Bibliographic Cite	Literature Type	Level of Evidence	Purpose	Population	Intervention and Outcome Measures	Results/ Recommendations	Study Limitiations
Alsaleh K, Ho D, Rosas-Arellano MP, et al. Radiographic assessment of degenerative lumbar spinal stenosis: is MRI superior to CT? Eur Spine J. 2017;26(2):362-7.	Comparative Study	Moderate	To determine the reliability and dependability of magnetic resonance imaging (MRI) and computerized tomography (CT) in the assessment of lumbar spinal stenosis and correlate the qualitative assessment to both a quantitative assessment and functional outcome measures.	54 Canadian patients (mean age was 65 years, age range 23–85 years) who were referred for surgical consultation of degenerative lumbar spinal stenosis and had CT and MRI performed	One orthopaedic spine surgeon and one neuroradiologist reviewed the CTs and MRIs in a randomized order. They were blinded to the radiology report associated with the scan, as well as the patient's identifying information. The evaluation was performed using both T2 weighted sagittal and axial images. There were 23 (43 %) males and 31 (57 %) females. For all patients, the CT and MRI were performed within 1 year of each other (mean = 8.1 months, SD = ± 5.42 months). In most cases the CT scan was performed prior to the MRI.	Almost perfect intra-observer reliability for MRI was achieved by the two expert reviewers (k = 0.91 for surgeon and k = 0.92 for neuro-radiologist). For CT, substantial intra-observer agreement was found for the surgeon (k = 0.77) while the neuro-radiologist was higher (k = 0.96). For both CT and MRI the standardized qualitative assessment used by the two expert reviewers had a better inter-observer reliability than that between the expert reviewers and the general reporting radiologist, who did not utilize a standardized assessment system. When the qualitative assessment was compared directly, CT overestimated the degree of stenosis 20–35% of the time ($p<0.05$) while MRI overestimated the degree of stenosis 2–11 % of the time ($p<0.05$). No correlation was found between qualitative and quantitative analysis with functional status. Authors concluded that MRI is a more reliable imaging tool than CT for the diagnosis and gradation of lumbar spinal stenosis. Both experience of the reader and the standardization of a qualitative assessment are influential to the reliability. CT in general is prone to over estimation of spinal stenosis severity and therefore MRI is the modality of choice for pre-operative planning. Further work is required in order to recommend which is the modality of choice for a primary diagnostic screening test.	The study limitations were small sample size and no comments was made about how indeterminate results were handled
Berg L, Thoresen H, Neckelmann G, et al. Facet arthropathy evaluation: CT or MRI? Eur Radiol. 2019; 29(9):4990-4998.	Multicenter retrospective study	Low	To assess the reliability of lumbar facet arthropathy evaluation with computed tomography (CT) or magnetic resonance imaging (MRI) in patients with and without lumbar disc prosthesis and to estimate the reliability for individual CT and MRI findings indicating facet arthropathy.	8-year followup imaging of 114 of 173 patients randomized to disc prosthesis surgery or non-operative treatment in a prospective multicenter trial in 2004–2007. Patients eligible for the trial were 25–55 years, had chronic LBP, and had disc degeneration at L4/L5 and/or L5/S1 on MRI. Patients were excluded if they had disc degeneration at any higher level (L1–L4) or had spondylolysis, spondylolisthesis, arthritis, osteoporosis, prior fracture L1-S1, prior spinal fusion, deformity, or symptomatic disc herniation / spinal stenosis.	Metal-artifact reducing CT and MRI protocols were performed at follow-up of 114 chronic back pain patients treated with (n = 66) or without (n = 48) lumbar disc prosthesis. Three experienced radiologists independently rated facet joint space narrowing, osteophyte/hypertrophy, erosions, subchondral cysts, and total grade facet arthropathy at each of the three lower lumbar levels on both CT and MRI, usingWeishaupt et al's rating system. Overall kappa and (due to low prevalence) prevalence- and bias-adjusted kappa were calculated to assess interobserver agreement.	Interobserver agreement on total grade facet arthropathy was moderate at all levels with CT (kappa 0.47–0.48) and poor to fair with MRI (kappa 0.20–0.32). Mean prevalence- and bias-adjusted kappa was lower for osteophyte/hypertrophy versus other individual findings (CT 0.58 versus 0.79–0.86, MRI 0.35 versus 0.81–0.90), higher with CT versus MRI when rating osteophyte/hypertrophy (0.58 versus 0.35) and total grade facet arthropathy (0.54 versus 0.31), and generally similar at levels with versus levels without disc prosthesis. The authors conclude that interobserver agreement on facet arthropathy was moderate with CT and better with CT than with MRI. Disc prosthesis did not influence agreement. A more reliable grading of facet arthropathy requires a more consistent evaluation of osteophytes/hypertrophy.	Limitations were the fixed, moderate sample size and the low prevalence of some findings, implying few presentations of these findings to rate.
