Bibliographic Cite	PMID Link	Literature Type	Level of Evidence	Purpose	Population	Intervention and Outcome Measures	Results/ Recommendations	Study Limitiations
Culvenor AG, Oiestad BE, Hart HF, et al. Prevalence of knee osteoarthritis features on magnetic resonance imaging in asymptomatic uninjured adults: A systematic review and meta-analysis. Br J Sports Med. 2019; 53(20):1268-1278.	29886437	systematic review and meta-analysis	Moderate level of evidence	To provide summary estimates of the prevalence of MRI features of osteoarthritis in asymptomatic uninjured knees.	Studies reporting the prevalence of MRI features of knee OA in asymptomatic adult knees (ie, mean age ≥18 years with no knee symptoms during any activity) with no history of injury or surgery were included.	Primary outcomes were individual MRI features assessed semiquantitatively and included in the definition of MRI-defined knee OA12: (i) cartilage defects, defined as partial-thickness or full-thickness cartilage lesions; (ii) meniscal tears, defined as high signal extending to an articular surface; (iii) BMLs, defined as areas of ill-delineated signal within trabecular bone (hypointense on T1-weighted images, hyperintense on T2-weighted fat-suppressed images); and (iv) osteophytes, defined as the presence of osteocartilagenous protrusions at articular margins. Secondary outcomes were other MRI features previously associated with knee OA (defined in detail in the online appendix eMethods 2): effusion-synovitis, subchondral cysts, ligament tears, subchondral sclerosis/attrition and infrapatellar fat pad synovitis/oedema. Two authors (AGC, HFH) independently assessed all titles and abstracts of identified reports for eligibility. Two reviewers independently assessed risk of bias. Summary estimates were calculated using random effects meta-analysis (and stratified by mean age: <40 vs ≥40 years). Meta-regression explored heterogeneity.	A total of 46 cross-sectional and 17 longitudinal studies involving a total of 4,751 individuals (5,397 knees) were included in the review. Out of 13 possible points on the risk of bias scoring criteria, 5 studies scored 0–4 points, 26 scored 5–7 points, 25 scored 8–10 points and 7 scored 11–13 points. The overall pooled prevalence of cartilage defects was 24% (95% CI 15% to 34%) and meniscal tears was 10% (7% to 13%), with significantly higher prevalence with age: cartilage defect <40 years 11% (6%to 17%) and ≥40 years 43% (29% to 57%); meniscal tear <40 years 4% (2% to 7%) and ≥40 years 19% (13% to 26%). The overall pooled estimate of bone marrow lesions and osteophytes was 18% (12% to 24%) and 25% (14% to 38%), respectively, with prevalence of osteophytes (but not bone marrow lesions) increasing with age. Significant associations were found between prevalence estimates and MRI sequences used, physical activity, radiographic osteoarthritis and risk of bias.	Limitations of this review include the heterogeneity between studies that remained unexplained by the variables examined. Unexplained factors, such as the inherent subjective nature of grading MRIs, irrespective of experience, may contribute to OA feature prevalence. The influence of BMI was unable to be assessed as half of the studies did not report BMI. Finally, the meta regression analyses relied on aggregated published data, which may have underestimated the association of MRI features with older age and female sex.
Decary S, Ouellet P, Vendittoli PA, et al. Diagnostic validity of physical examination tests for common knee disorders: An overview of systematic reviews and meta-analysis. Phys Ther Sport. 2017;23:143-55.	27693100	Meta-Analysis; Review	Moderate level of evidence	To systematically review evidence on the diagnostic validity of physical examination tests for the diagnosis of knee disorders.	To be included, articles needed to 1- be a systematic review or a meta-analysis, 2- report on the diagnostic properties of at least one physical test for at least one knee disorder and 3- be written in English or French. 17 articles and 16662 patients were ultimately included.	Seventeen reviews were included with mean AMSTAR score of 5.5 ± 2.3. Based on six SR, only the Lachman test for ACL injuries is diagnostically valid when individually performed (Likelihood ratio (LR+):10.2, LR-:0.2). Based on two SR, the Ottawa Knee Rule is a valid screening tool for knee fractures (LR-:0.05). Based on one SR, the EULAR criteria had a post-test probability of 99% for the diagnosis of knee osteoarthritis. Based on two SR, a complete physical examination performed by a trained health provider was found to be diagnostically valid for ACL, PCL and meniscal injuries as well as for cartilage lesions.	Many SR and MA are of low to moderate quality, which warrants caution from clinicians when reading these reviews for clinical guidance. However, a few methodologically sound reviews provide high-quality evidence for ACL and meniscal injuries. The evidence suggests that clinicians may diagnose or exclude an ACL injury with the Lachman test, exclude a knee fracture using the Ottawa Knee Rule and make a diagnosis of knee OA based on the results of the American College of Rhumatology and EULAR rules. For other knee disorders (meniscal injury, PFP, PCL injury and others), the available evidence does not demonstrate that tests used individually are diagnostically valid. Globally, very few clinical tests, when performed individually, can diagnose or exclude a knee disorder. Based on limited and low-quality evidence, the combination of history elements and physical tests may be more diagnostically valid. In the context of increasing healthcare costs, the development of clinical prediction rules comprising history elements and physical examination tests from methodologically sound diagnostic studies are necessary to further advance the diagnosis of knee disorders.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75% The limitations of this SR include the difficulty to combine the point estimates of SR and MA, there is also a wide range and heterogeneity of the evidence presented in the review. Many SR and MA are of low to moderate quality, which warrants caution from clinicians when reading these reviews for clinical guidance.
