Brain and Spine 4 (2024) 102787

Contents lists available at ScienceDirect

Brain and Spine

journal homepage: www.journals.elsevier.com/brain-and-spine

Towards a standardized reporting of the impact of magnetic resonance imaging on the decision-making of thoracolumbar fractures without neurological deficit: Conceptual framework and proposed methodology

Mohamed M. Aly[a,b,*], Sebastian F. Bigdon[c], Ulrich J.A. Speigl[d], Gaston Camino-Willhuber[e], Saleh Baeesa[f], Klaus J. Schnake[g,h]

[a] Department of Neurosurgery, Prince Mohammed Bin Abdulaziz Hospital, Riyadh, Saudi Arabia
[b] Department of Neurosurgery, Mansoura University, Mansoura, Egypt
[c] Department of Orthopedic Surgery and Traumatology, Inselspital, University Hospital, University of Bern, Bern, Switzerland
[d] Department of Orthopedics, Trauma Surgery and Plastic Surgery, University of Leipzig, Leipzig, Germany
[e] Traumatology Department, Asuncion Clinica, Basque Country, Spain
[f] Neuroscience department, King Faisal Specialist Hospital, Jeddah, Saudi Arabia
[g] Center for Spinal and Scoliosis Surgery, Malteser Waldkrankenhaus St. Marien, Erlangen, Germany
[h] Department of Orthopedics and Traumatology, Paracelsus Private Medical University Nuremberg, Nuremberg, Germany

A R T I C L E I N F O

Handling Editor: Prof F Kandziora

Keywords:
Magnetic resonance imaging
Thoracolumbar fractures, indications
Computed tomography
Diagnostic accuracy
Neurologically intact
Posterior ligamentous complex

A B S T R A C T

Introduction: A recent meta-analysis showed that only four prior studies have shown that magnetic resonance imaging (MRI) can change the fracture classification in 17% and treatment decisions in 22% of cases. However, previous studies showed a wide methodological variability regarding the study population, the definition of posterior ligamentous complex (PLC) injury, and outcome measures.
Research question: How can we standardize the reporting of the impact of MRI for neurologically intact patients with thoracolumbar fractures?
Material and methods: All available literature regarding the impact of MRI on thoracolumbar fracture classification or decision-making were reviewed. Estimating the impact of MRI on the TLFs' classification is an exercise of analyzing the CTs' accuracy for PLC injury against MRI as a *''Gold standard''and should follow* standardized checklists such as the Standards for the Reporting of Diagnostic Accuracy Studies. Additionally, specific issues related to TLFs should be addressed.
Results: A standardized approach for reporting the impact of MRI in neurologically intact TLF patients was proposed. Regarding patient selection, restricting the inclusion of neurologically intact patients with A- and B-injuries is crucial. Image interpretation should be standardized regarding imaging protocol and appropriate criteria for PLC injury. The impact of MRI can be measured by either the rate of change in fracture classification or treatment decisions; the cons and pros of each measure is thoroughly discussed.
Discussion and conclusion: We proposed a structured methodology for examining the impact of MRI on neurologically intact patients with TLFs, focusing on appropriate patient selection, standardizing image analysis, and clinically relevant outcome measures.

1. Introduction

Thoracolumbar fractures (TLFs) are the most frequently encountered traumatic spinal fractures, accounting for 60–70% of all cases. (Wood et al., 2014), (Bigdon et al., 2022) Several classification schemes have been presented to guide TLF treatment decisions over the decades (Rosenthal et al., 2018). The older TLFs' classifications were based on the extent of osseous injuries, as shown by conventional radiography or computed tomography (CT). (Denis, 1983)' (Magerl et al., 1994) In the last two decades, the integrity of the posterior ligamentous complex (PLC) has been recognized in predicting delayed kyphosis or back pain. (Holdsworth) Consequently, the Thoracolumbar Injury Classification

* Corresponding author. Department of Neurosurgery, Prince Mohamed Ben Abdulaziz Hospital, P.O Box 54146, Riyadh, 11514, Saudi Arabia.
E-mail addresses: drmoali26@gmail.com, ali@pmah.med.sa (M.M. Aly).

https://doi.org/10.1016/j.bas.2024.102787
Received 2 January 2024; Received in revised form 20 February 2024; Accepted 15 March 2024
Available online 21 March 2024

System (TLICS) and the new AOSpine classification systems have included PLC integrity as a significant element in fracture classification or decision-making. (Lee et al., 2005)., (Vaccaro et al., 2013)

The diagnostic assessment of PLC integrity remains a highly debated topic (Qureshi et al., 2019). CT can only asses the PLC integrity indirectly and was associated with as high as 30% risk of missed PLC injury. (Leferink et al., 2002), (Petersilge et al.) Conversely, Magnetic resonance imaging (MRI) permits direct visualization of PLC structures with high sensitivity and is considered the reference standard (Aly et al., 2022a). The additional expense, longer scanning time, and restricted availability of MRI at some trauma centers prevent its routine usage for TLFs (Khurana et al., 2018). More importantly, there is not enough evidence that MRI can modify TLF management to make its usage warranted despite logistical limitations (Qureshi et al., 2019). In clinical practice, some spine surgeons order MRIs selectively based on their perceived "uncertainty" of PLC status, while others order MRIs in the majority of TLFs to avoid overlooking PLC injury (Leferink et al., 2002), (Petersilge et al.) (Aly et al., 2022b),

A recent meta-analysis reported that MRI can change TLFs classification in 17% and treatment decisions in 22% of cases. (Aly et al., 2022b), (Aly et al., 2024), (Pizones et al., 2011), (Winklhofer et al., 2013), (Durmaz et al., 2021) The considerable variability in the reported impact of MRI could be due to the methodological challenges in analyzing the impact of MRI on TLFs (Aly et al., 2024). Previous research comprised a heterogeneous population regarding the fracture types, neurological status, and treatment modalities (Aly et al., 2024). The impact of MRI was measured by either the rate of change in fracture classification or the change in treatment decisions after MRI compared to CT alone. However, the lack of a consensus-based definition of PLC injury on CT and MRI may cast doubts on the reliability of evaluating the change in fracture classification (Van Middendorp et al., 2013). In addition, a change in fracture classification does not necessarily impact the treatment decisions, for instance, for patients with neurological deficits or unstable burst fractures (Aly et al., 2022b). Furthermore, different metrics were used to assess MRI's impact on treatment decisions, including surgeons' decisions, TLICS, or Thoracolumbar AOSpine injury severity Score (TLAOSIS). (Pizones et al., 2011), (Winklhofer et al., 2013), (Aly et al., 2022b), (Durmaz et al., 2021), (Vaccaro et al., 2016) The primary goal of this study is to propse a standardized methodology for measuring the impact of MRI on TLF decision-making that allows for comparing studies findings and synthesis of higher levels of evidence.

2. Conceptual framework

2.1. Internal and external validity

Estimating the impact of MRI on the TLFs' classification is an exercise of analyzing the CTs' accuracy for PLC injury against MRI as a ''Gold Standard''. It thus should follow the Standards for Reporting Diagnostic Accuracy Studies (STARD). (Bossuyt et al., 2015), (Whiting et al., 2011) The risk of bias should be reported according to QUADAS-2 in four domains: risk patient selection, time and flow domain, index test, and reference test. (Whiting et al., 2011)., (Schmidt and Factor, 2013)

Validity is the degree to which the study's conclusions accurately match the findings among comparable poulation outside the study. Internal validity can be ascertained at the study level by minimizing the study's bias or imprecision. (Fig. 1). (Patino and Ferreira, 2018),(Pavlou et al., 2021) Conversely, the external validity reflects the degree of generalization of findings at the population level. External validity relies

