REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

RESSONÂNCIA MAGNÉTICA NA DIFERENCIAÇÃO ENTRE FRATURAS VERTEBRAIS BENIGNAS POR INSUFICIÊNCIA E FRATURAS POR COMPRESSÃO VERTEBRAL MALIGNAS: REVISÃO FOCADA E SÍNTESE QUANTITATIVA EXPLORATÓRIA

MAGNETIC RESONANCE IMAGING FOR DIFFERENTIATING BENIGN INSUFFICIENCY-RELATED FROM MALIGNANT VERTEBRAL COMPRESSION FRACTURES: A FOCUSED REVIEW AND EXPLORATORY QUANTITATIVE SYNTHESIS

RESONANCIA MAGNÉTICA PARA DIFERENCIAR FRACTURAS VERTEBRALES BENIGNAS POR INSUFICIENCIA DE FRACTURAS VERTEBRALES MALIGNAS POR COMPRESIÓN: REVISIÓN FOCALIZADA Y SÍNTESIS CUANTITATIVA EXPLORATORIA

Renan Moraes Faria Monteiro Belém
Médico radiologista, Instituto Goiano de Radiologia - IGR, Brasil
E-mail: renanbelem@hotmail.com

George Harrison Ferreira de Carvalho
Doutor em Ciências Médicas (UnB), pós-doutorando (UnB), Brasil
E-mail: georgeharrisonfc@gmail.com

Resumo

As fraturas de compressão vertebral decorrentes de osteoporose ou de outras formas de insuficiência óssea podem apresentar características semelhantes às fraturas produzidas pela substituição maligna da medula óssea. Esta revisão focada com síntese quantitativa exploratória avaliou características da ressonância magnética empregadas na diferenciação entre fraturas vertebrais benignas e malignas, com síntese descritiva dos estudos que forneceram dados reconstruíveis de acurácia para medidas baseadas em deslocamento químico ou na técnica Dixon. Foram realizadas buscas na MEDLINE/PubMed, rastreamento de citações e buscas direcionadas por DOI até 14 de setembro de 2026. Foram elegíveis estudos com adultos portadores de fraturas de compressão vertebral que utilizaram a ressonância magnética como teste índice e determinaram a natureza benigna ou maligna por histopatologia, acompanhamento clínico e de imagem ou padrão de referência

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

composto. Quatro estudos, compreendendo 201 observações no nível vertebral, das quais 83 malignas e 118 benignas, forneceram dados 2 × 2 reconstruíveis para técnicas quantitativas de ressonância magnética sensíveis à gordura. Entre os estudos, a sensibilidade variou de 95,0% a 100,0%, com mediana de 95,4%, e a especificidade variou de 85,7% a 100,0%, com mediana de 92,4%. Devido ao pequeno número de estudos e à heterogeneidade das técnicas, limiares, padrões de referência e unidades de análise, não foi calculada uma estimativa metanalítica combinada. Os sinais convencionais mais consistentemente associados à malignidade foram massa paravertebral ou epidural, comprometimento do pedículo ou de outros elementos posteriores, substituição difusa da medula óssea e convexidade do contorno vertebral posterior. Padrão medular em faixa, sinal do líquido, retropulsão de fragmento ósseo e fraturas consolidadas em outros níveis favoreceram benignidade. Técnicas quantitativas de deslocamento químico e Dixon demonstraram capacidade discriminatória complementar, embora a variabilidade dos limiares, os delineamentos retrospectivos, os padrões de referência heterogêneos e o agrupamento no nível da lesão reduzam a certeza dos achados. A interpretação deve integrar morfologia, gordura medular, difusão e contexto clínico, sem depender de um único sinal.

Palavras-chave: ressonância magnética; fratura de compressão vertebral; fratura por insuficiência; fratura vertebral maligna; deslocamento químico.

