






Diagnostic accuracy of clinical tests for cam or pincer morphology in individuals with suspected FAI syndrome: a systematic review

Rahel Caliesch ^{1,2}, Martin Sattelmayer ¹, Stephan Reichenbach ^{3,4}, Marcel Zwahlen ³, Roger Hilfiker ^{1,5}

To cite: Caliesch R, Sattelmayer M, Reichenbach S, et al. Diagnostic accuracy of clinical tests for cam or pincer morphology in individuals with suspected FAI syndrome: a systematic review. *BMJ Open Sport & Exercise Medicine* 2020;**6**:e000772. doi:10.1136/bmjsem-2020-000772

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjsem-2020-000772>).

Accepted 6 April 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY. Published by BMJ.

¹Physiotherapy, HES-SO Valais-Wallis, Leukerbad, Valais-Wallis, Switzerland

²Department of Health Professions, Bern University of Applied Sciences, Bern, Switzerland

³Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

⁴Department of Rheumatology, Immunology, and Allergology, Bern University Hospital, Bern, Switzerland

⁵Graduate School for Health Sciences, University of Bern, Bern, Switzerland

Correspondence to
Rahel Caliesch;
rcaliesch@gmail.com

ABSTRACT

Objectives To determine the diagnostic accuracy of clinical tests for cam or pincer morphology in individuals with suspected femoroacetabular impingement (FAI) syndrome and to evaluate their clinical utility.

Design A systematic review of studies investigating the diagnostic accuracy of clinical tests for cam and pincer morphology.

Data sources PubMed, Embase, CINAHL and SPORTDiscus.

Eligibility criteria for selecting studies Studies investigating the diagnostic accuracy of clinical tests for cam, pincer or mixed morphology in symptomatic patients. Patients had to undergo an index test and a reference test able to identify cam or pincer morphology. Study results have to allow the calculation of true or false positives and/or negatives to calculate sensitivity, specificity, likelihood ratios (LR) and post-test probabilities.

Results Eight studies were included, investigating 17 tests and two test combinations. The studies reported a low specificity for all tests, ranging from 0.11 to 0.56. Sensitivity ranged from 0.11 to 1.00, with high sensitivities for the flexion-adduction-internal rotation (FADIR), foot progression angle walking (FPAW) and maximal squat tests. We estimated that negative test results on all of these three tests would result in a negative LR of 0.15. However, we judged the studies to provide low-quality evidence.

Conclusion There is low-quality evidence that negative test results reduce the post-test probability of cam or mixed morphologies and that consecutive testing with the FADIR, FPAW and maximal squat tests might be used as a clinical test combination. We would not recommend their use to confirm the diagnosis of FAI syndrome.

PROSPERO registration number CRD42018079116.

INTRODUCTION

The diagnosis of patients with hip and groin pain is challenging, as there may be multiple underlying aetiologies.¹ Femoroacetabular impingement (FAI) is a common cause of symptoms, with a prevalence ranging from 18% to 94%.² FAI is characterised by an abnormal morphology of the proximal femur

What is already known

- There is only limited evidence on diagnostic test accuracy for clinical tests to diagnose cam, pincer or mixed morphologies. The most current systematic reviews on clinical tests searched studies up to August 2014 and one review on flexion-adduction-internal rotation searched up to January 2017.

What are the new findings

- There is still only low-quality evidence for the diagnostic test accuracy of clinical tests for the detection of cam, pincer or mixed morphologies. A combination of three tests might be the best strategy to exclude a cam or mixed morphology, but this should be confirmed in a future study by a multivariable logistic regression model.

or the acetabulum. This results in premature contact between the femoral head or neck and the acetabulum during hip flexion and rotation, which in turn may lead to labral tears and degeneration of the acetabular cartilage.^{3,4}

Three types of anatomical morphologies are known to result in FAI: the cam and pincer morphologies or a mixed form of both.³ Early recognition of and intervention for FAI syndrome is needed to reduce sequelae such as osteoarthritis of the hip.³ Physiotherapy and activity modification may improve patient-reported outcomes and hip-related quality of life.^{5–7} Therefore, it is important to use adequate diagnostic tests. Arthroscopy, magnetic resonance arthrography (MRA), MRI, CT and radiography are currently used to diagnose cam or pincer morphology. However, these techniques are either invasive, time consuming or expensive.

To avoid unnecessary costs and invasive techniques, several clinical tests have been

proposed for the diagnosis of FAI morphology. There are already systematic reviews of the diagnostic accuracy of such clinical tests^{18–11}; however, they did not distinguish between symptomatic and asymptomatic participants, or between cam/pincer morphology and labral tears. According to a consensus meeting in 2016, the term ‘FAI syndrome’ requires a triad of *symptoms, clinical signs* and *imaging findings*; in other words, patients have to complain about pain, show a positive clinical test and demonstrate positive imaging findings of some kind of cam or pincer morphology. Additionally, they may or may not exhibit labral or articular cartilage damage.¹² Patients included in several previously published studies were asymptomatic or presented only with labral tears without FAI morphology, which is not in accordance with the official definition of an FAI syndrome. Symptomatic labral tears without accompanying cam or pincer morphology are not considered to represent FAI syndrome. Tests conducted in symptomatic patients (‘diagnostic setting’) or in asymptomatic people (‘screening setting’) are performed in two different contexts. Therefore, diagnostic accuracy should be assessed separately in symptomatic and asymptomatic individuals.

