Bibliographic Cite	PMID Link	Literature Type	AMSTAR Appraisal	Level of Evidence	Purpose	Population	Intervention and Outcome Measures	Results/ Recommendations	Study Limitations
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	well developed	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 Rheumatology 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.
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	moderately well developed	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 for detection of patellar versus trochlear defects and similarly 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	moderately well developed	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%).
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	moderately well developed	low level 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%

Citation	ID	Type	Development	Level of evidence	Purpose	Methods	Results	Conclusions	Limitations
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	well developed	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 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	moderately well developed	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 meta-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	well developed	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.
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	Well developed study	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).