Abstract

Vertebral compression fractures caused by osteoporosis or other forms of bone insufficiency may closely resemble fractures produced by malignant marrow replacement. This focused review with exploratory quantitative synthesis evaluated magnetic resonance imaging features used to distinguish benign from malignant vertebral compression fractures and descriptively summarized studies reporting reconstructible accuracy data for chemical-shift or Dixon-based measurements. MEDLINE/PubMed searches, citation tracking, and DOI-directed searches were performed through 14 September 2026. Eligible studies evaluated adults with vertebral compression fractures, used MRI as the index test, and established benign or malignant status through histopathology, clinical and imaging follow-up, or a composite reference standard. Four studies comprising 201 vertebral-level observations, including 83 malignant and 118 benign observations, supplied reconstructible 2 × 2 data for quantitative fat-sensitive MRI. Across studies, sensitivity ranged from 95.0% to 100.0%, with a median of 95.4%, and specificity ranged from 85.7% to 100.0%, with a median of 92.4%. Because of the small number of studies and heterogeneity in techniques, thresholds, reference standards, and analytic units, no pooled meta-analytic estimate was calculated. Conventional signs most consistently associated with malignancy were paraspinal or epidural mass, pedicle or other posterior-element involvement, diffuse marrow replacement, and convex posterior vertebral contour. A band-like marrow pattern, fluid sign, retropulsion of a bone fragment, and healed fractures at other levels favored benignity. Quantitative chemical-shift and Dixon techniques provided useful adjunctive discrimination, but threshold variability,

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

retrospective designs, mixed reference standards, and lesion-level clustering limited certainty. MRI interpretation should integrate morphology, marrow fat, diffusion behavior, and clinical context rather than rely on a single sign.

Keywords: magnetic resonance imaging; vertebral compression fracture; insufficiency fracture; malignant vertebral fracture; chemical-shift imaging.

Resumen

Las fracturas por compresión vertebral causadas por osteoporosis u otras formas de insuficiencia ósea pueden presentar características similares a las fracturas producidas por sustitución maligna de la médula ósea. Esta revisión focalizada con síntesis cuantitativa exploratoria evaluó las características de la resonancia magnética utilizadas para diferenciar fracturas vertebrales benignas y malignas y resumió descriptivamente los estudios con datos reconstruibles de exactitud para mediciones basadas en desplazamiento químico o técnica Dixon. Se realizaron búsquedas en MEDLINE/PubMed, seguimiento de citas y búsquedas dirigidas por DOI hasta el 14 de septiembre de 2026. Cuatro estudios, con 201 observaciones a nivel vertebral, 83 malignas y 118 benignas, proporcionaron datos 2 × 2 reconstruibles para técnicas cuantitativas sensibles a la grasa. La sensibilidad de los estudios varió del 95,0% al 100,0%, con una mediana del 95,4%, y la especificidad varió del 85,7% al 100,0%, con una mediana del 92,4%. Debido al pequeño número de estudios y a la heterogeneidad de las técnicas, umbrales, estándares de referencia y unidades de análisis, no se calculó una estimación metaanalítica combinada. Los signos convencionales más asociados con malignidad fueron masa paravertebral o epidural, afectación del pedículo o de otros elementos posteriores, sustitución difusa de la médula ósea y convexidad del contorno vertebral posterior. Un patrón medular en banda, el signo del líquido, la retropulsión de un fragmento óseo y las fracturas consolidadas en otros niveles favorecieron benignidad. La interpretación debe integrar morfología, grasa medular, difusión y contexto clínico, sin depender de un único signo.

Palabras clave: resonancia magnética; fractura por compresión vertebral; fractura por insuficiencia; fractura vertebral maligna; desplazamiento químico.

1. Introduction

Vertebral compression fractures are common in older adults, yet their cause may remain uncertain when acute marrow edema obscures residual fat and produces signal abnormalities resembling tumor infiltration. The distinction has direct consequences: a benign insufficiency fracture is generally managed with

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

analgesia, osteoporosis treatment, mobilization, or selected vertebral augmentation, whereas a metastatic fracture can change cancer staging and prompt biopsy, systemic therapy, radiotherapy, or surgical stabilization. Conventional radiography provides an initial assessment of spinal alignment and osseous abnormalities, while computed tomography offers more precise characterization of lesion location, cortical changes, bone destruction, and mineralization (Vasilevska Nikodinovska; Kaur; Lalam, 2024). MRI is particularly valuable because it depicts marrow replacement, edema, posterior-element disease, and epidural or paraspinal soft tissue.

The diagnostic challenge arises from overlap. Acute osteoporotic fractures may demonstrate extensive T1 hypointensity, T2 or STIR hyperintensity, and enhancement. Conversely, some malignant fractures retain focal fat, lack a large soft-tissue component, or become difficult to characterize after treatment. Early conventional MRI studies therefore evaluated combinations of morphology and signal distribution rather than a single decisive feature (Yuh et al., 1989; Baker et al., 1990). Subsequent work assessed diffusion-weighted imaging, apparent diffusion coefficients, in-phase and opposed-phase chemical-shift imaging, and multi-echo Dixon fat fraction.