The overall aim of this systematic review was to examine the diagnostic accuracy of clinical tests for cam, pincer or mixed morphology in symptomatic patients. Specific aims were to evaluate (1) the sensitivity, specificity and likelihood ratios (LR) of clinical tests for cam or pincer morphology, (2) the clinical utility of these tests, and (3) how clinical tests can be combined to increase clinical utility.

METHODS

The review protocol was registered in the international Prospective Register of Systematic Reviews.

The Cochrane Handbook for Diagnostic Test Accuracy Reviews was used for the conduct of this study and the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) checklist was used for reporting.^{13 14}

The search was conducted from inception to February 2019 in the following electronic databases: PubMed, Embase, CINAHL and SPORTDiscus (see online supplementary file 6 for the PubMed search strategy). Reference lists of included studies were checked for additional articles, and searches for papers cited in the included articles were performed using Google Scholar.

Selection criteria

Study design

Studies were included if: (1) they investigated the diagnostic accuracy of clinical tests for cam, pincer or mixed morphology of the hip; (2) they included patients with symptoms such as groin, hip or buttock pain; (3) all patients underwent both an index test and a reference test that was able to identify cam or pincer morphology; and (4) the results allowed for the calculation of true or false positives and/or negatives. Studies including

asymptomatic individuals and those classifying participants with labral tears only as true positives were excluded. No restrictions were set regarding the language of publication, study settings or the ages or previous surgical histories of included patients.

Index tests

Included studies had to describe a clinical test that was intended to identify cam, pincer or mixed morphology of the hip joint. All available clinical tests for FAI were accepted as index tests (ie, flexion-abduction-external rotation (FABER), flexion-adduction-internal rotation (FADIR), maximal squat test, and others).

Target condition

Patients had to show imaging findings of cam or pincer morphology,¹² and could also exhibit labral or articular cartilage damage. Patients presenting with acetabular labral tears without any cam or pincer morphology were not classified as true positives.

Reference tests

MRA, MRI, CT and radiography were considered to be adequate reference tests.¹² See online supplementary table 3S for a short description of the diagnostic accuracies of all reference tests used in the included studies and online supplementary table 4S for the reliability of plain radiography for measuring alpha angles.

Screening process

Titles and abstracts were independently screened by two reviewers (RC and RH), and consensus was sought. The same procedure was applied for full texts of included references; in case of disagreement, a third reviewer decided.

Data extraction and quality assessment

True positives, false positives, false negatives and true negatives were extracted from the publications or calculated from sensitivity, specificity and prevalence. If data were not available, we contacted authors by email.

The risk of bias and concerns regarding applicability of included articles was assessed by two independent reviewers (RC and RH) using the Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2) tool.¹⁵ For this, RevMan V.5.3 software was used.¹⁶ Evidence quality was graded with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach for diagnostic tests¹⁷ and evidence profile tables were created.

Statistical analysis and data synthesis

Diagnostic 2x2 tables were used to calculate sensitivity, specificity, disease prevalence and LRs for positive and negative test results (LR+, LR-). For each index test, we plotted the calculated sensitivities and specificities with their 95% CIs on forest plots for visual judgement of variation in test accuracy across studies. We planned to pool data if two or more studies reported data for the same

test, but this did not occur. Positive and negative LR_s were used to calculate post-test probabilities. An LR₊ or LR₋ of 1 indicates no shift of the likelihood of disease, and values close to 1 indicate only small changes. The higher the LR₊ above 1, the larger the increase of the probability of disease, while the lower the LR₋ below 1, the larger the decrease in the probability of the disease.¹⁸ To increase the changes from pretest to post-test probabilities (ie, to increase the diagnostic value), we combined several index tests by multiplying the corresponding LR_s.¹⁹ Changes in post-test probabilities for the individual tests and for the test combination are presented using a plot showing the relationship between pretest and post-test probabilities for different disease prevalence (ie, pretest probabilities) (see figure 4).

Amendments to the protocol

No meta-analysis could be performed because no test results with both sensitivity and specificity values were reported in more than one study. The execution and the positivity criteria of index tests were not precisely described in every included study, as required by the protocol. However, we did not exclude such manuscripts since this would have resulted in a small number of analysed studies. We did exclude studies that reported ultrasound as index test because ultrasound corresponds to imaging techniques and not to clinical tests. In a future study we plan to assess the diagnostic accuracy of ultrasound.

RESULTS

Search results

We screened 4091 titles and abstracts, and 21 full texts. Four studies were found by reference screening. Eight studies were included in this review. The study flow diagram, with reasons for exclusions, is presented in figure 1.

Description of included studies

The characteristics of included studies are shown in the online supplementary table 1. The studies were performed in Canada,^{20 21} the USA^{22–26} and France.²⁷ All studies were prospective. Seventeen clinical tests and two test combinations were reported. Three studies investigated the FABER test,^{21–23} but only one of these reported specificity.²³ Two studies examined the FABER distance test^{24 26} (only one reported specificity).²⁶ Two studies investigated the Stinchfield test^{22 23} (only one reported specificity).²³ Two studies investigated the sensitivity of the posterior impingement test (PIT)^{21 22} (none reported specificity) and four studies investigated the FADIR test^{21 22 24 25} (only one reported specificity). For the remaining tests, there was only one study per test. In total, both sensitivity and specificity could be calculated for nine tests in five studies. For the other 10 clinical tests, only sensitivity could be calculated.

