Evidence tables for review question: How effective are radiological imaging techniques in the diagnosis of spinal metastases, direct malignant infiltration of the spine or associated spinal cord compression?

Allan 2009
Allan L, Baker L, Dewar J, et al. Suspected malignant cord compression – Improving time to diagnosis via a hotline: A prospective audit. British Journal of Cancer, 100, 1867-1872, 2009

Study details

Country/ies where study was carried out	UK
Study type	Retrospective cohort study
Study dates	MSCC hotline established in 2004
Inclusion criteria	Criteria for the suspected MSCC hotline: patient known to have, or strongly suspected to have, cancer; new severe nerve root pain (unilateral or bilateral) and/or new severe localised vertebral pain, especially thoracic; and any new difficulty in walking. The comparison group came from Clinical Research and Audit Group (CRAG) audit data of people with MSCC from several Scottish centres.
Exclusion criteria	None reported
Patient characteristics	N=424 Patients with suspected malignant spinal cord compression referred via hotline and patients with malignant spinal cord com-pression (Clinical Research and Audit group) Gender [number of male, female]: not reported Age, mean (SD): not reported Myeloma versus other cancer types [percentage with myeloma]: 6% had myeloma in the MSCC hotline group – but not reported in the comparison group. Functional status/fitness for treatment [percentage who were ambulant]: Not reported. Ambulant rates at diagnosis of MSCC were reported – see outcomes below.
Intervention(s)/control	Hotline group: The referring GP or a hospital doctor speaks directly with a senior clinician on a dedicated phone number. After further discussion, usually between the hotline clinician and the patient's oncologist, the hotline clinician decides whether an MRI is required within 24 h, or the hotline clinician or the patient's oncologist may arrange to examine the patient before determining whether an urgent scan is required. An MRI slot was reserved at the end of each day's list for hotline referrals (if not used by mid-day it was re-allocated to an urgent inpatient or an outpatient). Patients with probable MCC presenting after this time were scanned first thing the next morning. An ad hoc on-

	call service was available at weekends and public holidays. Scans were immediately reported on a dedicated proforma by radiologists with a particular interest in MRI. MRI evidence of MCC was considered to be present if there was any extension into the epidural space with impingement, displacement or compression of the cord with or without cord signal change. The results were immediately communicated to the clinical team caring for the patient. Control group: a national Clinical Research and Audit Group (CRAG) prospective audit (324 cases of MCC) whose diagnosis followed usual care.
Duration of follow-up	From symptoms until results of diagnostic MRI.
Sources of funding	Macmillan Cancer Relief
Sample size	Hotline group N=100 (n=44 with MSCC); comparison group N=324 (all with MSCC)

Outcomes

Outcome	Hotline for suspected MSCC, Baseline, n=44	Comparison group (CRAG audit), Baseline, n=324
Time from GP or a hospital doctor referring the patient to diagnosis, days, median (IQR)	1 (0 to 21)	15 (3 to 66)
Number of patients walking (unaided or with assistance) at time of MSCC diagnosis	n=34	n=175

Critical appraisal – ROBINS-I

Section	Question	Answer
1. Bias due to confounding	Risk of bias judgement for confounding	Serious *(Analysis not controlled for confounders)*
2. Bias in selection of participants into the study	Risk of bias judgement for selection of participants into the study	Serious *(No information about patients discussed on the hotline but not referred for diagnostic tests)*
3. Bias in classification of interventions	Risk of bias judgement for classification of interventions	Low
4. Bias due to deviations from intended interventions	Risk of bias judgement for deviations from intended interventions	Low
5. Bias due to missing data	Risk of bias judgement for missing data	Low
6. Bias in measurement of outcomes	Risk of bias judgement for measurement of outcomes	Serious *(Unclear how the baseline timepoint was decided for the comparison group)*

Section	Question	Answer
7. Bias in selection of the reported result	Risk of bias judgement for selection of the reported result	Low
Overall bias	Risk of bias judgement	Critical. Risk of bias due to confounding, selection of participants, and measurement of outcomes.
Overall bias	Directness	Directly applicable

Bacher 2021
Bacher S, Hajdu S, Maeder Y, et al. Differentiation between benign and malignant vertebral compression fractures using qualitative and quantitative analysis of a single fast spin echo T2-weighted Dixon sequence. European Radiology, 31, 9418-942, 2021

Study details

Country/ies where study was carried out	Switzerland.
Study type	Retrospective cohort study
Study dates	July 2014 – June 2020.
Inclusion criteria	Consecutive patients undergoing spine MRI (at a single institution) prior to cementoplasty for acute vertebral compression fractures.
Exclusion criteria	• No MRI prior to cementoplasty or 1.5T MRI • No T2-weighted Dixon sequence • History of hematologic neoplasia • Benign vertebral compression fractures but history of malignancy; or malignancy detected ≥9-month follow-up; or no follow-up data ≥9 months available.
Patient characteristics	N=95 Consecutive patients undergoing spine MRI (at a single institution) prior to cementoplasty for acute vertebral compression fractures. Age, mean (SD), years: benign 76 (12); malignant 63 (12) Gender [number of male, female]: benign – male n=23; female n=40; malignant – male n=20; female n=12. Myeloma versus other cancer types [number with myeloma]: 0/95 (0%). Patients with a history of haematologic neoplasms were excluded. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	MRI – single sagittal fast spin echo (FSE) T2-weighted Dixon sequence.
Reference standard(s)	Best value comparator (including biopsy results). Vertebral compression fractures were categorised as benign or malignant based on a best valuable comparator consisting of a consensus reading performed by three observers after the end of readings of all available medical records, radiographs, CT, MRI, bone scans and PET-CT studies, and biopsy data (biopsy of target vertebra performed during cementoplasty).

