



**CONGRESS OF NEUROLOGICAL SURGEONS SYSTEMATIC REVIEW AND
EVIDENCE-BASED GUIDELINES ON THE EVALUATION AND TREATMENT OF
PATIENTS WITH THORACOLUMBAR SPINE TRAUMA: CLASSIFICATION OF
INJURY**

Sponsored by: Congress of Neurological Surgeons and the Section on Disorders of the Spine and Peripheral Nerves in collaboration with the Section on Neurotrauma and Critical Care

Endorsed by: Joint Guidelines Committee of the American Association of Neurological Surgeons (AANS) and the Congress of Neurological Surgeons (CNS)

Andrew T. Dailey, MD,¹ Paul M. Arnold, MD,² Paul A. Anderson, MD,³ John H. Chi, MD, MPH,⁴ Sanjay S. Dhall, MD,⁵ Kurt M. Eichholz, MD,⁶ James S. Harrop, MD,⁷ Daniel J. Hoh, MD,⁸ Sheeraz Qureshi, MD, MBA,⁹ Craig H. Rabb, MD,¹⁰ P. B. Raksin, MD,¹¹ Michael G. Kaiser, MD¹² and John E. O'Toole, MD, MS¹³

1. Department of Neurosurgery, University of Utah, Salt Lake City, Utah
2. Department of Neurosurgery, University of Kansas School of Medicine, Kansas City, Kansas
3. Department of Orthopedics and Rehabilitation, University of Wisconsin, Madison, Wisconsin
4. Department of Neurosurgery, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts
5. Department of Neurological Surgery, University of California, San Francisco, San Francisco, California

6. St. Louis Minimally Invasive Spine Center, St. Louis, Missouri
7. Departments of Neurological Surgery and Orthopedic Surgery, Thomas Jefferson University, Philadelphia, Pennsylvania
8. Lillian S. Wells Department of Neurological Surgery, University of Florida, Gainesville, Florida
9. Department of Orthopaedic Surgery, Weill Cornell Medical College, New York, New York
10. Department of Neurosurgery, University of Utah, Salt Lake City, Utah
11. Division of Neurosurgery, John H. Stroger, Jr. Hospital of Cook County and Department of Neurological Surgery, Rush University Medical Center, Chicago, Illinois
12. Department of Neurosurgery, Columbia University, New York, New York
13. Department of Neurological Surgery, Rush University Medical Center, Chicago, Illinois

Correspondence:

Andrew T. Dailey, MD

Department of Neurosurgery, University of Utah

Professor of Neurosurgery and Orthopedics

175 North Medical Drive East

Salt Lake City, UT 84132

Email: andrew.dailey@hsc.utah.edu

Keywords: Classification, fracture, thoracolumbar, vertebrae

Abbreviations

AO - Arbeitsgemeinschaft für Osteosynthesefragen (Association for the Study of Internal Fixation)

CT - computerized tomography

LSC - Load Sharing Classification

MRI - magnetic resonance imaging

PLC - posterior ligamentous complex

STSG - Spine Trauma Study Group

TLICS - Thoracolumbar Injury Classification and Severity Scale

TLISS - Thoracolumbar Injury Severity Score

No part of this article has been published or submitted for publication elsewhere.

ABSTRACT

Background: Classification systems should enhance communication between clinicians with varying degrees of experience about the severity of an injury or disease process, reliably guide treatment, and predict the outcome of various treatment options. Many classification systems have been developed, but no single classification system has been universally accepted, as early attempts were prone to pattern recognition of fracture types, and therefore, the interobserver reliability was low.

Objective: The authors tried to determine 1) whether there are classification systems for fractures of the thoracolumbar spine that have been shown to be valid and reliable, and 2) when treating patients, whether employing a particular classification system affects clinical outcomes.

Methods: The literature search yielded 932 abstracts, of which the task force selected 52 articles for full-text review. Of these, 32 were rejected for not meeting inclusion criteria or for being off topic. Twenty studies were selected for inclusion in this systematic review.