Studies provided data for 1666 hips. The prevalence of FAI morphology ranged from 10% to 64% (see table 1).

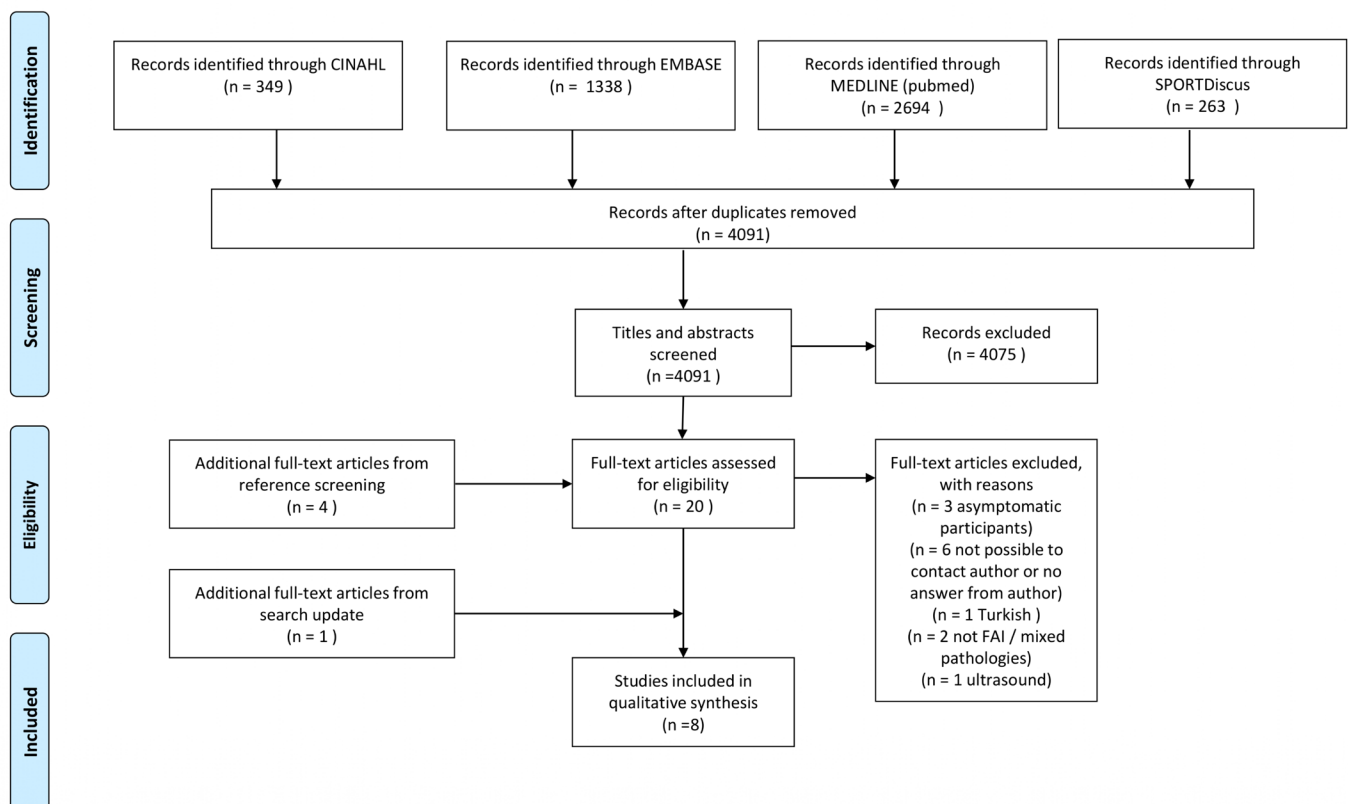


Figure 1 Study flow diagram. FAI, femoroacetabular impingement.

Table 1 Overview of results and quality of evidence of tests with sensitivity and specificity data