	For vertebral compression fractures categorized as benign according to the best valuable comparator, a follow-up of nine months or more was required, in particular to avoid false negative results of biopsy. Vertebral compression fractures were therefore considered benign if fulfilling all of the following criteria: no current or past history of malignancy, no positive biopsy result (biopsy could be absent or negative), no malignancy found at clinical and imaging follow-up of nine months or more. Vertebral compression fractures were considered malignant if the best valuable comparator based on all data available was suggestive of a malignant origin.
Duration of follow-up	Patients diagnosed with benign vertebral compression fractures were only included if they had follow-up data of greater than 9 months.
Sources of funding	None.
Results	See Appendix L

Critical appraisal – QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	High – *potential incorporation bias as MRI was part of the composite reference standard*
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Dearnaley 2022
Dearnaley D, Hinder V, Hijab A, et al. Observation versus screening spinal MRI and pre-emptive treatment for spinal cord compression in patients with castration-resistant prostate cancer and spinal metastases in the UK (PROMPTS): an open-label, randomised, controlled, phase 3 trial. Lancet: Oncology, 23, 501-513, 2022

Study details

Country/ies where study was carried out	UK
Study type	Randomised controlled trial (RCT)
Study dates	February 2013 to April 2017
Inclusion criteria	Eligible patients were aged 18 years and older, had a confirmed pathological diagnosis of prostate adenocarcinoma or a clinical diagnosis of prostate cancer with osteoblastic bone metastases and a serum prostate specific antigen (PSA) concentration of 100 ng/mL or higher at any time between diagnosis and randomisation. Other inclusion criteria were the presence of asymptomatic spinal metastasis, castration-resistant state (defined as PSA >5 ng/dL and more than 50% increase above the nadir during treatment with a luteinising hormone-releasing hormone analogue or after orchidectomy), PSA concentration of more than 5 ng/mL within 21 days before randomisation, life expectancy of 6 months or longer, and Eastern Cooperative Oncology Group (ECOG) performance status of 0–2.
Exclusion criteria	Presence of any back pain or neurological symptoms from spinal metastases, previous spinal MRI within 12 months from trial entry, previous external beam radiotherapy to the vertebrae or spinal surgery to treat SCC, and any contraindication for MRI.
Patient characteristics	N=410 Patients with metastatic castration resistant prostate cancer with bone involvement. Age at randomisation, years (median, IQR): MRI group: 74.3 (68.0–79.3) Control group: 74.2 (68.5–79.3)
Intervention(s)/control	Intervention: screening spinal MRI (in men with metastatic castration resistant prostate cancer with bone involvement) to detect and treat asymptomatic spinal cord compression Control: No MRI
Duration of follow-up	36 months
Sources of funding	Cancer Research UK
Sample size	Total: 420 Intervention: 210 Control: 210

Outcomes

Outcome	**MRI screening, 24 month, n=210**	**Control, 24 month, n=210**	**Relative effect**
Overall survival (event is death from any cause)	172/210	174/210	Adj HR 0.98

Outcome		MRI screening, 24 month, n=210		Control, 24 month, n=210	Relative effect
					(0.79 to 1.21)[1]
Neurological and functional status – clinical spinal cord compression	19/210	26/210	Adj HR 0.61 (0.35 to 1.08)[1]		
Neurological and functional status – persistent neurological functional deficit (Frankel score A-D)	15/210	23/210	RR 0.73 (0.42 to 1.28)		

1. Adjusted for time since development of castration-resistant prostate cancer, time since start of continuous hormone treatment, ECOG performance status (0, 1, and 2), and natural logarithm of PSA concentration.

Outcome	Mean (95% CI) difference MRI screening – Control (Change from baseline to 12 months)
Quality of life – EQ-5D-5L – health state today (range 0 to 100, higher scores are better)	-1.5 (-5.7 to 2.7)
Pain – Brief Pain Index – severity (range 0 to 10, lower scores are better)	0.4 (-0.2 to 0.9)

Critical appraisal – Cochrane RoB 2

Section	Question	Answer
Domain 1: Bias arising from the randomisation process	Risk of bias judgement for the randomisation process	Low
Domain 2a: Risk of bias due to deviations from the intended interventions (effect of assignment to intervention)	Risk of bias for deviations from the intended interventions (effect of assignment to intervention)	Low
Domain 3. Bias due to missing outcome data	Risk-of-bias judgement for missing outcome data	Low
Domain 4. Bias in measurement of the outcome	Risk-of-bias judgement for measurement of the outcome	Low
Domain 5. Bias in selection of the reported result	Risk-of-bias judgement for selection of the reported result	Low
Overall bias and Directness	Risk of bias judgement	Low
Overall bias and Directness	Overall Directness	Directly applicable

Husband 2001
Husband D, Grant K, Romaniuk C. MRI in the diagnosis and treatment of suspected malignant spinal cord compression. British Journal of Radiology 74, 15-23, 2001

Study details

Country/ies where study was carried out	UK.
Study type	Prospective cohort study
Study dates	Not reported.
Inclusion criteria	Consecutive patients with suspected malignant spinal cord compression undergoing MRI at a single institution.
Exclusion criteria	Not reported.
Patient characteristics	N=280 patients undergoing MRI for suspected malignant spinal cord compression Age, mean (SD), years: range 23 – 89 (median 67). Mean and SD not reported. Gender [number of male, female]: male n=158; female n=122. Myeloma versus other cancer types [percentage with myeloma]: Not reported. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	Plain radiograph and neurological examination
Reference standard(s)	Radiological follow-up (MRI carried out as soon as possible usually the next day)
Duration of follow-up	Not reported.
Sources of funding	Not reported.
Results	Diagnostic accuracy – plain radiograph plus neurological examination (N=280) TP 89, FP 2, FN 112, TN 87 (from which sensitivity, specificity, likelihood ratios as well as PPV and NPV were calculated)

Critical appraisal – QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low

Section	Question	Answer
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High *(composite index test of X-ray & neurological examination)*
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Kato 2015
Kato S, Hozumi T, Yamakawa K, et al. META: an MRI-based scoring system differentiating metastatic from osteoporotic vertebral fractures. Spine Journal, 15, 1563-70, 2015