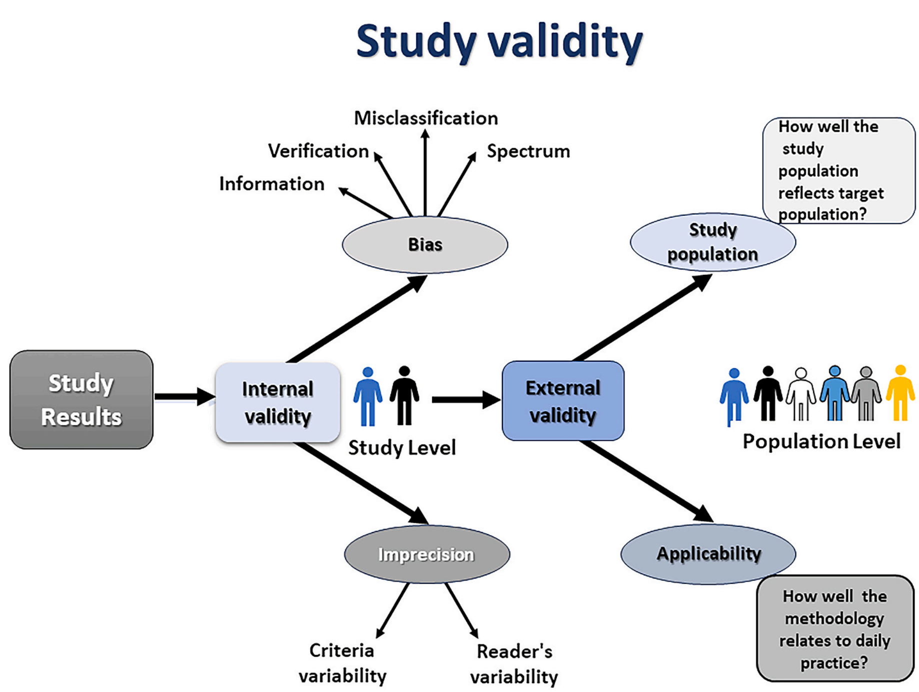

Fig. 1. Factors potentially impacting the internal and external validity of the study
The degree to which the study's results accurately reflect findings among comparable patients outside the study is referred to as validity. The study's validity can be ascertained at the study level (**internal validity**) or the population level (**external validity**). **Bias** and **imprecision** might undermine a study's internal validity. Any systematic deviation of accuracy measures from the actual value caused by a systematic error is referred to as **bias**. **Imprecision** is the random variation that occurs when several parameter estimates are made by multiple readers (interobserver reliability) or the same reader (intra-observer reliability). **Imprecision** is a crucial issue in image-based studies, which might be related to readers' backgrounds, experience, or the inconsistent definition of classifying criteria.
External validity examines whether the findings of a study can be generalized to the population level and relies on the study's population and applicability. The degree to which the study population is representative of the target population with TLFs is a significant determinant of the study's generalizability. The degree to which the study's patient demographics, selection and utilization of the index test, and test interpretation align with the reader's practice dictates the study's applicability. (Pavlou et al., 2021)

on how well the study population is representative of the target population. It also depends on how the study's methodology aligns with real-world practice, i.e., its applicability(Patino and Ferreira, 2018), (Pavlou et al., 2021).

3. PICO framework

A standardized framework for reporting MRI's impact on TLFs is based on the PICO model; PICO stands for P (population), I (index test), C (comparator or reference standard), and O (outcomes (Fig. 2). (Pavlou et al., 2021)

3.1. Domain 1: Population

3.1.1. Inclusion criteria

The main proposed inclusion criteria are (Aly et al., 2022b), (Qureshi et al., 2019):1) all consecutive traumatic thoracic or lumbar fractures (T1-T5); 2) Patients who had CT and MRI within ten days of the accident since the MRI sensitivity for detecting high signal intensity can drop after that time frame (Benedetti et al., 2000); 3) AO A-type or B-type injuries according to the new AOSpine classification. A0: isolated posterior element fractures should not be included, as they represent minor injuries. C-type injuries should be excluded since they are accurately classified with CT only (Vaccaro et al., 2013), (Rajasekaran et al., 2017); 4) neurologically intact patients (N0), those with transient radiculopathy (N1), or when neurological assessment is unattainable (Nx); (Aly et al., 2024) 5) Single or multi-level injuries; for multi-level injuries, only fractures with the highest AO classification should be included; (Vaccaro et al., 2013) 6) patients aged 18 to 65.

3.1.2. Exclusion criteria

The main proposed exclusion criterion could be (Aly et al., 2022b), (Qureshi et al., 2019): 1) Poor image quality due to motion artifacts or inadequate images that do not fully span the injury region; 2)pathological fractures due to tumor infection, among other reasons; 4) osteoporotic fractures (T score ≤2.5 by "dual-energy-X-ray- absorptiometry" (DEXA) or no-energy or low-energy trauma), 5) previous surgery or trauma to the thoracic or lumbar spine.

3.2. Domain 2 and 3: Index and reference tests (image interpretation)

Analysis of MRI's impact on TLFs relies heavily on the quality of CT and MRI image interpretation (Pavlou et al., 2021). The accuracy of image interpretation relies on numerous variables, including imaging acquisition protocol, the validity of PLC classifying criteria, reviewers' experience, consensus training, and image analysis.(Aly et al., 2023a)

3.2.1. Imaging protocol and technical considerations

The MRI spine trauma protocol includes axial and sagittal T2-WI and T1-WI and sagittal short tau inversion recovery (STIR). (Lee et al., 2007)˒ (Crosby et al., 2011) Sagittal STIR is the most sensitive to detect T2-WI hyperintensity or ''black stripe discontinuity''. There is insufficient evidence to recommend 3 T over 1.5 scanners or whole-spine MRI over a focused MRI protocol. (Aly et al., 2021a), (Khurana et al., 2019) The CT protocol should include a soft tissue algorithm in axial and sagittal planes and a bone algorithm (not bone window) in three imaging planes. The slice thickness for axial images of soft tissue and the bone algorithm was 5 mm and 1.5 mm, respectively, and 3 mm or less for reformatted images. (Aly et al., 2021a)˒ (Sixta et al., 2012) Optimizing the device-specific sequence parameters is highly recommended, which proves more challenging for MRI than CT, given the significant number of parameters. (Benedetti et al., 2000)., (Tins, 2010)

3.2.2. Criteria for PLC injury in CT and MRI

We propose that PLC incompetence be defined as ''black stripe discontinuity'' due to a supraspinous ligament (SSL) or ligamentum flavum (LF) disruption, best seen in Sagittal STIR or axial T2 images. conversely, a high signal intensity (HSI) due to interspinous ligament (ISL) edema, facet joint effusion, or complete lack of abnormal SI should be considered as competent PLC (Fig. 3). (Aly et al., 2023a), (Pizones et al., 2012a) PLC injury on CT can be defined by at least one of the following CT finding: spinous process fracture, interspinous widening, laminar fracture (Aly et al., 2021b), and facet diastasis. (Aly et al., 2023b) (Khurana et al., 2018), (Aly et al., 2023b)Recently, two studies have proposed CT criteria for PLC injury based on the number of positive findings as follows: (Khurana et al., 2018), (Aly et al., 2021a) Disrupted PLC, at least two positive findings, indeterminate (AKA M1modifier),

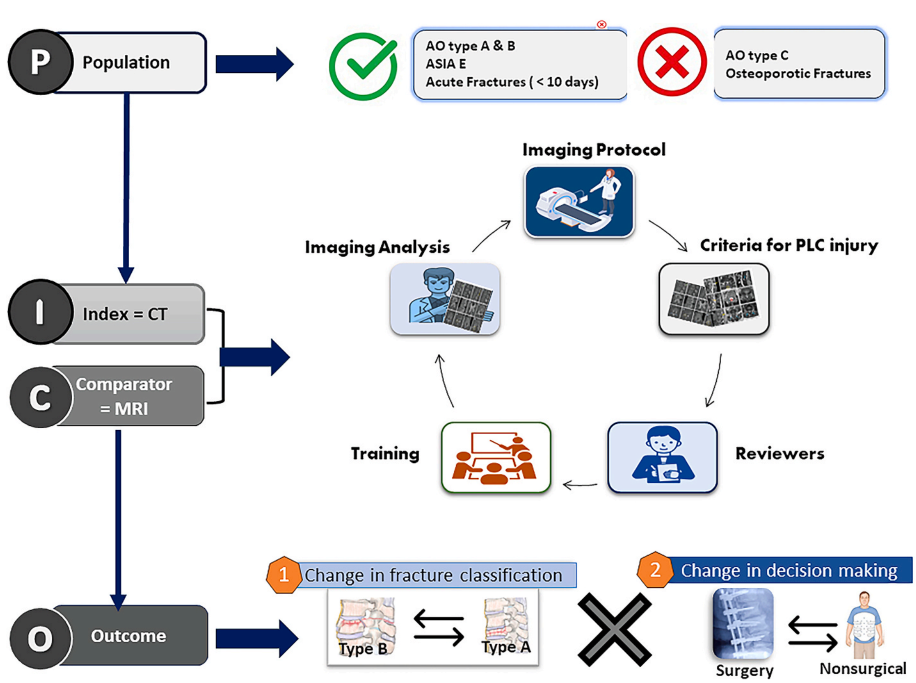

Fig. 2. The proposed PICO framework for reprting MRI's impact on thoracolumbar fractures PICO stands for P (population), I (index test), C (comparator or reference standard), and O (outcomes. PLC, Posterior ligamentous complex. (Pavlou et al., 2021