Brusko GD, Perez-Roman RJ, Tapamo H, et al. Preoperative SPECT imaging as a tool for surgical planning in patients with axial neck and back pain. Neurosurg Focus. 2019; 47(6):E19.	Retrospective study	Low	To describe the largest series to date of patients with axial neck or back pain who underwent preoperative SPECT imaging for surgical planning, and to examine the effectiveness of surgical intervention based on patient-reported improvements during a 1-year follow-up period.	A total of 23 patients underwent spine surgery on a hypermetabolic level following identification on SPECT imaging. The average age at time of surgery was 60.0 ± 11.0 years; 15 patients (65.2%) were male and 8 were female. All patients with a diagnosed neoplasm, infections, inflammatory spondyloarthropathy, history of trauma, or any pathological fracture were excluded.	Patients who underwent spine surgical intervention for spondylosis with primary symptoms of axial neck or back pain and who had evidence of hypermetabolic foci on spinal SPECT imaging were included. Only those patients who subsequently underwent surgery on a spinal level associated with increased radiotracer uptake were included in the analysis. Patient baseline and demographic information, and data pertaining to SPECT imaging, surgical planning, and postoperative care were collected and analyzed.	A total of 53 spinal levels were treated, with an average of 2.30 levels treated per patient. All patients underwent fusion surgery, either lumbar (n = 14), with interbody fusion most commonly used (64.2%); or cervical (n = 9), with anterior cervical discectomy and fusion (66.6%) being the most common. The average length of hospital stay was 3.45 ± 2.32 days. One patient developed a wound infection postoperatively, requiring readmission. At the 3-month follow-up, 18 patients (78.3%) reported clinical improvement in pain. Eleven patients (47.8%) reported complete symptom resolution at the 6-month follow-up. At 1 year postoperatively, 19 patients (82.6%) reported significant relief of their symptoms following surgery. The authors conclude that SPECT imaging may be a useful adjunct to guide surgical planning, resulting in substantial clinical improvement following surgery.	The authors note several important limitations to this study: "First, our series is small, with just 23 patients included. However, traditional management of axial neck or back pain is nonsurgical and thus, our series is the largest one to date examining surgical outcomes. Second, a subset of patients had involvement of multiple spinal levels. Therefore, the inability to treat all degenerated levels and the progressive nature of the osteoarthritic disease may be responsible for a lack of clinical improvement in these select patients over time. Third, biases related to retrospective reviews must be taken into account, and the results of our single-surgeon, single-institution series may not be generalizable to other patient populations.
Chang MY, Lee SH, Ha JW, et al. Predicting bone marrow edema and fracture age in vertebral fragility fractures using MDCT. AJR Am J Roentgenol. 2020; 215(4):970-977.	Single center retrospective	Low	To evaluate whether CT features can predict bone marrow edema (BME) on MRI and fracture age in vertebral fragility fractures.	Patients who had undergone both spine CT and spine MRI within 7 days were recruited. 189 thoracolumbar compression fractures in 103 patients (14 men, 89 women; mean age, 76 years). All patients were included, regardless of outpatient, inpatient, or emergency department setting.	The presence and extent of BME were assessed on MRI to divide fractures into those with and without BME. The group with BME was then classified for subgroup analysis into fractures with extensive BME (comprising 50% or more of the vertebral body) and those with BME comprising less than 50% of the vertebral body. On CT, five features (presence of cortical or endplate fracture line, presence of trabecular fracture line, presence of condensation band, change in trabecular attenuation, and width of paravertebral soft-tissue change) were analyzed.	All five CT findings were predominantly seen in fractures with BME ($p < 0.001$). Elevated trabecular attenuation, presence of a cortical or endplate fracture line, and paravertebral soft-tissue width showed excellent diagnostic indication for fractures with BME (ROC AUCs: 0.990, 0.976, and 0.950, respectively). In the subgroup with extensive BME, paravertebral soft-tissue width was significantly higher, whereas the change in trabecular attenuation was lower compared with those with BME comprising less than 50% of the vertebral body ($p < 0.001$). When BME was present, fracture age was not significantly different between the two subgroups, and only greater trabecular attenuation elevation was predictive of older fracture age on linear mixed model analyses ($p < 0.001$). Interobserver agreement was good for the trabecular fracture line factor and excellent for all other factors. The authors conclude that CT features accurately correlate with the presence and extent of BME in vertebral fragility fractures. Elevation of trabecular attenuation was the only significant image predictor of fracture age.	First, its retrospective study design including fracture age analysis was the main limitation of this study. Second, multiple myeloma can be present with normal bone marrow signal intensity. VCFs from multiple myeloma can appear benign in 38% of cases. Even symptomatic acute or subacute compression fracture in patients with multiple myeloma may not show BME. Finally, to simplify CT analyses, authors evaluated vertebral bodies only with fractures that were confirmed on MRI. Thus, they could not evaluate the false-positive rate of CT diagnosis of VCFs.