Study details

Country/ies where study was carried out	Japan.
Study type	Retrospective cohort study
Study dates	April 2004 – September 2011.
Inclusion criteria	• Patients with radiologically apparent collapse of thoracolumbar vertebra due to metastatic vertebral fractures or osteoporotic vertebral fractures. • Evaluated using MRI within 60 days of possible injury. • 100 cases of metastatic vertebral fracture were selected at random from database. Diagnosis confirmed either by positive biopsy results or by malignant radiographic changes (progressive expansion of vertebral signal intensity change or spinal canal invasion) observed for more than 60 days after their first presentation associated with a clinical diagnosis of malignancy at other sites. • 100 cases of osteoporotic vertebral fractures were selected at random from database. Diagnosis confirmed by eventual reduction of vertebral signal intensity change and remission of clinical symptoms observed for more than 60 days.
Exclusion criteria	• Metastatic vertebral fractures associated with haematologic disorders, including multiple myeloma and malignant lymphoma • Previously diagnosed metastatic vertebral fractures that had already received irradiation (due to potential to affect MRI appearance).
Patient characteristics	N=200 Patients with radiologically apparent collapse of thoracolumbar vertebra due to metastatic vertebral compression fractures or osteoporot-

	ic vertebral fractures Age, mean (SD), years: metastatic vertebral fractures 64 (11); osteoporotic vertebral fractures 73 (11). Gender [number of male, female]: metastatic vertebral compression fractures – male n=43; female n=57; osteoporotic vertebral fractures – male n=27; female n=73. Myeloma versus other cancer types [number with myeloma]: 0/200 (0%). Patients with metastatic fractures associated with haematological disorders were excluded. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	MRI. All images obtained using: Magnetom Avanto 1.5T; Signa Hde 1.5T; or Intera Achieva 1.5T.
Reference standard(s)	Biopsy result / surgical pathology (or clinical and radiological follow-up if no surgery/biopsy was done)
Duration of follow-up	60 days
Sources of funding	None.
Results	See Appendix L

Critical appraisal – Critical appraisal – QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear *(not reported if reference standard results interpreted without knowledge of the results of the index test)*
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Kim 2000

Kim J, Learch T, Colletti P, et al. Diagnosis of vertebral metastasis, epidural metastasis, and malignant spinal cord compression: are T(1)-weighted sagittal images sufficient? Magnetic Resonance Imaging 18, 819-24, 2000

Study details	
Country/ies where study was carried out	USA
Study type	Retrospective cohort study
Study dates	Not reported.
Inclusion criteria	Consecutive patients undergoing MRI for clinically suspected malignant spinal cord compression at a single institution.
Exclusion criteria	Not reported.
Patient characteristics	N=57 Consecutive patients undergoing MRI for clinically suspected malignant spinal cord compression Age, median (range), years: 49 (22 – 87). Gender [number of male, female]: male n=31; female n=26. Myeloma versus other cancer types [number with myeloma]: n=10 studies in patients with multiple myeloma. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	T1-weighted sagittal images alone (a subset of the complete studies – see reference standard information).
Reference standard(s)	Complete magnetic resonance studies of the spinal column (a complete study consisting of T1-weighted sagittal images, T2-weighted sagittal images, and T1- and/or T2-weighted axial images). This was divided into 'external standard' which is the agreement in diagnosis between radiologists on examination of complete imaging study for each patient and 'internal standard' which is the individual radiologist's diagnosis based on complete imaging study for each patient
Duration of follow-up	Not reported.
Sources of funding	Not reported.
Results	Diagnostic findings in cases reviewed (n=94 MRI studies in 57 patients) Vertebral metastasis n=72; epidural metastasis n=28; cord compression n=22. Sensitivity and specificity (ability to detect each diagnostic parameter) of T1-weigted sagittal images alone (in comparison to 'external standard' that is agreement in diagnosis between radiologists on examination of complete imaging study for each patient): Vertebral metastasis – sensitivity 87% (249/288) 95% CI 82 – 90; specificity 83% (73/88) 95% CI 74 – 89. Cord compression – sensitivity 70% (62/88) 95% CI 60 – 79; specificity 97% (278/288) 95% CI 90 – 99. Epidural metastasis – sensitivity 46 (51/112) 95% CI 37 – 55; specificity 89 (236/264) 95% CI 85 – 93. Sensitivity and specificity (ability to detect each diagnostic parameter) of T1-weigted sagittal images alone (in comparison to 'internal standard' that is individual radiologist's diagnosis based on complete imaging study for each patient): Vertebral metastasis – sensitivity 91% (242/265) 95% CI 87 – 94; specificity 80% (89/111) 95% CI 72 – 87. Cord compression – sensitivity 62% (62/100) 95% 52 – 71; specificity 96% (266/276) 95% CI 93 – 98. Epidural metastasis – sensitivity 48 % (49/102) 95% 39 – 58; specificity 89% (244/274) 95% CI 85 – 92.

Critical appraisal - QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	High *(57 patients but 94 MRI studies, unclear how many studies per patient)*
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	High. (*Results for 3 radiologists pooled - giving artificially narrow confidence intervals for sensitivity and specificity)*
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	High *(Risk of incorporation bias of index test in reference standard)*
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Unclear. *(Some patients had contrast enhanced MRI)*

Kim 2020

Kim S, Lee J. (2020) Diagnostic performance of F-18 FDG PET or PET/CT for differentiation of benign from malignant vertebral compression fractures; A meta-analysis. World Neurosurgery, 137: e626-e633, 2020

Study details

Country/ies where study was carried out	South Korea
Study type	Systematic review of diagnostic accuracy studies
Study dates	Included studies were published between 2008 and 2018
Inclusion criteria	Studies F-18 FDG PET or PET/CT used to differentiate benign and malignant VCFx; sufficient data available to reassess the sensitivity

	and specificity of F-18 FDG PET or PET/CT for the differentiation of malignant VCFxs or the absolute numbers had been provided of the true-positive, true-negative, false-positive, and false-negative data; and no data overlap.
Exclusion criteria	Duplicated studies were excluded, as were review articles, case reports, conference papers, and letters that did not contain the original data.
Patient characteris-tics	5 studies included in review (N=274) Age: mean age across studies ranged from 60 to 72 years Sex Male/female 182/92
Index test(s)	• FDG-PET • FDG-PET-CT Interpretation was based o SUV-max with cut-off ranging from 2 to 4.75 across studies. FDG dose ranged from 370 to 555 MBq.
Reference stand-ard(s)	• Composite reference standard of biopsy, clinical follow-up and repeat imaging
Duration of follow-up	Not reported
Sources of funding	Not reported
Other information	5 studies included and assessed with QUADAS-2. None were at high risk of bias. All were at unclear risk of bias for patient selection and some details of the imaging test were unclear in most of studies. All were at low risk of bias for reference standard and flow & timing.