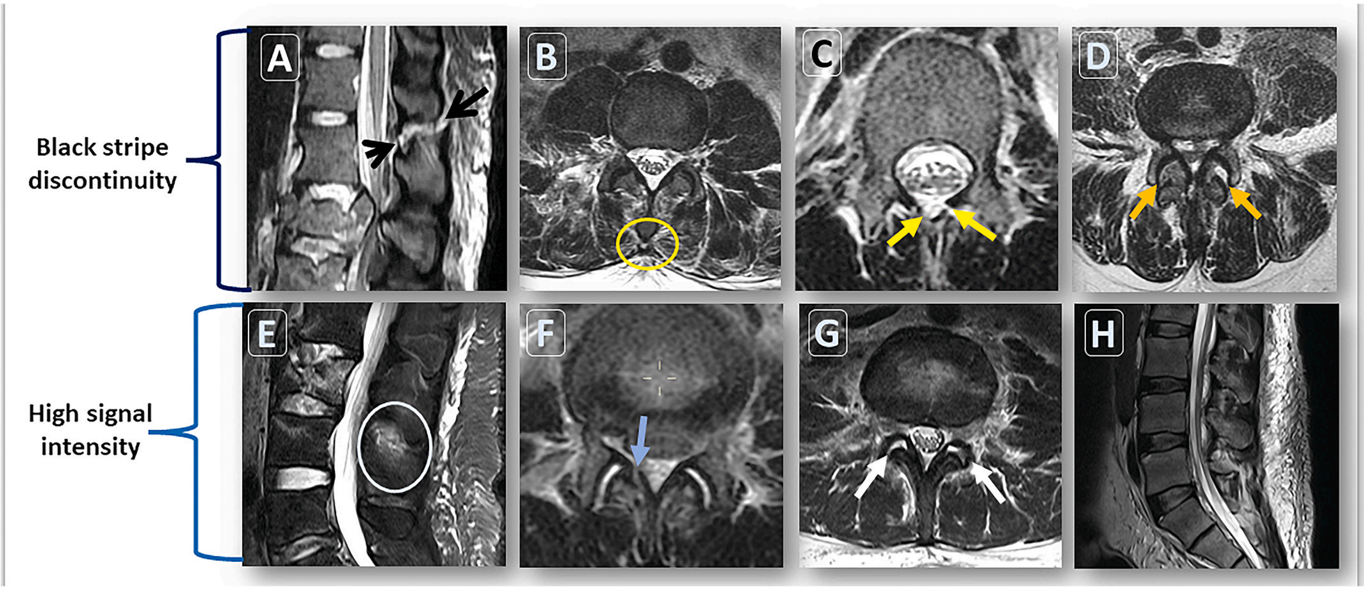

Fig. 3. MRI criteria for thoracolumbar posterior ligamentous complex injury
(A-D) Black stripe discontinuity; (A) Sagittal STIR images are the best sequence to detect the discontinuous black stripe due to supraspinous ligament rupture (black arrow) or ligamentum flavum rupture (black arrow head); **(B)** Axial T2-weighted images show supraspinous ligament stripped from the spinous process tip (green ring); **(C)** Axial T2 image shows diffusely macerated horizontal ligamentum flavum tear (two green arrows); **(D)** Axial T2 images showing bilateral dislocated facets (two green arrows).
(E–H) High signal intensity; (E) Sagittal STIR images show ill-defined high signal intensity indicating interspinous ligament edema (white ring); **(F)** Axial T2 image shows vertical ligamentum flavum tear: unilateral focal ligamentum flavum tears associated with vertical laminar fracture (blue arrow); **(G)** Axial T2-weighted images show high signal intensity due to bilateral facet joint effusion (two white arrows).; **(H)** Sagittal STIR images show no high signal intensity with no black stripe discontinuity, indicating an intact posterior ligamentous complex. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.) Aly et al., 2023a

single positive finding, intact PLC, negative CT for all findings (Fig. 4).. (Aly et al., 2021a)

3.2.3. *Selection of reviewers*

An odd number of reviewers, such as three, will enable the use of *'majority rule'';* approval of two of the three reviewers' reading, to provide the final dataset for accuracy analysis. In contrast, the disagreement between 2 reviewers requires a consensus reading that does not reflect real-life practice. It is recommended to include reviewers of varying degrees of experience, such as attending, fellows, and residents, to enhance the generalizability of findings (Crosby et al., 2011), (Obuchowski and Bullen, 2022). We recommend including at least one spine surgeon and one fellowship-trained radiologist to account for different perspectives in image analysis (Rihn et al., 2010).

3.2.4. *Consensus training*

The consensus pretraining aims to standardize the imaging interpretation and minimize interobserver variability. (Obuchowski and Bullen, 2022) This can be done by providing an illustrated imaging manual or video tutorial that details a standardized definition for imaging findings, the best imaging plane/sequence, and diagnostic pitfalls. (Aly et al., 2023a) The newly proposed imaging algorithm for PLC injury in CT and MRI might be invaluable (Fig. 1). (Aly et al., 2023a) Providing a data set for training, assessing the results for each reviewer, and providing feedback is highly recommended.(Obuchowski and Bullen, 2022) Additionally, an ongoing discussion among the reviewers about the challenges they encounter during image review is highly encouraged.

3.2.5. *Image interpretation protocol*

Each reviewer should review all deidentified images while blinded to all clinical data or other image readings, i.e., blinded reviews. (Bossuyt et al., 2015) When examining the CT or MRI images, knowledge of the other modality's results may bias the interpretation, the so-called diagnostic review bias (Pavlou et al., 2021).To reduce recall bias, CT and MRI images should be presented to the reviewers randomly in two separate sessions with at least four-week intervals. Each reviewer should interpret images independently because consensus reads do not represent either reviewer and are thus not generalizable to clinical practice. (Obuchowski and Bullen, 2022) At least one reviewer should be assigned to review all the images for two rounds to calculate intraobserver reliability. (Aly et al., 2021a) In the case of multiple fractures, the fracture with the highest AO classification will be selected for analysis by one reviewer to avoid confusion. (Vaccaro et al., 2013) Multiplanar image viewer, such as picture archiving and communication systems (PACS), is mandatory as they improve the accuracy of CT and MRI interpretation. (Fig. 1). (Aly et al., 2023a)

3.3. *Domain 4: Outcome and outcome measures*

MRI's impact on TLF management can be measured by its diagnostic impact: change in PLC status or fracture classification or therapeutic impact; a change in treatment decision from surgery to conservative or vice versa. (Aly et al., 2024)

3.3.1. *Diagnostic impact: Rate of change in fracture classification*

The rate of change in AOSpine classification can be quantified as the transition from type A to type B injuries or vice versa due to changes in PLC status after MRI. (Aly et al., 2024) The following should not be counted as a change in fracture classifications: (Aly et al., 2024) 1) a change in classification from A-injuries subtype to another, for example, A1 to A3 due to disclosure of posterior vertebral wall injury or A3 to A4 due to disclosure of second endplate injury. MRI is not routinely utilized in clinical practice to detect bone fractures; 2) identifying mild compression fractures or bone marrow edema as their influence on decision-making is negligible; 3) identifying or grading intervertebral