Ekedahl H, Jonsson B, Annertz M, et al. The 1-Year Results of Lumbar Transforaminal Epidural Steroid Injection in Patients with Chronic Unilateral Radicular Pain: The Relation to MRI Findings and Clinical Features. Am J Phys Med Rehabil. 2017;96(9):654-62.	Clinical Trial	Moderate	In patients with chronic radicular pain, we aimed to evaluate subgroup differences in 1-yr response to transforaminal epidural steroid injection.	71 Swedish patients, with chronic (>3-mo symptom duration) lumbar radicular pain; excluding TESI within last year, bilateral leg pain, central spinal stenosis, previous lumbar fusion surgery, neurological disease, cancer, pregnancy, autoimmune disease, and any known allergy to treatment agents.	To evaluate subgroup differences in 1-yr response to 1-yr response to transforaminal epidural steroid injection (TESI) by relating MRI findings and clinical test results, baseline (BL) characteristics, the number of TESIs performed, and conservative treatment to reduction in leg pain (primary outcome) and self-reported disability (co-primary outcome) and to evaluate potential predictors in relation to lumbar surgery post TESI (secondary outcome). Magnetic resonance images were analyzed by a well-experienced radiologist blinded to all clinical information except the level of TESI.	High-grade subarticular nerve compression predicted the 1-yr improvement in both visual analogue scale leg pain (P = 0.046) and Oswestry Disability Index (P = 0.027). Low age (P < 0.001), short duration of leg pain (P =0.015), and central/subarticular disc herniation (P = 0.017) predicted improvement in Oswestry Disability Index.	Limitations included: Patients with indeterminate results from the diagnostic test were excluded, or no comment was made about how indeterminate results were handled; Single reader, or no inter-reader reliability was calculated; Small sample size, single center, applicability, lack of control group. There are certain limitations to this study. Firstly, because of low frequency in the high-grade subarticular nerve compression group, the results must be cautiously interpreted. Secondly, TESI was performed under fluoroscopy but without contrast medium, which potentially can lead to inaccurate needle placement and, consequently, have an impact on the results. Lastly, selection bias in the inclusion process could have occurred because, in some few cases, surgery was decided upon without a previous TESI.
El Barzouhi A, Vleggeert-Lankamp CL, Lycklama ANGJ, et al. Magnetic resonance imaging interpretation in patients with sciatica who are potential candidates for lumbar disc surgery. PLoS ONE. 2013 8(7):e68411.	Observational study	High	The objective of this study was to assess MRI observer variation in patients with sciatica who are potential candidates for lumbar disc surgery	599 patients were screened for the study, 395 patients considered eligible for inclusion underwent MRI of whom 283 patients were randomized and 112 not. Mean age of the study population was 43.2 years. Patients for this study were patients with 6 to 12 weeks of sciatic symptoms being so severe that they were eligible for surgery according to their family practitioners and were therefore referred to a neurologist. Patients were excluded if they were presenting with cauda equina syndrome, insufficient strength to move against gravity, identical complaints in the previous 12 months, previous spine surgery, pregnancy, severe coexisting disease or if they were not between 18 to 65 years of age. All participants who were not meeting one or more of the aforementioned exclusion criteria underwent MRI.	Two neuroradiologists and one neurosurgeon independently evaluated all MRIs. A four point scale was used for both probability of disc herniation and root compression, ranging from definitely present to definitely absent. Multiple characteristics of the degenerated disc herniation were scored. For inter-agreement analysis absolute agreements and kappa coefficients were used. Kappa coefficients were categorized as poor (,0.00), slight (0.00– 0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) and excellent (0.81–1.00) agreement. MRI scans were performed in all 9 participating hospitals using standardized protocols tailored to a 1.5 Tesla scanner. Sagittal T1 and axial T1 spin echo images of the lumbar spine were acquired. In addition, T2 weighted sagittal and axial series were obtained. When all three observers finished reading the images they repeated the MRI evaluation for ten percent of the evaluated images to provide intra-observer reliability data. To assess the intra- and inter-observer reliability, the authors used percentages of absolute agreement and kappa coefficients.	Excellent agreement was found on the affected disc level (kappa range 0.81-0.86) and the nerve root that most likely caused the sciatic symptoms (kappa range 0.86-0.89). Interobserver agreement was moderate to substantial for the probability of disc herniation (kappa range 0.57-0.77) and the probability of nerve root compression (kappa range 0.42-0.69). Absolute pairwise agreement among the readers ranged from 90-94% regarding the question whether the probability of disc herniation on MRI was above or below 50%. Generally, moderate agreement was observed regarding the characteristics of the symptomatic disc level and of the herniated disc. CONCLUSION: The observer variation of MRI interpretation in potential candidates for lumbar disc surgery is satisfactory regarding characteristics most important in decision for surgery. However, there is considerable variation between observers in specific characteristics of the symptomatic disc level and herniated disc.	Low risk of bias: *patient selection increases the likelihood of disease relative to guideline target population selection bias wilol likely increase concordance *patient selection based on impression of the neurologist which is atypical in our target population * non US * large number of excluded patients *unclear if screening was consecutive. * no reference standard however primary outcome is inter-rater reliability An important limitation of the study is the number of observers, in particular the inclusion of only one non-radiologist, which limits the statistical power of the observer variation. Relatively homogenous study sample.
