Diagnostic accuracy of clinical tests of the hip: a systematic review with meta-analysis

Michael P Reiman,[1] Adam P Goode,[1] Eric J Hegedus,[2] Chad E Cook,[3] Alexis A Wright[2]

► Additional materials are published online only. To view these files please visit the journal online (http://bjsm.bmj.com/content/early/recent)

[1]Community and Family Practice, Duke University School of Medicine, Durham, North Carolina, USA
[2]Physical Therapy, High Point University, High Point, North Carolina, USA
[3]Physical Therapy, Walsh University, North Canton, Ohio, USA

Correspondence to
Michael P Reiman, Duke University School of Medicine, Community and Family Practice, 2200 W. Main, Durham, North Carolina 27705, USA; michael.reiman@duke.edu

ABSTRACT

Background Hip Physical Examination (HPE) tests have long been used to diagnose a myriad of intra-and extra-articular pathologies of the hip joint. Useful clinical utility is necessary to support diagnostic imaging and subsequent surgical decision making.

Objective Summarise and evaluate the current research and utility on the diagnostic accuracy of HPE tests for the hip joint germane to sports related injuries and pathology.

Methods A computer-assisted literature search of MEDLINE, CINHAL and EMBASE databases (January 1966 to January 2012) using keywords related to diagnostic accuracy of the hip joint. This systematic review with meta-analysis utilised the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) guidelines for the search and reporting phases of the study. Der-Simonian and Laird random effects models were used to summarise sensitivities (SN), specificities (SP), likelihood ratios and diagnostic OR.

Results The employed search strategy revealed 25 potential articles, with 10 demonstrating high quality. Fourteen articles qualified for meta-analysis. The meta-analysis demonstrated that most tests possess weak diagnostic properties with the exception of the patellar-pubic percussion test, which had excellent pooled SN 95 (95% CI 92 to 97%) and good specificity 86 (95% CI 78 to 92%).

Conclusion Several studies have investigated pathology in the hip. Few of the current studies are of substantial quality to dictate clinical decision-making. Currently, only the patellar-pubic percussion test is supported by the data as a stand-alone HPE test. Further studies involving high quality designs are needed to fully assess the value of HPE tests for patients with intra- and extra-articular hip dysfunction.

INTRODUCTION

Sports related hip injuries are common among athletes of all ages who participate in sports that are associated with trauma or repetitive strains at the hip joint.[1 2] Hip specific injuries may include sports hernia,[3] labral tears,[4] pathological fractures,[4] avascular necrosis[4] and trochanteric pain syndrome.[5] Associated disorders that the athlete may present for medical assessment/evaluation are and femoroacetabular impingement syndrome[6] and osteoarthritis.[7]

With the evolution of improved diagnostic imaging and advanced surgical techniques, examination of the coxofemoral (hip) joint and periarticular structures as a primary pain source for hip related pain/dysfunction has received a significant increase in attention.[8] Despite the increased attention, differential diagnosis of the hip joint continues to pose a diagnostic dilemma, particularly given pain in the hip region is often difficult to localise to a specific pathological structure.[9 10] Although limited information exists in support of diagnostic utility, emphasis on patient history, clinical examination findings, MRI, arthrogram and anaesthetic intra-articular injection pain response is currently advocated for determining the presence of intra-articular hip joint pathology.[11] Combined use of these examination processes by healthcare practitioners has been only marginally effective. Authors have reported that patients visit, on average, 3.3 healthcare providers before being correctly diagnosed with a hip labral tear over a period of 21 months.[12]

Further complicating the diagnostic challenge for hip joint pathologies is the complex regional anatomy and biomechanics of the hip joint.[13] The concept of overlapping and multifarious referral pain patterns, along with the recognition of the mechanical relationship between the hip and spine,[14–21] has led to various differential diagnostic pathways.[9 22–25]

In an attempt to improve the level of consensus among specialists treating hip pathology, a common language and protocol, although specific to labral pathology,[9] has been described, including commonly used hip physical examination (HPE) tests.[26] Lacking in this detailed description of the hip clinical examination is the diagnostic accuracy and subsequent quality of such tests. To our knowledge, a comprehensive review of the diagnostic accuracy of HPE tests does not exist, specifically for sports related hip injuries. Consequently, the purpose of this study was to conduct a systematic review and meta-analysis of the literature regarding the diagnostic accuracy of HPE tests relevant to patients injured in sport related activities and with diagnoses affiliated with intra- and extra-articular pathology of the hip with appropriate criterion reference standards in cohort, case control and/or cross sectional design studies.

METHODS
Study design

The *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) guidelines were utilised during the search and reporting phase of this systematic review/meta-analysis. The PRISMA statement includes a 27-item checklist that is designed to be used as a basis for reporting systematic review of randomised trials.[27] The PRISMA checklist and flow diagram were created to be used prospectively during the creation of

systematic reviews and meta-analyses and was used as such in this systematic review.

Search strategy
Identification and selection of the literature
A systematic, computerised search of the literature in MEDLINE, CINAHL and EMBASE databases (the search strategy is shown in Appendix I) was concluded in January 2012. 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 hand searched by one of the authors (MPR) and included publications, posters, abstracts, or conference proceedings. To identify relevant articles, titles and abstracts of all identified citations were independently screened by 2 reviewers (AAW and EJH). Full-text articles were retrieved if the abstract provided insufficient information to establish eligibility or if the article had passed the first eligibility screening.

Selection criteria
All articles examining HPE tests specific to the hip were eligible for this study. A HPE test was operationally defined as a stand-alone, clinical test (eg, special test) representative of a pathological condition of the hip joint. An article was further eligible if it met all of the following criteria: 1) patients presented with hip or groin pain, 2) a cohort, case control, and/or cross sectional design was used, 3) inclusion of at least one clinical examination test used to evaluate intra-or extra-articular pathology, 4) was compared against an acceptable criterion reference, 5) and reporting of diagnostic accuracy of the measures (eg, sensitivity (SN) and specificity (SP), or was present and 6) the article was in English.

An article was excluded if 1) the pathology was associated with a condition that was isolated elsewhere (eg, lumbar spine) but referred pain to the hip, 2) the studies omitted values of either SN or SP, 3) if the clinical examination test was performed under any form of anaesthesia or in cadavers, 4) if those studies that used instrumentation that was not readily available to all clinicians, 5) only individual physical clinical tests were included and 6) if studies were performed on infants/toddlers.

All criteria were independently applied by 2 reviewers (AAW and EJH) to the full text of the articles that passed the first eligibility screening. In case of disagreement, a third author (MPR) was consulted to discuss and solve the disagreement.

Quality assessment
The original Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS 1) was used to analyse the quality of the study (Appendix II). QUADAS 1 consists of 14 items with each having a 'yes/no/unclear' answer option. A 'yes' score indicated sufficient information, with bias considered unlikely. A 'no' score indicated sufficient information, but with potential bias from inadequate design or conduct. An 'unclear' score indicated that insufficient information was provided in the article or the methodology was unclear. The total score was the count of all of the criteria that scored 'yes', which was valued as '1' whereas, 'no' and 'unclear' scores carried a zero score value. The maximum attainable score on the criteria list was 14. The methodological quality of each of the studies was independently assessed by two additional reviewers (MPR and CEC). Disagreements among the reviewers were discussed and resolved with consensus. Inter-rater reliability

was configured with weighted κ. Qualitatively, studies that exhibit higher QUADAS values are associated with less risk of design bias than those of lower values. We stratified studies as high quality/low risk of bias if the QUADAS score was 10 or greater, and low quality/high risk of bias of the study score less than 10 on the QUADAS since this dichotomising stratification level had been utilised previously.[28]

A second iteration of QUADAS (QUADAS 2) was introduced in 2011.[29] The QUADAS 2 involves qualitative scoring and at present has yielded poor agreement among tool users. The tool primarily is used to determine whether there are risks of bias within a study and whether the applicability of the study is appropriate. We opted not to use the QUADAS 2 since we could not yield a total score and since the interrater reliability of the tool appears to be questionable.

