all but five studies[5 64 77 78 80] involving 10 of the 28 tests listed in table 3. Certainly, a high pretest probability can influence the likelihood of reporting a positive finding on the tests and this may be a reason we found high SN in the tests that were eligible for meta-analysis.
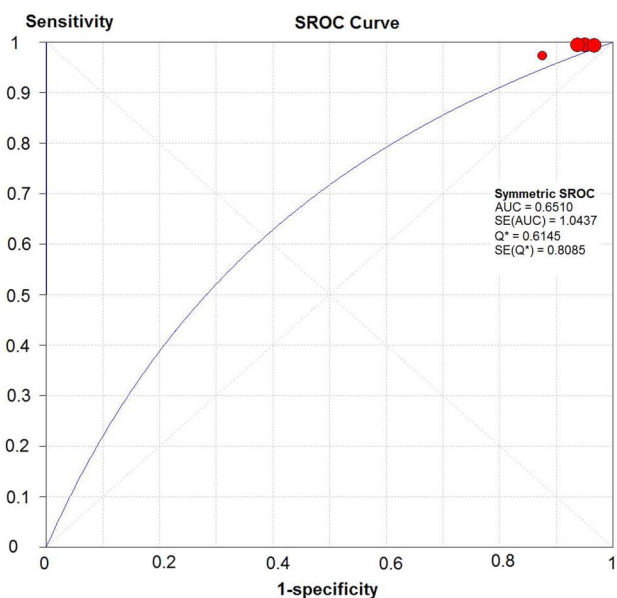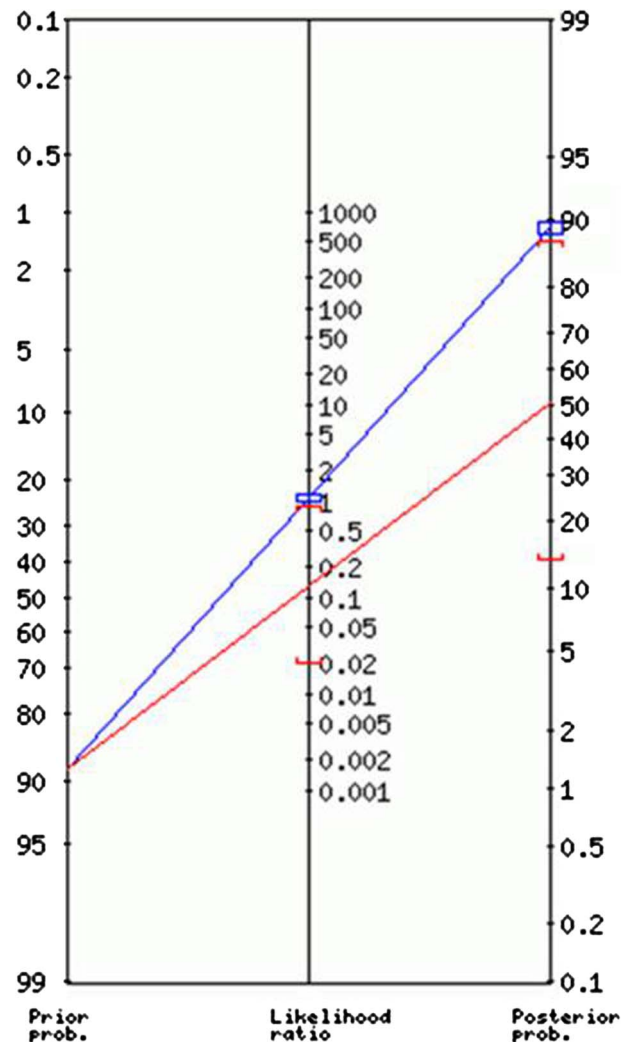

Figure 4 Nomogram demonstrating pretest to post-test probability changes (with CIs) after FADDIR test performance (surgery reference standard). Blue (top) line is post-test probability of diagnosis existing and red (bottom) line is post-test probability of diagnosis not existing.