Test name (author)	SN (95% CI)	SP (95% CI)	LR+ (95% CI)	LR- (95% CI)	Prevalence	Unit of analysis	QoE
FPAW (Ranawat <i>et al</i>) ²⁵	0.61 (0.52 to 0.70)	0.56 (0.45 to 0.66)	1.386 (1.05 to 1.83)	0.696 (0.52 to 0.94)	0.558	Patients	SN ⊕⊕⊕○ Moderate
							SP ⊕⊕⊕○ Moderate
Maximal squat (Ayeni <i>et al</i>) ²⁰	0.75 (0.57 to 0.89)	0.41 (0.27 to 0.57)	1.278 (0.93 to 1.75)	0.605 (0.30 to 1.21)	0.41	Hips	SN ⊕⊕⊕○ Moderate
							SP ⊕⊕⊕○ Moderate
Pain predominantly in F/IR (Nogier <i>et al</i>) ²⁷	0.70 (0.62 to 0.77)	0.44 (0.33 to 0.55)	1.245 (1.01 to 1.54)	0.684 (0.49 to 0.96)	0.639	Patients	SN ⊕⊕○○ Low
							SP ⊕⊕○○ Low
FADIR (f90 add IR) (Ranawat <i>et al</i>) ²⁵	0.96 (0.91 to 0.99)	0.11 (0.06 to 0.20)	1.079 (0.99 to 1.17)	0.364 (0.12 to 1.08)	0.558	Patients	SN ⊕○○○ Very low
							SP ⊕⊕○○ Low
FABER distance (Trindade <i>et al</i>) ²⁶	0.85 (0.79 to 0.90)	0.38 (0.33 to 0.42)	1.36 (1.23 to 1.50)	0.41 (0.28 to 0.59)	0.28	Patients	SN ⊕⊕○○ Low
							SP ⊕⊕⊕○ Moderate
IROP (Maslowski <i>et al</i>) ²³	1 (0.48 to 1)	0.16 (0.06 to 0.29)	1.184 (0.83 to 1.44)	0 (0.03 to 7.87)	0.1	Patients	SN ⊕○○○ Very low
							SP ⊕○○○ Very low
Scour (Maslowski <i>et al</i>) ²³	0.8 (0.28 to 0.99)	0.40 (0.26 to 0.56)	1.333 (0.81 to 2.20)	0.5 (0.08 to 2.99)	0.1	Patients	SN ⊕○○○ Very low
							SP ⊕○○○ Very low
Stinchfield (RSLR) (Maslowski <i>et al</i>) ²³	0.6 (0.15 to 0.95)	0.36 (0.22 to 0.51)	0.931 (0.44 to 1.97)	1.125 (0.36 to 3.53)	0.1	Patients	SP ⊕○○○ Very low
							SN ⊕○○○ Very low
FABER (Maslowski <i>et al</i>) ²³	0.6 (0.15 to 0.95)	0.2 (0.10 to 0.35)	0.75 (0.36 to 1.56)	2 (0.59 to 6.79)	0.1	Patients	SN ⊕○○○ Very low
							SP ⊕○○○ Very low

FABER, flexion-abduction-external rotation; f90 add IR, flexion 90-adduction-internal rotation; FADIR, flexion-adduction-internal rotation; F/IR, flexion internal rotation; FPAW, foot progression angle walking; IROP, internal rotation over pressure; LR, likelihood ratio; QoE, Quality of Evidence; RSLR, resisted straight leg raise; SN, sensitivity; SP, specificity.

Two studies^{22 24} included only participants with confirmed FAI syndrome, while from another study²¹ we extracted data from participants with confirmed FAI morphology and excluded asymptomatic participants so as to conform to our inclusion criteria.

The reference tests used in the analysed studies were radiography,^{21 22 24–27} MRI/MRA²⁰ or both radiography and MRI/MRA.²³ Criteria for index and/or reference test positivity were not clearly stated in three out of eight studies.^{21 23 25}

Descriptions of included index tests

Fourteen of the analysed index tests were pain provocation tests: the maximal squat test,²⁰ the FABER test,^{21–23} the log roll test,²² the resisted straight leg raise test (Stinchfield test),^{22 23} the FADIR test (anterior impingement test, flexion 90-adduction-internal rotation test (f90 add IR)),^{21 22 24 25} the PIT,^{21 22} the scour test,²³ the internal rotation over pressure (IROP) test,²³ the flexion plus IR pain test,²⁷ the foot progression angle walking (FPAW) test,²⁵ the IR pain test,²¹ the f120 add IR,²¹ the flexion 90-adduction-compression test (f90 add C)²¹ and the f120

add C.²¹ Five tests were range of motion (ROM) tests: two studies used the FABER distance test^{24 26} to compare the unaffected and affected hips in terms of loss of distance between the knee and examination table; there were also two passive hip ROM tests²¹ and two combinations of an ROM test plus a pain provocation test.

Quality assessment and GRADE

Judgements of risks of bias and concerns regarding applicability were made using the QUADAS-2 tool, as shown in figure 2. Four studies showed a high risk of bias or concerns regarding applicability in at least one domain. Only one study²⁰ was classified as having no risk of bias and no applicability concerns. In two studies,^{22 24} the investigators were informed before the testing that only participants with confirmed cam or pincer morphology were included. Ratings with GRADEpro demonstrated moderate-quality evidence for both the sensitivity and specificity of the FPAW test²⁵ and the maximal squat test,²⁰ as well as for the specificity of the FABER distance test.²⁶ All other index tests showed low to very-low-quality evidence according to the GRADEpro rating. A short

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Ayeni 2014	+	+	+	+	+	+	+
Clohisy 2009	-	-	?	+	-	?	+
Maslowski 2010	+	-	?	?	+	+	+
Nogier 2010	+	+	?	?	+	+	?
Philippon 2007	-	-	-	+	-	-	-
Ranawat 2017	?	?	+	+	+	?	+
Ratzlaff 2015	+	?	+	+	-	?	-
Trindade 2018	+	?	?	+	+	?	?

- High ? Unclear + Low

Figure 2 Results of the Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2) tool.

overview is presented in [table 1](#); for the detailed grading of evidence see online supplementary file 5.

Accuracy

The test accuracy of each index test is presented using forest plots ([figure 3](#)). Sensitivity could be calculated for each of the index tests, and ranged from 0.11 (IR ROM and f90 IR, 95% CI 0.07 to 0.18) to 1.00 (IROP, 95% CI 0.48 to 1.00). The lowest sensitivity occurred in ROM tests, ranging from 0.11 to 0.22,²¹ and in PIT, ranging from 0.18 to 0.21.^{21 22}

Specificity could be calculated for nine tests in five studies. The FPAW had the highest specificity at 0.56 (95% CI 0.45 to 0.66).²⁵ The FADIR had the lowest specificity at 0.11 (95% CI 0.06 to 0.20).²⁵

A summary of the accuracy data is shown in [table 1](#).