A previous research synthesis identified several individual MRI features with large diagnostic odds ratios but also emphasized limited assessment of reader reliability and substantial methodological variation (Thawait et al., 2012). Newer quantitative studies justify an updated, focused synthesis. The present review evaluates how MRI differentiates benign insufficiency-related vertebral compression fractures from malignant fractures and quantitatively summarizes reconstructible diagnostic accuracy for fat-sensitive chemical-shift and Dixon measurements.

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

2. Materials and Methods

2.1. Review Design and Reporting

This study was designed as a focused evidence review with an exploratory quantitative synthesis. Its reporting structure was informed by diagnostic-test-accuracy principles and the QUADAS-2 domains (Whiting et al., 2011), but it was not intended to constitute an exhaustive systematic review. The protocol was not prospectively registered. Study screening and data extraction were performed by one reviewer; the absence of independent duplicate assessment was considered a potential source of selection and extraction bias.

2.2. Eligibility Criteria

Eligible reports for the exploratory quantitative synthesis included adults with one or more vertebral compression fractures; compared benign fractures attributed to osteoporosis or insufficiency with malignant fractures caused primarily by metastasis or hematologic malignancy; evaluated a quantitative fat-sensitive MRI index test based on in-phase/opposed-phase chemical shift or Dixon fat fraction; and used histopathology, longitudinal clinical and imaging follow-up, or a defined composite reference standard. Inclusion in the pooled analysis additionally required verified true-positive, false-negative, true-negative, and false-positive counts, or sample sizes and accuracy measures that allowed these cells to be reconstructed without unresolved ambiguity. Reviews, editorials, case reports, nonvertebral studies, treatment studies, reader-reproducibility studies without reconstructible diagnostic tables, and computational classification studies outside the prespecified index tests were not pooled. Seminal studies and reviews of conventional MRI findings were retained for narrative clinical context.

2.3. Search and Study Selection

A focused PubMed search was performed through 14 September 2026 using the expression: (“vertebral compression fracture”) AND (“chemical shift” OR “opposed phase” OR Dixon OR “fat fraction”) AND (malignant OR metastatic OR neoplastic)

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

AND (MRI OR “magnetic resonance”). The search retrieved 24 records. Reference-list screening and DOI-directed searches were subsequently used to locate potentially relevant reports not retrieved by the focused electronic strategy. An author-curated supplementary bibliography containing 59 Portuguese and English language candidate records was also examined as an auxiliary source for citation tracking. This bibliography was not treated as an independent bibliographic database, and its records may have overlapped with those retrieved from PubMed. Because exhaustive multidatabase screening, formal deduplication, and independent duplicate study selection were not performed, these searches should not be interpreted as constituting a definitive systematic-review dataset. The quantitative synthesis was restricted to four studies for which complete or reliably reconstructible 2 x 2 data had been verified.

Potentially relevant reports were examined for population, index test, reference standard, and availability of diagnostic contingency data. Studies were retained in the narrative synthesis when they contributed clinically relevant information on conventional or advanced MRI findings. Inclusion in the exploratory quantitative synthesis additionally required complete or reliably reconstructible numbers of true-positive, false-negative, true-negative, and false-positive findings. Reports were not included quantitatively when they evaluated a different index test, included a nonvertebral population, focused primarily on reader reproducibility or computational classification, duplicated a previously reported population, or did not permit unambiguous reconstruction of a single 2 × 2 table. Donners et al. (2020) was retained for narrative context but not included quantitatively because its published accuracy estimates were reader-specific and rounded, preventing unequivocal reconstruction of one prespecified contingency table. Because formal counts for every stage of screening were not prospectively recorded, a PRISMA-DTA flow diagram was not constructed (McInnes et al., 2018).

2.4. Data Extraction and Quality Assessment

Extracted variables were publication year, MRI technique, analytic unit, benign and malignant sample sizes, threshold, reference standard, and diagnostic 2 by 2

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

counts. Malignancy was treated as the target condition. Risk of bias was judged across patient selection, index test, reference standard, and flow and timing. Thresholds selected retrospectively from the same dataset were considered at risk of index-test bias; composite standards without uniform histopathology were considered potentially biased.