Data abstraction
One reviewer (MPR) independently extracted information and data regarding study population, setting, HPE test performance, hip pathology, diagnostic reference-standard, number of true positives, false positives, false negatives and true negatives for meta-analysis. Sensitivity, SP, positive likelihood ratio (LR+), and negative likelihood ratio (LR-) were also calculated and/or reported to determine clinical utility of the HPE tests. Sensitivity is defined as the percentage of people who test positive for a specific disease among a group of people who have the disease. Specificity is the percentage of people who test negative for a specific disease among a group of people who do not have the diagnosis/disorder. A 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 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. Post-test probability can be altered to a minimal degree (LR+'s of 1 to 2, or LR-'s of .5 to 1), to a small degree (LR+'s of 2 to 5 and LR-'s of .2 to .5), to a moderated degree (LR+'s of 5 to 10, LR-'s of .1 to .2) and to a significant and almost conclusive degree (LR+'s greater than 10, LR-'s less than 0.1).[30]

Meta-Analysis
Studies were explored for statistical pooling where ≥ 2 studies examined the same index test and diagnosis with the same reference standard. Der-Simionian and Laird[31] random effects models, which incorporate both between and within study heterogeneity into summary estimates, were used to produce summary estimates of SN, SP, LR+, LR-, and diagnostic OR (DOR) for those studies with the same reference standard. When appropriate (ie, >2 homogenous studies), the joint distribution of SN and SP were analysed with the Moses-Shapiro-Littenberg linear model methods to draw sROC and calculate the area under the curve (AUC) and Q*as measures of test accuracy.[32] Heterogeneity was tested with χ^2 tests and Cochrane-Q to estimate between-study heterogeneity, of SN and SP and likelihood ratios respectfully, with values of p<0.10 indicating significant heterogeneity. Publication bias was not formally tested due to limitations of the tests with less than 10 studies.[33] No significant threshold effects were found using Spearman correlation coefficients. Cell counts of zero are common in diagnostic accuracy studies and when cell counts of

zero were encountered, 0.5 was added to all four cells as suggested by Cox.[34] All analyses were conducted by one of the authors (APG) in Meta-DiSc version 1.4.[35]

RESULTS

Selection of studies

A computerised search, along with reference checking, yielded a total of 25 studies for inclusion in the review (figure 1). A total of 127 out of 152 articles were excluded for not reporting of both SN and SP. Multiple studies were performed on cohorts of subjects with defined or suspected pathology, limiting reporting of SP. Out of the 25 total studies investigated, 20 studies examined intra-articular and/or fracture pathologies (two studies for hip osteoarthritis, 12 studies for impingement/labral tear/intra-articular pathology, five studies for

fracture of the hip or femur and one study for avascular necrosis), while five studies investigated extra-articular pathologies (three studies for gluteal tendinopathy, and only one each for the diagnoses of sports related chronic groin pain and leg complaints in endurance athletes due to vascular causes.

Quality scores

The weighted κ between testers for the overall score using QUADAS was 0.68 (95% CI 0.31 to 0.73). For the individual items of the QUADAS, items 2, 5, 6, 7, 8, 9, 11, and 14 had 100% agreement, items 3 and 12 had 93% agreement, items 1 and 4 had 90% agreement, item 10 had 87% agreement, and item 13 had 78% agreement between raters. κ values in the range of 0.41 to 0.60, 0.61 to 0.80, and 0.81 to 1.00 are labelled as strength of agreement as 'moderate', 'substantial', and 'almost perfect' respectively.[36] Using our arbitrary stratification of the QUADAS, the assessment of the 25 articles retained for this review indicated that 10 articles were of high quality/low risk of bias and the remaining 15 articles had QUADAS scores below 10 out of 14 points, yielding low quality/high risk of bias (tables 1–7).

RESULTS OF INDIVIDUAL DIAGNOSTIC CLINICAL TESTS

The definition of each HPE test was variable among the studies. Therefore, in order to allow the clinician and researcher the ability to compare HPE diagnostic values, each HPE was grouped according to how it was performed. All similarly performed HPE tests were then grouped and compared statistically when appropriate. Additionally, the reliability and diagnostic accuracy of each test is listed to allow the reader to discern their clinical applicability.

Intra-articular pathology and/or fracture

Hip osteoarthritis

Two studies met inclusion criteria for the diagnosis of hip osteoarthritis (table 1).[37 38] Trendelenburg's test, resisted hip abduction,[37] and FABER's test[38] were investigated and were considered high quality studies. Of the three tests the resisted hip abduction test yielded small post-test probability influence (LR+ 3.5), the highest of the three.

Impingement/labral/intra-articular pathology

Twelve studies met inclusion criteria for diagnosis of impingement/labral/intra-articular pathology (table 2).[11 39–49] The FABER test was utilised in three studies.[11 39 40] The SN values
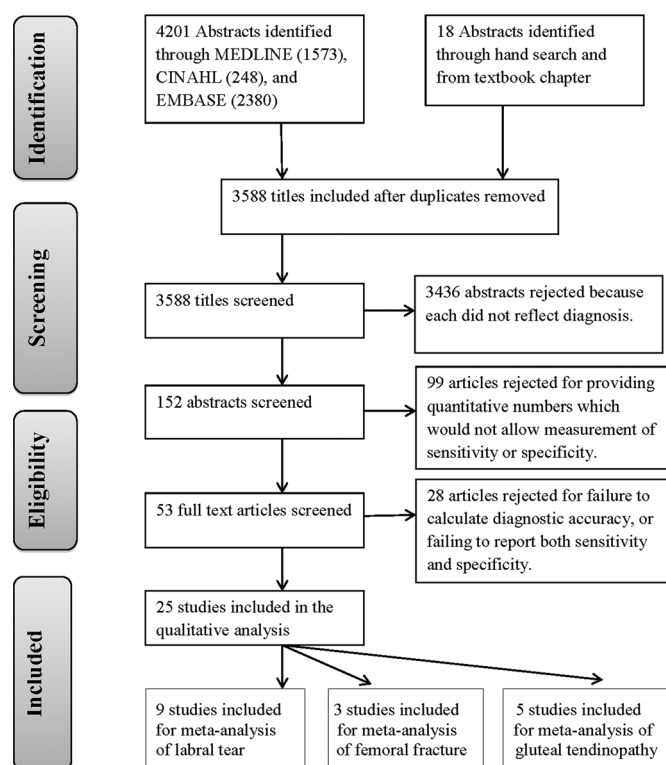

Figure 1 Diagram of study flow.

Identification
- 4201 Abstracts identified through MEDLINE (1573), CINAHL (248), and EMBASE (2380)
- 18 Abstracts identified through hand search and from textbook chapter

Screening
- 3588 titles included after duplicates removed
- 3588 titles screened → 3436 abstracts rejected because each did not reflect diagnosis.

Eligibility
- 152 abstracts screened → 99 articles rejected for providing quantitative numbers which would not allow measurement of sensitivity or specificity.
- 53 full text articles screened → 28 articles rejected for failure to calculate diagnostic accuracy, or failing to report both sensitivity and specificity.

Included
- 25 studies included in the qualitative analysis
- 9 studies included for meta-analysis of labral tear
- 3 studies included for meta-analysis of femoral fracture
- 5 studies included for meta-analysis of gluteal tendinopathy

Table 1 Summary of articles reporting on the diagnostic accuracy of OSTs for pathologies of the hip: hip osteoarthritis

Test, authors	Subjects	Age (mean, SD)	Gender	Pathology	Symptom duration	SN/SP (95% CI)	LR+/LR-	Q	Criterion standard SN/SP (95% CI)	Reliability
Trendelenburg's sign										
Youdas et al[37]	40 subjects	50.4, 7.2 years (controls); and 53.4, 9.0 years; (pathology)	10 F in each group	Radiographic evidence for OA	NR	55 (NR)/70 (NR)	1.83/0.82	10	Radiograph NR	0.63 ICC intra-rater
Resisted Hip Abduction										
Youdas et al[37]	40 subjects	50.4, 7.2 years (controls); and 53.4, 9.0 years; (pathology)	10 F in each group	Radiographic evidence for OA	NR	35 (NR)/90 (NR)	3.5/0.72	10	Radiograph NR	0.97 ICC intra-rater
FABER's Test										
Sutlive et al[38]	72 subjects	58.6, 11.2 years (control); 61.1, 12.7 years; (pathology)	40 F (control); 7 F (pathology)	Radiographic evidence of OA	From > 6 months to > 5 years	57 (34 to 77)/71 (56 to 82)	1.9/0.61	12	Radiograph NR	0.47 κ inter-rater

F, female; ICC, Intra-class correlation; IR, internal rotation; LR+, positive likelihood ratio; LR-, negative likelihood ratio; M, male; NR, not reported; OST, orthopaedic special tests; Q, Quality Assessment of Diagnostic Accuracy Studies scores (QUADAS); SN, sensitivity (%); SP, specificity (%); κ, kappa reliability statistic.