Drew BT, Redmond AC, Smith TO, et al. Which patellofemoral joint imaging features are associated with patellofemoral pain? Systematic review and meta-analysis. Osteoarthritis Cartilage. 2016;24(2):224-36.	26471209	systematic review and meta-analysis	moderate level of evidence	To review the association between patellofemoral joint (PFJ) imaging features and patellofemoral pain (PFP).	Studies were eligible if they used magnetic resonance imaging (MRI), computed tomography (CT), ultrasound (US) or X-ray (XR) to compare PFJ features between a PFP group and an asymptomatic control group in people <45 years of age. A pooled meta-analysis was conducted and data was interpreted using a best evidence synthesis. Forty studies (all moderate to high quality) describing 1043 people with PFP and 839 controls were included (1882 patients).The mean age was 27.0 years (range: 14-40.7years), with 74.3% women in the case group and 69.0% in the control group.	Two features were deemed to have a large standardised mean difference (SMD) based on meta-analysis: an increased MRI bisect offset at 0degree knee flexion under load (0.99; 95% CI: 0.49, 1.49) and an increased CT congruence angle at 15degree knee flexion, both under load (1.40 95% CI: 0.04, 2.76) and without load (1.24; 95% CI: 0.37, 2.12). A medium SMD was identified for MRI patella tilt and patellofemoral contact area. Limited evidence was found to support the association of other imaging features with PFP. A sensitivity analysis showed an increase in the SMD for patella bisect offset at 0degree knee flexion (1.91; 95% CI: 1.31, 2.52) and patella tilt at 0degree knee flexion (0.99; 95% CI: 0.47, 1.52) under full weight bearing.	Certain PFJ imaging features were associated with PFP. Future interventional strategies may be targeted at these features.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75%. Risk of bias -one or more key results were based on studies with a majority having a high risk of bias. Study limitations included considerable clinical heterogeneity was present in the studies utilising XR and US. In addition, lack of blinding in the studies raises the concern of confirmation bias. Finally, the reliability of the imaging assessment was reported in fewer than half the included studies (internal validity).

Duncan ST, Khazzam MS, Burnham JM, et al. Sensitivity of standing radiographs to detect knee arthritis: a systematic review of Level I studies. Arthroscopy. 2015;31(2):321-8.	25312767	Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't; Review	high level of evidence	to perform a systematic review of the available literature to define the level of quality evidence for determining the sensitivity and specificity of different radiographic views in detecting knee osteoarthritis and to determine the impact of different grading systems on the ability to detect knee osteoarthritis.	The sytematic review included only studies in the English language, those limited to humans, those identifying primary osteoarthritis, those using either the standing AP or the 45 degrees PA radiographic view, those reporting the sensitivity and specificity of each radiographic view with a minimum radiographic view of one compartment examined, and those confirming chondromalacia grade using the gold standard of arthroscopy. Exclusion criteria specified those studies failing to report sensitivity and specificity, those failing to confirm chondromalacia using arthroscopy and those not relating to primary tibiofemoral osteoarthritis. 6 studies and 970 patients were ultimately included. One study examined using 2 different radiographic grading systems to detect osteoarthritis. The study found that the Kellgren-Lawrence system was more sensitive for the severe osteoarthritis in the medial compartment, with 95% sensitivity versus 83% for the JSN classification, but this was not statistically or clinically significant. The JSN classification was more specific for the medial compartment, with 96% specificity versus 59% for the Kellgren-Lawrence system, which was statistically significant (P < .01).	All 6 studies examined the medial compartment of the knee. The sensitivity of the standing AP radiograph to detect significant medial compartment arthritis ranged from 3% to 95%. For the 45 degrees flexion PA view, the sensitivity ranged from 6% to 86%. Three studies directly compared the standing AP view versus the 45 degrees flexion PA view for the medial compartment. All 3 studies found the 45 degrees flexion PA view to be more sensitive at detecting severe osteoarthritis, but only one study18 found a statistically significant difference of 61% between the 2 views (P < .01). For the lateral compartment of the knee, 4 studies reported on the sensitivity of the standing AP knee radiographs to detect severe osteoarthritis, which ranged from 16% to 42%. For the 45 degrees flexion PA view, 3 studies reported the sensitivity to range from 6% to 83%. For the 3 studies that directly compared the standing AP view to the 45 degrees flexion PA view for detecting severe lateral compartment arthritis, most found the 45 degrees PA view to be more sensitive at detecting severe arthritis.	The diagnosis and treatment of patients with knee pain can be challenging. The use of knee radiographs to help diagnose and guide treatment has been practiced for years. Standing knee radiographs, especially the 45 degrees flexion PA view, are sensitive for detecting severe osteoarthritis of the tibiofemoral joint. Using the amount of JSN as a guide for the amount of osteoarthritis can help to both rule in and rule out the presence of severe osteoarthritis. Further studies are needed to help determine the optimal diagnostic tests to evaluate osteoarthritis of the patellofemoral joint and mild osteoarthritis involving the tibiofemoral joint.	Risk of bias - one or more key results were based on studies with a majority having a high risk of bias The exclusion criteria used, study design and quality of the included studies potentially limited the overall number of studies examined. Limiting the review to only those studies that qualified as Level I evidence may have also excluded quality studies, including those that examined the patellofemoral joint. As such, the ability of radiographs to detect patellofemoral arthritis falls outside the scope of this review.