Outcomes

Outcome	**Pooled results N=274**
Sensitivity (95% CI)	0.96 (0.82 to 0.99)
Specificity (95% CI)	0.77 (0.56 to 0.89)
Positive likelihood ratio (95% CI)	4.1 (2.1 to 8)
Negative likelihood ratio (95% CI)	0.05 (0.01 to 0.23)
Area under the curve (95% CI)	0.94 (0.92 to 0.96)

Critical appraisal - ROBIS checklist

Section	**Question**	**Answer**
Study eligibility criteria	Concerns regarding specification of study eligibility criteria	Low
Identification and selection of studies	Concerns regarding methods used to identify and/or select studies	Low
Data collection and study appraisal	Concerns regarding methods used to collect data and appraise studies	Low

Section	Question	Answer
Synthesis and findings	Concerns regarding the synthesis and findings	Low
Overall study ratings	Overall risk of bias	Low
Overall study ratings	Applicability as a source of data	Fully applicable

Laufer 2009
Laufer I, Lis E, Pisinski L, et al. The accuracy of [(18) F] fluorodeoxyglucose positron emission tomography as confirmed by biopsy in the diagnosis of spine metastases in a cancer population. Neurosurgery, 64, 107-4, 2009

Study details

Country/ies where study was carried out	USA.
Study type	Retrospective cohort study
Study dates	1996 - 2005.
Inclusion criteria	• Patients with a previous diagnosis of cancer undergoing CT guided biopsy of suspected malignant spinal column lesion (initially identified via MRI). • Patients who underwent FDG-PET scan within 6 weeks of initial CT guided biopsy.
Exclusion criteria	• Radiotherapy of chemotherapy initiated before biopsy. • Clear radiographic and clinical discitis/ osteomyelitis.
Patient characteristics	N=82 patients with a previous diagnosis of cancer undergoing CT guided biopsy of suspected malignant spinal column lesion Age, mean (SD), years: 56. SD not reported. Gender [number of male, female]: male n=41; female n=49. Myeloma versus other cancer types [number with myeloma]: n=5/82. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	FDG-PET.
Reference standard(s)	CT guided biopsy.
Duration of follow-up	Not reported.
Sources of funding	Not reported.
Results	Insufficient detail to extract diagnostic accuracy data for FDG-PET <u>Biopsy results (n=82):</u>

	Positive on biopsy n=74 Negative on biopsy n=8. Biopsy failure rate 1/82 Diagnostic yield: 81/82

Critical appraisal - QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Li 2019

Li K, Huang L, Lang Z, et al. Reliability and Validity of Different MRI Sequences in Improving the Accuracy of Differential Diagnosis of Benign and Malignant Vertebral Fractures: A Meta-Analysis. American Journal of Roentgenology, 213, 427-436, 2019

Study details

Country/ies where study was carried out	China
Study type	Systematic review of diagnostic accuracy studies
Study dates	Jan 2000 to Sep 2016
Inclusion criteria	Studies related to the differential diagnosis of benign and malignant vertebral fractures by MRI and reference standard (histopathologic diagnosis or clinical follow-up examination)
Exclusion criteria	Abstracts, reviews or conference papers. Studies performed on cadavers or animals, the sample size was less than 20, raw data was not complete, patients were not examined with MRI and reference standard, the trial was not double-blind and repeated studies.
Patient characteristics	N=895 Patients with vertebral fractures Age: mean ranged from 54.6 to 69 years across the studies Sex: female n=486; male n=409.

Index test(s)	• MRI chemical shift imaging • MRI conventional sequences plus diffusion weighted imaging • MRI conventional sequences • MRI conventional sequences plus contrast enhanced images
Reference standard(s)	Histopathologic diagnosis (from surgery or biopsy) or clinical & radiological follow-up
Duration of follow-up	Not reported
Sources of funding	Not reported
Results	See Appendix L
Other information	18 studies included and assessed with QUADAS-2. None were at high risk of bias. Flow and timing was unclear in 8/12 studies and some details of MRI were unclear in 12/18 studies.

Critical appraisal - ROBIS checklist

Section	Question	Answer
Study eligibility criteria	Concerns regarding specification of study eligibility criteria	Low
Identification and selection of studies	Concerns regarding methods used to identify and/or select studies	Low
Data collection and study appraisal	Concerns regarding methods used to collect data and appraise studies	Low
Synthesis and findings	Concerns regarding the synthesis and findings	Low
Overall study ratings	Overall risk of bias	Low
Overall study ratings	Applicability as a source of data	Fully applicable

Maeder 2018
Maeder Y, Dunet V, Richard R, et al. Bone Marrow Metastases: T2-weighted Dixon Spin-Echo Fat Images Can Replace T1-weighted Spin-Echo Images. Radiology, 286, 948-959, 2018

Study details

Country/ies where study was carried out	Switzerland.
Study type	Retrospective cohort study
Study dates	September 2014 - April 2016.
Inclusion criteria	Consecutive patients undergoing whole spine MRI for suspected vertebral metastases at a single institution.

Exclusion criteria	• History of haematological neoplasia • Spinal osteosynthesis • MRI performed with 1.5T scanner • Incomplete MRI sequences
Patient characteristics	N=121 consecutive patients undergoing whole spine MRI for suspected vertebral metastases Age, mean (SD), years: 61.4 (11.8). Gender [number of male, female]: male n=63; female n=58. Myeloma versus other cancer types [number with myeloma]: 0/121 (0%). Patients with a history of haematological neoplasia were excluded. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	• MRI - sagittal SE Dixon T2-weighted fat-only and water-only imaging. • MRI - sagittal SE T1-weighted and SE Dixon T2-weighted water-only images.
Reference standard(s)	• Best value comparator (including biopsy result where possible). This consisted of consensus reading of all examinations by two musculoskeletal radiologists (performed 1 month after the end of all readings), as well as review of all available medical data. These data included clinical, histologic (spinal bone biopsy data, available for 30 patients), biologic, and imaging data.
Duration of follow-up	≥ 8 months after imaging (mean 15.2 months).
Sources of funding	Not reported.
Other information	All imaging performed with 3-T scanner. A total of three contiguous sagittal stacks with nonenhanced fast. SE T1-weighted and fast SE Dixon T2-weighted sequences of the entire spine from the base of the skull to the last sacral piece were performed for all patients. Four sets of images were routinely reconstructed from the Dixon T2 sequences: in-phase, out-of phase, Dixon T2-weighted water-only, and Dixon T2-weighted fat-only images, of which only the latter two were considered for our study. Additional sequences performed on a case-by case basis whenever indicated were considered only for the best valuable comparator. They included contrast material–enhanced fat-suppressed T1-weighted sequences on the sagittal and axial planes, as well as axial fat suppressed T2-weighted sequences.
Results	See Appendix L