Graves JM, Fulton-Kehoe D, Jarvik JG, et al. Early imaging for acute low back pain: one-year health and disability outcomes among Washington State workers. Spine. 2012 37(18):1617-27.	Prospective cohort study	Moderate	To evaluate the association of early imaging and health and disability status 1 year following acute low back injury, among a population-based sample of Washington State workers' compensation claimants.	1226 participants, 18.6% received early MRI. Most (77.9%) had mild/major sprains and 22.1% had radiculopathy. Participants were restricted to adults (older than 18 yr of age) with an accepted claim, with at least 4 missed workdays due to injury, and no hospitalization following the injury. Of the 4354 claimants identifi ed by the D-RISC study, 49.3% agreed to participate, 27.1% could not be contacted, 20.9% declined, and 8.8% were ineligible due to language limitations, lacked disability compensation, or were excluded for other reasons (e.g., incomplete records).	Used administrative claims and interview data. Multivariable regression methods were used to estimate change in health outcome scores, the relative risk of disability at 1 year, and the rate of recovery 1 year after injury. The Roland-Morris Disability Questionnaire (RDQ) consists of 24 yes/no items to assess the physical disability due to LBP The 36-Item Short Form Health Survey (SF-36v2) is a generic health profi le that measures dimensions of health-related quality of life, each scored from 0 (worst) to 100 (best Workers' pain intensity was evaluated at each interview, using an adaptation of the pain intensity subscale of the Graded Chronic Pain Scale.	Of 1226 participants, 18.6% received early MRI. Most (77.9%) had mild/major sprains and 22.1% had radiculopathy. Participants with early MRI differed significantly at baseline in pain, function, and psychosocial variables. After adjusting for covariates, early imaging was not associated with substantial differences in 1-year health outcomes for sprains or radiculopathy. For workers with mild/major sprain, early imaging was associated with a 2-fold increase in the likelihood of work disability benefits at 1 year (adjusted relative risk: 2.03, 95% confidence interval: 1.33-3.11). Early imaging was not associated with an increased risk of long-term disability for workers with radiculopathy (adjusted relative risk: 1.31, 95% confidence interval: 0.84-2.05). For both groups, early MRI was associated with longer disability duration (P < 0.001). The authors conclude that, among workers with LBP, early MRI is not associated with better health outcomes and is associated with increased likelihood of disability and its duration. These associations warrant further testing in a randomized controlled trial. Our findings suggest that adherence to evidence-based guidelines is an important factor in ensuring that workers receive the highest quality care for occupational injuries.	More than 70% of workers with early MRI completed the baseline interview after receiving imaging. This could contribute to biased interview responses, if the respondent received and was influenced by the imaging results. Finally, although the authors statistically adjusted for multiple individual-level variables, residual confounding may remain. Future research should address this topic using a randomized controlled design, which could validate our findings.