Table 2 Summary of articles reporting on the diagnostic accuracy of OSTs for pathologies of the hip: impingement/labral/intra-articular tests

Test, authors	Subjects	Age (mean, SD)	Gender	Pathology	Symptom Duration	SN/SP (95% CI)	LR+/LR-	Q	Criterion standard SN/SP (95% CI)	Reliability
FABER Test – Intra-Articular Pathology										
Maslowski et al[39]	50 subjects	60.2 years	30 F	Variable: labral tear; FAI, arthritic changes, AVN	NR	81 (57 to 96)/25 (9 to 48)	1.1/0.72	7	NR	NR
Martin et al[11]	105 subjects	42, 15 years	24 F	Variable: labral tear, FAI, arthritic changes, dysplasia	1.9 years (mean)	60 (41 to 77)/18 (7 to 39)	0.73/2.2	9	MRA; 87 (84 to 90)/64 (54 to 74)[72]	NR
Troelsen et al[40]	18 subjects	Range 32 to 56 years	16 F	All had previous peri-acetabular osteotomy; all had dysplasia; post-op: labral tear	NR	42 (NR)/75 (NR)	1.7/0.77	9	MRA; 87 (84 to 90)/64 (54 to 74)[72]	NR
Scour Test										
Maslowski et al[39]	50 subjects	60.2 years	30 F	Variable: labral tear; FAI, arthritic changes, AVN	NR	50 (26 to 74)/29 (12 to 51)	0.70/1.72	7	NR	NR
Internal Rotation with Overpressure										
Maslowski et al[39]	50 subjects	60.2 years	30 F	Variable: labral tear; FAI, arthritic changes, AVN	NR	91 (68 to 99)/18 (5 to 40)	1.1/0.5	7	NR	NR
Resisted Straight Leg Raise Test										
Maslowski et al[39]	50 subjects	60.2 years	30 F	Variable: labral tear; FAI, arthritic changes, AVN	NR	59 (34 to 82)/32 (14 to 55)	0.87/1.28	7	NR	NR
Impingement (FADDIR) Test – Labral Tear/Intra-Articular Pathology										
Beaule et al[41]	30 subjects	40.7 years	13 F	FAI, labral tear	> 3 months	100/all (+) (FAI); 99 (NR)/25 (NR) (labral tear)	NA/NA (FAI); 1.3/0.04 (labral tear)	9	MRA, 87 (84 to 90)/64 (54 to 74)[72]; CT scan, 92 to 97/87 to 100 [78, 79]	NR
Keeney et al[43]	101 subjects	37.6 years	71 F	Variable: Labral tear; chondral defect; synovitis	21.6 months	99 (NR)/5 (NR)	1.0/0.2	8	MRA, 87 (84 to 90)/64 (54 to 74)[72]; Arthroscopy: NA	NR
Leunig et al[42]	23 subjects	40.2 years	14 F	Variable: Labral tear; dysplasia; arthritic changes	3.5 years (mean)	97 (NR)/13 (NR)	1.1/0.23	7	Surgery: NA	NR
Martin et al[11]	105 subjects	42.15 years	24 F	Variable: labral tear, FAI, arthritic changes, dysplasia	1.9 years (mean)	78 (59 to 89)/10 (3 to 29)	0.86/2.2	9	MRA, 87 (84 to 90)/64 (54 to 74)[72]	NR
Sink et al[44]	35 subjects	16 years	30 F	Variable: FAI, labral tear; cartilage damage	3 months to 3 years	100/all (+) (FAI); 97 (NR)/4 (NR) (labral tear)	NA/NA (FAI) 1.0/0.75 (labral tear)	10	MRI, 66 (59 to 73)/79 (67 to 91); MRA, 87 (84 to 90)/64 (54 to 74)[72]	NR
Troelsen et al[40]	18 subjects	Range 32–56 years	16 F	All had previous peri-acetabular osteotomy; all had dysplasia; post-op: labral tear	NR	59 (NR)/75 (NR)	2.4/0.55	9	MRA, 87 (84 to 90)/64 (54 to 74)[72]	NR
Impingement Provocation Test (Postero-Inferior Labrum Tear)										
Leunig et al[42]	23 subjects	40, 2 years	14 F	Variable: Labral tear; dysplasia; arthritic changes	3.5 years (mean)	97 (NR)/11 (NR)	1.1/0.27	7	Surgery: NA	NR
Flexion-Internal Rotation Test										
Chan et al[45]	30 subjects	41 years	13 F	Variable: Labral tear; AVN	NR	98 (NR)/8 (NR)	1.1/0.25	11	MRA 87 (84 to 90)/64 (54 to 74)[72]	NR
Chan et al[45]	17 subjects	Subjects were a subset of above subjects (n=30)				97 (NR)/25 (NR)	1.3/0.12	11	Arthroscopy; NA	NR
Hase and Ueo[46]	10 subjects	28.7 years	7 F	Labral tear	2.9 years (mean)	94 (NR)/13 (NR)	1.1/0.46	8	Arthroscopy; NA	NR
Petersilge et al[47]	10 subjects	38.4 years	5 F	Variable: labral tear, dysplasia, ganglion cyst	NR	94 (NR)/17 (NR)	1.1/0.35	9	Arthroscopy; NA	NR
Internal Rotation-Flexion-Axial Compression Test										
Narvani et al[48]	18 subjects	30.5, 8.5 years	5 F	Labral tear	NR	75 (19 to 99)/43 (18 to 72)	1.3/0.58	8	MRA 87 (84 to 90)/64 (54 to 74)[72]	NR
Thomas Test										
McCarthy and Busconi[49]	59 subjects	37 years	32 F	Variable: labral tear, synovitis, loose bodies, arthritic changes	NR	89 (NR)/92 (NR)	11.1/0.12	10	Arthroscopy: NA	NR

ARS, acetabular rim syndrome; AVN, avascular necrosis; F, female; OST, orthopaedic special tests; FABER, flexion, abduction, external rotation test; FADDIR, flexion-adduction-internal rotation impingement test; LR+, positive likelihood ratio; LR-, negative likelihood ratio; MRA, magnetic resonance angiogram; NA, not applicable; NR, not reported; Q, Quality Assessment of Diagnostic Accuracy Studies scores (QUADAS); SN, sensitivity (%); SP, specificity (%).

Table 3 Summary of articles reporting on the diagnostic accuracy of OSTs for pathologies of the hip: fracture of hip or femur

Test, authors	Subjects	Age (mean, SD)	Gender	Pathology	Symptom Duration	SN/SP (95% CI)	LR+/LR-	Q	Criterion standard SN/SP (95% CI)	Reliability
Patellar-Pubic Percussion Test										
Adams and Yarnold[50]	41 subjects	NR	NR	Femoral neck, inter-trochanteric, trochanteric and acetabular fracture	NR	94 (NR)/95 (NR)	20.4/0.06	8	Radiograph 90 to 95 (NR)/68 to 100 (NR)[80]	89% inter-rater agreement
Bache and Cross[51]	100 subjects	78.6 years	82 F	Femoral neck fracture	NR	91 (NR)/82 (NR)	5.1/0.11	8	Radiograph 90 to 95 [80] (NR)/68 to 100 (NR)	NR
Tiru et al[52]	290 subjects	72, 6.8 years	236 F	Femoral neck fracture	NR	96 (87 to 99)/86 (49 to 98)	6.7/0.75	10	Radiograph 90 to 95 (NR)/68 to 100 (NR)[80]; Bone scan 91(NR)/100 (NR) [81]; MRI 100 (NR)/100 (NR)[81]; CT; NR	NR
Stress Fracture (Fulcrum) Test										
Johnson et al[53]	7 subjects	19.8 years	4 F	Proximal 1/3 femoral shaft stress fracture	NR	93 (NR)/75 (NR)	3.7/0.09	5	Bone scan 91(NR)/100 (NR) [81]; Radiograph 90 to 95 (NR)/68 to 100 (NR)[80]	NR
Kang et al[54]	6 subjects	Range 19–23 years	6 F	Femoral shaft stress fracture	NR	88 (NR)/13 (NR)	1.0/0.92	7	Radiograph 90 to 95 (NR)/68 to 100 (NR) [80]; Bone scan 91(NR)/100 (NR) [81]; MRI 100 (NR)/100 (NR) [81]	NR

F, female; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported; OST, orthopaedic special tests; Q, Quality Assessment of Diagnostic Accuracy Studies scores; SN, sensitivity (%); SP, specificity (%).