Critical appraisal - QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low

Section	Question	Answer
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	High *(Potential for incorporation bias)*
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Perry 2018
Perry M, and Sebro R. Accuracy of Opposed-phase Magnetic Resonance Imaging for the Evaluation of Treated and Untreated Spinal Metastases. Academic Radiology, 25, 877-882, 2018

Study details

Country/ies where study was carried out	USA.
Study type	Retrospective cohort study
Study dates	January 2006 - November 2016.
Inclusion criteria	Patients undergoing opposed-phase MRI of the cervical, thoracic, or lumbar spine.
Exclusion criteria	• Patients whose MRI studies did not include opposed-phase sequences. • Patients whose MRI studies were motion-degraded. • Patients with lesions confirmed as osteomyelitis or spondylodiscitis.
Patient characteristics	N=101 Patients undergoing opposed-phase MRI of the cervical, thoracic, or lumbar spine. Gender [number of male, female]: male n = 39; female n=62. Age, mean (SD), years: 57.7 (14.1) Myeloma versus other cancer types [number with myeloma]: n=13. Functional status/fitness for treatment [percentage who were ambulant]: Not reported. n=136 lesions identified from n=120 opposed phase MRI studies. Benign lesions n=25 Untreated metastases n=25 Treated spinal metastases n=86 (radiation n=19; chemotherapy only n=67).
Index test(s)	Opposed phase MRI.

	All examinations performed on 1.5-T or 3-T systems. Opposed-phase gradient recalled-echo images were performed in the sagittal plane, 1.5-T (repetition time [TR] 140–350 ms, echo time [TE] out-of-phase 2.204–2.54 ms, TE in-phase 4.373–5.04 ms, slice thickness 4 mm, interslice gap 0.4–0.8 mm); 3-T (TR 4.82–173 ms, TE out-of-phase 1.24– 1.28 ms, TE in-phase 2.48–2.56 ms, slice thickness 4 mm, interslice gap 0.4–0.8 mm). All images were reviewed using a GE Centricity Picture Archiving and Communication System (PACS) workstation. Region of interest (ROI) measurements were obtained using the PACS ellipse ROI markup tool. The largest possible ROI was placed over the lesion and the mean SI was recorded on out-of-phase and in-phase sequences. An approximately similar sized ROI (same area) was placed on the out-of-phase and in-phase sequences. Care was taken to avoid vessels and vertebral body cortex when obtaining ROIs. The SIR of out-of-phase SI to the inphase SI was then calculated (SIR = mean lesion SI on out-of-phase MRI sequence/mean lesion SI on in-phase MRI sequence).
Reference standard(s)	• Biopsy result / surgical pathology. • Clinical and Radiological follow up (if no surgery/biopsy done) up to 2 years. Lesions were determined to be spinal metastases if there was histologic confirmation of spinal metastases from percutaneous or surgical biopsies; if there was progression of disease (increase in size of lesion) on subsequent imaging over 2 years; or if there was decrease in size or imaging appearance of the lesion in response to chemotherapy or radiation therapy. Lesions were categorized as benign if there was no evidence of change in size or imaging appearance of the lesion for at least 2 years or if there was histologic confirmation that they were benign and there was no history of malignancy. Lesions confirmed as osteomyelitis or spondylodiscitis were excluded from the study.
Duration of follow-up	≥ 2 years (for benign lesions).
Sources of funding	Not reported.
Results	See Appendix L

Critical appraisal – QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Unclear
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review ques-	Low

Section	Question	Answer
	tion?	
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Phadke 2001
Phadke D, Lucas D, Madan S. Fine-needle aspiration biopsy of vertebral and intervertebral disc lesions: specimen adequacy, diagnostic utility, and pitfalls. Archives of Pathology and Laboratory Medicine, 125, 1463-8, 2001

Study details

Country/ies where study was carried out	USA.
Study dates	January 1994 – February 2000.
Inclusion criteria	Not reported.
Exclusion criteria	Not reported.
Patient characteristics	N=78 Patients undergoing CT guided fine needle aspiration biopsy for vertebral and intervertebral lesions. Included patients with and without a known primary malignancy. Age, mean (SD), years: Not reported. Gender [number of male, female]: male n=29; female n=49. Myeloma versus other cancer types [number with myeloma]: Not reported. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	CT guided fine needle aspiration biopsy. N=36 patients with vertebral lesions had a history of malignancy at another site. In these cases, FNAB was performed on radiologically suspected or detected lesions to rule out metastasis. In the other 42 cases of both vertebral and intervertebral lesions, FNAB was performed as a part of the workup in patients presenting with signs and symptoms related to the spine and abnormal radiologic findings.
Reference standard(s)	A cytopathologist classified the biopsy as: Positive for malignancy, Suspicious for malignancy, Normal cellular elements present, with no evidence of malignancy, Unsatisfactory/inadequate for diagnosis or Benign neoplastic lesion.
Duration of follow-up	Not reported.
Sources of funding	Not reported.
Results	Vertebral lesions with a clinical history of malignancy – cytologic diagnosis (n=36) Positive for malignancy: n=24/36.