Harris JD, Brophy RH, Jia G, et al. Sensitivity of magnetic resonance imaging for detection of patellofemoral articular cartilage defects. Arthroscopy. 2012;28(11):1728-37.	22749495	Review	Moderate level of evidence	To identify the sensitivity, specificity, and accuracy of MRI in diagnosis of patellofemoral chondral defects of the knee, using arthroscopy as the reference gold standard.	Inclusion criteria included English-language studies in humans that report the diagnostic ability of MRI to identify and characterize AKS-confirmed chondral defects in the PF joint (patella and/or femoral trochlea). Thus a prerequisite study component was that patients had both MRI and AKS, with MRI preceding AKS. Minimum MRI magnet strength was 1.5 T. Exclusion criteria were any non–English-language studies, basic science studies, animal model studies, or biomechanical or surgical/technical studies. Level V evidence was excluded. Studies that grouped PF and tibiofemoral defect results were excluded if separate PF results were not reported. If the MRI magnet strength was less than 1.5 T, the study was excluded. Any study that analyzed MRI outcomes after cartilage surgery was excluded. Ultimately 13 studies with 596 patients were included.	Thirteen studies were included in this analysis. There were 8 Level I studies, Level II studies, and 3 Level III studies. For the patella and trochlea, the sensitivity of MRI to detect chondral pathology ranged from 0% to 95% and 62% to 100%, respectively. Within all studies that performed a direct comparison between patellar and trochlear defects, MRI was more sensitive in detection of patellar (87%) versus trochlear (72%) defects. For the patella and trochlea, the specificity of MRI ranged from 62% to 100% and 81% to 97%, respectively. Within all studies that performed a direct comparison between patellar and trochlear defects, MRI was similarly specific for patellar (86%) and trochlear (89%) defects. For the patella and trochlea, the accuracy of MRI in identifying and characterizing chondral defects ranged from 72% to 98% and 74% to 93%, respectively. Within all studies that performed a direct comparison between patellar and trochlear defects, MRI was similarly accurate for patellar (84%) and trochlear (83%) defects. Interobserver agreement was substantial to almost perfect for both patellar and trochlear defects.	MRI is a highly sensitive, specific, and accurate noninvasive diagnostic modality for the detection of chondral defects in the PF compartment of the knee, using arthroscopy as the reference gold standard. Although there was wide variability in the statistical parameters assessed, MRI was more sensitive specific and accurate for patellar and trochlear defects. Interobserver reliability is substantial to near perfect in the assessment of these lesions, without a significant difference between patellar and trochlear defects.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75% There is also heterogeneity in classification systems, reporting of results, patient populations, and defect size and depth, as well as other intra-articular knee diagnoses (anterior cruciate ligament tear, meniscus tear). This significant heterogeneity precluded performance of not only a meta-analysis but also any significant statistical comparisons across different studies. Limitations of this systematic review are reliant on the biases in the studies analyzed. Thus the level of evidence of this diagnostic review is only as high as the lowest of the studies analyzed, Level III (analysis of nonconsecutive patients). The use of arthroscopy as the gold standard for the confirmation of chondral lesions, though necessary, is another significant limitation of this review.
Karel YH, Verkerk K, Endenburg S, et al. Effect of routine diagnostic imaging for patients with musculoskeletal disorders: A meta-analysis. European Journal of Internal Medicine. 2015;26(8):585-95.	26186812	Meta-Analysis; Review	Moderate level of evidence	The increasing use of diagnostic imaging has led to high expenditures, unnecessary invasive procedures and/or false-positive diagnoses, without certainty that the patients actually benefit from these imaging procedures. This review explores whether diagnostic imaging leads to better patient-reported outcomes in individuals with musculoskeletal disorders.;	Trials were eligible when: 1) a diagnostic imaging procedure was compared with any control group not getting or not receiving the results of imaging; 2) the population included individuals suffering from musculoskeletal disorders, and 3) if patient-reported outcomes were available. Primary outcome measures were pain and function. No exclusion criteria. Ultimately 11 studies with 2777 patients were included.	For the improvement in pain on short and long-term follow-up, pooling the studies with low back pain patients resulted in a significant effect in favor of no imaging on the short [SMD 0.17 (95% CI: 0.04– 0.31)] and long term [SMD 0.13 (95% CI: 0.02–0.24)] but the effect size was below 0.2, while the trials with patients with knee complaints found no difference on the long term [SMD 0.02 (95% CI: −0.14– 0.18)]. Heterogeneity was small (I2 = 39%) at short-term follow-up and not present at long-term follow-up. When all trials were pooled, no significant and clinically relevant differences were found on the short term [SMD 0.10 (95% CI: −0.08–0.29)]. On long-term follow-up data showed borderline significant results in favor of no imaging [SMD 0.09 (95% CI: 0.00–0.18)] but the effect size remained below 0.2. In the short-term analysis there were four studies and in the long-term analysis there were five studies with a primary care population. Effect sizes for both the short term [SMD 0.15 (95% CI: 0.01–0.30)] and long term [SMD 0.11 (95% CI: 0.01–0.20)] resulted in borderline significant effects in favor of no imaging but the effect size was below 0.20. Pooling only the trials using radiography (n = 3) as imaging method resulted in a significant effect in favor of no imaging but a SMD below 0.2 [SMD 0.15 (95% CI: 0.03–0.26)], whereas pooling the trials with MRI (n = 8) found no difference [SMD 0.07 (95% CI: −0.05–0.18)]. Overall improvement showed a significant but clinically irrelevant result in favor of the no imaging group (RR 1.15, 95% CI: 1.03–1.28). Sensitivity analysis showed that excluding two trials with high risk of bias did not change the results (RR 1.13, 95% CI: 1.01–1.27). Four studies were performed in primary care; pooling these studies did not alter the results (RR 1.15, 95% CI: 1.03–1.28).	The results indicate that it is unlikely that the use of routine diagnostic imaging in all patients leads to better patient-reported outcome measures. Imaging has its place in health care where serious conditions are suspected or when surgery is considered. Diagnostic imaging can be considered in patients with low back pain to rule out a serious underlying condition in the presence of red flags and in subacute/chronic low back pain patients who show no improvement. Clinical decision rules should be used by clinicians in patients with traumatic knee complaints. In non-traumatic knee complaints diagnostic imaging should be used if conservative treatment fails. This review strengthens the available evidence that routine referral to diagnostic imaging by general practitioners for patients with knee and low back pain yields little to no benefit.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75%. Risk of bias was present in a considerable percentage of the included studies (45%).