Results: There are at least 12 different classification systems that have been used over the years. Early attempts, such as the Denis and AO Comprehensive Classification systems, were often

developed based on the experience of a single surgeon or a small group and were prone to pattern recognition of fracture types with low interobserver reliability. More recently developed systems, including the Thoracolumbar Injury Classification and Severity Scale (TLICS) or the AO Spine Thoracolumbar Spine Injury Classification System, focus not only on description of the fracture but also focus on prognosis and treatment, and these systems generally have higher inter- and intraobserver reliability.

Conclusion: The authors recommend using a thoracolumbar trauma classification scheme that uses readily available clinical data, such as the TLICS/TLISS or the AO Spine Thoracolumbar Spine Injury Classification System. However, there is insufficient evidence to recommend a universal classification system that can guide treatment and affect outcomes of these injuries.

RECOMMENDATIONS

Question 1

Are there classification systems for fractures of the thoracolumbar spine that have been shown to be internally valid and reliable (i.e., do these instruments provide consistent information between different care providers)?

Recommendation 1

A classification scheme that uses readily available clinical data (e.g., computed tomography scans with or without magnetic resonance imaging) to convey injury morphology, such as Thoracolumbar Injury Classification and Severity Scale or the AO Spine Thoracolumbar Spine Injury Classification System, should be used to improve characterization of traumatic thoracolumbar injuries and communication among treating physicians.

Strength of Recommendation: Grade B

Question 2

In treating patients with thoracolumbar fractures, does employing a formally tested classification system for treatment decision-making affect clinical outcomes?

Recommendation 2

There is insufficient evidence to recommend a universal classification system or severity score that will readily guide treatment of all injury types and thereby affect outcomes.

Strength of Recommendation: Grade Insufficient

INTRODUCTION

Goals and Rationale

This clinical guideline was created to improve patient care by outlining the appropriate information-gathering and decision-making processes involved in the evaluation and treatment of patients with thoracolumbar spine trauma. The surgical management of these patients often takes place under a variety of circumstances and by various clinicians. This guideline was created as an educational tool to guide qualified physicians through a series of diagnostic and treatment decisions in an effort to improve the quality and efficiency of care.

Classification systems are designed to achieve several goals and are prevalent throughout all medical fields, particularly the surgical specialties. Classification systems should enhance communication among clinicians with varying degrees of experience about the severity of an injury or disease process, reliably guide treatment, and predict the outcome of various treatment options. Although classification systems should be inclusive, it is more important for them to be reliable, reproducible, and practical in the clinical setting, and based on interpretation of clinical

and radiographic data. Finally, classification systems should suggest to the clinician a prognosis regarding the injury pattern observed.¹⁻⁵

Clinicians treating patients with thoracolumbar trauma need a classification system that can predict whether a fracture pattern is stable or unstable, and if unstable, whether the injury warrants surgical intervention to prevent late deformity or neurological deficit.^{2,6} Accumulating information of a particular injury type within a classification system would allow a clinician researcher to then study treatment options for that particular injury. In this way, classification systems are critical to research.

The initial attempt at classification dates to 1929 when Bohler used morphological description in conjunction with a presumed mechanism to identify 5 fracture types, including compression fractures, flexion-distraction injuries, extension fractures, shear fractures, and rotational or torsional injuries.⁷ Watson-Jones⁸ further categorized fracture types with a morphologic classification system that recognized the importance of the posterior ligaments. Seven fracture types with 3 major patterns were introduced, including compression fractures, comminuted fractures, and fracture dislocation.⁸ Watson-Jones⁸ recognized the concept of spinal instability and the importance of the posterior ligaments in maintaining stability. This system was followed by Nicoll's description⁹ of injury patterns in 166 fractures in coal miners. All patients had similar mechanism of injury, buried in a mine with a "hyperflexion" injury.⁹ Four injury patterns were recognized: anterior wedge fractures, lateral wedge, fracture-dislocation, and isolated fractures of the neural arch. The concept of stability was introduced with the exception of fracture-dislocation being termed stable, and injury to the posterior ligamentous structure was deemed

important in determining stability. When treating potentially unstable injuries, Nicoll⁹ remarked on the importance of an anterior fusion, thereby restoring the load-bearing column of the spine. None of these abovementioned classification schemes underwent validation or even wider description in the literature beyond the original articles.