The single largest increase in post-test probability (in studies of low bias) after use of a particular test was only 6.1% when utilising the painful squat test by Ayeni *et al*[80] Most other tests investigated in this review provided very small percentage increases in post-test probability. In fact, in meta-analysis calculations (table 4), there were minimal to no increases in post-test probabilities, ranging from 1% decline in probability when using the FADDIR test (MRA reference standard), no change using the FADDIR test (surgical reference standard; figure 2), and 3% increase using the Flexion IR test with MRA reference standard. While the post-test probability could change significantly with, for example, the FADDIR test to rule out the pathology existing, the wide range CIs limit the accuracy of interpretation of these findings (figure 4). Additionally, having these test performed on participants of high suspicion of pathology results in scarce true negative findings in these studies.

Ascertaining results of the tests investigated in this review strongly suggests limitations in their clinical utility. The results of these tests appear to minimally change the clinician's treatment strategy, and the participants do not appear to be better off as a result of performing these tests. The current scope of literature investigating these tests is narrow in its focus (examining

Figure 3 Summary receiver operating characteristic curve for FADDIR test (surgery reference standard).

Review

participants with high suspicion of pathology prior to test performance) and, therefore, results in test results of high SN, poor SP and limited post-test probability in determining the diagnosis of FAI/ALT. In fact, a clinician practising in an orthopaedic/ sports clinic seeing participants with high suspicion of FAI/ALT would benefit minimally by performing these tests.

An additional concern regarding diagnostic accuracy studies is the potential for bias. Only one of the included studies in this review had low risk of bias.[80] This study investigated a newly described test (bilateral lower extremity squat to maximum depth). Using a sample size of 76 participants, a SN of 75 and a SP of 41, this test was found to also provide the greatest shift in pretest to post-test probability (6.1%). As mentioned previously, the large majority of these tests had much greater pretest probabilities prior to test implementation than this test. This study utilised the more potentially biased reference standard of MRI and/or MRA.

Although MRI/MRA is rightfully not considered the 'gold standard' for these studies, the use of surgery alone leads to biased participant sampling. Determination of the clinical utility of these tests requires their implementation in future studies across the spectrum of participants with and without undetermined hip pathology. The reference standard for these studies would thus be the more imprecise MRI/MRA. Although the diagnostic accuracy of MRI/MRA currently is limited,[13] limiting image reading to those with specialist training,[83] utilisation of precise, tailored protocols[84] and improving imaging technology could afford suitability for MRI/MRA as acceptable reference standards for these participants. Implementation of these tests across this broad continuum of hip participants is paramount to determine their actual clinical utility.

Limitations

Limitations of this study include: limiting the search strategy to only those articles written in English, one author pulling the data points, lack of comparison of participant inclusion and exclusion across the studies and studies performed in settings of high pretest probability. Limiting to English language could potentially miss some studies to be included in our review. Only one author pulling the data points increases the risk of potential error, although this author (MPR) has pulled many of these data points for a similar review of both intra-articular and extra-articular hip pathology.[85] Since many of the included studies were published prior to prospective guidelines for diagnostic accuracy studies, the robustness of reporting inclusion/exclusion criteria was highly variable among the studies and very difficult to comprehensively describe in this review. Having these studies

performed in hip surgeon clinical settings is a limitation of the current literature that is uncontrollable by the authors of this review.

CONCLUSION

Owing to the low quality and biased sampling of patients with high probability of disease, hip physical examination tests do not appear to currently provide the clinician any significant value in altering probability of disease with their use. Currently, only the FADDIR and Flex-IR tests are supported by the data as valuable screening tests for FAI/LT pathology. Further studies involving high quality designs across a wider spectrum of hip pathology patients are necessary to discern the confirmed clinical utility of these tests.