	Suspicious for malignancy: n=0/36. Normal cellular elements present, with no evidence of malignancy: n=5/36. Unsatisfactory/inadequate for diagnosis: n=4/36. Benign neoplastic lesions: n=2/36. Acute inflammatory process: n=1/36. <u>Vertebral lesions without a clinical history of malignancy – cytologic diagnosis (n=30)</u> Positive for malignancy: n=11/30. Suspicious for malignancy: n=0/30. Normal cellular elements present, with no evidence of malignancy: n=3/30. Unsatisfactory/inadequate for diagnosis: n=11/30. Benign neoplastic lesions: n=5/30. <u>Intervertebral disc lesions – cytologic diagnosis (n=12)</u> Positive for malignancy: n=0/12. Suspicious for malignancy: n=0/12. Normal cellular elements present, with no evidence of malignancy: n=1/12. Unsatisfactory/inadequate for diagnosis: n=6/12. Benign neoplastic lesions: n=0/12. Acute inflammatory process: n=3/12. Degenerative disc disease: n=2/12. <u>Vertebral and intervertebral disc lesions – comparison of radiologic impression with cytologic diagnosis (n=48)</u> Malignant on radiology (n=9) – malignant on cytology n=7; normal cellular elements with no evidence of malignancy n=2. Indeterminate on radiology (n=30) – malignant on cytology n=13; benign n=6; normal cellular elements with no evidence of malignancy n=4; unsatisfactory/inadequate for diagnosis n=6; suspicious for malignancy n=1. Benign on radiology (n=9) – benign on cytology n=4; normal cellular elements with no evidence of malignancy n=1; unsatisfactory/inadequate for diagnosis n=4.

Critical appraisal – QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Unclear *(no details of inclusion/ exclusion criteria)*
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low

Section	Question	Answer
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear *(limited details of diagnostic criteria used by cytopathologist)*
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Razek 2009
Razek A, Sherif, F. Diagnostic accuracy of diffusion tensor imaging in differentiating malignant from benign compressed vertebrae. Neuroradiology, 61, 1291-1296, 2019

Study details

Country/ies where study was carried out	Egypt.
Study type	Retrospective cohort study
Study dates	Not reported.
Inclusion criteria	Patients with untreated compressed vertebrae undergoing MRI at a single institution.
Exclusion criteria	Patients whose imaging was of a poor quality.
Patient characteristics	N=45 Patients with untreated compressed vertebrae undergoing MRI Age, mean (SD), years: 56.14 (7.9). Gender [number of male, female]: male n=22; female n=22. Myeloma versus other cancer types [number with myeloma]: Not reported. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	MRI – All patients were examined using routine T1- and T2-weighted MR imaging and DTI of the spine.All MR images were performed using a 1.5-Tesla scanner. Images were analysed by two neuroradiologists who were blinded to the clinical presentation and final histopathological results..
Reference standard(s)	• Biopsy result / surgical pathology. Final diagnosis done with biopsy, performed 10 – 18 days after MRI.

Duration of follow-up	Not reported.
Sources of funding	None.
Results	See Appendix L

Critical appraisal – QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Schmeel 2018
Schmeel F, Luetkens J, Feist A, et al. Quantitative evaluation of T2* relaxation times for the differentiation of acute benign and malignant vertebral body fractures. European Journal of Radiology, 108, 59-65, 2018

Study details

Country/ies where study was carried out	Germany.
Study type	Retrospective cohort study
Study dates	February 2015 – March 2018.
Inclusion criteria	Consecutive patients with a suspected acute vertebral compression fracture or known primary malignancy and suspected pathologic vertebral compression fracture. • > 18 years • acute onset of back pain (less than 1 month from admission) • presence of an acute benign (osteoporotic and/or post-traumatic) or malignant vertebral compression fracture as determined on routine clinical spine MRI • histopathologic confirmation of vertebral compression fracture obtained from direct bone biopsy.
Exclusion criteria	• Pregnancy

	• Contraindications to MRI (such as non-MR conditional cardiac pacemaker) • Prior bisphosphonate treatment • metallic instrumentation of the spine segment under investigation.
Patient characteristics	N=37 Consecutive patients with a suspected acute vertebral com-pression fracture or known primary malignancy and suspected pathological vertebral compression fracture. Gender [number of male, female]: male n=17; female n=20. Age, mean (SD), years: 64.8 (16.5) Myeloma versus other cancer types [number with myeloma]: n=7. Functional status/fitness for treatment [percentage who were ambulant]: Not reported. Patients were divided into two groups according to underlying pathology. Group 1 (n=19) patients with acute osteoporotic and/or benign vertebral compression fractures. Diagnosis of benign vertebral compression fractures was established on the basis of direct biopsy and histopathologic con-firmation of bone specimens obtained during vertebroplasty or spinal instrumentation. Group 2 (n=18) patients with neoplastic vertebral compression fractures due to hematological malignancies (n=8) or vertebral metastasis (n=17). Diagnosis of malignant vertebral compression fractures was established on the basis of direct bone biopsy and subsequent histopathological confirmation of bone specimens obtained via CT-guidance, surgery and/or spinal instrumentation.
Index test(s)	MRI - T2*-weighted. All imaging was performed on a clinical 3.0-Tesla whole-body MR imager. Routine clinical MRI of the spine included at least a sagittal T1-weighted spin-echo (450–750/6-12 [repetition time ms (TR)/echo time ms (TE)]) and T2-weighted turbo spin-echo sequence (3000-5000/80-120 [TR/TE]) as well as a sagittal T2 spectral-attenuated-inversion-recovery (SPAIR)-weighted turbo spin-echo sequence (3000-5000/80-120 [TR/TE]).
Reference standard(s)	• Biopsy result / surgical pathology
Duration of follow-up	Not reported.
Sources of funding	None.
Results	See Appendix L

Critical appraisal - QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low

Section	Question	Answer
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Schmeel 2021
Schmeel F, Enkirch S, Luetkens J, et al. Diagnostic Accuracy of Quantitative Imaging Biomarkers in the Differentiation of Benign and Malignant Vertebral Lesions: Combination of Diffusion-Weighted and Proton Density Fat Fraction Spine MRI. Clinical Neuroradiology, 31, 1059-1070, 2021

Study details

Country/ies where study was carried out	Germany.
Study type	Prospective cohort study
Study dates	June 2018 - September 2019
Inclusion criteria	Consecutive patients with untreated vertebral bone marrow lesions (benign and malignant) undergoing MRI. Presence of at least one vertebral bone marrow lesion with ≥1cm in size as determined on routine clinical spine MRI or at least one of the following indications: • clinically suspected acute vertebral fracture and acute onset of back pain (≤1 month from admission) • suspected osseous metastasis or malignant spine disease • and/or persisting localized back pain without typical discogenic radiation for more than 3 months.
Exclusion criteria	• Contraindication for MRI (such as nonconditional cardiac pacemaker) • previous or concurrent chemotherapy (including angiogenesis inhibitors) and/or radiotherapy • bisphosphonate and/or growth colony-stimulating factor treatment • previous surgery and metallic implants in the spine segment under investigation.
Patient characteristics	N=55 Consecutive patients with untreated vertebral bone marrow lesions (benign and malignant) undergoing MR Age, mean (SD), years: 68 (14) Gender [number of male, female]: male n=25; female n=30. Myeloma versus other cancer types [number with myeloma]: Patients with myeloma were included however the number of patients is not reported. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.