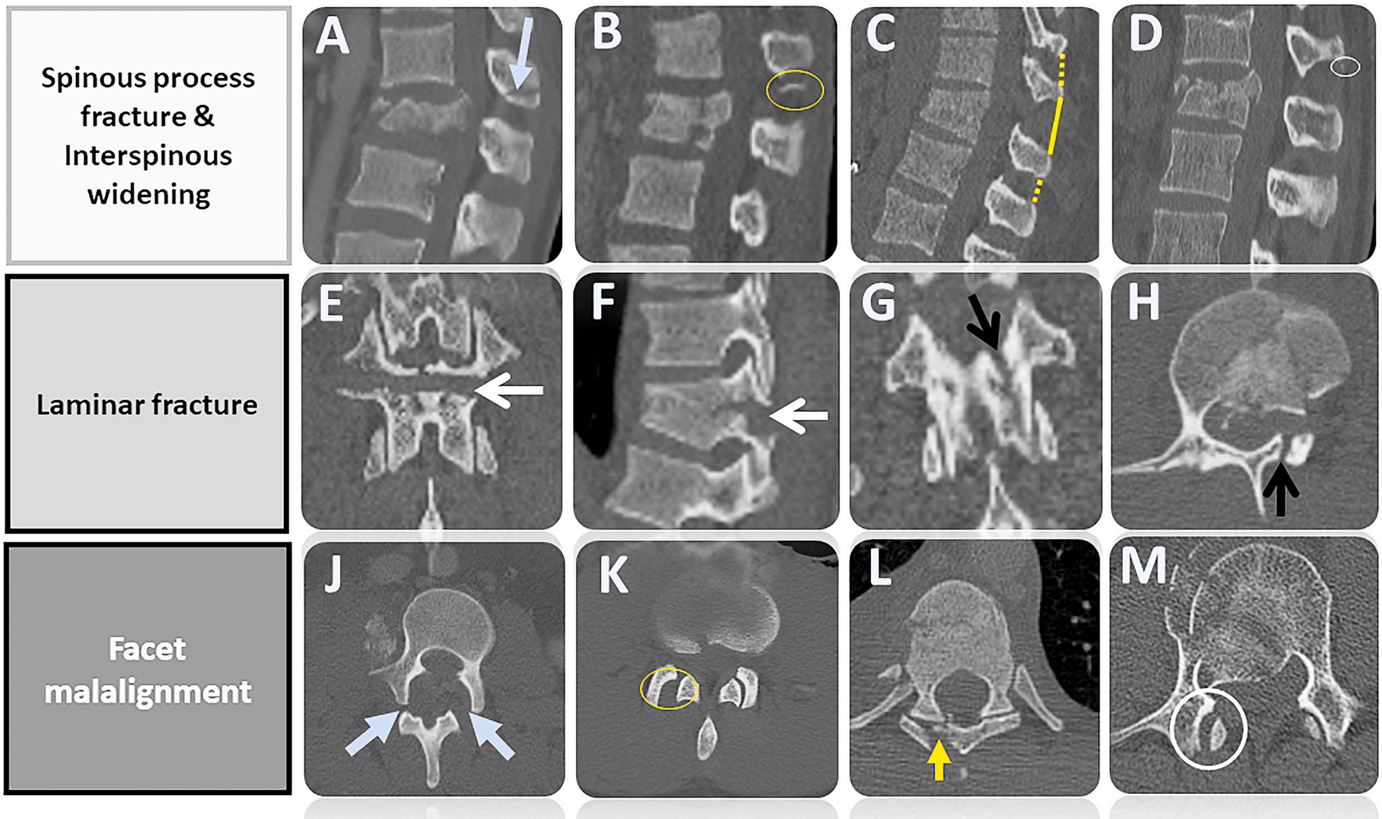

Fig. 4. CT findings of thoracolumbar posterior ligamentous complex injury
(A-D) Midsagittal CT image showing **(A)** transverse spinous process fracture (white arrow), **(B)** avulsion spinous process fracture (green circle), **(C)** interspinous widening (solid green line) compared to the level above and below (2 dashed green lines); **(D)** avulsion spinous process fracture (white circle); **(E)** Coronal reconstruction images showing horizontal lamina fracture (white arrow); **(F)** sagittal CT image showing horizontal laminar fracture (white arow); **(G)** Coronal reconstruction images showing the vertical orientation of the laminar fracture (black arrow); **(H)** Axial CT image showing a vertical fracture of the left side of the lamina (black arrow); **(J–M)** Axial CT axial image showing **(J)** bilateral facet dislocation (2 white arrows); **(K)** right Facet Subluxation (green circle) ; **(L)** displaced fracture of the left superior articular facet (green arrow); **(M)**Facet joint widening (>3 mm with the preserved alignment of articular surfaces (white circle). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.) Aly et al., 2023a

disc injury, as its role in decision-making is still debatable (Kanezaki et al., 2018).

3.3.2. *Therapeutic impact: rate of change in treatment decision*

The rate of change in the decision-making is defined by a shift from conservative to surgical or vice versa due to MRI's change in PLC status. (Aly et al., 2024) The following outcome measures can be used for treatment decision change: 1) the rate of change in prospective treatment decisions after MRI compared to a decision based on CT alone. (Pizones et al., 2012b); (Winklhofer et al., 2013) In a retrospective design, the reviewers should comment on their treatment decision based on the CT and the MRI while blinded to the actual treatments; 2) the rate of change in TLICS-based treatment recommendations after the MRI. The rate of change in the decision can be defined as any switch between the three categories or between*''Conservative/Grey zone''* and *''Surgery''*; 3) the rate of change in TL AOSIS-based treatment recommendations after MRI. The rate of change in the decision can be defined as any switch between the three categories or between*''Conservative/Grey zone''* and *''Surgery''*; (Aly et al., 2022b) 3)Comparing MRI-based decisions with real-world decision-making is highly discouraged given the substantial individual variations in decision-making.

A change in surgical approaches, such as posterior-only vs. 360° fusion, fusion vs non-fusion instrumentation, the extent of instrumentation, or the extent of decompression in patients with neurologic deficits, should not be counted as a change in the treatment decision (Aly et al., 2024).There are no agreed-on indications for these decisions, and they usually rely on surgeons' preferences, making them challenging to standardize. (Aly et al., 2024)

3.3.3. *Prediction of when MRI may change classification or decision-making*

To guide the indications of MRI, it is essential to identify the clinical or radiological predictors of upgrading from type A to type B. (Aly et al., 2024) Potential variables include 1) Indeterminate PLC status; 2) AO fracture subtype (A1-A4); 3)Spine region as thoracic (T1-T10), thoracolumbar junction (T11-L2), or lo lumbar (L3-L5); 4)TLICS/TLAOSIS score (Aly et al., 2022b).

4. Discussion

4.1. *Key findings and interpretation*

Insufficient evidence exists regarding whether MRI sufficiently influences TLFs' decision-making to justify the increased cost and time delay (Aly et al., 2024), (Qureshi et al., 2019). The lack of evidence may be attributed to the intrinsic bias in image-based diagnostic accuracy studies and those unique to TLFs, such as the wide variations in definitions for PLC injury and the outcome metrics for MRI's impact. (Aly et al., 2024) To our knowledge, this is the first study to present a structured methodology for examining the effects of MRI on neurologically intact patients with TLFs.

4.2. Study design

Prospective studies are generally preferred over retrospective designs because they are less prone to recall and sampling bias. (Obuchowski and Bullen, 2022) However, retrospective image-based studies are associated with a low risk of recall bias. A prospective study design might help minimize verification and sampling bias. A retrospective, well-designed study can be as effective as a prospective design if it has a high verification rate and includes a balanced cohort regarding fracture types, mode of treatment, and neurological deficits (Obuchowski and Bullen, 2022). Multicenter studies may have significant advantages over single-center studies, particularly involving diverse settings and geographical areas. The diversity of study populations and observers would reduce selection bias and improve the generalizability of results. Multicenter studies would account for differences between facilities in imaging protocols and CT or MRI machines (Obuchowski and Bullen, 2022).

4.3. Patients selection

4.3.1. Patient population

The key to appropriate patient selection is consecutive patient sampling and adhering to strict eligibility criteria that target TLFs' population whose MRI may affect the decision-making (Elfil and Negida, 2017). Consecutive sampling, including subjects who meet the predefined eligibility criteria for a specified period, has proven more accurate than nonconsecutive sampling (Rutjes et al., 2006). We propose including only neurologically intact patients with TLFs because patients with neurological deficits are universally treated surgically, and MRI will not alter their treatment decision. (Aly et al., 2022b)' (Rajasekaran et al., 2017) C-type injuries may be excluded because they could be accurately classified using CT or even X-ray; hence, MRI has a negligible impact on their treatment decisions. (Aly et al., 2022b)' (Rajasekaran et al., 2017) By excluding those two patient groups, a more homogeneous sample of the target population can be obtained.