REFERENCES

1 Ganz R, Parvizi J, Beck M, et al. Femoroacetabular impingement: a cause for osteoarthritis of the hip. Clin Orthop Relat Res 2003;417:112–20.
2 McCarthy JC, Noble PC, Schuck MR, et al. The Otto E. Aufranc Award: the role of labral lesions to development of early degenerative hip disease. Clin Orthop Relat Res 2001;393:25–37.
3 McCarthy JC, Busconi B. The role of hip arthroscopy in the diagnosis and treatment of hip disease. Orthopedics 1995;18:753–6.
4 Lewis CL, Sahrmann SA, Moran DW. Anterior hip joint force increases with hip extension, decreased gluteal force, or decreased iliopsoas force. J Biomech 2007;40:3725–31.
5 Narvani AA, Tsiridis E, Kendall S, et al. A preliminary report on prevalence of acetabular labrum tears in sports patients with groin pain. Knee Surg Sports Traumatol Arthrosc 2003;11:403–8.
6 Montgomery SR, Ngo SS, Hobson T, et al. Trends and demographics in hip arthroscopy in the United States. Arthroscopy 2013;22:661–5.
7 Lee YK, Ha YC, Yoon BH, et al. National trends of hip arthroscopy in Korea. J Korean Med Sci 2014;29:277–80.
8 Colvin AC, Harrast J, Harner C. Trends in hip arthroscopy. J Bone Joint Surg Am 2012;94:e23.
9 Reiman MP, Mather RC III, Hash TW II, et al. Examination of acetabular labral tear: a continued diagnostic challenge. Br J Sports Med 2013;48:311–19.
10 Byrd JW, Jones KS. Diagnostic accuracy of clinical assessment, magnetic resonance imaging, magnetic resonance arthrography, and intra-articular injection in hip arthroscopy patients. Am J Sports Med 2004;32:1668–74.
11 McCarthy JC, Busconi BD, Owens BD. Assessment of the painful hip. In: MCCarthy JC, ed. Early hip disorders. New York: Springer, 2003:7–8.
12 Tibor LM, Sekiya JK. Differential diagnosis of pain around the hip joint. Arthroscopy 2008;24:1407–21.
13 Smith TO, Hilton G, Toms AP, et al. The diagnostic accuracy of acetabular labral tears using magnetic resonance imaging and magnetic resonance arthrography: a meta-analysis. Eur Radiol 2010;21:863–74.
14 Smith TO, Simpson M, Ejindu V, et al. The diagnostic test accuracy of magnetic resonance imaging, magnetic resonance arthrography and computer tomography in the detection of chondral lesions of the hip. Eur J Orthop Surg Traumatol 2013;23:335–44.
15 Ayeni OR, Wong I, Chien T, et al. Surgical indications for arthroscopic management of femoroacetabular impingement. Arthroscopy 2012;28:1170–9.
16 Hetaimish BM, Khan M, Crouch S, et al. Consistency of reported outcomes after arthroscopic management of femoroacetabular impingement. Arthroscopy 2013;29:780–7.
17 Ayeni OR, Belzile EL, Musahl V, et al. Results of the PeRception of femOroaCetabular impingEment by Surgeons Survey (PROCESS). Knee Surg Sports Traumatol Arthrosc 2014;22:906–10.
18 Register B, Pennock AT, Ho CP, et al. Prevalence of abnormal hip findings in asymptomatic participants: a prospective, blinded study. Am J Sports Med 2012;40:2720–4.
19 Schmitz MR, Campbell SE, Fajardo RS, et al. Identification of acetabular labral pathological changes in asymptomatic volunteers using optimized, noncontrast 1.5-T magnetic resonance imaging. Am J Sports Med 2012;40:1337–41.

What are the new findings

► Few hip clinical tests, as investigated in current studies, actually make a significant change in post-test probability for the potential of femoroacetabular impingement/ acetabular labral tear (FAI/ALT) existing.

► These clinical tests, when utilised in clinical settings of high pretest probability, will provide the practising clinician with limited to no assistance in determination of the presence or absence of FAI/ALT.

► Prospective studies examining clinical utility of these tests in patients with/without various hip pathologies are not available, but suggested.

20 Silvis ML, Mosher TJ, Smetana BS, et al. High prevalence of pelvic and hip magnetic resonance imaging findings in asymptomatic collegiate and professional hockey players. Am J Sports Med 2011;39:715–21.