Likelihood ratios

The LRs of the nine tests were poor. No LR+ was higher than 1.4, and no LR- was smaller than 0.3. For two tests (the FABER and Stinchfield²³), the LR+ and LR- pointed in the opposite direction, with an LR+ <1 and an LR- >1. The IROP test had a sensitivity of 100%, which results in an LR- of 0.00; however, the CI of the sensitivity ranged from 48% to 100%. None of the tests were able to relevantly shift the post-test probability. [Figure 4](#) illustrates the small changes of post-test probabilities depending on the varying prevalence of cam or pincer morphology.

To increase the ability to rule out a cam or pincer morphology, we combined the three clinical tests with the highest possible sensitivity and smallest LR (ie, the FADIR, FPAW and maximal squat tests). Chaining of these tests resulted in an LR- of 0.15. In a population with a prevalence of 55%, the probability of a person having a cam or pincer morphology with three negative

tests would decrease to 0.15 ([figure 4](#)). If the three tests were positive this would result in an LR+ of 1.91 and an increase of the post-test probability to 0.70 ([figure 4](#)).

DISCUSSION

This systematic review examined the current literature on clinical tests for the detection of cam or pincer morphology in individuals suspected of having FAI syndrome. Eight out of 4091 studies were included, and these reported on 17 clinical tests and two test combinations. Because of the insufficient number of studies per test, a meta-analysis could not be performed. There are three main findings: (1) there is only low-quality evidence; (2) no single test effectively rules in a cam or mixed morphology; (3) the FADIR, FPAW and the maximal squat test showed the best sensitivities and should be combined to cautiously rule out a cam or mixed morphology, but the validity of this combination should be tested with a multivariable regression model.

Specificity could only be calculated for nine tests. Overall results showed low specificity for all tests, ranging from 0.11 to 0.56. This indicates that these clinical tests might not be appropriate to rule in a cam or pincer morphology. High sensitivity was found for some pain provocation tests (the FADIR, FPAW, maximal squat) and for the FABER distance test. The interpretation of the FABER distance test, however, is questionable because the positivity criterion is a loss of distance between the lateral aspect of the knee and the examination table compared with the unaffected side. This requires the unaffected hip to be free of a cam or pincer morphology, but this can only be determined with imaging studies. Hence, this test is not applicable in physiotherapy practice. The lowest values for sensitivity were from a study for which we only had an abstract.²¹ No test reached a sensitivity above 0.75 in that study. Detailed information on test execution and criteria for a positive test were unavailable, and therefore findings from that study should be interpreted with caution.

LRs could be calculated for nine tests. The LRs only allow for small changes from pretest to post-test probabilities. However, the combination of three negative test results in the FADIR, FPAW and maximal squat tests yielded an LR- of 0.15. Unexpected results were obtained for the Stinchfield and FABER tests, where the LR+ was below 1 and the LR- above 1. These two tests were investigated in the same study,²³ with a prevalence of FAI morphology of 10%. All subjects were suspected to have intra-articular pathology, which might explain the high false positive rate.

A higher suspicion of cam and pincer morphology may result in a higher sensitivity of clinical tests, because evaluators might rate the test as positive in cases where the result is less clear; this would in turn decrease the number of false negatives. Three studies^{21 22 24} included only patients with confirmed FAI deformities. In two studies,^{22 24} the raters were aware of this fact, while in one