Kopkow C, Freiberg A, Kirschner S, et al. Physical examination tests for the diagnosis of posterior cruciate ligament rupture: a systematic review. J Orthop Sports Phys Ther. 2013;43(11):804-13.	24175598	Review	Moderate level of evidence	To summarize and evaluate research on the accuracy of physical examination tests for diagnosis of posterior cruciate ligament (PCL) tear. To confirm the diagnosis, arthroscopy, arthrotomy, and magnetic resonance imaging (MRI) were used as reference standards. MRI was considered a valid reference standard because recent literature has shown excellent correlation between MRI and arthroscopic as well as arthrotomy findings for the diagnosis of PCL injuries.	All study designs for diagnostic accuracy were considered eligible if they compared the results of physical examination tests performed in the context of a physical examination of the PCL with those of a reference standard. Studies on patients of any age and in any clinical setting were included. Studies that assessed the diagnostic accuracy of physical examination tests to assess a PCL rupture, which was defined as the target condition, were included. PCL rupture could be acute or chronic, as well as partial or complete. The authors excluded studies if they evaluated physical (index) tests under anesthesia or intraoperatively or postoperatively. Studies on animals and cadavers were also excluded. Studies were excluded from this systematic review if they did not name or describe a physical examination test or did not reference a source that did. Studies were also excluded if they reported the overall accuracy of a group of tests or if the diagnostic accuracy data of individual tests could not be extracted. If authors made use of generic terms, such as physical examination, to denote an unspecified combination of physical tests, these studies were also excluded. Ultimately 11 studies with 369 patients were included.	A total of 11 different physical examination tests were evaluated: posterior drawer test, quadriceps active test, recurvatum test, posterior sag sign, varus/valgus test at 0°, reverse Lachman test, dynamic posterior shift, reverse pivot shift, reverse Lachman end point, and valgus and varus tests at 30° of flexion. These tests were compared to an accepted reference standard in all included studies. Results for sensitivity and specificity of physical examination tests were heterogeneous, which was statistically significant (visually assessed using forest plots and statistically using chi-square tests [plots/data not shown]; α = .05). Reliability was not assessed in any of the included studies; therefore, no values were available for reporting. The posterior drawer test was the most frequently studied test, with sensitivity data reported in 8 studies (range from 22%-100%) and specificity data in only 1 study (98%; CI 90-100). The quadriceps active test seemed to be the most specific of the evaluated tests, although only 3 studies evaluated this test with 2 of the 3 studies reporting the data needed to calculate specificity (96% and 100%) and all 3 studies to calculate sensitivity (range from 53%-98%). The posterior sag sign was evaluated in 5 studies and seemed to be the most sensitive physical examination test (range 46%-100%). However, data to calculate specificity were only available from a single study (100%; CI 95-100).	Based on the results, the quadriceps active test seems to be the most specific test and the posterior sag sign the most sensitive test to help in the diagnosis of a potential PCL injury, although this conclusion is based on a few studies of low methodological quality. Presently, most physical examination tests have not been evaluated sufficiently, and, at this stage, determining the most appropriate tests for assessing the integrity of the PCL is difficult. Thus, there is a strong need for further research in this area.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75% Risk of bias - one or more key results were based on studies with a majority having a high risk of bias High risk of bias of existing studies regarding patient selection, index tests and reference standards, and flow and timing. Heterogeneity was due to different patient populations, study design, missing description of index tests, or lack of blinding. Most of the included studies were not performed recently, and there is a lack of similar studies published in the more recent literature. Most of the included studies provided data solely to calculate sensitivity; therefore, the calculation of specificity was not always possible. A meta-analysis could not be performed because of the low number of included studies and their heterogeneity, which also prevented a subgroup analysis, a common problem in the context of diagnostic test accuracy studies.