He X, Zhao L, Guo X, et al. Differential diagnostic value of 18F-FDG PET/CT for benign and malignant vertebral compression fractures: Comparison with magnetic resonance imaging. Cancer Manag Res. 2018; 10:2105-2115.	Single center retrospective	Low	To evaluate the differential diagnostic value of 2-[fluorine-18]-fluoro-2-deoxy-D-glucose (18F-FDG) positron emission tomography (PET)/computed tomography (CT) for benign and malignant vertebral compression fractures (VCFs), where the diagnostic accuracy of 18F-FDG PET/CT was compared with MRI.	87 patients (55 males and 32 females; mean age: 68 years; age range: 60–79 years) with 116 VCFs were retrospectively evaluated. Forty-two cases did not have any primary malignancy, whereas 45 cases had histories of malignant tumors	MRI was performed in all the 87 patients, whereas 18F-FDG PET/CT was executed in 51 patients. Three malignant features (convex posterior cortex, epidural mass formation, and pedicle enhancement) from MRI and the maximum standardized uptake value (SUVmax) from 18F-FDG PET/CT were evaluated in benign and malignant VCFs, respectively. Sensitivity, specificity, positive predictive value, and negative predictive value of MRI and 18F-FDG PET/CT were compared in the differentiation of malignant from benign VCFs.	Results showed that the sensitivity and specificity for predicting malignant VCFs were 75.6% and 77.3% for convex posterior cortex, 82.9% and 813% for epidural mass formation, and 85.7% and 70.8% for pedicle enhancement. 18F-FDG PET/CT demonstrated higher sensitivity (100%) but lower specificity (38.9%) as compared to MRI with regard to differentiation between benign and malignant VCFs. A significant difference in the SUVmax values was observed between the benign and malignant fractures (2.9 ± 1.0 vs 5.0 ± 1.8, P < 0.01). Besides the value of SUVmax, it has been noticed that the FDG uptake pattern differed in malignant and benign fractures. The authors conclude that significant MRI findings such as convex posterior cortex, epidural mass formation, and pedicle enhancement are highly suggestive of malignancy. 18F-FDG PET/CT reliably differentiated the fractures of malignant from benign based on both SUVmax and 18F-FDG uptake pattern. In a situation where MRI findings are not diagnostic, 18F-FDG PET/CT provides additional information as it has high sensitivity and is semiquantitative.	Firstly, the optimal cutoff value of SUVmax as suggested is generated from a single-center retrospective study. Such a SUVmax value cannot be used as a general threshold for differentiating the lesions of malignant from benign. The recommended SUVmax threshold for practice required multicenter study with a large population. Additionally, in this study, the MR imaging protocol primarily used T1 and T2 sequences for the evaluation of VCFs, and additional sequences (such as DWI) were not routinely performed in spine MRI.
Jacobs JC, Jarvik JG, Chou R, et al. Observational study of the downstream consequences of inappropriate MRI of the lumbar spine. J Gen Intern Med. 2020; 35(12):3605-3612.	Retrospective matched cohort study	Low	To determine the downstream consequences of early imaging (MRI ordered in the first 6 weeks fo new episodes of uncomplicated non-specific low back pain.	This study used data from electronic health records of primary care clinics of the U.S. Department of Veterans Affairs. Included were patients seeking primary care for nonspecific low back pain without a red flag condition or an encounter for low back pain in the prior 6 months (N = 405,965).	Patients with early MRI were matched to patients without an early MRI using coarsened exact matching (CEM). Individuals exposed to an early MRI were assigned a value of 1. Unexposed individuals were assigned a weight that accounts for the number of exposed and unexposed observations within that strata and the total number of exposed and unexposed observations that were matched. Outcomes were lumbar surgery, prescription opioid use, acute health care costs, and last pain score recorded within 1 year of the index visit.Outcomes were measured from 43 to 365 days post-index visit, and costs were also measured for the first 42 days.	Early MRI was found to be associated with more back surgery (1.48% vs. 0.12% in episodes without early MRI), greater use of prescription opioids (35.1% vs. 28.6%), a higher final pain score (3.99 vs. 3.87), and greater acute care costs ($8082 vs. $5560), p < 0.001 for all comparisons. The association between early imaging and increased utilization was apparent even in a setting largely unaffected by incentives of fee-for-service care. Reduced imaging cost is only part of the motivation to improve adherence with guidelines for the use of MRI. Early scans are associated with excess surgery, higher costs for other care, and worse outcomes, including potential harms from prescription opioids.	Reliance on data gathered in normal clinical care and the potential for residual confounding despite the use of coarsened exact matching weights to adjust for baseline differences.