Table 4 Summary of articles reporting on the diagnostic accuracy of OSTs for pathologies of the hip: avascular necrosis

Test, authors	Subjects	Age (mean, SD)	Gender	Pathology	Symptom duration	SN/SP (95% CI)	LR+/LR-	Q	Criterion standard SN/SP (95% CI)	Reliability
Restricted/Painful Movement										
Joe et al[55] (extension < 15 degrees)	176 subjects	NR	NR	HIV infection with AVN	NR	19 (0 to 38)/92 (89 to 95)	2.38/0.88	10	MRI 99 (NR)/99 (NR)[82]	NR
Joe et al[55] (external rotation < 60 degrees)						38 (14 to 61)/73 (68 to 77)	0.48/0.85			
Joe et al[55] (pain with internal rotation)						13 (0 to 29)/86 (83 to 90)	0.93/1.01			

AVN, avascular necrosis; F, female; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported; OST, orthopaedic special tests; Q, Quality Assessment of Diagnostic Accuracy Studies scores (QUADAS); SN, sensitivity (%); SP, specificity (%).

for this test ranged from 42 to 81%, while the SP values ranged from 18 to 75%. Maslowski et al[39] was the only study to investigate the scour, internal rotation with overpressure, and resisted straight leg raise tests. Six studies investigated the FADDIR test.[11 40–44] All of these articles were available for meta-analysis. The SN values for this test ranged from 59 to 100%, and the SP values ranged from 4 to 75%. The various pathologies captured in these studies are described in table 2. Leunig et al[42] was the only study investigating the impingement provocation test. Three studies investigated the flexion-internal rotation test.[45–47] All three studies were available for meta-analysis. The SN values for these tests ranged from 94 to 98%, while the SP values ranged from 8 to 25%. One study comprised of 18 subjects with labral tear and various other pathologies investigated the internal rotation-flexion-axial compression test.[48] One study of 59 subjects with various intra-articular pathologies (labral tear, loose bodies, chondral defect and arthritic changes) investigated the Thomas test.[49] All of the studies in this category were of low quality/high risk of bias with the exception of one high quality/low risk of bias

study.[49] In general, these tests demonstrated greater SN than SP. The Thomas Test[49] demonstrated value as both a screen and diagnostic test (SN 89%; SP 92%) with a LR+ of 11.1, indicating the ability to alter post-test probability significantly.

Fracture of hip or femur
Five studies met inclusion criteria for diagnosis of fracture of the hip or femur (table 3).[50–54] Only one of the five studies was of high quality according to our definition. This study investigated the patellar-pubic percussion test (PPPT)[52] as did two other studies.[50 51] All three studies found that the PPPT moderately influenced post-test probability as a stand-alone test with LR+ values ranging from 5.1 to 20.4, and LR- ranging from 0.06 to 0.75. The remaining two studies investigated the stress fracture (fulcrum) test.[53 54] All three were included in meta-analysis.

Avascular necrosis
Only one study examined the diagnostic accuracy of HPE tests for avascular necrosis of the hip and was limited to one high quality study on 176 subjects infected with HIV (table 4).[55]

Review

Table 5 Summary of articles reporting on the diagnostic accuracy of OSTs for pathologies of the hip: gluteal tendinopathy

Test, authors	Subjects	Age (mean, SD)	Gender	Pathology	Symptom Duration	SN/SP (95% CI)	LR+/LR-	Q	Criterion standard SN/SP (95% CI)	Reliability
Trendelenburg's Sign										
Bird et al[56]	24 subjects	Range 36–75 years	24 F	GMed tear and/or tendinitis, partial tear	NR	73 (NR)/77 (NR)	3.15/0.35	10	MRI SN 93 (NR)[83]	0.68 κ intra-tester
Woodley et al[57]	40 subjects	54.4, 9.5 years	37 F	Variable: GMed/GMin partial tear, tendinosis, bursitis, arthritic changes	Variable: < 1 to > 5 years	23 (5 to 57)/94 (53 to 100)	3.6/0.82	12	MRI SN 93 (NR)[83]	NR
Lequesne et al[58]	17 subjects	68.1, 10.8 years	16 F	GMed/GMin tear and/or tendinitis, bursitis	13, 10.5 months (mean)	97/ (NR)96 (NR)	24.3/0.03	10	MRI SN 93 (NR)[83]; Surgery; NA	NR
Resisted Hip Abduction										
Bird et al[56]	24 subjects	Range 36–75 years	24 F	GMed tear and/or tendinitis, partial tear	NR	73 (NR)/46 (NR)	1.35/0.59	10	MRI SN 93 (NR)[83]	0.63 κ intra-tester
Lequesne et al[58]	17 subjects	68.1, 10.8 years	16 F	GMed/GMin tear and/or tendinitis, bursitis	13, 10.5 months (mean)	71 (NR)/97 (NR)	23.7/0.30	10	MRI SN 93 (NR)[83] Surgery; NA	NR
Resisted Internal Rotation										
Bird et al[56]	24 subjects	Range 36–75 years	24 F	GMed tear and/or tendinitis, partial tear	NR	55 (NR)/69 (NR)	1.77/0.66	10	MRI SN 93 (NR)[83]	0.03 κ intra-tester
Resisted External Derotation Test										
Lequesne et al[58]	17 subjects	68.1, 10.8 years	16 F	GMed/GMin tear and/or tendinitis, bursitis	13, 10.5 months (mean)	88 (NR)/97.3 (NR)	32.6/0.12	10	MRI SN 93 (NR)[83]	NR

F, female; GMed, gluteus medius; GMin, gluteus minimus; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NR, not reported; OST, orthopaedic special tests; Q, Quality Assessment of Diagnostic Accuracy Studies scores (QUADAS); SN, sensitivity (%); SP, specificity (%); κ, kappa reliability statistic.

Table 6 Summary of articles reporting on the diagnostic accuracy of OSTs for pathologies of the hip: sports related chronic groin pain

Test, authors	Subjects	Age (mean, SD)	Gender	Pathology	Symptom duration	SN/SP (95% CI)	LR+/LR-	Q	Criterion standard Criterion Standard SN/SP (95% CI)	Reliability
Single Adductor Test										
Verrall et al[59]	89 Australian Rules football players	NR	89 M	Bone marrow oedema	NR	30 (NR)/91 (NR)	3.3/0.66	7	MRI 78 (NR)/88 (NR) [84]	NR
Squeeze Test										
Verrall et al[59]	89 Australian Rules football players	NR	89 M	Bone marrow oedema	NR	43 (NR)/91 (NR)	4.8/0.63	7	MRI 78 (NR)/88 (NR) [84]	NR
Bilateral Adductor Test										
Verrall et al[59]	89 Australian Rules football players	NR	89 M	Bone marrow oedema	NR	54 (NR)/93 (NR)	7.7/0.49	7	MRI 78 (NR)/88 (NR) [84]	NR

LR+, positive likelihood ratio; LR-, negative likelihood ratio; M, male; NR, not reported; OST, orthopaedic special tests; Q, Quality Assessment of Diagnostic Accuracy Studies scores (QUADAS); SN, sensitivity (%); SP, specificity (%).

The SN (range of 13 to 88%) and SP (range of 34 to 92%) was variable depending on the test, but none of the findings reported LR+ ratio greater than 2.38 (extension < 15 degrees) or a LR- less than 0.35 (exam complex) indicating that at best, there are only small alterations in post-test probability for avascular necrosis (table 4).

Extra-articular pathology

Gluteal tendinopathy

Three studies met inclusion criteria for diagnosis of gluteal tendinopathy (table 5).[56–58] All of the studies were of high quality/low risk of bias by our definition despite presenting a maximum of 40 subjects enrolled. Both passive (SN range of 43 to 53%; SP of 86%) and active hip internal rotation (SN of 31 and SP of 86%) appeared to be more specific than sensitive for assessment of gluteal tendinopathy. The resisted external derotation test[58] demonstrates high SN and SP of 88 and 97.3%, respectively.