Risk of bias was assessed qualitatively using the four QUADAS-2 domains: patient selection, index test, reference standard, and flow and timing. Applicability concerns were considered separately for patient selection, the index test, and the reference standard. Each domain was classified as low, high, or unclear. Retrospective or insufficiently described enrollment was classified as unclear or high risk in the patient-selection domain. Thresholds derived after inspection of the same dataset were classified as high risk for the index test. Composite reference standards and incomplete reporting of the interval between MRI and final diagnosis were classified conservatively as unclear risk. These judgments were qualitative because a duplicate independent QUADAS-2 assessment was not performed.

2.5. Quantitative and Narrative Synthesis

Diagnostic accuracy was summarized descriptively at the vertebral-lesion level. For each included study, sensitivity and specificity were calculated from the extracted true-positive, false-negative, true-negative, and false-positive counts. Ninety-five percent confidence intervals were calculated using the Wilson score method. Because only four studies provided reliably reconstructible 2 × 2 data and because they differed in MRI technique, positivity threshold, patient selection, reference standard, and unit of analysis, a formal pooled meta-analytic estimate was not considered sufficiently robust. Therefore, no bivariate hierarchical, hierarchical summary receiver operating characteristic, or random-effects pooled model was fitted.

The quantitative synthesis presents the range and median of the study-level sensitivity and specificity estimates, together with a forest plot without a pooled summary estimate. The medians were used only as descriptive measures and should not be interpreted as meta-analytic estimates. Chemical-shift signal-intensity ratios

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

and Dixon-derived fat fractions were interpreted as related methods for evaluating residual marrow fat, but they were not assumed to be diagnostically equivalent. Differences in technique and positivity thresholds were examined qualitatively.

Descriptive calculations and graphical displays were generated from the extracted 2 × 2 data. These data were cross-checked against the sample sizes and diagnostic accuracy measures reported in the original publications. Recalculated study-level sensitivity and specificity were then compared with the values reported in Table 1.

3. Results

3.1. Study Selection and Characteristics

The focused PubMed strategy retrieved 24 records. The supplementary author-curated bibliography contained 59 candidate records and was used only for auxiliary citation tracking; overlap between these sources was possible and was not formally quantified. After assessment of the available reports, four studies met the additional requirements for inclusion in the exploratory quantitative synthesis by providing complete or reliably reconstructible diagnostic 2 × 2 data: Erly, Oh, and Outwater (2006), Ragab et al. (2009), Yilmaz Ovali et al. (2017), and Kim et al. (2017). Together they included 201 vertebral-level observations, of which 83 were malignant and 118 benign. The first three used in-phase and opposed-phase signal behavior; Kim et al. used multi-echo Dixon fat fraction. Positivity thresholds were not uniform, ranging from signal-intensity ratios near 0.80-0.96 to a fracture fat-fraction cutoff of 5.26%. Donners et al. (2020), which evaluated Dixon-derived fat fraction in 69 vertebral fractures, was retained for narrative context but excluded from the descriptive 2 × 2 synthesis because the reported sensitivity and specificity were reader-specific and rounded. The presence of additional candidate studies is acknowledged, and the findings are presented as an exploratory descriptive synthesis rather than a definitive systematic-review dataset.

Table 1. Studies included in the quantitative synthesis. SIR = signal-intensity ratio; FF = fat fraction.

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

Study	MRI metric and cutoff	TP	FN	TN	FP	Sensitivity	Specificity
Erly et al. (2006)	Opposed/in-phase SIR, 0.80	19	1	26	3	95.0%	89.7%
Ragab et al. (2009)	Signal loss/SIR	19	1	20	0	95.0%	100.0%
Yilmaz Ovali et al. (2017)	Opposed/in-phase SIR, 0.96	19	0	42	7	100.0%	85.7%
Kim et al. (2017)	Six-echo Dixon FF, 5.26%	23	1	19	1	95.8%	95.0%

Supplementary Table 1. Eligibility framework for the exploratory quantitative synthesis.