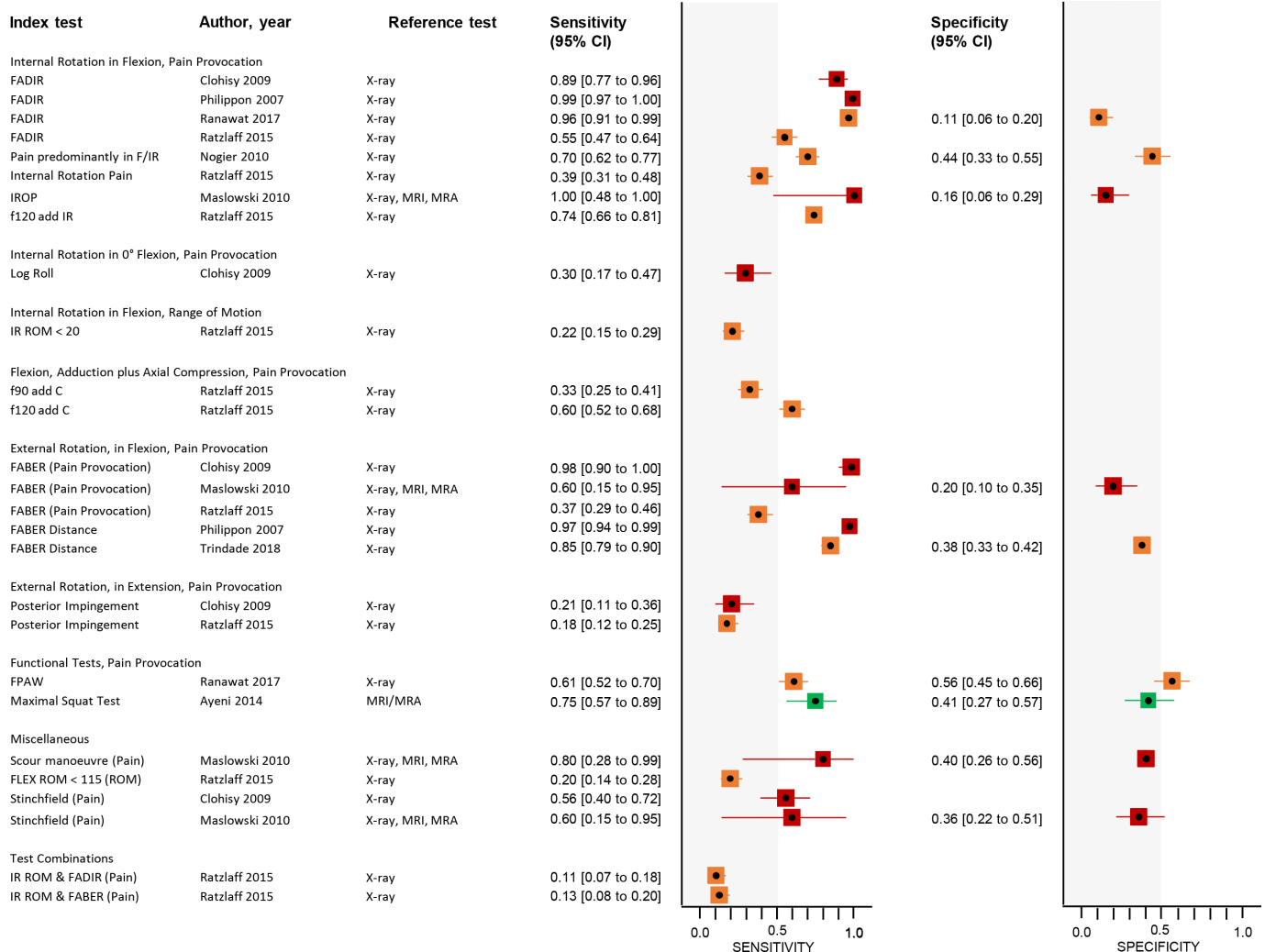


Figure 3 Forest plots of included clinical tests (see online supplementary file 4 for true positives, true negatives, false positives and false negatives). Red point estimates and CIs in the forest plot indicate high risk of bias (at least one item at high risk of bias), orange indicates unclear risk of bias (at least one unclear risk of bias and no high risk of bias), green indicates all risk of bias items at low risk. add, adduction; AIT, anterior impingement test; C, compression; f90, flexion 90°; f120, flexion 120°; FABER, flexion-abduction-external rotation; FADIR, flexion-adduction-internal rotation; FLEX, flexion; FPAW, foot progression angle walking; IR, internal rotation; IROP, internal rotation over pressure; MRA, magnetic resonance arthrography; ROM, range of motion; RSLR, resisted straight leg raise.

study the manuscript was unclear regarding the blinding status.

The reference test in most of the included studies was radiography, though one study²³ used three different imaging techniques (MRI, MRA, X-ray) and one²⁰ used MRI/MRA. Further, the included studies showed varying criteria for positivity of the reference test. They defined different alpha angle values for the diagnosis of cam morphology, as well as varying positivity criteria for pincer morphology, making them difficult to compare. Of the three tests proposed for our test combination, the maximal squat test was compared with MRI/MRA (head-neck offset <9 mm or alpha angle >55°), while the FPAW and the FADIR were compared with radiography (alpha angle >60°).

It is known that cam or pincer morphology can lead to labral and cartilage damage. Both types of damage

are considered to be risk factors for early degenerative processes and osteoarthritis of the hip joint, due to reduced hip joint motions, elevated contact pressures and shear stress caused by cam and pincer deformities.^{3 28–31} It is important to recognise that lesions of the labrum can occur as a consequence of impingement but are not present in all cases.³² Thus, studies including participants who have only labral lesions are not appropriate for assessing the accuracy of tests for cam or pincer morphology. There is an association between cam morphology and the development of osteoarthritis, whereas pincer morphology (in contrast to acetabular dysplasia) does not seem to be a risk factor for osteoarthritis.^{31 33}

It is not possible to make a general statement on whether sensitivity or specificity is more important. This depends on the context in which we apply a test. In

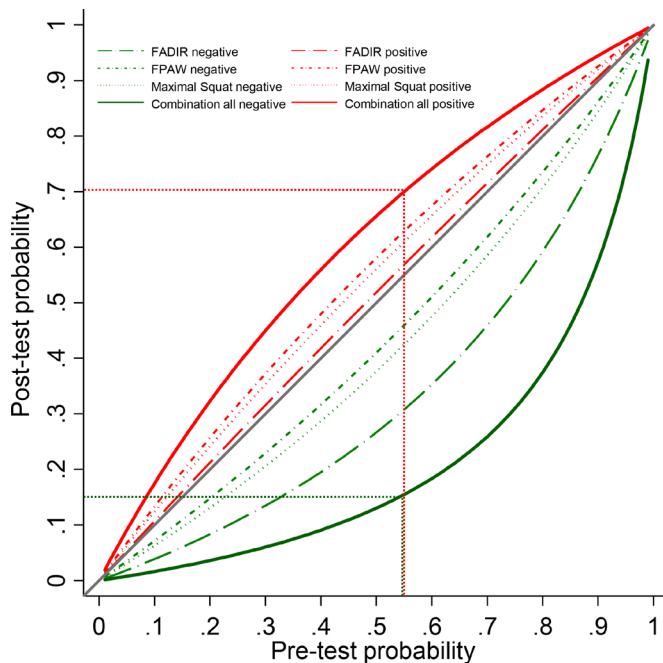


Figure 4 Post-test probabilities depending on varying prevalence (pretest probability). FADIR, flexion-adduction-internal rotation; FPAW, foot progression angle walking.

the context of professional athletes, the sensitivity of a test should be high, so as not to miss potential cam or pincer morphologies. A diagnosis of such morphologies will have consequences on the athlete's training or competing behaviour and might even have an impact on the pursuit of his career. In contrast, in a general population screening process we want the specificity to be high, so as not to have too many false positives, the impact of missing one case in that population is less serious.