Sports related chronic groin pain

Only one study met inclusion criteria for diagnosis of sports related chronic groin pain investigating the single adductor

test, squeeze test, and bilateral adductor test (table 6).[59] The bilateral adductor test was found to be most diagnostic of sports related chronic groin pain with reported SP of 93% with a LR+ of 7.7, whereas the squeeze test and single adductor test reported SP of 91% and LR+ of 4.8 and 3.3 respectively.[59]

Leg complaints in endurance athletes due to vascular causes

Only one study met inclusion criteria for assessment of leg complaints in endurance athletes due to vascular causes (table 7).[60] This study was of low quality/high risk of bias as per our a priori QUADAS criteria. Assessment for a femoral bruit with the hip extended was found to be most diagnostic of leg complaints in these athletes with a SP of 94% and LR+ of 6.0. Assessment of a femoral bruit with the hip flexed, and the SI joint gapping test appear to be valuable as SN tests.

META-ANALYSIS

Table 8 provides diagnostic properties and total sample sizes of the 14 studies included in the meta-analysis. One study[43] investigated the FADDIR test with both MRA and arthroscopy as a reference standard. Two other studies[56 58] investigated both

Table 7 Summary of articles reporting on the diagnostic accuracy of OSTs for pathologies of the hip: Leg Complaints in Endurance Athletes due to Vascular Causes

Test, authors	Subjects	Age (mean, SD)	Gender	Pathology	Symptom Duration	SN/SP (95% CI)	LR+/LR-	Q	Criterion standard SN/SP (95% CI)	Reliability
Femoral Bruit with Extended Hip										
Schep et al[60]	80 subjects	31.4, 7.0 years	14 F	Vascular cause of leg pain	NR	36 (NR)/94 (NR)	6.0/0.68	6	Validated Clinical Classification 88 (NR)/83 (NR)[85]	0.56 κ intertester
Femoral Bruit with Flexed Hip										
Schep et al[60]	80 subjects	31.4, 7.0 years	14 F	Vascular cause of leg pain	NR	76 (NR)/65 (NR)	2.2/0.37	6	Validated Clinical Classification 88 (NR)/83 (NR)[85]	0.56 κ intertester
Normal SI Joint Gapping Test										
Schep et al[60]	80 subjects	31.4, 7.0 years	14 F	Vascular cause of leg pain	NR	90 (NR)/26 (NR)	1.2/0.38	6	Validated Clinical Classification 88 (NR)/83 (NR)[85]	0.56 κ intertester
Cycling Test (Ankle Pressure <107 mm Hg)										
Schep et al[60]	80 subjects	31.4, 7.0 years	14 F	Vascular cause of leg pain	NR	53 (NR)/85 (NR)	3.5/0.56	6	Validated Clinical Classification 88 (NR)/83 (NR)[85]	0.56 κ intertester
Cycling Test (Ankle Brachial Index <0.54)										
Schep et al[60]	80 subjects	31.4, 7.0 years	14 F	Vascular cause of leg pain	NR	43 (NR)/1.0 (NR)	Infinite/0.57	6	Validated Clinical Classification 88 (NR)/83 (NR)[85]	0.56 κ intertester
Cycling Test (Ankle difference >23 mm Hg)										
Schep et al[60]	80 subjects	31.4, 7.0 years	14 F	Vascular cause of leg pain	NR	73 (NR)/95 (NR)	14.6/0.28	6	Validated Clinical Classification 88 (NR)/83 (NR)[85]	0.56 κ intertester

F, female; LR+, positive likelihood ratio; LR-, negative likelihood ratio; M, male; NR, not reported; OST, orthopaedic special tests; Q, Quality Assessment of Diagnostic Accuracy Studies scores (QUADAS); SN, sensitivity (%); SP, specificity (%).

Table 8 Pooled diagnostic properties and for the diagnosis of labral tear, femoral fracture and gluteal tendinopathy

Diagnostic test	Number studies sample size (n)	SN (95% CI)	SP (95% CI)	-LR (95% CI)	+LR (95% CI)	DOR (95% CI)
Labral Tear						
FADDIR (MRA)	4 (n=128) [40 41 43 44]	94 (88 to 97)*	8 (2 to 20)*	0.48 (0.20 to 1.16)	1.02 (0.96 to 1.08)	4.28 (0.63 to 29.09)
FADDIR (Arthroscopy)	2 (n=157) [42, 43]	99 (95 to 100)	7 (0 to 34)	0.15 (0.01 to 2.24)	1.06 (0.92 to 1.21)	7.27 (0.42 to 125.60)
Flexion IR	3 (n=42) [45–47]	96 (82 to 100)	17 (12 to 54)	0.27 (0.03 to 2.34)	1.12 (0.83 to 1.51)	4.23 (0.38 to 47.28)
Femoral Fracture						
Patellar-Pubic Percussion	3 (n=782) [50–52]	95 (92 to 97)	86 (78 to 92)	0.07 (0.03 to 0.13)	6.11 (3.73 to 10.03)	96.42 (36.34 to 255.87)
Gluteal Tendinopathy						
Trendelenburg	3 (n=78) [56–58]	61 (46 to 75)*	92 (83 to 97)	0.25 (0.02 to 3.15)*	6.83 (1.64 to 28.46)*	26.46 (1.92 to 365.23)*
Resisted Hip Abduction	2 (n=79) [56–58]	71 (51 to 87)	84 (71 to 93)*	0.37 (0.20 to 0.69)	5.50 (0.13 to 241.25)*	13.24 (0.36 to 484.13)*

DerSimoninian-Laird random-effects models used throughout.
*indicates those properties demonstrating significant heterogeneity (p<0.10).
DOR, diagnostic OR; FADDIR, flexion, adduction, internal rotation test; LR+, positive likelihood ratio; LR-, negative likelihood ratio; SN, sensitivity (%); SP, specificity (%).

the Trendelenburg and resisted hip abduction test. Figures 2 and 3 display the ROC Curve and Forest Plot for the PPPT. Of the studies selected for inclusion in the systematic review, five tests met eligibility for analysis of heterogeneity and potential pooling of data: FADDIR for labral tear, Flexion IR for labral tear, and PPPT for femoral fracture and Trendelenburg and resisted hip abduction both for gluteal tendinopathy. A total of six articles were available for meta-analysis addressing the FADDIR test for labral tear.[11 40–44] A total of three articles each per test were available for meta-analysis addressing the flexion internal rotation test for labral tear[45–47] and the PPPT for femoral fracture.[50–52]

Only the FADDIR test and Flexion IR test were available for meta-analysis relative to diagnosis of hip labral tear. The FADDIR test was analysed separately for the reference standards of MRA and arthroscopy. One study used both MRA and arthroscopy as a reference standard; therefore the analysis was performed separately for each.[43] One study used injection response as the reference standard for labral tear and impingement and was not included in either analysis.[11] The summary SN was excellent for both the FADDIR test 99 (95% CI 95 to 100) and the Flexion IR test 96 (95% CI 82 to 100) whereas the SP of both tests was poor. The DORs for both test were weak and contained the null value; therefore unlikely to provide any diagnostic discriminative ability for labral tear. Data were extracted from three of the studies for the FABERs test for intra-articular pathology; however, all three used a different criterion standard (MRA, injection, pain improvement) and were therefore not eligible for further analyses.
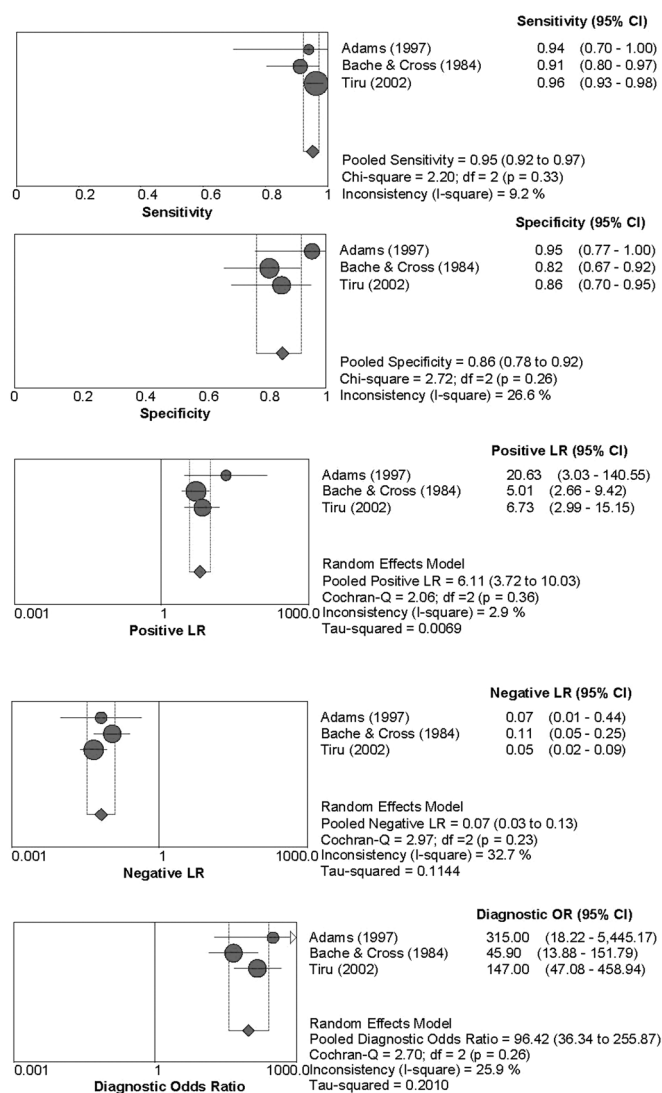
Figure 2 Pooled sensitivity, specificity, negative liklihood ratio, positive likelihood ratio and diagnostic OR with 95% CI for the patella percussion test.
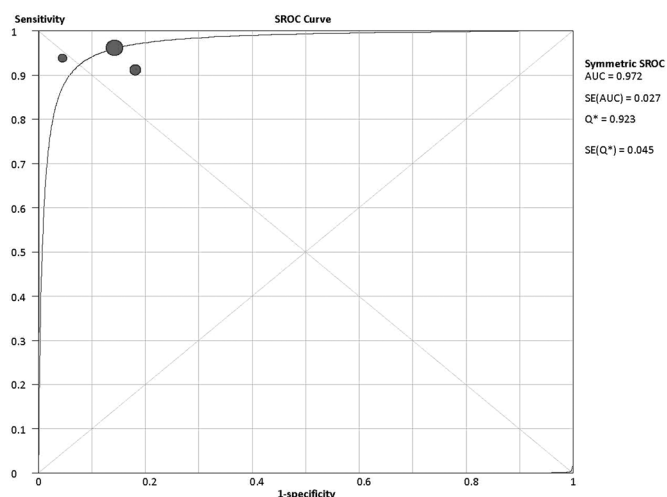