Index test(s)	MRI - sagittal DWI (single-shot spin-echo echo-planar with multi-slice short TI inversion recovery fat suppression) and CSE-based MRI (gradient-echo 6-point modified Dixon) in addition to routine clinical spine MRI at 1.5T or 3.0T.
Reference standard(s)	• Biopsy result / surgical pathology • Clinical and radiological follow up (if no surgery/biopsy done)
Duration of follow-up	≥ 6 months.
Sources of funding	Not reported.
Results	See Appendix L

Critical appraisal - QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Shi 2017

Shi Y, Li X, Zhang X, et al. Differential diagnosis of hemangiomas from spinal osteolytic metastases using 3.0 T MRI: comparison of T1-weighted imaging, chemical-shift imaging, diffusion-weighted and contrast-enhanced imaging. Oncotarget, 8, 71095-71104, 2017

Study details

Country/ies where study was carried out	China.
Study type	Retrospective cohort study
Study dates	October 2013 - November 2015.
Inclusion criteria	• history of primary malignancy confirmed by needle biopsy or pathological examination following surgery • patients with spinal lesions who underwent conventional MRI at 3T as well as DWI with ADC values, chemical-shift imaging, and contrast-enhanced imaging

	• CT scanning on the corresponding vertebrae • ≥ 6 months follow-up with either MR or CT imaging • No radiation and chemotherapy history.
Exclusion criteria	• spinal lesions complicated with fracture • lesions without a complete MRI examination • lesions of osteoblastic metastases.
Patient characteristics	N=53 Consecutive patients with spinal haemangiomas or cancer patients with spinal metastases Spinal haemangioma group (n=27): n=33 lesions Age, mean (SD), years: 60.62 (8.23). Gender [number of male, female]: male n=16; female n=11. Myeloma versus other cancer types [percentage with myeloma]: NA. Functional status/fitness for treatment [percentage who were ambulant]: Not reported. Cancer group (n=26) n=71 lesions Age, mean (SD), years: 54.33 (10.66). Gender [number of male, female]: male n=9; female n=17. Myeloma versus other cancer types [number with myeloma]: 0/26. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	MRI - T1-weighted imaging with and without fat suppression, chemical-shift, diffusion-weighted imaging, and enhanced imaging at 3.0 T MRI.
Reference standard(s)	• Biopsy result / surgical pathology • Clinical and radiological follow up (if no surgery/biopsy done) of at least 6 months
Duration of follow-up	6 – 24 months.
Sources of funding	• National Natural Science Foundation of China (Grant No.81471640) • National Natural Science Foundation of China (Grant No. 81371715) • Beijing Health System High Level Health • Technical Personnel Training Plan (No. 2013-3-083).
Results	See Appendix L

Critical appraisal - QUADAS-2

Section	Question	Answer

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low.
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Spinnato 2018
Spinnato P, Bazzocchi A, Facchini G, et al. Vertebral Fractures of Unknown Origin: Role of Computed Tomography-Guided Biopsy. International Journal of Spine Surgery, 12, 673-679, 2018

Study details

Country/ies where study was carried out	Italy.
Study type	Retrospective cohort study
Study dates	Not reported.
Inclusion criteria	Patients with 1 or more non-traumatic vertebral fracture of unknown aetiology.
Exclusion criteria	Not reported.
Patient characteristics	N=32 Patients undergoing CT guided biopsy for vertebral fractures of unknown origin Age, mean (SD), years: 57.1 (23.3). NB included paediatric patients: 5/32 were under 10 years of age. Not reported how many were younger than 16. Gender [number of male, female]: male n=13; female n=19. Myeloma versus other cancer types [number with myeloma]: Not reported. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	CT guided biopsy.
Reference standard(s)	Histopathologist’s assessment of sample adequacy and diagnosis.
Duration of follow-up	Not reported.
Sources of funding	Not reported.
Results	Biopsy specimen of diagnostic standard n=26/32 (Osteopenia n=8/26; multiple myeloma lesions n=6/26; oste-

	omyelitis n=4/26; eosinophilic granuloma n=2/26; lung cancer metastases n=2/26; kidney cancer metastasis n=1/26; mastocytosis n=1/26; Paget's disease n=1/26; dysmielopoiesis secondary to a specific systemic disease n=1/26). Need for second biopsy n=4/32. Complication rate n=0/32.

Critical appraisal - QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	High. *Included paediatric patients: 5/32 were under 10 years of age. Not reported how many were younger than 16..*
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear *(No detail on criteria for sample adequacy or histopathology criteria)*
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Suh 2018
Suh C, Yun S, Jin W, et al. Diagnostic Performance of In-Phase and Opposed-Phase Chemical-Shift Imaging for Differentiating Benign and Malignant Vertebral Marrow Lesions: A Meta-Analysis. American Journal of Roentgenology 211, W1-W10, 2018

Study details

Country/ies where study was carried out	South Korea
Study type	Systematic review of diagnostic accuracy studies
Study dates	October 2017

Inclusion criteria	Studies in patients with vertebral BMLs or VCFs where MRI was used to differentiate between benign and malignant, histo-pathologic result or best-value comparator used as reference standard, an original article and enough data to make a 2x2 table
Exclusion criteria	Case reports or case series; review articles, guidelines, consensus statements, letters, editorials, clinical trials and conference abstracts; studies where opposed-phase images were the index test; studies with insufficient data and studies with overlapping populations
Patient characteristics	N=591 in 12 studies Age, mean, years (SD): ranged from 45 to 68 across studies, SD not reported Sex: female n=293; male n=277 (where reported in 11 studies). Myeloma versus other cancer types [number with myeloma]: Not reported. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	MRI chemical shift imaging
Reference standard(s)	Histopathologic result of biopsy or surgery or best-value comparator (for example clinical or radiological follow-up)
Duration of follow-up	Ranged from 1 to 20 months
Sources of funding	Not reported
Results	See Appendix L
Other information	8/12 included studies were prospective, 4/12 retrospective. Risk of bias assessed using QUADAS 2. 2/12 studies were at high risk of bias due to exclusion criteria and pre-designated cut-off value respectively – others at low risk of bias.