21 Abe I, Harada Y, Oinuma K, et al. Acetabular labrum: abnormal findings at MR imaging in asymptomatic hips. Radiology 2000;216:576–81.

22 Aydingoz U, Ozturk MH. MR imaging of the acetabular labrum: a comparative study of both hips in 180 asymptomatic volunteers. Eur Radiol 2001;11:567–74.

23 Lecouvet FE, Vande Berg BC, Malghem J, et al. MR imaging of the acetabular labrum: variations in 200 asymptomatic hips. AJR Am J Roentgenol 1996;167:1025–8.

24 Cotten A, Boutry N, Demondion X, et al. Acetabular labrum: MRI in asymptomatic volunteers. J Comput Assist Tomogr 1998;22:1–7.

25 Nepple JJ, Brophy RH, Matava MJ, et al. Radiographic findings of femoroacetabular impingement in national football league combine athletes undergoing radiographs for previous hip or groin pain. Arthroscopy 2012;28:1396–403.

26 Kapron AL, Anderson AE, Aoki SK, et al. Radiographic prevalence of femoroacetabular impingement in collegiate football players: AAOS exhibit selection. J Bone Joint Surg Am 2011;93:e111(1–10).

27 Khanna V, Caragianis A, Diprimio G, et al. Incidence of hip pain in a prospective cohort of asymptomatic volunteers: is the cam deformity a risk factor for hip pain? Am J Sports Med 2014;42:793–7.

28 Ergen FB, Vudali S, Sanverdi E, et al. CT assessment of asymptomatic hip joints for the background of femoroacetabular impingement morphology. Diagn Interv Radiol 2014;20:271–6.

29 Chakraverty JK, Sullivan C, Gan C, et al. Cam and pincer femoroacetabular impingement: CT findings of features resembling femoroacetabular impingement in a young population without symptoms. AJR Am J Roentgenol 2013;200:389–95.

30 Joo JH, Lee SC, Ahn HS, et al. Evaluation of the alpha angle in asymptomatic adult hip joints: analysis of 994 hips. Hip Int 2013;23:395–9.

31 Jung KA, Restrepo C, Hellman M, et al. The prevalence of cam-type femoroacetabular deformity in asymptomatic adults. J Bone Joint Surg Br 2011;93:1303–7.

32 Hack K, Di Primio G, Rakhra K, et al. Prevalence of cam-type femoroacetabular impingement morphology in asymptomatic volunteers. J Bone Joint Surg Am 2010;92:2436–44.

33 Bixby SD, Kienle KP, Nasreddine A, et al. Reference values for proximal femoral anatomy in adolescents based on sex, physis, and imaging plane. Am J Sports Med 2013;41:2074–82.

34 Yuan BJ, Bartelt RB, Levy BA, et al. Decreased range of motion is associated with structural hip deformity in asymptomatic adolescent athletes. Am J Sports Med 2013;41:1519–25.

35 Klingenstein GG, Zbeda RM, Bedi A, et al. Prevalence and preoperative demographic and radiographic predictors of bilateral femoroacetabular impingement. Am J Sports Med 2013;41:762–8.

36 Reichenbach S, Leunig M, Werlen S, et al. Association between cam-type deformities and magnetic resonance imaging-detected structural hip damage: a cross-sectional study in young men. Arthritis Rheum 2011;63:4023–30.

37 Reichenbach S, Juni P, Werlen S, et al. Prevalence of cam-type deformity on hip magnetic resonance imaging in young males: a cross-sectional study. Arthritis Care Res 2010;62:1319–27.

38 Sutter R, Dietrich TJ, Zingg PO, et al. How useful is the alpha angle for discriminating between symptomatic patients with cam-type femoroacetabular impingement and asymptomatic volunteers? Radiology 2012;264:514–21.

39 Sutter R, Dietrich TJ, Zingg PO, et al. Femoral antetorsion: comparing asymptomatic volunteers and patients with femoroacetabular impingement. Radiology 2012;263:475–83.