Strengths of this systematic review include the facts that wherever possible, we clearly stated the types of clinical tests investigated—ROM, pain provocation or imaging—and precisely described the positivity criteria. We considered only studies that included symptomatic participants. This was done to meet the official definition of FAI syndrome, where symptoms are mandatory and asymptomatic individuals are not diagnosed with this condition.¹² The guidelines of the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy and the PRISMA-DTA checklist were followed to ensure sound scientific practice. Data extraction, estimation of risk of bias and grading of evidence were performed by two independent reviewers. The findings of this review were presented visually in forest plots to provide a simple, quick and informative overview of the test accuracy of clinical tests. Furthermore, a test combination was designed to help practitioners apply the findings to clinical practice. **Figure 4** permits the quick identification of post-test probabilities for different prevalences and tests. In comparison to a review published by Reiman *et al*,¹⁰ this report has several advantages. First, data of two additional studies were analysed. Second, only studies with symptomatic participants were included. Third, FAI

deformities were clearly differentiated from labral tears alone.

A limitation of this review is that the proposition of chaining clinical tests might result in an overestimation of the post-test probability, if the combined tests are not fully independent.³⁴ The value of this test combination should be evaluated in a new study with a multivariable logistic regression model. Additionally, three^{21 22 24} out of eight studies included only cases and hence, there was no clinical uncertainty, which introduces high risk of bias.

There were several limitations of the included studies. Most of the studies had a high risk of bias and rather low statistical precision. Different diagnostic criteria were used for the radiographic definition of cam or pincer morphology, as mentioned above, and in some cases, there was no clear statement of the diagnostic criteria. A further limitation is that the diagnostic test accuracy was not reported separately for cam, pincer or mixed morphology. In our proposed test combination, we included one test (maximal squat) from a study that diagnosed cam morphology, and two tests (the FADIR and FPAW) from a study that diagnosed FAI, defined as cam, pincer or mixed morphology. The inclusion of patients with only pincer morphology would probably lower the diagnostic test accuracy of the tests (see ref 35). Therefore, our suggestion is valid for the detection of cam or mixed morphologies. We cannot make a recommendation for the detection of pure pincer morphology.

There is a need for studies with larger numbers of participants, clear definitions of the diagnostic criteria of the reference tests and clear distinctions between patient subgroups (ie, those with cam morphology only, pincer morphology only or mixed-type morphologies) and between those with or without labral tears. Symptomatic patients with acetabular labral tears alone should not be considered as having FAI syndrome. Future studies should always include cases and non-cases so that sensitivity and specificity can be calculated, and the risk of bias should be reduced, especially by blinding the assessors concerning the patient's morphology.

There is only low-quality evidence that negative test results reduce the post-test probability of cam or mixed morphologies to a moderate amount and that consecutive testing with the FADIR, FPAW and maximal squat tests might be used as a clinical test combination. Due to the low specificity of clinical tests, we would not recommend their use to confirm the diagnosis of FAI syndrome. But so far, we do not have strong information about the interpretability of these test results, that is, there is too high uncertainty due to low-quality evidence and high risk of bias. Therefore, further adequately designed studies in larger populations and with different patient settings are required.

Acknowledgements We thank Professor Dr Amir Tal and Professor Dr Jan Taeymans of the Bern University of Applied Sciences, Department of Health Professions, Bern, Switzerland, for their very valuable comments on the presentation of the proposal of this thesis.

Contributors RC and RH designed the study, performed the literature search, extracted and analysed the data, drafted the first version of the manuscript, edited the manuscript and approved the final manuscript. MS, SR and MZ contributed to the data analysis, edited the manuscript and approved the final version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Rahel Caliesch <http://orcid.org/0000-0003-0664-5793>