Report	Population compatible	Eligible fat-sensitive MRI index	Reference standard reported	Reconstructible 2 × 2 data	Quantitative decision
Erly et al. (2006)	Yes	Yes	Yes	Yes	Included
Ragab et al. (2009)	Yes	Yes	Yes	Yes	Included
Yilmaz Ovali et al. (2017)	Yes	Yes	Yes	Yes	Included
Kim et al. (2017)	Yes	Yes	Yes	Yes	Included
Donners et al. (2020)	Yes	Yes	Yes	No - reader-specific rounded estimates	Narrative synthesis only

Reports used exclusively for narrative clinical context were not considered quantitatively eligible unless their diagnostic contingency data could be reconstructed without unresolved assumptions.

3.2. Descriptive Diagnostic Performance

Across the four included studies, study-level sensitivity ranged from 95.0% to 100.0%, with a median of 95.4%. Specificity ranged from 85.7% to 100.0%, with a median of 92.4%. Three false-negative and 11 false-positive classifications were reported among the 201 vertebral-level observations included in the descriptive

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

synthesis. These totals were not used to calculate a pooled diagnostic accuracy estimate because the studies differed in patient populations, MRI techniques, positivity thresholds, reference standards, and the degree of clustering of multiple vertebrae within individual patients. Study-level sensitivity and specificity with Wilson 95% confidence intervals are displayed in Figure 1. The consistently high study-level estimates suggest that quantitative assessment of residual marrow fat may provide clinically useful information. However, the small number of studies and their methodological differences prevent derivation of a universal diagnostic threshold or definitive summary accuracy estimate.