Holdsworth¹⁰ later developed a 2-column classification system. The anterior column consisted of the vertebral body and the disc (everything anterior to the posterior longitudinal ligament) and the posterior column, including the facet joints, the neural arch, and the posterior ligaments (intraspinous, supraspinous, and ligamentum flavum). Fracture types included compression fractures, fracture-dislocations, rotational injuries, extension injuries, shear injuries, and the new concept burst fractures. Injuries that involved just the anterior column were inherently stable injuries, while those that also caused disruption of the posterior ligaments were unstable. Kelly and Whitesides¹¹ generally agreed with the 2-column theory and described 11 cases. Stable fractures were anterior and lateral wedge fractures, and the stable burst. Those considered unstable were flexion dislocation, flexion rotation (slice), and the unstable burst with instability indicated if there were damage to both anterior and posterior columns.¹¹ Although a step forward in classification systems, neither system was ever validated by any group outside of the original authors.

Since Bohler's initial attempt to classify patterns of thoracolumbar fractures according to radiographic appearance and proposed mechanism of injury, many classification systems have been developed.⁵⁻¹⁸ However, no single classification system has been universally accepted, because early attempts were prone to pattern recognition of fracture types, and therefore

interobserver reliability was low. In addition, these early classification systems were often developed based on the experience of a single surgeon or a small group, and this often further limited the reproducibility and reliability. More recent attempts have focused not only on description of the fracture but also on prognosis and treatment. These systems have attempted to provide an injury severity score to help guide the clinician determine an acceptable treatment plan.^{5,18} Although these classification systems have been shown to have reproducibility within the group developing the system (internal reliability) and by groups outside the original developers (external reliability), they have not been universally accepted. As a result, new classification schemes continue to be developed and published, and attempts are made to validate these novel schemes.^{6, 17} The authors addressed the questions regarding whether the currently available classification systems for thoracolumbar spine injuries 1) are valid and reliable, and 2) affect clinical outcomes.

METHODS

The guidelines task force initiated a systematic review of the literature relevant to the diagnosis and treatment of patients with thoracolumbar trauma. Through objective evaluation of the evidence and transparency in the process of making recommendations, this evidence-based clinical practice guideline was developed for the diagnosis and treatment of adult patients with thoracolumbar injury. These guidelines are developed for educational purposes to assist practitioners in their clinical decision-making processes. Additional information about the methods used in this systematic review is provided in the [introduction and methodology chapter](#).

Literature Search

The task force members identified search terms/parameters, and a medical librarian implemented the literature search, consistent with the literature search protocol (see Appendix I), using the National Library of Medicine PubMed database and the Cochrane Library (which included the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effect, the Cochrane Central Register of Controlled Trials, the Health Technology Assessment Database, and the National Health Service Economic Evaluation Database) for the period from January 1, 1946, to March 31, 2015, using the search strategies provided in Appendix I.

RESULTS

The literature search yielded 932 abstracts. Task force members reviewed all abstracts yielded from the literature search and identified the literature for full-text review and extraction, addressing the clinical questions, in accordance with the literature search protocol (Appendix I). Task force members identified the best research evidence available to answer the targeted clinical questions. When level I, II, or III literature was available to answer specific questions, the task force did not review level IV studies.

The task force selected 52 articles for full-text review. Of these, 32 were rejected for not meeting inclusion criteria or for being off topic. Twenty articles were selected for inclusion in this systematic review (Appendix II).

Inclusion/Exclusion Criteria

Articles were retrieved and included only if they met specific inclusion/exclusion criteria. These criteria were also applied to articles provided by guideline task force members who

supplemented the electronic database searches with articles from their own files. To reduce bias, these criteria were specified before conducting the literature searches.