40 Hartofilakidis G, Bardakos NV, Babis GC, et al. An examination of the association between different morphotypes of femoroacetabular impingement in asymptomatic subjects and the development of osteoarthritis of the hip. J Bone Joint Surg Br 2011;93:580–6.

41 Kang AC, Gooding AJ, Coates MH, et al. Computed tomography assessment of hip joints in asymptomatic individuals in relation to femoroacetabular impingement. Am J Sports Med 2010;38:1160–5.

42 Weir A, de Vos RJ, Moen M, et al. Prevalence of radiological signs of femoroacetabular impingement in patients presenting with long-standing adductor-related groin pain. Br J Sports Med 2011;45:6–9.

43 Allen D, Beaule PE, Ramadan O, et al. Prevalence of associated deformities and hip pain in patients with cam-type femoroacetabular impingement. J Bone Joint Surg Br 2009;91:589–94.

44 Schmitz MR, Bittersohl B, Zaps D, et al. Spectrum of radiographic femoroacetabular impingement morphology in adolescents and young adults: an EOS-based double-cohort study. J Bone Joint Surg Am 2013;95:e90.

45 Leibold MR, Huijbregts PA, Jensen R. Concurrent criterion-related validity of physical examination tests for hip labral lesions: a systematic review. J Man Manip Ther 2008;16:E24–41.

46 Burgess RM, Rushton A, Wright C, et al. The validity and accuracy of clinical diagnostic tests used to detect labral pathology of the hip: a systematic review. Man Ther 2011;16:318–26.

47 Tijssen M, van Cingel R, Willemsen L, et al. Diagnostics of femoroacetabular impingement and labral pathology of the hip: a systematic review of the accuracy and validity of physical tests. Arthroscopy 2012;28:860–71.

48 Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8:336–41.

49 Swartz MK. The PRISMA statement: a guideline for systematic reviews and meta-analyses. J Pediatr Health Care 2011;25:1–2.

50 Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. BMC Med 2004;2:23.

51 Wilczynski NL, Haynes RB. Optimal search strategies for detecting clinically sound prognostic studies in EMBASE: an analytic survey. J Am Med Inform Assoc 2005;12:481–5.

52 Doust JA, Pietrzak E, Sanders S, et al. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. J Clin Epidemiol 2005;58:444–9.

53 Akobeng AK. Understanding systematic reviews and meta-analysis. Arch Dis Child 2005;90:845–8.

54 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.

55 Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The evidence-based medicine working group. JAMA 1994;271:703–7.

56 Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56:1129–35.

57 DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.

58 Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000;53:1119–29.

59 Cox D. The analysis of binary data. London: Methuen, 1970.

60 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.

61 Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Stat Med 2002;21:1237–56.

62 Zamora J, Abraira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006;6:31.

63 Petersilge CA, Haque MA, Petersilge WJ, et al. Acetabular labral tears: evaluation with MR arthrography. Radiology 1996;200:231–5.

64 Nogier A, Bonin N, May O, et al. Descriptive epidemiology of mechanical hip pathology in adults under 50 years of age. Prospective series of 292 cases: clinical and radiological aspects and physiopathological review. Orthop Traumatol Surg Res 2010;96(8 Suppl):S53–8.

65 Aprato A, Masse A, Faletti C, et al. Magnetic resonance arthrography for femoroacetabular impingement surgery: is it reliable? J Orthop Traumatol 2013;14:201–6.

66 Barton C, Salineros MJ, Rakhra KS, et al. Validity of the alpha angle measurement on plain radiographs in the evaluation of cam-type femoroacetabular impingement. Clin Orthop Relat Res 2011;469:464–9.

67 Domayer SE, Ziebarth K, Chan J, et al. Femoroacetabular cam-type impingement: diagnostic sensitivity and specificity of radiographic views compared to radial MRI. Eur J Radiol 2011;80:805–10.

68 Keeney JA, Peelle MW, Jackson J, et al. Magnetic resonance arthrography versus arthroscopy in the evaluation of articular hip pathology. Clin Orthop Relat Res 2004;163–9.

69 Laude F, Sariali E, Nogier A. Femoroacetabular impingement treatment using arthroscopy and anterior approach. Clin Orthop Relat Res 2009;467:747–52.