Jarvik J, Gold L, Comstock B, et al. Association of early imaging for back pain with clinical outcomes in older adults. JAMA. 2015; 313(11):1143-53.	Large cohort study	Moderate	To compare function and pain at the 12-month follow-up visit among older adults who received early imaging with those who did not receive early imaging after a new primary care visit for back pain without radiculopathy.	5239 patients 65 years or older (mean age 74.3) with a new primary care visit for back pain (2011-2013) in 3 US health care systems. We matched controls 1:1 using propensity score matching of demographic and clinical characteristics, including diagnosis, pain severity, pain duration, functional status, and prior resource use.	Primary outcome measure was back or leg pain-related disability measured by the modified Roland-Morris Disability Questionnaire (score range, 0-24; higher scores indicate greater disability) 12 months after enrollment.	RESULTS: Patients with early MRI/CT vs controls had statistically significant but not clinically meaningful differences o 2 measures: the early MRI/CT group had lower 6-month leg pain numeril rating scale scores (difference, −0.58 [95%CI, −1.07 to −0.089]) and higher 12-month Euro Qol5D-visual analog scale scores (difference,4.04 [95%CI, 0.92 to 7.15]). DELAYED DIAGNOSIS: The authors did not observe differences in proportions of patients with cancer diagnoses in the next year among patients receiving early imaging vs controls (Table 5). Among patients who underwent early imaging, only 1 of 16:30 (0.06%) had cancer (lymphoma) diagnosed on the early imaging study (lymph-adenopathy seen on MRI). In contrast, patients who underwent imaging diagnostics early had more fractures detected (2% in the early radiograph group vs 0.6% in the no early or no radiograph group; 0.9% in the early MRI/CT group vs 0% in the no early or no MRI/CT group). IMPLICATIONS: Our study demonstrates that older adults who had spine imaging within 6 weeks of a new primary care visit for back pain had pain and disability over the following year that was not different frommatched patients who did not undergo early imaging.	Potential for confounding by indication; ie, patients receiving early imaging had worse prognoses than patients not getting early imaging. The authors tried to minimize confounding through propensity matching. analysis. Second, our data on pain duration is limited by the overlap of the pain duration categories. Third, our baseline measures were administered up to 3 weeks after the index visit and thus could reflect responses to therapy since the index visit. The authors assumed that all index back pain. Fourth, patients who are more likely to ask for early imaging might also be more likely to use resources subsequently.
Kim Y, Lee E, Lee JW, et al. Clinical and imaging characteristics of patients with extreme low back pain or sciatica referred for spinal injection. Neuroradiology. 2019; 61(8):881-889.	Single center retrospective	Low	To analyze the causes of pain, imaging characteristics, and therapeutic effect of spinal injection in patients with extreme low back pain or sciatica.	381 consecutive patients with extreme low back pain or sciatica visiting our spinal intervention center between January and December 2017. Mean age of 67 years; 58% had spinal stenosis, 25% had herniated disc, 13% had facet osteoarthritis, 4% had osteoporotic compression fracture.	Clinical and imaging characteristics were analyzed. The treatment response, defined as a numerical pain rating scale decrease of ≥ 30%, was measured. Fisher's exact test was performed to identify the association between the injection response and subsequent lumbar surgery rate.	A herniated intervertebral disc was the most common disorder in patients < 50 years of age, while spinal stenosis was the most common in patients ≥ 50 years of age. The majority of lumbar pathologies occurred below L3/4. Spinal injection was found to be effective in 44.2% of cases. Those who responded to the injection showed a significantly lower rate of lumbar surgery within 6 months (P = 0.004). The authors conclude that those with extreme low back pain or sciatica had clinical and imaging characteristics similar to those with typical low back pain referred for spinal injection. Spinal injection could be an effective method of pain control for patients with extreme low back pain or sciatica.	This was a retrospective study conducted in a single center, which limits the generalizability of results. Results would not be applicable to all extreme low back pain, since spinal injection was requested only after red flags were excluded. Authors note it was difficult to measure the placebo effect. Injection techniques may vary across hospitals, and results may change accordingly. Finally, pathologies in the lumbar spine frequently involve multiple diseases and multiple sites simultaneously; however, authors analyzed only one culprit lesion per patient.