Leblanc MC, Kowalczuk M, Andruszkiewicz N, et al. Diagnostic accuracy of physical examination for anterior knee instability: a systematic review. Knee Surg Sports Traumatol Arthrosc. 2015;23(10):2805-13.	25763847	Review	Moderate level of evidence	To determine the diagnostic accuracy of Lachman, pivot shift and anterior drawer tests versus gold standard diagnosis (magnetic resonance imaging or arthroscopy) for anterior cruciate ligament (ACL) insufficiency cases. In addition, the hypothesis was made that the diagnostic accuracy would be lower for acute injuries, partial ruptures and examinations conducted in the awake clinical setting.	Studies were eligible for inclusion if they primarily assessed the diagnostic accuracy of physical examination (Lachman, pivot shift or anterior drawer tests) relative to MRI or arthroscopy as a gold standard for diagnosis. The study population included all patients with anterior knee instability secondary to ACL insufficiency. The inclusion criteria were as follows: (1) patients with a knee injury, (2) physical diagnosis with at least one physical test (clinic or EUA), (3) correlation with a gold standard (MRI, arthroscopy, arthrotomy), (4) in vivo human studies, (5) adults and (6) studies published in English or French. Exclusion criteria included: (1) review articles, (2) knee dislocation with multiple ligamentous injuries, (3) studies on specific injuries other than primary ACL, (4) no specification of the physical diagnostic test used, (5) studies which only discussed surgical techniques and (6) publications published prior to the year 2000. Systematic reviews and biomechanical (non-human) studies were excluded. Ultimately 8 studies and 1196 patients were included.	For combined (partial and complete) ruptures, the pooled sensitivity was 89 % (95 % CI 76–98 %) for the Lachman test and 79 % (95 % CI 63–91 %) for the pivot shift test. For complete ruptures, the pooled sensitivity was 96 % (95 % CI 90–100 %) for the Lachman test and 86 % (95 % CI 68–99 %) for the pivot shift test. For partial ruptures, the pooled sensitivity was lower and more variable: 68 % (95 %N10 CI 25–98 %) for the Lachman test and 67 % (95 % CI 47–83 %) for the pivot shift test. Only two studies provided complete data on both ACL-deficient knees and non-ACL-deficient knees. They reported specificity for the anterior drawer test of 83.9 % (95 % CI not available) and 57.0 % (95 % CI 0.48, 0.67). Only one study reported on specificity for the Lachman and pivot shift tests, at 78.1 % (95 % CI 0.61, 0.89) and 86.4 % (95 % CI not available), respectively. Due to insufficient data, pooled sensitivity results were not calculated for the anterior drawer test, EUA and chronicity. Only two studies reported sensitivity results for combined or complete ruptures using the anterior drawer test, and only one reported results for partial ruptures for the anterior drawer test. Also, only two studies used EUA. One reported a sensitivity of 100 % (largest 95 % CI 0.87, 1.00) for Lachman and pivot shift physical examinations in all settings including combined (partial and complete) and individual rupture types. The second reported on complete ruptures, with a sensitivity of 100 % (95 % CI 0.88, 1.00) for the Lachman and 82 % (95 % CI 0.63, 0.94) for the pivot shift. Data regarding the effect of the chronicity of the lesion were also too scarce to analyze.	The key finding of this systematic review was that although both Lachman and pivot shift tests are sensitive in diagnosing ACL ruptures, the clinical setting (awake vs. nonawake) and extent of injury (partial vs. complete rupture) have an impact on diagnostic accuracy. The current literature did not contain sufficient data to calculate pooled specificity; therefore, no clear recommendation regarding diagnostic accuracy of the physical examination for ACL insufficient knees could be made. Given the advances in the resolution of MRI and concomitant capability for diagnosing ACL ruptures, the possibility of conducting a diagnostic accuracy study for physical examination of ACL ruptures is now available and could greatly improve the understanding of the true accuracy of these physical diagnostic tests.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75% Risk of bias - one or more key results were based on studies with a majority having a high risk of bias There are inherent biases in the included studies, and the included studies were of different observational designs with diverse patient populations, resulting in a large inter-study heterogeneity. Most studies did not report sufficient data to allow a complete diagnostic accuracy analysis.
Meserve BB, Cleland JA, Boucher TR. A meta-analysis examining clinical test utilities for assessing meniscal injury. Clin Rehabil. 2008;22(2):143-61.	18212035	Meta-Analysis; Review	Moderate level of evidence	To systematically review the most recent literature with meta-analysis to summarize the accuracy of clinical tests for assessing meniscal lesions of the knee.	Studies were eligible for inclusion if they investigated the diagnostic accuracy of at least one clinical test for identifying meniscal lesions of the knee and used arthrotomy or arthroscopy as reference standards. 11 studies with 1071 patients were ultimately included.	Three tests - joint line tenderness, McMurray's and Apley's - were compared in the meta-analysis. The methodological quality of the studies was found to have a significant effect on both the test sensitivities and specificities. Summary receiver operating characteristic (ROC) curves, sensitivity values, mean likelihood ratios and diagnostic odd ratios (DOR) uniformly show joint line tenderness (DOR = 10.98) to be the best ;common' test, followed by McMurray's (DOR = 3.99) and Apley's (DOR = 2.2). Thessaly's test reported the strongest DOR of 227, but samples were smaller (n = 410), than those for joint line tenderness (n = 1354), McMurray's (n = 1232) and Apley's (n = 479).	Methodological quality varied from poor to fair among studies, affecting test performance. Future studies should, where possible, utilize larger samples of individuals without meniscal lesions to better estimate test specificity and thus more accurately identify optimal clinical tests.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75% Risk of bias - one or more key results were based on studies with a majority having a high risk of bias Spectrum bias was common, where subjects often had only meniscal type injuries, affecting the internal validity of many studies. Therefore, high diagnostic odd ratios and favourable likelihood ratios must be interpreted with caution. Verification bias, also known as 'test referral bias' is another potential source of error in the data. Variation between primary studies in both index and reference test thresholds is an obvious limitation to the current literature.