Figure 3 Summary symmetrical receiver operating curve (sROC) for the three studies for patella precussion test. Area under the curve (AUC) and Q* and their SE are provided as measures of test accuracy.

For the PPPT, the meta-analytic summary estimates of this test were good-to-excellent for SN 95% (95% CI 92 to 97%), SP 86% (95% CI 78 to 92%) and the DOR 96.42 (95% CI 36.34 to 255.87). The AUC (0.97) and Q* statistic (0.92) both indicate excellent accuracy of the PPPT. Data were also extracted for two studies for the fulcrum test for stress fracture; however, both of these studies used different reference standards (ie, radiography, bone scan or MRI) and were therefore not eligible for further summary analyses.

The meta-analytic summary estimates for the Trendelenburg test were good for SN 61:95% CI 46 to 75%), good-to-excellent for SP 92% (95% CI 83 to 97%) and the DOR 26.46 (95% CI 1.92 to 365.23), while those for the resisted hip abduction test were good for SN 71:95% (CI 51 to 87%), SP 84% (95% CI 71 to 93%) and the DOR 13.24 (95% CI 0.36 to 484.13).

DISCUSSION

Our study reviewed a broad spectrum of HPE tests designed to detect intra and extra-articular hip pathology. Similar to what others have found in the spine,[61] shoulder[28] and knee,[62] the majority of HPE tests for the hip are deficient for contributing to post-test probability for a dedicated hip specific diagnoses. The majority of stand-alone HPE tests do not demonstrate high levels of SN and/or SP,[10 28 62] thus questioning their clinical utility as stand-alone tests. Only five tests (in 14 studies) were included in the meta-analysis, and of those the PPPT was the only one to significantly alter post-test probability. Indeed, additional high quality studies are needed before most HPE tests can be recommended as stand-alone diagnostic tests.

There are a number of reasons for the poor clinical utility of the majority of the HPE tests. First, the overlap in signs, symptoms and pathomechanics between many intra-articular pathologies,[63–66] as well as changes associated with disease progression,[65] may lead to misdiagnoses.[12] Second, the majority of the HPE tests for the intra-articular pathologies of hip OA, labral tear and avascular necrosis were highly SN with poor SP. Consequently, a positive test means very little in the diagnosis of a condition such as a labral pathology since the same tests will also be positive in other disparate conditions. Only the Thomas test was found to substantially improve probability for diagnosis of labral tears. One of the primary predisposing factors to labral tear is femoroacetabular impingement,[63] [64] which involves abutment between the femoral head and acetabular rim in an adducted and internally rotated position of the hip.[63 64] Although the Thomas test does not reproduce this position, or account for the other primary etiological factors (capsular laxity, dysplasia, trauma, or degeneration)[63 64] it does recreate hip extension, which has been shown to recreate the greatest forces on the hip joint.[67] Additional support for utilisation of the Thomas test for labral tear testing is the fact that the majority of labral tears are in the anterior portion of the hip joint.[68] Last, there is a risk that the lower quality of studies fails to fully discriminate the true utility of the HPE tests and further studies, that exhibit higher quality, may yield findings that are notably different than those identified in this review.

Of the studies examining diagnostic accuracy of HPE tests for gluteal tendinopathy, only the resisted external derotation test demonstrated the ability to modify the post-test probability of a gluteal tendinopathy diagnosis. Clinical features of gluteal tendinopathy include pain reproduction with passive elongation of the involved tendons, as well as active contraction of these same tendons.[58] This test replicated both of these

clinical features, therefore likely improving its diagnostic accuracy. Many study design related shortcomings limit the interpretability and generalisabity of these tests. A prime example is the previously described Lequesne et al article.[58] Despite the low bias found on the QUADAS, the sample size included only 17 patients limiting the external generalisability.

The HPE test demonstrating the strongest diagnostic value was the PPPT which yielded both excellent SN and SP in three different studies. Previous investigation has also supported the tuning fork and stethoscope as a valid measure of fracture assessment in multiple other bones.[69] This finding was particularly the case in transverse fractures, which likely have sufficient space created by the fracture to decrease the sound the tuning fork produces.[69] The strong diagnostic value of this test provides the clinician with an increased sense of its clinical utility, especially in light of ease of application, access to radiology, and cost-effectiveness in prescription of radiology.

Worth noting is the dichotomy between a quality study and the results of a study. The two must be considered together when advocating the use of HPE tests for clinical practice. In our study, we used the original QUADAS tool (QUADAS 1) and a cut off of 10 or higher to define high quality. Others[28 61 62 70] have used this value to define higher quality studies. The majority of studies scored below 10 and had notable deficiencies associated with bias, that were captured by the QUADAS 1 tool. Despite the fact that many of the studies included scored as 'high quality/low risk of bias' in this review, there are several notable factors which may question the quality of these studies. Many of the studies had very low sample sizes and compared against control groups that were of healthy individuals, all elements which are known to influence study bias,[71] and which are not part of the QUADAS 1 tool. Although we are uncertain if the QUADAS 2 tool improves the capacity of identifying biases beyond QUADAS 1 it is worth noting that a better measure of study bias toward overall outcome is needed as is a mechanism to better define high versus low quality.

Another consideration is the limited capacity of a single test for making a definitive clinical decision.[61] Clustering tests does appear to provide more promising findings and the process of clustering to produce a preponderance of evidence of the existence of a hip specific diagnosis is more closely associated to actual clinical examination. One such study, eliminated from this review with our exclusion criteria, was the clinical prediction rule for detecting hip OA.[38] This study appears to be the only one of any clinical value for diagnosing hip OA. Future studies should investigate the diagnostic capabilities of these stand-alone tests when used in clusters.

Finally, it is important to consider the accepted criterion standard used in diagnostic accuracy studies from which to compare HPE tests. Not all criterion standards utilised in these studies were of equal value. Magnetic resonance arthrography is suggested as the preferred imaging criterion reference for labral tear pathology.[10 72] CT and MRI are therefore less ideal imaging standards. The use as intra-articular injections has also been suggested as the less preferable alternative.[11] With respect to hip OA, plain film radiographs, are confirmatory for moderate to advanced hip OA. Plain radiographs are less useful in demonstrating early OA joint changes.[73] Additionally, to the authors knowledge, the diagnostic accuracy of hip OA findings on radiographs has not been determined. In light of this uncertainty, and inconsistent use of true criterion standards, judgment of the clinical utility of each of

the HPE tests investigated in this study therefore requires careful consideration.