Konstantinou K, Dunn KM, Ogollah R, et al. Prognosis of sciatica and back-related leg pain in primary care: the ATLAS cohort. Spine J. 2018;18(6):1030-40.	Prospective cohort study	Low	To describe the prognosis and prognostic factors in primary care patients with low back-related leg pain and sciatica.	The present study included adults visiting their family doctor with back-related leg pain in the United Kingdom.	Information about pain, function, psychological, and clinical variables, was collected. Good outcome was defined as 30% or more reduction in disability (Roland-Morris Disability Questionnaire). Participants completed the questionnaires, underwent clinical assessments, received a magnetic resonance imaging scan, and were followed-up 12 months later. Mixed-effects logistic regression evaluated the prognostic value of six a priori defined variable sets (leg pain duration, pain intensity, neuropathic pain, psychological factors, clinical examination, and imaging variables). A combined model, including variables from all models, examined independent effects. The National Institute for Health Research funded the study. There are no conflicts of interest.	RESULTS: A total of 609 patients were included. At 12 months, 55% of patients improved in both the total sample and the sciatica group. For the whole cohort, longer leg pain duration (odds ratio [OR] 0.41; confidence interval [CI] 0.19-0.90), higher identity score (OR 0.70; CI 0.53-0.93), and patient's belief that the problem will last a long time (OR 0.27; CI 0.13-0.57) were the strongest independent prognostic factors negatively associated with improvement. These last two factors were similarly negatively associated with improvement in the sciatica subgroup. The present study provides new evidence on the prognosis and prognostic factors of back-related leg pain and sciatica in primary care. Just over half of patients improved at 12 months. Patient's belief of recovery timescale and number of other symptoms attributed to the pain are independent prognostic factors. These factors can be used to inform and direct decisions about timing and intensity of available therapeutic options.	indirectness and high risk of bias due to non-randomization
Lee, HS, Lee, JH, Lee, JH. A comparison of dynamic views using plain radiographs and thin-section three-dimensional computed tomography in the evaluation of fusion after posterior lumbar interbody fusion surgery. Spine J. 2013;13(10):1200-7	Observational Study	Low	This study aimed to evaluate the usefulness of dynamic flexion-extension radiographs as a method for evaluating fusion, by comparing it with three-dimensional thin-section computed tomography (CT).	108 Korean patients. INCLUSION CRITERIA: Patients diagnosed with severe spinal stenosis and Grade I and Grade II spondylolisthesis, underwent PLIF surgery, with follow-up by dynamic plain radiographs, functional rating scale, and 3D thin-section CT for 1 year after surgery. EXCLUSION CRITERIA: Patients who were pregnant or had malignant tumors, hepatitis, abnormal blood test results, abnormal liver function, or metabolic bone disease were excluded. BASELINE CHARACTERISTICS OF STUDY POPULATION: The participating patients had an average age of 64.3 years (37–86 years) with 40 men (37.04%) and 68 women (63.0%).	REFERENCE STANDARD: To evaluate the usefulness of dynamic flexion-extension radiographs as a method for evaluating fusion, by comparing it with three-dimensional thin-section computed tomography (CT). The patients were classified into fusion or nonfusion groups based on CT. In Group A, 78 (81.3%) levels belonged to the CT fusion group and 18 (18.7%) levels to the CT nonfusion group, in Group B, 51 (89.5%) and 6 (10.5%) levels, and in Group C, 2 (40%) and 3 (60%) levels, respectively. Correlation between plain radiographs and CT groups was analyzed.	The mean translation values for the degenerative type were 11.3 degrees preoperatively and 3.54 degrees postoperatively, whereas those for the lytic type were 9.29degrees preoperatively and 2.74degrees postoperatively. No fusion was observed in 19 levels (47.5%) in the degenerative type and 2 levels (28.6%) in the lytic type cases. However, there were no statistical correlations in the fusion results between the two types of spondylolisthesis cases (p=.4503). There were no correlations between the groups for the incidence of osteoporosis (p=.4488). There was a poor agreement on average (59.0%) and a statistical nonagreement (Kappa=0.151), indicating that diagnostic agreement using plain radiographs has a low level of significance. On the other hand, in 3D-CT, there was greater than 91% interobserver and interobserver agreement, with mean kappa values of 0.5 or higher, indicating a significantly high level of objective diagnostic agreement.There was no statistically significant difference between the radiography-assessed fusion and nonfusion groups in their clinical assessment values (VAS, ODI, and FRI) (p=.220, .730, and .092), and similarly, CT fusion and nonfusion groups did not show significant differences (p=.941, .960, and .973). CONCLUSION: For an objective and accurate assessment of fusion after PLIF surgery, it would be more appropriate to look for interbody bridging bone formation at 12 months by 3D thin-section CT rather than dynamic flexion-extension radiographs that lead to a high rate of false-positive and false-negative findings and show poor agreement with 3D-CT diagnoses.	STUDY LIMITATIONS: Non-consecutive recruitment Readers were not blinded or no comment was made about the blinding of the readers Small sample size from single center. For the coronal and sagittal CT images, the formation of a bridging bone was classified as fusion, but an analysis of the fusion area was not conducted. In addition, when using cages, the differences in local bone quality precluded identical conditions.