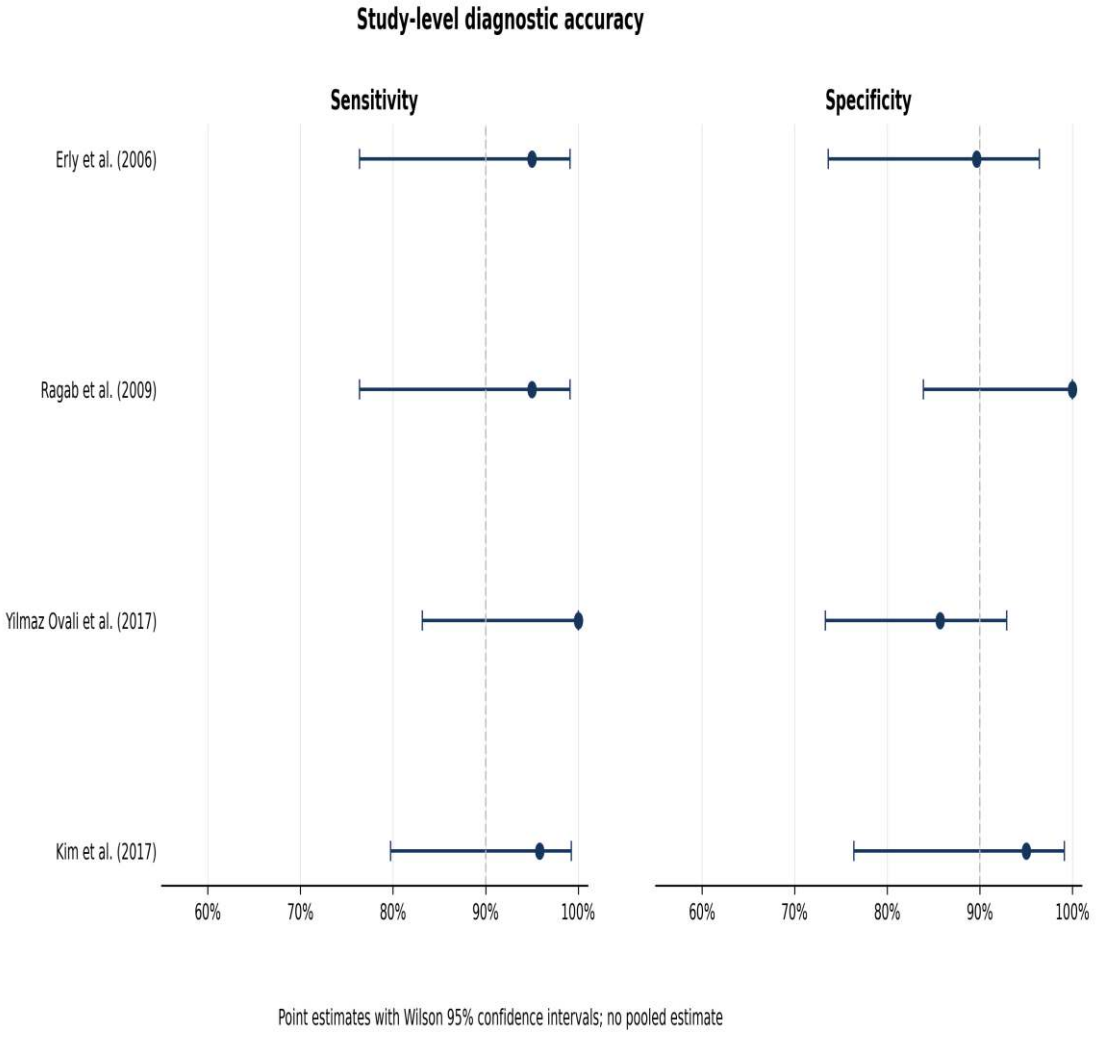


Figure 1. Study-level sensitivity and specificity with Wilson 95% confidence intervals. Estimates are displayed descriptively; no pooled meta-analytic estimate was calculated.

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

3.3. Conventional and Advanced MRI Findings

Morphologic evidence of extraosseous tumor was the most persuasive conventional indicator of malignancy. A paraspinal or epidural mass, pedicle or other posterior-element involvement, diffuse replacement of normal marrow, and a smoothly convex posterior vertebral border consistently increased suspicion of malignant fracture. Malignant collapse may also coexist with additional noncharacteristic marrow lesions. However, posterior-element signal abnormality is not completely specific because biomechanical stress and inflammation may extend into the pedicles of benign fractures (Mauch et al., 2018).

Benignity was favored by a linear or band-like zone of abnormal signal parallel to an endplate, visible fracture line, retropulsion of a bone fragment rather than expansile convexity, a fluid sign, and other healed compression fractures. Baur et al. (2002) described the fluid sign as strongly associated with acute osteoporotic fracture, although it is not universally present. Enhancement alone is insufficient because both granulation tissue and tumor may enhance.

Diffusion-weighted imaging has shown inconsistent performance across studies because qualitative high signal depends on sequence design and T2 shine-through, while measured ADC values vary with b values, susceptibility, marrow composition, and lesion age. Fat-sensitive approaches have a clearer biological rationale: benign edema generally preserves intermixed marrow fat and loses signal on opposed-phase imaging, whereas malignant cellular replacement reduces that signal loss. Exceptions include fat-containing or sclerotic metastases and treated marrow, particularly after radiotherapy (Erly; Oh; Outwater, 2006; Yilmaz Ovali et al., 2017).

3.4. Risk of Bias and Applicability

Overall, the included studies raised moderate to substantial concerns regarding risk of bias and applicability. Patient enrollment was frequently retrospective or incompletely consecutive. Thresholds were commonly optimized after inspection of the study data, which may inflate apparent accuracy.

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

Histopathology was not ethically or clinically appropriate for every benign fracture, leading to composite reference standards and potential incorporation bias when follow-up MRI contributed to final classification. Several studies analyzed multiple vertebrae from the same patient without accounting for within-patient correlation. Applicability was further limited by the inclusion of trauma, infection, myeloma, or previously treated lesions alongside the intended comparison of osteoporotic insufficiency fracture and untreated malignant fracture.

Table 2A. QUADAS-2 risk-of-bias judgments.

Study	Patient selection	Index test	Reference standard	Flow and timing
Erly et al. (2006)	Unclear	High	Unclear	Unclear
Ragab et al. (2009)	Unclear	High	Unclear	Unclear
Yilmaz Ovali et al. (2017)	Unclear	High	Unclear	Unclear
Kim et al. (2017)	Unclear	High	Unclear	Unclear

Table 2B. QUADAS-2 applicability concerns.

Study	Patient selection	Index test	Reference standard
Erly et al. (2006)	High	Low	Low
Ragab et al. (2009)	High	Low	Low
Yilmaz Ovali et al. (2017)	High	Low	Low
Kim et al. (2017)	High	Low	Low

High risk in the index-test domain reflects data-derived or insufficiently prespecified diagnostic thresholds. High applicability concern for patient selection reflects heterogeneous benign comparators, inclusion of hematologic malignancy or treated lesions in some samples, and lesion-level rather than exclusively patient-level analysis. The evaluated index tests and target-condition definitions were considered consistent with the focused review question.