4.3.2. Mitigating potential sampling bias

Due to the absence of unequivocal MRI indications for TLFs, surgeons' preferences and logistical considerations have dominated patient selection and led to specific sampling bias. (Aly et al., 2024) Verification bias happens when only a portion of TLFs patients who did CT will be verified by MRI depending on CT results (Fig. 5). (Pavlou et al., 2021), (Schmidt and Factor, 2013)For instance, patients with more severe fractures on CT are more likely to undergo an MRI for PLC assessment than those with less severe injuries who receive conservative treatment based on CT findings. Therefore, patients with neurological deficits and surgical patients are usually overrepresented, potentially underestimating the impact of MRI on decision-making. (Aly et al., 2022c) Spectrum bias occurs when the study's sample is restricted to more severe TLFs due to referral filters to tertiary trauma centers or a higher spectrum of injury in some geographical regions (Fig. 6). (Pavlou et al., 2021) Spectrum bias inflates the accuracy measures and reduces the generalizability of findings to different settings. (Pavlou et al., 2021), (Schmidt and Factor, 2013).

4.3.3. Should osteoporotic fractures be excluded?

Given their high prevalence and low risk of PLC damage, osteoporotic fractures would falsely inflate CT sensitivity for PLC injury if included (Schnake et al., 2018). Excluding patients based on DEXA scan findings may be inappropriate since only a portion of patients with traumatic fractures undergo DEXA scans. (Aly et al., 2023a) Defining osteoporosis based on low Hounsfield units (HU), below 90–130 in axial CT images, may be helpful since it highly correlates with manifest osteoporosis. Ahern et al., 2021 Another approach is to exclude all patients with pre-existing osteoporosis, no-energy, or low-energy trauma since they either have osteoporotic fractures or have a very minimal risk of ligamentous injury. Nonetheless, high-energy fractures in osteoporotic patients should not be excluded because they are at substantial risk of ligamentous injury, which may be particularly challenging to detect in CT images (Aly et al., 2023a). High-energy trauma can be defined by a fall from a great height (>2 m above ground), a car accident, or mountain-related sports, while low-energy trauma is characterized by a fall from standing (Bigdon et al., 2022).

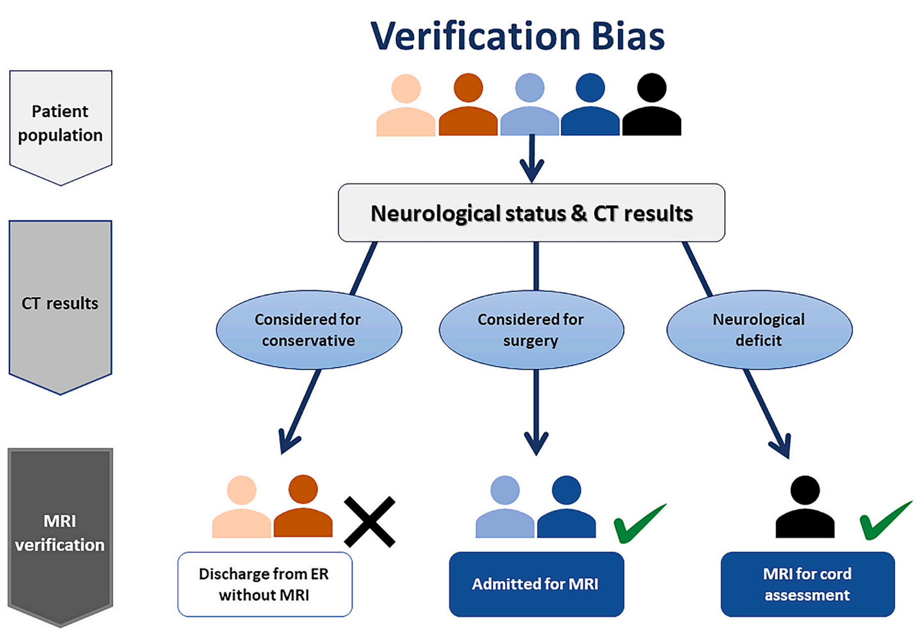

Fig. 5. An illustration of the underlying reason for verification bias
Verification bias happens when only a portion of TLFs patients who did CT will be verified by MRI depending on CT results. For instance, patients with more severe fractures on CT are more likely to undergo an MRI for PLC assessment than those with less severe injuries who receive conservative treatment based on CT findings. Therefore, patients with neurological deficits and surgical patients are overrepresented, potentially underestimating the impact of MRI on decision-making.

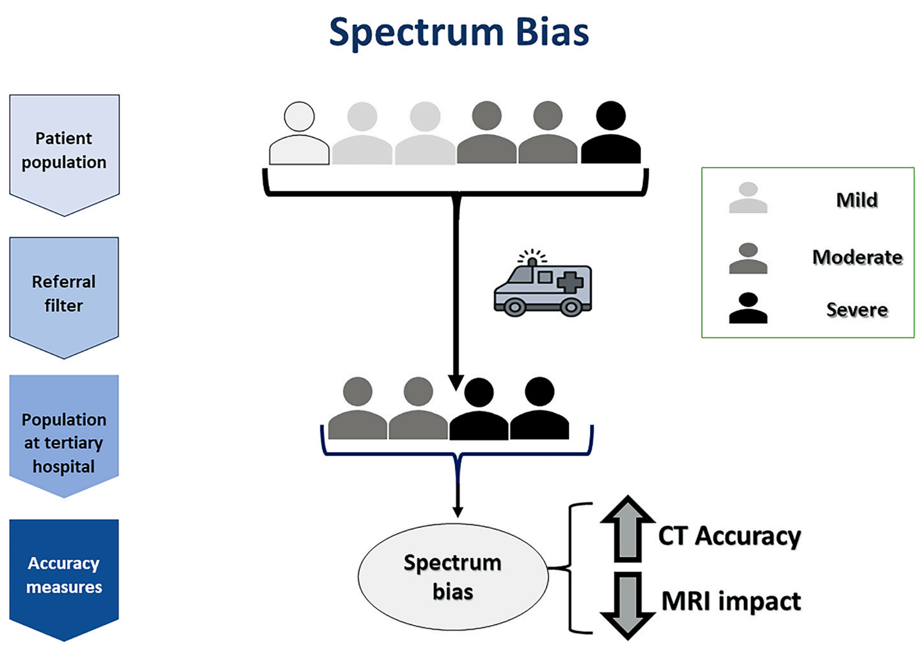

Fig. 6. An illustration of the underlying reason for spectrum bias
Spectrum bias occurs when the study's sample is restricted to more severe cases due to referral filters to tertiary trauma centers or a higher spectrum of injury in some geographical regions. Spectrum bias may inflate the accuracy measures and reduces the generalizability of findings to different settings.

4.4. The importance of standardized image interpretation

When evaluating the influence of MRI on TLF management, the accuracy and reliability of image interpretation are critical determinants of the findings' generalizability (Pavlou et al., 2021), (Aly et al., 2023a). Nonetheless, the interpretation of images may differ based on many factors, including imaging protocol, observers' experience, pretraining, and the validity of criteria for PLC injury. Attempts should be made to standardize all imaging variables while not limiting the results' application to other contexts. The value of high-quality consensus training cannot be overstated (Pavlou et al., 2021). Diagnostic accuracy studies are often criticized for overestimating the results, as they are conducted by experts in a research context and may not apply to real-world practice. (Bossuyt et al., 2015) In light of this, a substantial number of reviewers with diverse backgrounds must be chosen to ensure that the findings are applicable (Pavlou et al., 2021).