Nunes GS, Stapait EL, Kirsten MH, et al. Clinical test for diagnosis of patellofemoral pain syndrome: Systematic review with meta-analysis. Phys Ther Sport. 2013;14(1):54-9.	23232069	Meta-Analysis; Review	lowlevel of evidence	To investigate the diagnostic accuracy of clinical and functional tests used to diagnose PFPS through a systematic review.	The search identified 16,169 potential studies and five studies (496 patients) met the eligibility criteria. Inclusion criteria were studies evaluating the accuracy of clinical and functional tests for diagnosing PFPS were included. No limits regarding date of publication or language were established. Exclusion criteria: studies where the patients had undergone surgery in lower limbs affected by PFPS; studies evaluating the accuracy of diagnostic imaging tests; studies in which diagnosis was determined using questionnaires; studies in which the participants had other associated diseases (such as osteoarthritis and ligament injuries). The authors also excluded studies evaluating the accuracy of tests in individuals with chondromalacia patellae, because in this condition there is structural injury to the cartilage and it is thus not considered PFPS.	The 5 studies in this review analyzed 25 tests intending to accurately diagnose PFPS. Two tests were analyzed in two studies and were possible to perform a meta-analysis. Within the five studies included, one study had high methodological quality, two studies had good methodological quality and two studies had low methodological quality. Squatting was the most sensitive test (91%), with the lowest LR- (0.2) and highest PV- (74%). The vastus medialis coordination test had the best specificity among all tests (93%); the patellar tilt had the highest LR+ (5.4) and the active instability test had the highest PV+ (100%).	Future diagnostic studies should focus on the sample homogeneity and standardization of tests analyzed so future systematic reviews can determine with more certainty the accuracy of the tests for diagnosis of PFPS.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75%
Phelan N, Rowland P, Galvin R, et al. A systematic review and meta-analysis of the diagnostic accuracy of MRI for suspected ACL and meniscal tears of the knee. Knee Surg Sports Traumatol Arthrosc. 2016;24(5):1525-39.	26614425	Meta-Analysis; Review	Moderate level of evidence	To determine the diagnostic accuracy of magnetic resonance imaging (MRI) and ultrasound (US) in the diagnosis of anterior cruciate ligament (ACL), medial meniscus and lateral meniscus tears in people with suspected ACL and/or meniscal tears.	Studies were included if they met the following inclusion criteria: (1) prospective cohort or cross-sectional studies; (2) evaluated MRI and/or US in the diagnosis of ACL and/ or meniscal tears; (3) used arthroscopy or arthrotomy as the reference standard; and (4) reported findings that enabled the calculation of the number of true-positive, true-negative, false-positive and false-negative values for the diagnostic accuracy of both index tests. Studies that included patients of 13 years and older but were of a predominantly adult population were included. The following exclusion criteria were used: (1) retrospective design; (2) predominantly paediatric patients; (3) asymptomatic patient study group; (4) participants suspected of a specific pathology, e.g. bucket handle tear of the meniscus; and (5) evidence of verification bias, whereby the result of the index test may have excluded patients from undergoing the reference standard. Ultimately 21 studies with 1339 patients were included.	The results of Bayesian analysis showed that a positive finding on MRI doubles the probability of an ACL tear across all clinical settings from 35.7 % (95 % CI 25.9– 45.5 %) to 85.8 % (95 % CI 82.0–90.0 %). The summary estimates of sensitivity and specificity of MRI were 87 % (95 % CI 77–94 %) and 93 % (95 % CI 91–96 %), respectively, for ACL tears; 89 % (95 % CI 83–94 %) and 88 % (95 % CI 82–93 %), respectively, for medial meniscal tears; and 78 % (95 % CI 66–87 %) and 95 % (95 % CI 91–97 %), respectively, for lateral meniscal tears. The sensitivity of MRI for lateral meniscal tears is lower, than for ACL and medial meniscal tears, but the specificity was higher, 95 % (95 % CI 0.91–0.97). The ROC curve demonstrates wide variability in study findings for the sensitivity of the test. There were an insufficient number of studies that evaluated US to perform a meta-analysis.	This review highlights the lack of high-quality evidence in support of a common diagnostic test. While MRI will continue to play an import role in the management of ACL and meniscal injuries, surgeons should be aware of the level of evidence supporting its use when interpreting results and should question its applicability in the context of their clinical setting.	Risk of bias - one or more key results were based on studies with a majority having a high risk of bias The risk of bias in most studies is high or unclear in relation to the reference standard. Concerns regarding the applicability of patient selection are also present in most studies. Inclusion of studies with long time intervals between the index test and the reference standard is a potential source of weakness.
Smith BE, Thacker D, Crewesmith A, et al. Special tests for assessing meniscal tears within the knee: a systematic review and meta-analysis. Evid Based Med. 2015;20(3):88-97.	25724195	Meta-Analysis; Research Support, Non-U.S. Gov't; Review	Moderate level of evidence	To synthesise the most current literature on the diagnostic accuracy of special tests for meniscal tears of the knee in adults.	All studies examining the accuracy of special tests in diagnosing meniscal tears of the knee in adults (16 years of age or older) were included. The study must have had at least one clinical special test, must have reported specificity and sensitivity and been written in English. Special tests included McMurray's test, Apley's test, Thessaly's test or JLT. The tests must not have been carried out under anaesthetics or on cadavers or been part of a composite examination. Clinical diagnosis by MRI or arthroscopy surgery was considered the gold standard reference test. Studies were excluded due to participants not meeting the criteria, the study design not meeting the criteria and due to no outcome data being recorded. 9 studies and 1234 patients were ultimately included.	The methodological quality of the included studies was generally poor. Three special tests were included in the meta-analysis: McMurray's, JLT and Thessaly at 20° knee flexion. McMurray's had a pooled sensitivity of 61% (95% CI 45% to 74%) and a pooled specificity of 84% (95% CI 69% to 92%). JLT had a pooled sensitivity of 83% (95% CI 73% to 90%) and a pooled specificity of 83% (95% CI 61% to 94%). Thessaly 20° had a pooled sensitivity of 75% (95% CI 53% to 89%) and a pooled specificity of 87% (95% CI 65% to 96%). LR+ of 3.2, 4.0 and 5.6, and LR− of 0.52, 0.23 and 0.28 for McMurray's, JLT and Thessaly 20°, respectively. LR+ of between 0.2 and 0.5 indicate only small shifts in probability post-test. Two of the tests, JLT and Thessaly 20°, had a high heterogeneity I2 score, with McMurray's having a moderate between-study heterogeneity I2 score. These data, coupled with the relatively low shifts in probability with the likelihood ratios, show that the three tests analyzed will not accurately diagnose a torn meniscus. Apley's test had a combined (medial and lateral) sensitivity of 84% and 20% and specificity of 79% and 84%. Thessaly 5° had a combined (medial and lateral) sensitivity of 35% and 65% and specificity of 89% and 82%.	The results of this systematic review indicate that the accuracy of McMurray's, Apley's, JLT and Thessaly to diagnose meniscal tears remains poor. This conclusion must be taken with caution since frequent methodological design flaws exist within the included studies, most studies suffered from various biases, and between-study heterogeneity makes pooled data unreliable. The latest research surrounding meniscal tears within asymptomatic patients, and modern thinking with regard to pain and lack of efficacy for surgical treatment starts to challenge the need for such a diagnosis and use of special tests. This review cannot recommend the use of special tests for diagnosing meniscal tears. It is unclear if further research would considerably alter this conclusion.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75% Risk of bias - one or more key results were based on studies with a majority having a high risk of bias Limitation of the included studies is that all but one study used arthroscopy as the gold standard test, and it is thought that this also introduces verification bias. In general, wide variation in test procedures were applied to a wide variety of patients including different ages, sex ratios and duration of symptoms. There was also wide variation in how the special tests were performed. Another possible cause of heterogeneity between included studies is the differing prevalence rates within each sample.