4. Discussion

This synthesis indicates that MRI provides clinically useful discrimination between benign and malignant vertebral compression fractures, particularly when conventional morphology is combined with a quantitative measure of residual marrow fat. Across the four studies with reconstructible contingency data, sensitivity ranged

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

from 95.0% to 100.0%, and specificity ranged from 85.7% to 100.0%. These ranges describe the included studies and do not represent pooled meta-analytic estimates. The consistency of the point estimates is encouraging, but it should be interpreted in light of the small evidence base, data-derived thresholds, heterogeneous reference standards, and lesion-level clustering.

The descriptive findings should not be interpreted as validation of one universal threshold. Signal-intensity ratios are influenced by echo timing, magnetic-field strength, scanner implementation, region-of-interest placement, lesion sclerosis, and marrow treatment effects. Multi-echo Dixon fat fraction may reduce some technical limitations of two-point chemical-shift imaging, but its evidence base remains smaller. The strong diagnostic performance reported by Kim et al. (2017) therefore requires external validation before the 5.26% cutoff is generalized. Donners et al. (2020) likewise found that Dixon-derived fat fraction could discriminate malignant from osteoporotic fractures, but its optimized cutoff and reader-specific results reinforce the need for prospective external validation. The present findings summarize a shared fat-sensitive diagnostic principle rather than the performance of a standardized MRI sequence or cutoff.

The findings agree with the earlier meta-analysis by Thawait et al. (2012), which reported particularly large diagnostic odds ratios for opposed-phase signal-intensity ratio, low ADC under specified acquisition conditions, paraspinal mass, posterior-element involvement, pedicle involvement, and complete marrow replacement. That synthesis also found that a fluid sign, retropulsion, and a band-like signal pattern favored benign fractures. The present review adds later fat-sensitive evidence but deliberately avoids a new pooled estimate because only four studies supplied transparent, reconstructible contingency data and their methods were not sufficiently uniform.

Clinical interpretation should follow a layered approach. First, the reader should determine whether the collapse pattern and fracture line are mechanically plausible for insufficiency. Second, marrow distribution, pedicles, posterior elements, epidural space, and paraspinal tissues should be assessed. Third, opposed-phase

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

signal loss or Dixon fat fraction should be incorporated when conventional findings remain equivocal. DWI and contrast enhancement may add evidence but should not override discordant morphology or fat-sensitive findings. Finally, prior radiotherapy, known myeloma, marrow-stimulating medication, and treatment response must be considered before assigning etiology.

The most important research need is a prospective, multicenter diagnostic cohort using consecutive patients, prespecified thresholds, a standardized multiparametric MRI protocol, blinded readers of different experience levels, and a reference standard independent of the index MRI wherever feasible. Reporting should provide patient-level and lesion-level 2 × 2 tables and use clustered models when more than one vertebra is evaluated per patient. Such a design would permit a reliable hierarchical summary ROC model and clinically relevant subgroup analysis.

4.1. Limitations

This review has several limitations. Searches did not include subscription-only Embase or Web of Science interfaces, and screening and data extraction were performed by one reviewer without independent duplicate assessment. The protocol was not prospectively registered. Only four studies allowed unambiguous reconstruction of the primary contingency data. The included studies used different quantitative techniques and optimized different thresholds. Some benign groups included traumatic or infectious fractures, and some malignant groups included myeloma rather than metastasis alone. The synthesis was lesion-based and did not adjust for clustering of multiple vertebrae within individuals. To avoid overstating precision under these conditions, no formal pooled meta-analytic estimate was calculated.

In addition, the search process did not generate prospectively documented counts for deduplication, title and abstract screening, full-text assessment, or study-specific exclusion. The absence of a formal PRISMA-DTA flow diagram and a complete excluded-studies table limits the reproducibility of study selection. The QUADAS-2 assessment was qualitative and performed without duplicate

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

independent judgment. Finally, chemical-shift ratios and Dixon fat fractions were discussed together on the basis of a shared biological principle despite differences in acquisition and threshold definition.

Accordingly, the reported ranges and medians describe the included studies and should not be interpreted as definitive summary estimates for every clinical population.