4.5. Definition of PLC injury in CT and MRI

The variability in the definition of PLC injury on MRI may lead to an inaccurate estimate of CT's accuracy for PLCinjury, the so-called misclassification bias (Van Middendorp et al., 2013), (Aly et al., 2021a) Initially, PLC injury was defined as any HSI on T2 MRI, which later proved to overread PLC injury during surgery(Fig. 3). (Pizones et al., 2012b), (Rihn et al., 2010) Instead, ''black stripe discontinuity,'' due to SSL or LF disruption, yielded higher specificity in identifying PLC injury during surgery (Pizones et al., 2012b). The biomechanical evidence showing that SSL and LF are the key structures for PLC incompetence further supports using the black stripe discontinuity criterion (Wu et al., 2018). However, the use of HSI is still popular today and remains one of the most considerable biases in evaluating MRI's impact on TLFs. A possible reason for using the HSI criterion is that being more prevalent would increase the study's sample size. (Aly et al., 2023a) Additionally, identifying HSI may be less demanding than black stripe discontinuity, which requires a thorough assessment of each PLC component for complete vs. partial injury (Aly et al., 2021a). Aly et al. showed that using HSI instead of black stripe discontinuity would reduce the specificity of CT in detecting PLC injury from 91% to 66%, thereby illustrating the magnitude of misclassification bias (Aly et al., 2021a). It might be argued that PLC injury can occur as a continuum with increasing instability as the number of damaged PLC components increases. (Aly et al., 2022a) However, from a practical point of view, we need MRI dichotomous criteria for PLC injury to allow comparison between CT and MRI accuracy. In that context, black stripe discontinuity is most valid criterion based on its specificity for detecting PLC injury intraoperatively and the supporting biomechanical evidence. (Pizones et al., 2012b), (Wu et al., 2018)

Regarding CT, PLC injury has been identified based on one or more CT indicators of posterior element injury, such as spinous process fracture, interspinous widening, laminar fracture, and facet diastasis (Fig. 4). (Winklhofer et al., 2013), (Pizones et al., 2011) This criterion has shown only moderate accuracy (50–70%) in detecting PLC Injury compared to MRI. A possible explanation is that this criterion assumes that fractures with a single positive finding and those with four positive findings would have the same positive predictive value (PPV) for PLC injury (Aly et al., 2021a). Recently, two studies have proposed CT criteria for PLC injury based on the number of positive findings. (Khurana et al., 2018), (Aly et al., 2021a) At least two positive findings yielded high enough PPV (88–91%) for PLC injury on MRI, warranting its use as a criterion for PLC injury. A single CT finding was insufficient to rule out or rule in PLC injury and was considered an indeterminate PLC status (M1 modifier) (Aly et al., 2021a). One study showed the use of. CT findings increased the accuracy in detecting PLC injury from 66% to 91%. We concur that certain spine surgeons may employ standing X-rays to reveal ambiguous PLC injuries. However, the utility of standing X-ray in our standardized protocol is limited by the lack of a cut-off point for interspinous widening or kyphosis that denotes PLC damage (Aly et al., 2023a). X-ray utility is further restricted to patients who are in excruciating pain. X-rays would probably be more helpful in detecting delayed kyphosis for patients treated conservatively (De Gendt et al., 2021).

4.6. Impact of MRI on the management of thoracolumbar fractures

4.6.1. Diagnostic versus therapeutic impact of MRI on thoracolumbar fractures

The influence of MRI on the management of TLFs can be determined by its impact on the diagnosis of PLC injury (diagnostic impact) or treatment decisions (therapeutic impact). (Aly et al., 2024).. In TLFs with intact neurology, the decision-making process relies on two variables: the level of bone instability and the severity of the PLC injury. (Aly et al., 2022b) Changing the PLC injury status via MRI does not necessarily alter the treatment decision, as with unstable burst fractures. Therefore, MRI's therapeutic impact may be more clinically meaningful than its diagnostic impact. However, the assessment of therapeutic impact is limited by its reliance on the contentious decision-making process for burst fractures.(Aly et al., 2024) In an international survey among spine surgeons, 46% of European surgeons recommended surgical treatment for A4 fractures vs. 0% of North American surgeons (Vaccaro et al., 2016). Arguably, PLC damage and MRI are less critical for decision-making as the surgical burst fracture treatment threshold decreases. Simply put, MRI indications for PLC assessment may have been affected by regional disparities in the patterns for managing burst fractures.Vaccaro et al., 2016 We believe each outcome has virtues and demerits and should be reported.

4.6.2. Therapeutic impact of MRI on thoracolumbar fractures: Various metrics

Prior research has used different metrics for treatment decisions, including prospective surgeon's decisions, TLICS, or TLAOSIS recommendations (Aly et al., 2024). TLICS/TLAOSIS offers three-tiered recommendations, comprising surgical, conservative, and grey zone. (Lee et al., 2005)(Vaccaro et al., 2013) A change in treatment decision may be defined as the transition from the "Conservative/Grey zone" to the "Surgery" category or any transition between the three categories. (Pizones et al., 2011) (Winklhofer et al., 2013). We lean towards the latter definition since it takes into account the transition from the ''Grey Zone "to ''Conservative'. Burst fractures with indeterminate PLC injury fall under the TLICS ″grey zone'' and may receive unnecessary surgery. MRI ruling out PLC injury will lead to a transition from Grey Zone to conservative, which should be counted in MRI's impact. It should be emphasized that the rate of change in treatment decisions using this 3-tier recommendation will be higher than dichotomous prospective decisions or even changes in PLC status due to the presence of an additional category grey Zone. (Fig. 7). (Aly et al., 2024) TLAOSIS gives A4 fractures five points and A3 fractures three points, and indeterminate PLC one point instead of two in TLICS. Thus, TLICS and TLAOSIS advocate distinct treatments for A4M0 (Conservative and Grey Zone) and A4M1 (Grey Zone and surgery, respectively). This may explain why the rate of change in TLAOSIS and TLICS recommendations will differ. (Vaccaro et al., 2016) (Lee et al., 2005)

4.6.3. Impact of MRI on treatment outcomes

The proposed methodology assumes that PLC injury on MRI has a prognostic value based on AOSpine classification and TLICS. (Lee et al., 2005), (Vaccaro et al., 2013), (Vaccaro et al., 2016) No clinical studies have documented radiographic, neurological, or patient-reported outcomes of untreated PLC injuries. (Van Middendorp et al., 2013), (Alanay et al., 2004) It might be argued that the impact of MRI on managing TLFs should not rely solely on identifying PLC injury, but on the impact of identified PLC injury on operative or non-operative outcomes. However, examining the predictive value of MRI would be illusive as it varies according to the extent of PLC injuries regarding the number of PLC structures damaged, the extent of anterior column or disc injury, and affected spine region (Alanay et al., 2004) Furthermore, studying the natural history of PLC injuries will be challenging because most PLC injuries are treated surgically according to the current treatment algorithm by AOSpine and TLICS. (Lee et al., 2005)

4.7. Can we predict when MRI would significantly impact the decision-making of thoracolumbar fractures?

It is critical to determine the influence of MRI on TLF decision-making. It is also crucial to forecast when MRI will influence decision-making to establish cost-effective MRI indications that balance the added benefits with additional cost, effort, and treatment's delay (Aly et al., 2024). The impact of MRI on decision-making may vary in different spine regions due to their distinctive anatomy and

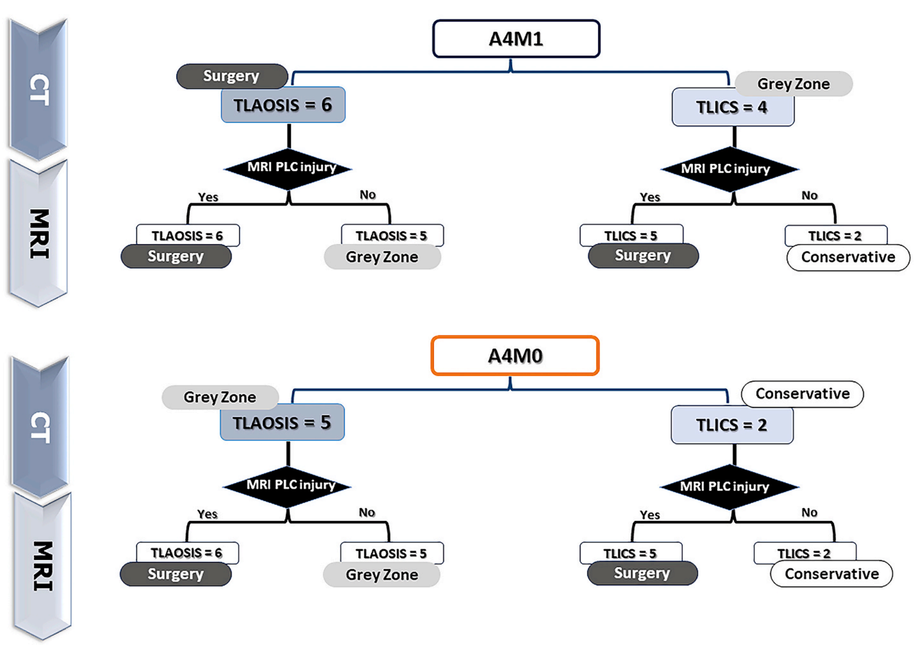

Fig. 7. Impact of MRI on treatment recommendation of Thoracolumbar injury score (TLICS) Vs. Thoracolumbar AOSpine injury severity score (TLAOSIS) for A4M0 & A4M1 fractures
Abbreviations, M1, M1 modifer for indetrminate PLC status, **M0,** No M1 modifier.