70 Myrick KM, Nissen CW. THIRD test: diagnosing hip labral tears with a new physical examination technique. J Nurse Pract 2013;9:501–5.

71 Peters CL, Schabel K, Anderson L, et al. Open treatment of femoroacetabular impingement is associated with clinical improvement and low complication rate at short-term followup. Clin Orthop Relat Res 2010;468:504–10.

72 Sink EL, Gralla J, Ryba A, et al. Clinical presentation of femoroacetabular impingement in adolescents. J Pediatr Orthop 2008;28:806–11.

73 Wang WG, Yue DB, Zhang NF, et al. Clinical diagnosis and arthroscopic treatment of acetabular labral tears. Orthop Surg 2011;3:28–34.

74 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.

75 Hananouchi T, Yasui Y, Yamamoto K, et al. Anterior impingement test for labral lesions has high positive predictive value. Clin Orthop Relat Res 2012;470:3524–9.

76 Leunig M, Werlen S, Ungersbock A, et al. Evaluation of the acetabular labrum by MR arthrography. J Bone Joint Surg Br 1997;79:230–4.

Review

77 Martin RL, Irrgang JJ, Sekiya JK. The diagnostic accuracy of a clinical examination in determining intra-articular hip pain for potential hip arthroscopy candidates. *Arthroscopy* 2008;24:1013–18.

78 Maslowski E, Sullivan W, Forster Harwood J, *et al.* The diagnostic validity of hip provocation maneuvers to detect intra-articular hip pathology. *PM R* 2010;2:174–81.

79 Troelsen A, Mechlenburg I, Gelineck J, *et al.* What is the role of clinical tests and ultrasound in acetabular labral tear diagnostics? *Acta Orthop* 2009;80:314–18.

80 Ayeni O, Chu R, Hetaimish B, *et al.* A painful squat test provides limited diagnostic utility in CAM-type femoroacetabular impingement. *Knee Surg Sports Traumatol Arthrosc* 2014;22:806–11.

81 Beaule PE, Zaragoza E, Motamedi K, *et al.* Three-dimensional computed tomography of the hip in the assessment of femoroacetabular impingement. *J Orthop Res* 2005;23:1286–92.

82 Guyatt G, Rennie D, Meade MO, *et al. Users' guides to the medical literature: a manual for evidence-based clinical practice*. 2nd edn. New York, NY: McGraw-Hill Education, 2008.

83 McGuire CM, MacMahon P, Byrne DP, *et al.* Diagnostic accuracy of magnetic resonance imaging and magnetic resonance arthrography of the hip is dependent on specialist training of the radiologist. *Skeletal Radiol* 2012;41:659–65.

84 Llopis E, Fernandez E, Cerezal L. MR and CT arthrography of the hip. *Semin Musculoskelet Radiol* 2012;16:42–56.

85 Reiman MP, Goode AP, Hegedus EJ, *et al.* Diagnostic accuracy of clinical tests of the hip: a systematic review with meta-analysis. *Br J Sports Med* 2013;47:893–902.

86 Smith TO, Hilton G, Toms AP, *et al.* The diagnostic accuracy of acetabular labral tears using magnetic resonance imaging and magnetic resonance arthrography: a meta-analysis. *Eur Radiol* 2011;21:863–74.

BJSM

Diagnostic accuracy of clinical tests for the diagnosis of hip femoroacetabular impingement/labral tear: a systematic review with meta-analysis

M P Reiman, A P Goode, C E Cook, P Hölmich and K Thorborg

Updated information and services can be found at:

http://bjsm.bmj.com/content/49/12/811

These include:

Supplementary Material	Supplementary material can be found at: **http://bjsm.bmj.com/content/suppl/2014/12/16/bjsports-2014-094302.DC1.html**
References	This article cites 82 articles, 17 of which you can access for free at: **http://bjsm.bmj.com/content/49/12/811#BIBL**
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections	Articles on similar topics can be found in the following collections BJSM Reviews with MCQs (154) Labral tears (12)

Notes