Articles that do not meet the following criteria were, for the purposes of this evidence-based clinical practice guideline, excluded. To be included as evidence in the guideline, an article had to be a report of a study that:

- Investigated patients with thoracolumbar injuries;
- Included patients ≥ 18 years of age;
- Enrolled $\geq 80\%$ of thoracolumbar injuries (studies with mixed patient populations were included if they reported results separately for each group/patient population);
- Was a full article report of a clinical study;
- Was not an internal medical records review, meeting abstract, historical article, editorial, letter, or commentary;
- Appeared in a peer-reviewed publication or a registry report;
- Enrolled ≥ 10 patients per arm per intervention (20 total) for each outcome;
- Included only human subjects;
- Was published in or after 1946 through March 31, 2015;
- Quantitatively presented results;
- Was not an in vitro study;
- Was not a biomechanical study;
- Was not performed on cadavers;
- Was published in English;

- Was not a systematic review, meta-analysis, or guideline developed by others* ;
- Was a case series (therapeutic study) where higher level evidence exists.

Rating Quality of Evidence

The guideline task force used a modified version of the North American Spine Society’s (NASS) evidence-based guideline development methodology. The NASS methodology uses standardized levels of evidence (Appendix III) and grades of recommendation (Appendix IV) to assist practitioners in easily understanding the strength of the evidence and recommendations within the guidelines. The levels of evidence range from level I (high quality randomized controlled trial) to level IV (case series). Grades of recommendation indicate the strength of the recommendations made in the guideline based on the quality of the literature. Levels of evidence have specific criteria and are assigned to studies prior to developing recommendations. Recommendations are then graded based upon the level of evidence. To better understand how levels of evidence inform the grades of recommendation and the standard nomenclature used within the recommendations, see Appendix IV.

Guideline recommendations were written using a standard language that indicates the strength of the recommendation. “A” recommendations indicate a test or intervention is “recommended”; “B” recommendations “suggest” a test or intervention; “C” recommendations indicate a test or intervention or “is an option.” “Insufficient”

* The guideline task force did not include systematic reviews, guidelines, or meta-analyses conducted by others. These documents are developed using different inclusion criteria than those specified in this guideline; therefore, they may include studies that do not meet the inclusion criteria specific in this guideline. In cases where these types of documents’ abstract suggested relevance to the guideline’s recommendations, the task force searched their bibliographies for additional studies.

statements clearly indicate that “there is insufficient evidence to make a recommendation for or against” a test or intervention. Task force consensus statements clearly state that “in the absence of reliable evidence, it is the task force’s opinion that” a test or intervention may be considered. Both the levels of evidence assigned to each study and the grades of each recommendation were arrived at by consensus of the workgroup employing up to three rounds of voting when necessary.

In evaluating studies as to levels of evidence for this guideline, the study design was interpreted as establishing only a potential level of evidence. For example, a therapeutic study designed as a randomized controlled trial would be considered a potential level I study. The study would then be further analyzed as to how well the study design was implemented and significant shortcomings in the execution of the study would be used to downgrade the levels of evidence for the study’s conclusions (see Appendix V for additional information and criteria).

Revision Plans

In accordance with the Institute of Medicine’s standards for developing clinical practice guidelines and criteria specified by the National Guideline Clearinghouse, the task force will monitor related publications following the release of this document and will revise the entire document and/or specific sections “if new evidence shows that a recommended intervention causes previously unknown substantial harm; that a new intervention is significantly superior to a previously recommended intervention from an efficacy or harms perspective; or that a recommendation can be applied to new populations.”¹⁹ In addition, the task force will confirm within 5 years from the date of publication that the

content reflects current clinical practice and the available technologies for the evaluation and treatment for patients with thoracolumbar trauma.

DISCUSSION

Question 1

Are there classification systems for fractures of the thoracolumbar spine that have been shown to be internally valid and reliable (ie, do these instruments provide consistent information between different care providers)?

Recommendation 1

A classification scheme that uses readily available clinical data (eg, computed tomography scan with or without magnetic resonance imaging) to convey injury morphology, such as Thoracolumbar Injury Classification and Severity Scale or the AO Spine Thoracolumbar Spine Injury Classification System, should be used to improve characterization of traumatic thoracolumbar injuries and communication among treating physicians.