Therefore, the extreme number of imperfect and occasionally disparate reference standards existing across many of the hip-related diagnoses is of diagnostic importance because the use of different reference standards has also been recognised as a large source of bias and can lead to widely different diagnostic accuracy values across studies[74 75] and may trend toward underestimating the value of a test.[76] The hip is laden with symptom-based diagnoses. Symptom-based diagnoses may look markedly different from person to person and are based on a collection of symptoms versus a true, well understood biological cause. It is our impression that FAI, sports related chronic groin pain, and OA all fall within this categorisation. This dilemma, which is not unique to the hip, creates controversy and discourse among research clinicians and is typically adjusted through selected meta-analytic statistical methods.[76 77] The models used in this meta-analysis, Der-Siminon and Laird random effects models[31] take into account heterogeneity produced from using different reference standards. If there were any differences between the estimates from different reference standards, we analysed them separately.

Limitations

This review is not without limitations. One limitation of this study is our use of stratified QUADAS scores to organise study quality. Although many studies have used QUADAS summary scores,[28 61 62 70] others have cautioned against the use of a dedicated quality score,[75] As noted, it is very likely that some of the scoring of studies that were ranked as high quality were likely inflated because QUADAS 1 does not have a quality item score for sample size, case control designs, or other areas which may greatly inflate or poorly represent true study quality. Our meta-analytical results for likelihood ratios and DORs were not statistically significant for most diagnostic tests and some had significant heterogeneity. The small number of studies and number of subjects within these studies is one possible reason for the lack of statistical significance and decreased precision of the pooled estimates. An additional limitation is the lack of comparison of patient inclusion and exclusion criteria across the studies. Because many of the studies are older and were published before prospective guidelines for diagnostic accuracy the robustness of reporting inclusion/exclusion criteria was highly variable among studies and very difficult to comprehensively describe in this review. Lastly, only one author pulled data points, which increases the risk of potential error.

CONCLUSIONS

In terms of individual HPE tests of the hip, the PPPT demonstrated strong diagnostic accuracy properties for ruling in/ruling out femur fracture; the resisted external derotation test shows promise in diagnosis of a gluteal tendinopathy and the Thomas test shows intriguing findings with respect to a labral tear. Caution should be used since a single HEP test may not yield diagnostic findings that are compelling.

Competing interests None.

Review

What is already known on this topic

► Hip physical examination (HPE) tests, in combination with imaging and a detailed clinical history are routinely used in clinical practice to detect various intra-and extra-articular hip pathologies.

► Hip joint clinical examination has traditionally focused on variable versions of HPE tests.

► Variable criterion reference standards have been utilised when investigating various hip pathologies, especially labral tear/impingement and fractures.

► Various levels of diagnostic accuracy have been reported for HPE tests.

What this study adds

► There is limited evidence to support the use of HPE tests as stand-alone clinical tests for the diagnosis of hip related pathology.

► Meta-analysis for the patellar-pubic percussion test shows that this test has discriminatory ability for diagnosis of femur fracture.

► Other individual HPE tests investigated in this study have limited discriminatory ability to diagnose the variable other hip pathologies due to various reasons including limited study quality and limitations in criterion reference standards in some cases.

Recommendations are as follows

► The patellar-pubic percussion test is a useful diagnostic test for hip fracture.

► Clinicians should not rely on a single HPE test when diagnosing other hip joint pathologies.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Smith DV,** Bernhardt DT. Hip injuries in young athletes. *Curr Sports Med Rep* 2010;**9**:278–83.
2. **Kang C,** Hwang DS, Cha SM. Acetabular labral tears in patients with sports injury. *Clin Orthop Surg* 2009;**1**:230–5.
3. **Minnich JM,** Hanks JB, Muschaweck U, *et al.* Sports hernia: diagnosis and treatment highlighting a minimal repair surgical technique. *Am J Sports Med* 2011;**39**:1341–9.
4. **Kovacevic D,** Mariscalco M, Goodwin RC. Injuries about the hip in the adolescent athlete. *Sports Med Arthrosc* 2011;**19**:64–74.
5. **Strauss EJ,** Nho SJ, Kelly BT. Greater trochanteric pain syndrome. *Sports Med Arthrosc* 2010;**18**:113–9.
6. **Keogh MJ,** Batt ME. A review of femoroacetabular impingement in athletes. *Sports Med* 2008;**38**:863–78.
7. **Molloy MG,** Molloy CB. Contact sport and osteoarthritis. *Br J Sports Med* 2011;**45**:275–7.
8. **Tibor LM,** Sekiya JK. Differential diagnosis of pain around the hip joint. *Arthroscopy* 2008;**24**:1407–21.
9. **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.
10. **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.
11. **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**:1013–8.
12. **Burnett RS,** Della Rocca GJ, Prather H, *et al.* Clinical presentation of patients with tears of the acetabular labrum. *J Bone Joint Surg Am* 2006;**88**:1448–57.
13. **Feeley BT,** Powell JW, Muller MS, *et al.* Hip injuries and labral tears in the national football league. *Am J Sports Med* 2008;**36**:2187–95.
14. **Offierski CM,** MacNab I. Hip-spine syndrome. *Spine* 1983;**8**:316–21.
15. **Ben-Galim P,** Ben-Galim T, Rand N, *et al.* Hip-spine syndrome: the effect of total hip replacement surgery on low back pain in severe osteoarthritis of the hip. *Spine* 2007;**32**:2099–102.
16. **Murata Y,** Utsumi T, Hanaoka E, *et al.* Changes in lumbar lordosis in young patients with low back pain during a 10-year period. *J Orthop Sci* 2002;**7**:618–22.
17. **Nakamura N,** Sugano N, Nishii T, *et al.* Scintigraphic image patterns in dysplastic coxarthrosis: evaluation with reference to radiographic findings in 210 hips. *Acta Orthop Scand* 2003;**74**:159–64.
18. **Yoshimoto H,** Sato S, Masuda T, *et al.* Spinopelvic alignment in patients with osteoarthrosis of the hip: a radiographic comparison to patients with low back pain. *Spine* 2005;**30**:1650–7.
19. **Takemitsu Y,** Harada Y, Iwahara T, *et al.* Lumbar degenerative kyphosis. Clinical, radiological and epidemiological studies. *Spine* 1988;**13**:1317–26.
20. **Itoi E.** Roentgenographic analysis of posture in spinal osteoporotics. *Spine* 1991;**16**:750–6.
21. **Watanabe W,** Sato K, Itoi E, *et al.* Posterior pelvic tilt in patients with decreased lumbar lordosis decreases acetabular femoral head covering. *Orthopedics* 2002;**25**:321–4.
22. **Brown MD,** Gomez-Marin O, Brookfield KF, *et al.* Differential diagnosis of hip disease versus spine disease. *Clin Orthop Relat Res* 2004;**3**:280–4.
23. **Anderson K,** Strickland SM, Warren R. Hip and groin injuries in athletes. *Am J Sports Med* 2001;**29**:521–33.
24. **Kelly BT,** Buly RL. Hip arthroscopy update. *HSS J* 2005;**1**:40–8.
25. **Margo K,** Drezner J, Motzkin D. Evaluation and management of hip pain: an algorithmic approach. *J Fam Pract* 2003;**52**:607–17.
26. **Martin HD,** Kelly BT, Leunig M, *et al.* The pattern and technique in the clinical evaluation of the adult hip: the common physical examination tests of hip specialists. *Arthroscopy* 2010;**26**:161–72.
27. **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.
28. **Hegedus EJ,** Goode A, Campbell S, *et al.* Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. *Br J Sports Med* 2008;**42**:80–92.
29. **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.
30. **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.
31. **DerSimonian R,** Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88.
32. **Walter SD.** Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Stat Med* 2002;**21**:1237–56.
33. **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.
34. **Cox D.** The analysis of binary data. London: Methuen 1970.
35. **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.
36. **Landis JR,** Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977;**33**:363–74.
37. **Youdas JW,** Madson TJ, Hollman JH. Usefulness of the Trendelenburg test for identification of patients with hip joint osteoarthritis. *Physiother Theory Pract* 2010;**26**:184–94.
38. **Sutlive TG,** Lopez HP, Schnitker DE, *et al.* Development of a clinical prediction rule for diagnosing hip osteoarthritis in individuals with unilateral hip pain. *J Orthop Sports Phys Ther* 2008;**38**:542–50.
39. **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.
40. **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**:314–8.
41. **Beaulé 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.