5. Conclusions

Quantitative fat-sensitive MRI measures showed consistently high diagnostic performance for differentiating benign insufficiency-related from malignant vertebral compression fractures in the four studies included in the descriptive synthesis. Study-level sensitivity ranged from 95.0% to 100.0%, and specificity ranged from 85.7% to 100.0%. However, the limited number of studies, heterogeneous thresholds, mixed reference standards, and lesion-level clustering prevented a sufficiently robust formal meta-analysis. No single imaging sign or quantitative threshold should be considered definitive. MRI interpretation should integrate fracture morphology, extraosseous soft-tissue findings, posterior-element involvement, marrow fat, diffusion behavior, treatment history, and clinical context. Prospective external validation with standardized acquisition and prespecified thresholds is required before a universal quantitative cutoff can be recommended.

References

Baker, Larry L., et al. “Benign versus Pathologic Compression Fractures of Vertebral Bodies: Assessment with Conventional Spin-Echo, Chemical-Shift, and STIR MR Imaging.” Radiology, vol. 174, no. 2, 1990, pp. 495-502. https://doi.org/10.1148/radiology.174.2.2296658.

Baur, Andreas, et al. “Acute Osteoporotic and Neoplastic Vertebral Compression Fractures: Fluid Sign at MR Imaging.” Radiology, vol. 225, no. 3, 2002, pp. 730-35. https://doi.org/10.1148/radiol.2253011413.

Donners, Ricardo, et al. “Dixon or DWI: Comparing the Utility of Fat Fraction and

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

Apparent Diffusion Coefficient to Distinguish between Malignant and Acute Osteoporotic Vertebral Fractures." European Journal of Radiology, vol. 132, 2020, article 109342. https://doi.org/10.1016/j.ejrad.2020.109342.

Erly, William K., E. S. Oh, and Eric K. Outwater. "The Utility of In-Phase/Opposed-Phase Imaging in Differentiating Malignancy from Acute Benign Compression Fractures of the Spine." American Journal of Neuroradiology, vol. 27, no. 6, 2006, pp. 1183-88.

Kim, Dong Hyun, et al. "Differentiation of Acute Osteoporotic and Malignant Vertebral Fractures by Quantification of Fat Fraction with a Dixon MRI Sequence." American Journal of Roentgenology, vol. 209, no. 6, 2017, pp. 1331-39. https://doi.org/10.2214/AJR.17.18168.

Mauch, John T., et al. "Review of the Imaging Features of Benign Osteoporotic and Malignant Vertebral Compression Fractures." American Journal of Neuroradiology, vol. 39, no. 9, 2018, pp. 1584-92. https://doi.org/10.3174/ajnr.A5528.

McInnes, Matthew D. F., et al. "Preferred Reporting Items for a Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement." JAMA, vol. 319, no. 4, 2018, pp. 388-96. https://doi.org/10.1001/jama.2017.19163.

Ragab, Yasser, et al. "Differentiation of Osteoporotic and Neoplastic Vertebral Fractures by Chemical Shift In-Phase and Out-of-Phase MR Imaging." European Journal of Radiology, vol. 72, no. 1, 2009, pp. 125-33. https://doi.org/10.1016/j.ejrad.2008.06.019.

Thawait, Shrey K., et al. "Research Synthesis: What Is the Diagnostic Performance of Magnetic Resonance Imaging to Discriminate Benign from Malignant Vertebral Compression Fractures?" Spine, vol. 37, no. 12, 2012, pp. E736-E744. https://doi.org/10.1097/BRS.0b013e3182458cac.

Vasilevska Nikodinovska, Violeta, et al. "Conventional Radiography and Computed Tomography." Imaging of Primary Tumors of the Osseous Spine, edited by Mohamed Fethi Ladeb and Filip Vanhoenacker, Springer Nature Switzerland, 2024, pp. 55-84. https://doi.org/10.1007/174_2023_465.

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 21/09/2026
Vol: 13.20
DOI: 10.66104/g0h0aw68
Pages: 1-17

Whiting, Penny F., et al. “QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies.” Annals of Internal Medicine, vol. 155, no. 8, 2011, pp. 529-36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.

Yilmaz Ovali, Gulgun, et al. “Benign versus Malignant Vertebral Compression, Chemical Shift MR Imaging, Is It Useful?” Iranian Journal of Radiology, vol. 14, no. 2, 2017, article e42016. https://doi.org/10.5812/iranjradiol.42016.

Yuh, William T. C., et al. “Vertebral Compression Fractures: Distinction between Benign and Malignant Causes with MR Imaging.” Radiology, vol. 172, no. 1, 1989, pp. 215-18. https://doi.org/10.1148/radiology.172.1.2740506.

CC BY