Liu P, Liang Y, Bian C, et al. Diagnostic accuracy of MR, CT, and ECT in the differentiation of neoplastic from nonneoplastic spine lesions. Asia Pac J Clin Oncol. 2020; 16(5):e192-e197.	Single center retrospective	Low	To provide guidance for appropriate imaging examinations for diagnosing spinal tumors or tumor-like lesions.	The study included 121 patients (62 men and 59 women, between 24 and 83 [57.2 ± 13.9] years of age) with supected spinal tumors. Exclusion criteria included incomplete imaging data, unclear pathological diagnoses, accompanying metabolic bone disease, spinal surgery, current treatment for infections, or history of antituberculosis treatment.	Each patient underwent ≥ 2 imaging examinations, including computerized tomography (CT), magnetic resonance (MR), and/or emission computed tomography (ECT). All patients were diagnosed by pathology after core needle or surgical biopsies. The results were compared with those of pathological examinations using paired chi squared tests, and comparedwith each other. Statistical indicators that tested the consistency of the results included McNemar's and kappa coefficients, as well as receiver operating characteristic curves.	The differences among MR, CT, ECT, and pathology were not significant. The kappa coefficient of MR, CT, and ECT was 46.1%, 36.0%, and 55.9%, respectively. The area under the curve of ECT, MR, and CT scans was 0.809, 0.705, and 0.704, respectively; and the differences among them were significant (P < .05). Post hoc multiple comparisons showed no significant differences among imaging examinations in terms of sensitivity, specificity, misdiagnosis rate, and coincidence rate (P > .05). However, significant differences were noted in the kappa coefficient and missed diagnosis rate (P < .05). The authors conclude that, although ECT was the most accurate imaging method, its high cost and large radiation dosage limit its widespread application. Furthermore, MR verified spinal tumors more effectively; however, CT excluded them more efficiently. In summary, when all factors are considered, MR is still the optimal modality for the diagnosis of spinal tumors, especially during the initial screening.	Study was limited by the number of cases in which ECT was not further divided into PET/CT and SPECT/CT imaging. Therefore, the difference between these two techniques is not clear. Moreover, the tumors were not differentiated into benign or malignant, and images may have shown some variance.
Tender GC, Davidson C, Shields J. Primary pain generator identification by CT-SPECT in patients with degenerative spinal disease. Neurosurg Focus. 2019; 47(6):E18.	Retrospective study	Low	To evaluate the degree of pain improvement in patients who underwent surgery, addressing primary pain generators identified by CT-SPECT	A total of 315 patients underwent diagnostic CT-SPECT between January 2014 and August 2018. Forty-eight patients underwent either cervical or lumbar fusion; there were 26 women (16 cervical, 10 lumbar) and 22 men (9 cervical, 13 lumbar).	The authors retrospectively reviewed all patients with chronic axial spine pain who underwent diagnostic CT-SPECT at their institution and analyzed pain improvement in those who underwent surgical treatment in order to determine whether CT-SPECT correctly identified the primary pain generator.	The overall axial spinal pain, as assessed through self-reporting of visual analog scale scores at 6 months postoperatively, improved from 9.04 ± 1.4 to 4.34 ± 2.3 (p = 0.026), with cervical fusion patients improving from 8.8 ± 1.8 to 3.92 ± 2.2 (p = 0.019) and lumbar fusion patients improving from 9.35 ± 0.7 to 4.87 ± 2.3 (p = 0.008). The authors conclude that CT-SPECT may offer a diagnostic advantage over current imaging modalities in identifying the primary pain generator in patients with axial spinal pain.	The authors note some limitations of this study: "1) The cost of the CT-SPECT scan was not discussed. It is our understanding that this study is not covered by some of the health insurance companies and/or hospitals, which may limit its applications. 2) There was a relatively low percentage of patients who underwent surgical intervention based on the CT-SPECT results. The large number of negative CT-SPECT scans in our study may be perceived as wasteful; however, these patients were clearly diagnosed as nonoperative candidates and could be directed to other services, such as pain management. 3) There was no control group to use for results comparison.

Wnuk NM, Alkasab TK, Rosenthal DI. Magnetic resonance imaging of the lumbar spine: determining clinical impact and potential harm from overuse. Spine J. 2018;18(9):1653-8.	Retrospective cohort study	Low	To determine the proportion of examinations with a detectable impact on patient care (actionable outcomes).	A total of 5,365 outpatient lumbar spine magnetic resonance (MR) examinations were conducted.	Actionable outcomes included (1) findings leading to an intervention making use of anatomical information such as surgery; (2) new diagnoses of cancer, infection, or fracture; or (3) following known lumbar spine pathology. Potential harm was assessed by identifying examinations where suspicion of cancer or infection was raised but no positive diagnosis made. METHODS: A medical record aggregation/search system was used to identify lumbar spine MR examinations with positive outcome measures. Patient notes were examined to verify outcomes. A random sample was manually inspected to identify missed positive outcomes. The proportion of actionable lumbar spine magnetic resonance imaging was 13%, although 93% were appropriate according to the American College of Radiology guidelines. Of 36 suspected cases of cancer or infection, 81% were false positives. Further investigations were ordered on 59% of suspicious examinations, 86% of which were false positives.	The proportion of lumbar spine MR examinations that inform management is small. The false-positive rate and the proportion of false positives involving further investigation are high. Further study to improve the efficiency of imaging is warranted.	imprecision and high risk of bias due to non-randomization