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

43. **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;**6**:163–9.

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

45. **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.

46. **Hase T,** Ueo T. Acetabular labral tear: arthroscopic diagnosis and treatment. *Arthroscopy* 1999;**15**:138–41.

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

48. **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.

49. **McCarthy JC,** Busconi B. The role of hip arthroscopy in the diagnosis and treatment of hip disease. *Orthopedics* 1995;**18**:753–6.

50. **Adams SL,** Yarnold PR. Clinical use of the patellar-pubic percussion sign in hip trauma. *Am J Emerg Med* 1997;**15**:173–5.

51. **Bache JB,** Cross AB. The Barford test. A useful diagnostic sign in fractures of the femoral neck. *Practitioner* 1984;**228**:305–8.

52. **Tiru M,** Goh SH, Low BY. Use of percussion as a screening tool in the diagnosis of occult hip fractures. *Singapore Med J* 2002;**43**:467–9.

53. **Johnson AW,** Weiss CB Jr, Wheeler DL. Stress fractures of the femoral shaft in athletes—more common than expected. A new clinical test. *Am J Sports Med* 1994;**22**:248–56.

54. **Kang L,** Belcher D, Hulstyn MJ. Stress fractures of the femoral shaft in women's college lacrosse: a report of seven cases and a review of the literature. *Br J Sports Med* 2005;**39**:902–6.

55. **Joe GO,** Kovacs JA, Miller KD, *et al.* Diagnosis of avascular necrosis of the hip in asymptomatic HIIV-infected patients: Clinical correlation of physical examination with magnetic resonance imaging. *J Back Musculoskelet Rehabil* 2002;**16**:135–9.

56. **Bird PA,** Oakley SP, Shnier R, *et al.* Prospective evaluation of magnetic resonance imaging and physical examination findings in patients with greater trochanteric pain syndrome. *Arthritis Rheum* 2001;**44**:2138–45.

57. **Woodley SJ,** Nicholson HD, Livingstone V, *et al.* Lateral hip pain: findings from magnetic resonance imaging and clinical examination. *J Orthop Sports Phys Ther* 2008;**38**:313–28.

58. **Lequesne M,** Mathieu P, Vuillemin-Bodaghi V, *et al.* Gluteal tendinopathy in refractory greater trochanter pain syndrome: diagnostic value of two clinical tests. *Arthritis Rheum* 2008;**59**:241–6.

59. **Verrall GM,** Slavotinek JP, Barnes PG, *et al.* Description of pain provocation tests used for the diagnosis of sports-related chronic groin pain: relationship of tests to defined clinical (pain and tenderness) and MRI (pubic bone marrow oedema) criteria. *Scand J Med Sci Sports* 2005;**15**:36–42.

60. **Schep G,** Bender MH, Schmikli SL, *et al.* Recognising vascular causes of leg complaints in endurance athletes. Part 2: the value of patient history, physical examination, cycling exercise test and echo-Doppler examination. *Int J Sports Med* 2002;**23**:322–8.

61. **Cook C,** Hegedus E. Diagnostic utility of clinical tests for spinal dysfunction. *Man Ther* 2011;**16**:21–5.

62. **Hegedus EJ,** Cook C, Hasselblad V, *et al.* Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis. *J Orthop Sports Phys Ther* 2007;**37**:541–50.

63. **Groh MM,** Herrera J. A comprehensive review of hip labral tears. *Curr Rev Musculoskelet Med* 2009;**2**:105–17.

64. **Kelly BT,** Weiland DE, Schenker ML, *et al.* Arthroscopic labral repair in the hip: surgical technique and review of the literature. *Arthroscopy* 2005;**21**: 1496–504.

65. **Kappe T,** Kocak T, Bieger R, *et al.* Radiographic risk factors for labral lesions in femoroacetabular impingement. *Clin Orthop Relat Res* 2011;**469**: 3241–7.

66. **Harris-Hayes M,** Royer NK. Relationship of acetabular dysplasia and femoroacetabular impingement to hip osteoarthritis: a focused review. *PM R* 2011;**3**:1055–1067.e1.

67. **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.

68. **Mintz DN,** Hooper T, Connell D, *et al.* Magnetic resonance imaging of the hip: detection of labral and chondral abnormalities using noncontrast imaging. *Arthroscopy* 2005;**21**:385–93.

69. **Moore MB.** The use of a tuning fork and stethoscope to identify fractures. *J Athl Train* 2009;**44**:272–4.

70. **Cook C,** Hegedus E, Hawkins R, *et al.* Diagnostic accuracy and association to disability of clinical test findings associated with patellofemoral pain syndrome. *Physiother Can* 2010;**62**:17–24.

71. **Replogle WH,** Johnson WD, Hoover KW. Using evidence to determine diagnostic test efficacy. *Worldviews Evid Based Nurs* 2009;**6**:87–92.

72. **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.

73. **Karachalios T,** Karantanas AH, Malizos K. Hip osteoarthritis: what the radiologist wants to know. *Eur J Radiol* 2007;**63**:36–48.

74. **Lijmer JG,** Mol BW, Heisterkamp S, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.

75. **Rutjes AW,** Reitsma JB, Di Nisio M, *et al.* Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;**174**:469–76.

76. **Walter SD,** Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. *J Clin Epidemiol* 1999;**52**:943–51.

77. **Walter SD,** Turner RM, Macaskill P, *et al.* Optimal allocation of participants for the estimation of selection, preference and treatment effects in the two-stage randomised trial design. *Stat Med* 2012;**31**:1307–22.

78. **Nishii T,** Tanaka H, Sugano N, *et al.* Disorders of acetabular labrum and articular cartilage in hip dysplasia: evaluation using isotropic high-resolutional CT arthrography with sequential radial reformation. *Osteoarthr Cartil* 2007;**15**:251–7.

79. **Yamamoto Y,** Tonotsuka H, Ueda T, *et al.* Usefulness of radial contrast-enhanced computed tomography for the diagnosis of acetabular labrum injury. *Arthroscopy* 2007;**23**:1290–4.

80. **Rosenberg ZS,** La Rocca Vieira R, Chan SS, *et al.* Bisphosphonate-related complete atypical subtrochanteric femoral fractures: diagnostic utility of radiography. *AJR Am J Roentgenol* 2011;**197**:954–60.

81. **Rubin SJ,** Marquardt JD, Gottlieb RH, *et al.* Magnetic resonance imaging: a cost-effective alternative to bone scintigraphy in the evaluation of patients with suspected hip fractures. *Skeletal Radiol* 1998;**27**:199–204.

82. **Lieberman JR,** Berry DJ, Mont MA, *et al.* Osteonecrosis of the hip: management in the 21st century. *Instr Course Lect* 2003;**52**:337–55.

83. **Cvitanic O,** Henzie G, Skezas N, *et al.* MRI diagnosis of tears of the hip abductor tendons (gluteus medius and gluteus minimus). *AJR Am J Roentgenol* 2004;**182**:137–43.

84. **Brukner P.** Is there a role for magentic resonance imaging in chronic groin pain? *Aust J Sci Med Sport* 2004;**7**:74.

85. **Schep G,** Schmikli SL, Bender MH, *et al.* Recognising vascular causes of leg complaints in endurance athletes. Part 1: validation of a decision algorithm. *Int J Sports Med* 2002;**23**:313–21.

Diagnostic accuracy of clinical tests of the hip: a systematic review with meta-analysis

Michael P Reiman, Adam P Goode, Eric J Hegedus, et al.

Br J Sports Med published online July 7, 2012
doi: 10.1136/bjsports-2012-091035

Updated information and services can be found at:

http://bjsm.bmj.com/content/early/2012/11/07/bjsports-2012-091035.full.html

---

These include:

Data Supplement	*"Web Only Data"* http://bjsm.bmj.com/content/suppl/2012/07/07/bjsports-2012-091035.DC1.html
References	This article cites 84 articles, 9 of which can be accessed free at: http://bjsm.bmj.com/content/early/2012/11/07/bjsports-2012-091035.full.html#ref-list-1
P<P	Published online July 7, 2012 in advance of the print journal.
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 (5 articles)

Notes

Advance online articles have been peer reviewed, accepted for publication, edited and typeset, but have not not yet appeared in the paper journal. Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/