biomechanics. (Aly et al., 2024), (Aly et al., 2022b) For instance, MRI has shown much less impact on decision-making for low lumbar fractures (L3-L5) than thoracic fractures (T1-T10). (Aly et al., 2022c), (AlRaddadi et al., 2024) One study found that M1 modifier; as defined by a single CT finding, may predict most upgrades from A to B-type injuries , indicating the need for further MRI for PLC assessment (Aly et al., 2022b).A3M1 exhibited a 20% upgrade rate to B-injuries compared to 0% for A3M0. Therfore, the combining the A fracture subtype with M1 modifier may guide MRI indications in individual patients (Aly et al., 2022b). The TLAOSIS ″Gey Zone'' predicted a higher chance of MRI upgrading from Type A to Type B than the conservative or surgical category. (Aly et al., 2022b)

5. Conclusion

There is insufficient evidence that MRI could substantially impact the decision-making in neurologically intact patients with thoracolumbar fractures. The scarcity of research could be attributable to methodological difficulties in assessing MRI's impact on thoracolumbar fractures. The methodological issues may be related to intrinsic bias in image-based studies and those linked to TLFs, such as the absence of consensus on criteria for PLC injury or metrics for MRI's impact. We proposed the first structured methodology for examining the impact of MRI on neurologically intact patients with TLFs. The proposed method focuses on appropriate patient selection, standardized CT/MRI image analysis, and using clinically meaningful outcome measures.

Declarations authors'

Contribution All authors contributed to the study conception and design. Mohamed Aly, Sebastian Bigdon, Ulrich Speigel, Gaston Camino, Saleh Baeesa, and Klaus Schnake performed material preparation, data collection, and analysis. The first author wrote the first draft of the manuscript; all authors commented on previous versions. All authors read and approved the final manuscript.

Ethics approval

Not applicable.

Availability of data and material (data transparency)

'Not applicable'.

Code availability (software application or custom code)

'Not applicable'.

Consent to publication and to participate

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We acknowledge Dr Klara Brunnhuber, Guideline Project Lead, ELSEVIER, for her help editing this manuscript.

References

Ahern, Daniel P., McDonnell, Jake M., Riffault, Mathieu, Evans, Shane, Wagner, Scott C., Vaccaro, Alexander R., Hoey, David A., Butler, Joseph S., 2021. A meta-analysis of the diagnostic accuracy of Hounsfield units on computed topography relative to dual-energy X-ray absorptiometry for the diagnosis of osteoporosis in the spine surgery population. The Spine Journal 21 (10), 1738–1749. ISSN 1529-9430. https://doi.org/10.1016/j.spinee.2021.03.008.

Alanay, A., Yazici, M., Acaroglu, E., et al., 2004. Course of nonsurgical management of burst fractures with intact posterior ligamentous complex: an MRI study. Spine 29, 2425–2431. https://doi.org/10.1097/01.brs.0000143169.80182.ac.

AlRaddadi, K.K., Abdulbaset, M., Al-Shoaibi, A.A., W, A., et al., 2024. Traumatic thoracic spine fracture: can we predict when MRI would modify the fracture classification or decision-making compared to CT alone? Eur. Spine J. https://doi.org/10.1007/s00586-024-08196-8.

Aly, M.M., Al-Shoaibi, A.M., Alzahrani, A.J., Al Fattani, A., 2021a. Analysis of the combined computed tomography findings improves the accuracy of computed tomography for detecting posterior ligamentous complex injury of the thoracolumbar spine as defined by magnetic resonance imaging. World Neurosurg 151. https://doi.org/10.1016/j.wneu.2021.04.106.

Aly, M.M., Al-Shoaibi, A.M., Al Fattani, A., AlJuzair, A.H., 2021b. Diagnostic value of various morphological features of horizontal and vertical laminar fractures for posterior ligamentous complex injury of the thoracolumbar spine as defined by magnetic resonance imaging. World Neurosurg 153, e290–e299. https://doi.org/10.1016/J.WNEU.2021.06.109.

Aly, M.M., Elemam, R.A., El-Sharkawi, M., Hurlbert, R.J., 2022a. Injury of the thoracolumbar posterior ligamentous complex: a bibliometric literature review. World Neurosurg 161, 21–33. https://doi.org/10.1016/j.wneu.2022.01.041.

Aly, M.M., Al-Shoaibi, A.M., Abduraba, Ali S., et al., 2022b. How often would MRI change the thoracolumbar fracture classification or decision-making compared to CT alone? Glob spine J, 21925682221089580. https://doi.org/10.1177/21925682221089579.

Aly, M.M., Al-Shoaibi, A.M., Abduraba, S., et al., 2022c. Traumatic low lumbar fractures: how often MRI changes the fracture classification or clinical decision-making compared to CT alone? Eur. Spine J. 31, 37–45. https://doi.org/10.1007/s00586-021-06987-x.

Aly, M.M., Al-Shoaibi, A.M., Aljuzair, A.H., et al., 2023a. A proposal for a standardized imaging algorithm to improve the accuracy and reliability for the diagnosis of thoracolumbar posterior ligamentous complex injury in computed tomography and magnetic resonance imaging. Glob spine J 13, 873–896. https://doi.org/10.1177/21925682221129220.

Aly, M.M., Al-Shoaibi, A.M., Abduraba, Ali S., et al., 2023b. Which morphological features of facet diastasis predict thoracolumbar posterior ligamentous complex injury as defined by magnetic resonance imaging? World Neurosurg 171, e276–e285. https://doi.org/10.1016/j.wneu.2022.12.010.

Aly, M.M., Soliman, Y., Elemam, R.A., Pizones, J., Alzahrani, A.E.S., 2024. How frequently MRI modifies thoracolumbar fractures' classification or decision-making? A systematic review and meta-analysis. Eur. Spine J. https://doi.org/10.1007/s00586-023-08087-4.

Benedetti, P.F., Fahr, L.M., Kuhns, L.R., Hayman, L.A., 2000. MR imaging findings in spinal ligamentous injury. Am. J. Roentgenol. 175, 661–665. https://doi.org/10.2214/ajr.175.3.1750661.

Bigdon, S.F., Saldarriaga, Y., Oswald, K.A.C., et al., 2022. Epidemiologic analysis of 8000 acute vertebral fractures : evolution of treatment and complications at 10 - year follow - up. J. Orthop. Surg. Res. 9, 1–12. https://doi.org/10.1186/s13018-022-03147-9.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., et al., 2015. Stard 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 351, 1–9. https://doi.org/10.1136/bmj.h5527.

Crosby, C.G., Even, J.L., Song, Y., et al., 2011. Diagnostic abilities of magnetic resonance imaging in traumatic injury to the posterior ligamentous complex: the effect of years in training. Spine J. 11, 747–753. https://doi.org/10.1016/j.spinee.2011.07.005.

De Gendt, E.E.A., Vercoulen, T.F.G., Joaquim, A.F., et al., 2021. The current status of spinal posttraumatic deformity: a systematic review. Glob spine J 11, 1266–1280. https://doi.org/10.1177/2192568220969153.

Denis, F., 1983. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. Spine 8, 817–831. https://doi.org/10.1097/00007632-198311000-00003.