Smith C, McGarvey C, Harb Z, et al. Diagnostic Efficacy of 3-T MRI for Knee Injuries Using Arthroscopy as a Reference Standard: A Meta-Analysis. AJR Am J Roentgenol. 2016;207(2):369-77.	27248283	Meta-Analysis	Moderate level of evidence	To assess the evidence for the diagnostic efficacy of 3-T MRI for meniscal and anterior cruciate ligament (ACL) injuries in the knee using arthroscopy as the reference standard and to compare these results with the results of a previous meta-analysis assessing 1.5-T MRI.	The online Cochrane Library, MEDLINE, and PubMed databases were searched. One hundred one studies were identified by the search strategy, and 13 studies were included in the review. All 13 studies had high methodologic integrity and low risk of bias using the QUADAS-2 tool. The studies included 1197 patients with a mean age of 41.9 years. Studies were included if 3-T MRI had been used to diagnose medial meniscal, lateral meniscal, or ACL injuries and if the MRI findings were correlated with arthroscopic findings; both prospective and retrospective studies were eligible for inclusion. Studies were excluded if MRI field strengths other than 3 T were used, if a new scanning protocol for 3-T MRI was used without inclusion of the results of a previously established control protocol, if nonhuman subjects were used, and if the full text of the article or a translation of the full text was not available in the English language. Case reports, review articles, and comments about existing studies were excluded.	Ten of the 13 studies were eligible for meta-analysis. The mean sensitivity and mean specificity of 3-T MRI for knee injuries by location were as follows: medial meniscus, 0.94 (95% CI, 0.91-0.96) and 0.79 (95% CI, 0.75-0.83), respectively; lateral meniscus, 0.81 (95% CI, 0.75-0.85) and 0.87 (95% CI, 0.84-0.89); and ACL, 0.92 (95% CI, 0.83-0.96) and 0.99 (95% CI, 0.96-1.00). The specificity of 3-T MRI for injuries of the lateral meniscus was significantly lower than that of 1.5-T MRI (p = 0.0013).	The results of this study show that 3-T MRI scanners have excellent diagnostic efficacy for ACL and meniscal injuries. However, the diagnostic studies published through 2013 do not provide any evidence that 3-T scanners are superior when compared with a previous me ta-analysis of studies performed using 1.5-T machines. In fact, the authors' analysis shows that the specificity of 3-T MRI is lower than that of 1.5-T MRI with regard to the diagnosis of lateral meniscal tears. Advances in technology and software developments may improve the diagnostic efficacy of 3-T MRI scanners in the future to a point at which it is greater than that of 1.5-T scanners.	Risk of bias - one or more key results were based on studies with a majority having a high risk of bias All studies except one have a high risk of bias. The limitations of this meta-analysis are dependent on the limitations of the studies included. Although 10 studies were able to have their data pooled for medial and lateral meniscal injuries, only three were suitable for ACL injuries. The results for ACL injuries are therefore more open to bias than the results for the meniscal injuries.