Critical appraisal - ROBIS checklist

Section	Question	Answer
Study eligibility criteria	Concerns regarding specification of study eligibility criteria	Low
Identification and selection of studies	Concerns regarding methods used to identify and/or select studies	Low
Data collection and study appraisal	Concerns regarding methods used to collect data and appraise studies	Low
Synthesis and findings	Concerns regarding the synthesis and findings	Low
Overall study ratings	Overall risk of bias	Low
Overall study ratings	Applicability as a source of data	Fully applicable

Taheri 2017

Taheri M, Mirzaei H, Shahhamzei S, et al. Comparison of chemical shift MR imaging findings between vertebral benign and metastatic lesions. International Journal of Cancer Management 10, e8661, 2017

Study details	
Country/ies where study was carried out	Iran.
Study type	Prospective cohort study
Study dates	2010 - 2012.
Inclusion criteria	Patients with vertebral focal lesions referred for routine MR imaging of the spine at a single institution (cervical, thoracic, lumbosacral imaging or any combination) • vertebral lesions with abnormal SI on conventional MRI or bone nuclear scan • previous history of malignancy and vertebral lesion • known metastatic vertebral lesions and new onset of acute back pain and tenderness over vertebral column) (less than 20 days).
Exclusion criteria	• Patients who had received radiotherapy. • Patients for whom adequate follow-up or documentation could not be obtained were excluded from the analysis.
Patient characteristics	N=51 Patients with vertebral focal lesions referred for routine MR imaging n=116 vertebral focal lesions Age, mean (SD), years: 52.61 (13.52) Gender [number of male, female]: male n=28; female n=23. Myeloma versus other cancer types [percentage with myeloma]: Not reported. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	MRI - dual-phase chemical shift MRI. MR imaging was performed for all patients using a 1.5-Tesla superconducting system. Also a phased array spine coil was used. The following pulse sequences were used for all patients: sagittal T1-weighted spin-echo (400-700/8-16 [repetition time (TR) msec/echo time (TE) msec]), sagittal T2-weighted fast spin-echo (TR/TE 2000-5000/80-100) fast multi-planar spoiled gradient-echo MR imaging. Chemical shift sequences for sagittal IP were obtained at RT/ET 100-165/4.2 and OP 100-165/2.4 with breath holding. The flip angle was 30°. For chemical shift MR imaging, the total imaging time was 40 - 50 seconds for the entire pulse sequence. Sagittal images with a 4-mm section thickness and a 1- mm section gap were obtained for all sequences. The field of view was 20 cm for cervical vertebrae, 34 cm for thoracic vertebrae, and 24 cm for lumbosacral vertebrae. The matrix was 256 - 192.
Reference standard(s)	• Biopsy result / surgical pathology • Clinical and radiological follow up (if no surgery/biopsy done) On the basis of final clinical diagnosis after follow-up, vertebral lesions were classified as either benign focal lesions or malignant lesions. Final diagnosis of malignant lesions was proved by biopsy in 27 lesions, 49 of them had known underlying malignancy and metastatic vertebral lesion and in 11 cases diagnosis was based on clinical basis.
Duration of follow-up	6 - 12 months.
Sources of funding	National Foundation of Iranian Elites, Tehran, Iran No. 15/3012 dated 22/7/1389 (14 October 2010).

Results	See Appendix L.

Critical appraisal - QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low

Zafar 2020
Zafar U, Malik A, Shahzad I, et al. Diagnostic accuracy of qualitative diffusion weighted MRI of spine in differentiating between benign and malignant vertebral fractures taking histopathology as gold standard. Pakistan Journal of Medical and Health Sciences 14, 390-392, 2020

Study details

Country/ies where study was carried out	Pakistan.
Study type	Prospective cohort study
Study dates	July 2016 - June 2017.
Inclusion criteria	• Patients with vertebral fractures on digital x ray of spine showing decreased vertebral body height, reduced disc intervertebral disc space or collapsed vertebra (reported by radiologist) • Aged between 16 and 60 years
Exclusion criteria	• Patients with history of caries spine • Patients with claustrophobia • Patients with prosthesis/metal implants.
Patient characteristics	N=280 Patients with vertebral fractures on digital x ray of spine showing decreased vertebral body height, reduced disc intervertebral disc space or collapsed vertebra referred for diffusion weighted magnetic resonance imaging Age, mean (SD), years: 42.61 (11.79). Gender [number of male, female]: male n=156; female n=124.

	Myeloma versus other cancer types [percentage with myeloma]: Not reported. Functional status/fitness for treatment [percentage who were ambulant]: Not reported.
Index test(s)	MRI - Plain and diffusion weighted imaging with a 1.5-T MR unit and a spine-array surface coil. The conventional MR imaging protocols included a sagittal T1- weighted turbo spin-echo sequence, sagittal T2- weighted turbo spin-echo sequences with and without fat suppression, and an axial T2-weighted turbo spin-echo sequence. An axial T1- weighted turbo spin-echo sequence and axial and sagittal fat-suppressed contrast material–enhanced T1-weighted sequences was performed. A single consultant radiologist reported the vertebral fracture as benign or malignant lesion without prior knowledge of biopsy results. Data was collected on structured proforma.
Reference stand-ard(s)	• Biopsy result / surgical pathology
Duration of follow-up	Not reported.
Sources of funding	Not reported.
Results	See Appendix L.

Critical appraisal - QUADAS-2

Section	Question	Answer
Patient selection: risk of bias	Could the selection of patients have introduced bias?	Low
Patient selection: applicability	Are there concerns that included patients do not match the review question?	Low
Index tests: risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low
Index tests: applicability	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low
Reference standard: risk of bias	Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Reference standard: applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low
Flow and timing: risk of bias	Could the patient flow have introduced bias?	Low