Durmaz, A., İlter, M.H., Tuzlali, H., 2021. Is magnetic resonance imaging needed for decision making diagnosis and treatment of thoracic and lumbar vertebral fractures? Eur. J. Orthop. Surg. Traumatol. https://doi.org/10.1007/S00590-021-03165-Z.

Elfil, M., Negida, A., 2017. Sampling Methods in Clinical Research ; an Educational Review, vol. 5, pp. 3–5.

Holdsworth F. Fractures, dislocations and fracture-dislocations of the spine. JBJSA 1970 D-51. P 5483077. (1970) Fractures, dislocations, and fracture-dislocations of the spine. J Bone Jt Surg Am 52:.

Kanezaki, S., Miyazaki, M., Ishihara, T., et al., 2018. Magnetic resonance imaging evaluation of intervertebral disc injuries can predict kyphotic deformity after posterior fixation of unstable thoracolumbar spine injuries. Medicine (Baltim.) 97. https://doi.org/10.1097/MD.0000000000011442.

Khurana, B., Prevedello, L.M., Bono, C.M., et al., 2018. CT for thoracic and lumbar spine fractures: can CT findings accurately predict posterior ligament complex injury? Eur. Spine J. 27, 3007–3015. https://doi.org/10.1007/s00586-018-5712-z.

Khurana, B., Karim, S.M., Zampini, J.M., et al., 2019. Is focused magnetic resonance imaging adequate for treatment decision making in acute traumatic thoracic and

lumbar spine fractures seen on whole spine computed tomography? Spine J. 19, 403–410. https://doi.org/10.1016/j.spinee.2018.08.010.
Lee, J.Y., Vaccaro, A.R., Lim, M.R., et al., 2005. Thoracolumbar injury classification and severity score: a new paradigm for the treatment of thoracolumbar spine trauma. In: Journal of Orthopaedic Science. Springer, Japan, pp. 671–675.
Lee, J.Y., Vaccaro, A.R., Schweitzer, K.M., et al., 2007. Assessment of injury to the thoracolumbar posterior ligamentous complex in the setting of normal-appearing plain radiography. Spine J. 7, 422–427. https://doi.org/10.1016/j.spinee.2006.07.014.
Leferink, V., Veldhuis, E., Zimmerman, K., et al., 2002. Classificational problems in ligamentary distraction type vertebral fractures: 30% of all B-type fractures are initially unrecognised. Eur. Spine J. 11, 246–250. https://doi.org/10.1007/s00586-001-0366-6.
Magerl, F., Aebi, M., Gertzbein, S.D., et al., 1994. A comprehensive classification of thoracic and lumbar injuries. Eur. Spine J. 3, 184–201. https://doi.org/10.1007/BF02221591.
Obuchowski, N.A., Bullen, J., 2022. Multireader Diagnostic Accuracy Imaging Studies : Fundamentals of Design and Analysis.
Patino, C.M., Ferreira, J.C., 2018. Internal and External Validity : Can You Apply Research Study Results to Your Patients ?, vol. 44, p. 6548.
Pavlou, A., Kurtz, R.M., Song, J.W., et al., 2021. Diagnostic Accuracy Studies in Radiology : How to Recognize and Address Potential Sources of Bias, vol. 2021.
Petersilge A, Masaryk J, Emery E Evaluation burst fractures : Radiology 49–54.
Pizones, J., Izquierdo, E., Álvarez, P., et al., 2011. Impact of magnetic resonance imaging on decision making for thoracolumbar traumatic fracture diagnosis and treatment. Eur. Spine J. 20 (Suppl. 3), 390–396. https://doi.org/10.1007/s00586-011-1913-4.
Pizones, J., Izquierdo, E., Sánchez-Mariscal, F., et al., 2012a. Sequential damage assessment of the different components of the posterior ligamentous complex after magnetic resonance imaging interpretation: prospective study 74 traumatic fractures. Spine 37. https://doi.org/10.1097/BRS.0b013e3182422b2b.
Pizones, J., Zúniga, L., Sánchez-Mariscal, F., et al., 2012b. MRI study of post-traumatic incompetence of posterior ligamentous complex: importance of the supraspinous ligament. Prospective study of 74 traumatic fractures. Eur. Spine J. 21, 2222–2231. https://doi.org/10.1007/s00586-012-2403-z.
Qureshi, S., Dhall, S.S., Anderson, P.A., et al., 2019. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: radiological evaluation. Clin. Neurosurg. 84, E28–E31. https://doi.org/10.1093/neuros/nyy373.
Rajasekaran, S., Vaccaro, A.R., Kanna, R.M., et al., 2017. The value of CT and MRI in the classification and surgical decision-making among spine surgeons in thoracolumbar spinal injuries. Eur. Spine J. 26, 1463–1469. https://doi.org/10.1007/s00586-016-4623-0.
Rihn, J.A., Yang, N., Fisher, C., et al., 2010. Using magnetic resonance imaging to accurately assess injury to the posterior ligamentous complex of the spine: a prospective comparison of the surgeon and radiologist. J. Neurosurg. Spine 12, 391–396. https://doi.org/10.3171/2009.10.SPINE08742.
Rosenthal, B.D., Boody, B.S., Jenkins, T.J., et al., 2018. Thoracolumbar burst fractures. Clin Spine Surg 31, 143–151. https://doi.org/10.1097/BSD.0000000000000634.
Rutjes, A.W.S., Reitsma, J.B., Di, Nisio M., et al., 2006. R Esearch Evidence of Bias and Variation in Diagnostic Accuracy Studies, pp. 1–12. https://doi.org/10.1503/cmaj.050090.
Schmidt, R.L., Factor, R.E., 2013. Understanding sources of bias in diagnostic accuracy studies. Arch. Pathol. Lab Med. 137, 558–565. https://doi.org/10.5858/arpa.2012-0198-RA.
Schnake, K.J., Blattert, T.R., Hahn, P., et al., 2018. Classification of osteoporotic thoracolumbar spine fractures: recommendations of the spine section of the German society for orthopaedics and trauma (DGOU). Global Spine J. 8, 46S–49S. https://doi.org/10.1177/2192568217717972.
Sixta, S., Moore, F.O., Ditillo, M.F., et al., 2012. Screening for thoracolumbar spinal injuries in blunt trauma: an eastern association for the surgery of trauma practice management guideline. J. Trauma Acute Care Surg. 73 https://doi.org/10.1097/TA.0b013e31827559b8.
Tins, B., 2010. Technical aspects of CT imaging of the spine. Insights Imaging 1, 349–359. https://doi.org/10.1007/s13244-010-0047-2.
Vaccaro, A.R., Oner, C., Kepler, C.K., et al., 2013. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. Spine 38, 2028–2037. https://doi.org/10.1097/BRS.0b013e3182a8a381.
Vaccaro, A.R., Schroeder, G.D., Kepler, C.K., et al., 2016. The surgical algorithm for the AOSpine thoracolumbar spine injury classification system. Eur. Spine J. 25, 1087–1094. https://doi.org/10.1007/s00586-015-3982-2.
Van Middendorp, J.J., Patel, A.A., Schuetz, M., Joaquim, A.F., 2013. The precision, accuracy and validity of detecting posterior ligamentous complex injuries of the thoracic and lumbar spine: a critical appraisal of the literature. Eur. Spine J. 22, 461–474. https://doi.org/10.1007/s00586-012-2602-7.
Whiting, P.F., Rutjes, A.W.S.S., Westwood, M.E., et al., 2011. QUADAS-2: arevised tool for quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155, 529–536.
Winklhofer, S., Thekkumthala-Sommer, M., Schmidt, D., et al., 2013. Magnetic resonance imaging frequently changes classification of acute traumatic thoracolumbar spine injuries. Skeletal Radiol. 42, 779–786. https://doi.org/10.1007/s00256-012-1551-x.
Wood, K.B., Li, W., Lebl, D.S., Ploumis, A., 2014. Management of thoracolumbar spine fractures. Spine J. 14, 145–164.
Wu, C.C., Jin, H.M., Yan, Y.Z., et al., 2018. Biomechanical role of the thoracolumbar ligaments of the posterior ligamentous complex: a finite element study. World Neurosurg 112, e125–e133. https://doi.org/10.1016/j.wneu.2017.12.171.