Smith TO, Drew BT, Toms AP, et al. Accuracy of magnetic resonance imaging, magnetic resonance arthrography and computed tomography for the detection of chondral lesions of the knee. Knee Surg Sports Traumatol Arthrosc. 2012;20(12):2367-79.	22270676	Review	Moderate level of evidence	To assess the diagnostic test accuracy of magnetic resonance imaging (MRI), magnetic resonance arthrography (MRA) and computed tomography arthrography (CTA) for the detection of chondral lesions of the patellofemoral and tibiofemoral joints.	Twenty-seven studies assessing 2,592 knees from 2,509 patients were included. Studies assessing the diagnostic test accuracy (sensitivity/specificity) of MRI or MRA or CTA for the assessment of adults with chondral (cartilage) lesions of the knee (tibiofemoral/patellofemoral joints) with surgical comparison (arthroscopic or open) as the reference test were included. Studies assessing cadaveric knees or animal models were excluded. Studies that did not use surgery as the reference standard or did not aim to assess the diagnostic accuracy (sensitivity/specificity) were excluded.	Overall, the specificity of radiological measurements was greater than their sensitivity for the detection of both patellofemoral and tibiofemoral joint lesions. The pooled meta-analysis indicated that MRA and CTA were superior in the detection of patellofemoral joint chondral lesions compared with MRI investigations. MRA reported a pooled sensitivity of 0.70 (95% CI: 0.57–0.81) and specificity of 0.99 (0.97–1.00), CTA sensitivity was 0.80 (95% CI: 0.70–0.88) and specificity 0.99 (95% CI: 0.95–1.00), whilst MRI reported a sensitivity of 0.74 (0.71–0.77) and a specificity of 0.95 (0.94–0.95). The sROC plot indicated superior diagnostic test accuracy for the detection of tibiofemoral over patellofemoral joint lesions with the tibiofemoral joint reported a sensitivity for 0.88 (95% CI: 0.86–0.89) and specificity of 0.82 (0.81–0.83), compared with 0.74 (95% CI: 0.71–0.77) and 0.95 (95% CI: 0.94–0.95) for patellofemoral joint sensitivity and specificity, respectively. Higher field strength MRI scanner and grade four lesions were more accurately detected compared with lower field-strength and grade one lesions. There appeared no substantial difference in diagnostic accuracy between the interpretation from musculoskeletal and general radiologists when undertaking an MRI review of tibiofemoral and patellofemoral chondral lesions.	Currently MRA, CTA and MRI can only be considered to be accurate for detecting the more advanced chondral lesions. The sensitivity for less severe lesions is limited. Further study to assess the diagnostic test accuracy of newer MR pulse sequences may be indicated to as the technology advances. Until then, there is little indication to replace the 'gold-standard' arthroscopic investigation with any of these radiological investigations.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75%; Risk of bias - one or more key results were based on studies with a majority having a high risk of bias; There was heterogeneity in methods of analysis and data presentation. Most of the included studies have a high risk of bias. A recurrent limitation to the studies was that the MRI results were available to the surgeons prior to the arthroscopic 'reference standard' procedure. Lack of detail provided by most included studies regarding their arthroscopic technique. Arthroscopy is operator-dependent and therefore the reliability of arthroscopy for the detection of chondral lesions may be affected by the training and experience of the orthopaedic surgeon who undertakes the reference standard.
Zarringam D, Saris DB, Bekkers JE. The value of SPECT/CT for knee osteoarthritis: A systematic review. Cartilage. 2021; 12(4):431-437.	31204483	Systematic review	low level of evidence	To review the added value of SPECT/CT in the diagnostic algorithm of knee osteoarthritis.	Any trial potentially focusing on the diagnostic value of SPECT/CT for knee osteoarthritis was identified as relevant. Exclusion criteria were case reports, animal studies, and cadaver studies. Furthermore, only studies with a patient population with knee osteoarthritis or possible knee osteoarthritis were included. Other pathology such as meniscal tears, chondral lesions, and anterior cruciate ligament lesions were excluded unless they were found in an osteoarthritis population.	The retrieved articles were screened for relevance on title and abstract. After a full text screening, relevant articles were assessed on risk of bias and applicability based on the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies–2).10 First, the research question was evaluated in general. Second, a flow diagram of the study was drawn. Third, 3 domains were assessed on bias and applicability according to the QUADAS-2, including patient selection, index test, and reference standard. The fourth domain, concerning patients flow and timing of the diagnostic tests, was only assessed on risk of bias. To answer our research question, articles were included based on the risk of bias and applicability.	In total, 9 trials were included. Results found that the use of SPECT/CT might objectify some clinical knee osteoarthritis symptoms. It could correlate with findings on plain radiography and magnetic resonance imaging. Furthermore, there is some evidence SPECT/CT gives additional information compared with these imaging modalities; however, superiority is not proven. The uptake on SPECT could predict the intraoperative macroscopic findings. Yet the clinical relevance remains unclear. The authors conclude that there is no strong evidence SPECT/CT should play a role in the diagnosing and decision-making processes of knee osteoarthritis. Yet there is evidence suggesting SPECT/CT might give additional information in the diagnosing process. More research would be of added value to answer this research question.	A limitation of this systematic review is that all studies with SPECT, without the CT component, were also analyzed due to the lack of literature. All these studies found correlations in the use of SPECT in some way. Since SPECT/CT is a more accurate diagnostic tool than SPECT, these correlations should be taken into account in favor of SPECT/CT.
Zhang M, Min Z, Rana N, et al. Accuracy of magnetic resonance imaging in grading knee chondral defects. Arthroscopy. 2013;29(2):349-56.	22906758	Meta-Analysis	low level of evidence	To determine the accuracy of routine magnetic resonance imaging (MRI) in the grading of knee cartilage lesions through a meta-analysis.	454 patients in 8 studies. The inclusion criteria were as follows: (1) the purpose of the study was to investigate the diagnostic accuracy of MRI in knee cartilage lesions; (2) patients were clinically suspected of having knee degeneration or traumatic damage; (3)arthroscopic outcome was used as a reference standard;(4) 6 articular surfaces (medial and lateral femoral condyle, medial and lateral tibial plateau, trochlea, and patella) were evaluated separately; and (5) the grading classification was definite and unified on MRI and arthroscopy. Studies with inadequate mapping of chondral defects and magnetic resonance arthrography or contrast-enhancement MRI were excluded. If there were overlapping patient populations (confirmed by contacting the corresponding author), the studies with the most complete data were included.	The overall sensitivity, specificity, diagnostic odds ratio, positive likelihood ratio, and negative likelihood ratio were 75% (95% confidence interval [CI], 62% to 84%), 94% (95% CI, 89% to 97%), 47 (95% CI, 18 to 122), 12.5 (95% CI, 6.5 to 24.2), and 0.27 (95% CI, 0.17 to 0.42), respectively.	The results showed that MRI was effective in discriminating normal morphologic cartilage from disease but was less sensitive in detecting knee chondral lesions (higher than grade 1). The negative results of MRI should not prevent a diagnostic arthroscopy.	Heterogeneity - one or more key results were highly variable with studies concluding opposite things or with I^2 statistic > 75%. Publication bias is found in this study because unpublished findings were not included (internal validity). Finally, the small number of studies decreases the power to detect true differences between groups (generalizability).