Martin Sattelmayer <http://orcid.org/0000-0001-8001-4776>

Stephan Reichenbach <http://orcid.org/0000-0001-8098-6567>

Marcel Zwahlen <http://orcid.org/0000-0002-6772-6346>

Roger Hifliker <http://orcid.org/0000-0001-8662-6116>

REFERENCES

- 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;16:24E–41.
- Mascarenhas VV, Rego P, Dantas P, et al. Imaging prevalence of femoroacetabular impingement in symptomatic patients, athletes, and asymptomatic individuals: a systematic review. *Eur J Radiol* 2016;85:73–95.
- 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.
- Tanzer M, Noiseux N. Osseous abnormalities and early osteoarthritis: the role of hip impingement. *Clin Orthop Relat Res* 2004;429:170–7.
- Griffin DR, Dickenson EJ, Wall PDH, et al. Hip arthroscopy versus best conservative care for the treatment of femoroacetabular impingement syndrome (UK fashion): a multicentre randomised controlled trial. *Lancet* 2018;391:2225–35.
- Mansell NS, Rhon DI, Meyer J, et al. Arthroscopic surgery or physical therapy for patients with femoroacetabular impingement syndrome: a randomized controlled trial with 2-year follow-up. *Am J Sports Med* 2018;46:1306–14.
- Palmer AJR, Ayyar Gupta V, Fernquest S, et al. Arthroscopic hip surgery compared with physiotherapy and activity modification for the treatment of symptomatic femoroacetabular impingement: multicentre randomised controlled trial. *BMJ* 2019;364:1185.
- 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.
- Pacheco-Carrillo A, Medina-Porqueres I. Physical examination tests for the diagnosis of femoroacetabular impingement. A systematic review. *Phys Ther Sport* 2016;21:87–93.
- Reiman MP, Goode AP, Cook CE, et al. Diagnostic accuracy of clinical tests for the diagnosis of hip femoroacetabular impingement/labral tear: a systematic review with meta-analysis. *Br J Sports Med* 2015;49:811.
- 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.
- Griffin DR, Dickenson EJ, O'Donnell J, et al. The Warwick agreement on femoroacetabular impingement syndrome (FAI syndrome): an international consensus statement. *Br J Sports Med* 2016;50:1169–76.
- McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;319:388–96.
- Macaskill P, Gatsonis C, Deeks J, et al. Chapter 10: Analysing and Presenting Results. In: *Cochrane Handbook for systematic reviews of diagnostic test accuracy*. The Cochrane Collaboration, 2010. <http://srdta.cochrane.org/>
- Leeftang MMG, Deeks JJ, Takwoingi Y, et al. Cochrane diagnostic test accuracy reviews. *Syst Rev* 2013;2:82.
- The Cochrane Collaboration. *Review manager (RevMan)*. Copenhagen: The Nordic Cochrane Centre, 2014.
- Brozek JL, Akl EA, Jaeschke R, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The grade approach to grading quality of evidence about diagnostic tests and strategies. *Allergy* 2009;64:1109–16.
- Cook C, Cleland J, Huijbregts P. Creation and critique of studies of diagnostic accuracy: use of the STARD and QUADAS methodological quality assessment tools. *J Man Manip Ther* 2007;15:93–102.
- Pewsnr D, Battaglia M, Minder C, et al. Ruling a diagnosis in or out with "SpPin" and "SnNOut": a note of caution. *BMJ* 2004;329:209–13.
- 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.
- Ratzlaff C, Li L, Wong H, et al. The prevalence and diagnostic accuracy of clinical exam tests in identifying radiographic femoroacetabular impingement (FAI): a prospective population-based study. *Osteoarthritis Cartilage* 2015;23:A170–1.
- Clohisy JC, Knaus ER, Hunt DM, et al. Clinical presentation of patients with symptomatic anterior hip impingement. *Clin Orthop Relat Res* 2009;467:638–44.
- 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.
- Philippon MJ, Maxwell RB, Johnston TL, et al. Clinical presentation of femoroacetabular impingement. *Knee Surgery, Sports Traumatology, Arthroscopy* 2007;15:1041–7.
- Ranawat AS, Gaudiani MA, Slullitel PA, et al. Foot progression angle walking test: a dynamic diagnostic assessment for femoroacetabular impingement and hip instability. *Orthop J Sports Med* 2017;5:2325967116679641.
- Trindade CAC, Briggs KK, Fagotti L, et al. Positive FABER distance test is associated with higher alpha angle in symptomatic patients. *Knee Surg Sports Traumatol Arthrosc* 2019;27:3158–61.
- 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:S53–8.
- Beaulé PE, Hynes K, Parker G, et al. Can the alpha angle assessment of CaM impingement predict acetabular cartilage delamination? *Clin Orthop Relat Res* 2012;470:3361–7.
- Kappe T, Kocak T, Reichel H, et al. Can femoroacetabular impingement and hip dysplasia be distinguished by clinical presentation and patient history? *Knee Surg Sports Traumatol Arthrosc* 2012;20:387–92.
- Lamontagne M, Kennedy MJ, Beaulé PE. The effect of CaM FAI on hip and pelvic motion during maximum squat. *Clin Orthop Relat Res* 2009;467:645–50.
- Agricola R, Heijboer MP, Bierma-Zeinstra SMA, et al. Cam impingement causes osteoarthritis of the hip: a nationwide prospective cohort study (check). *Ann Rheum Dis* 2013;72:918–23.
- Beck M, Kalhor M, Leunig M, et al. Hip morphology influences the pattern of damage to the acetabular cartilage: femoroacetabular impingement as a cause of early osteoarthritis of the hip. *J Bone Joint Surg Br* 2005;87:1012–8.
- Saber Hosnijeh F, Zuiderwijk ME, Versteeg M, et al. Cam deformity and acetabular dysplasia as risk factors for hip osteoarthritis. *Arthritis Rheumatol* 2017;69:86–93.
- Bachmann LM, ter Riet G, Clark TJ, et al. Probability analysis for diagnosis of endometrial hyperplasia and cancer in postmenopausal bleeding: an approach for a rational diagnostic workup. *Acta Obstet Gynecol Scand* 2003;82:564–9.
- Casartelli NC, Brunner R, Maffiuletti NA, et al. The FADIR test accuracy for screening CAM and pincer morphology in youth ice hockey players. *J Sci Med Sport* 2018;21:134–8.