Strength of Recommendation: Grade B

Question 2

In treating patients with thoracolumbar fractures, does employing a formally tested classification system for treatment decision-making affect clinical outcomes?

Recommendation 2

There is insufficient evidence to recommend a universal classification system or severity score that will readily guide treatment of all injury types and thereby affect outcomes.

Strength of Recommendation: Grade Insufficient

Before Denis' classification system, descriptions of thoracolumbar fractures were based on the appearance from plain x-ray and the presumed mechanism of injury recreated from the patient's description of the accident. With the advent of computed tomography (CT), advanced imaging could give a better anatomic image of a thoracolumbar injury and allow physicians to describe the injury in multiple planes with fine detail. As a result of better anatomic definition of the injury, Denis conceptually divided the spine into 3 columns, with the anterior column comprised of anterior longitudinal ligament, anterior annulus, and anterior vertebral body, the posterior column including the neural arch and the posterior ligamentous complex (PLC), and the middle column included the posterior portion of the vertebral body, the posterior longitudinal ligament, and the posterior annulus fibrosis.¹² Denis argued that the integrity of the middle column was the most important for stability, with disruption leading to potential neurological instability. Denis also described varying degrees of instability with mechanical instability that could lead to progressive kyphosis and pain, neurological instability, where the injury was severe enough to produce a neurological deficit, and most severely, a combination of both mechanical and neurological instability (third-degree instability).

In a series of 412 patients (53 of whom had a CT), Denis described four injury patterns: 1) compression fractures developed with failure of the anterior column in compression; 2) burst fractures developed when both the anterior and middle columns failed in compression and may or may not have led to neurological or mechanical instability; 3) seat belt injuries resulted from failure of the posterior and middle columns in distraction;

and 4) fracture dislocations (the most severe injuries) occurred as the result of failure of all three columns. These 4 types were then divided into 16 subtypes. The Denis classification provided level III evidence and became a popular scheme for the description of thoracolumbar fractures in trauma centers in North America. However, the system does not clearly identify injuries, which may or may not require operative intervention.

Often, clinicians thought that if 2 or more columns were involved then the patient needed surgical intervention.^{3,4} However, as McAfee¹⁵ quickly determined, there were burst fractures which were stable and could be treated nonoperatively, and those that were unstable and should be considered for surgical intervention.

Studies attempting to show the reliability of the Denis classification provide level II evidence that there is only moderate inter- and intraobserver reliability with this system. In a study reporting 31 consecutive cases of thoracolumbar trauma, Wood²⁰ found an interobserver reliability with a kappa value (κ) of 0.60 for the four types but only 0.17 for the subtypes. In addition, there was only an intraobserver reliability of 79% for type and 56% for subtype, leading the authors to question the utility of the classification system in wider studies of trauma populations. Oner²¹ showed that the use of CT or MRI yielded good reliability ($\kappa = 0.52-0.60$) at the most basic level of the type of fracture, but reduced to fair to moderate reliability ($\kappa = 0.39-0.45$) when trying to classify into one of the 16 subtypes.

As previously mentioned, McAfee used CT to better define patterns of instability. Based on observations from 100 consecutive CT scans, he simplified the 3-column theory, and based on the mode of failure of the middle column in axial compression, axial distraction, and translation, one could determine the pattern of injury, the severity of the deficit, the degree of instability, and the type of instrumentation need for correction.¹⁵ Six fracture types were identified, including wedge compression, stable burst, unstable burst, Chance, flexion distraction, and translation injuries. The mechanism of failure of the middle column helped determine whether the posterior elements were involved and if a burst fracture were stable or unstable. If there was disruption of the neural arch or facet joints, burst fractures were unstable. In addition, the mode of failure of the middle column could help define if compressive, distractive, or segmental instrumentation was needed.

McAfee's paper provides level III evidence, and while it describes a simpler classification than Denis', it was never tested for reliability or reproducibility.

In contradistinction to anatomic classifications such as Denis', Ferguson and Allen¹³ provided a mechanistic classification scheme by which the presumed mechanism of injury was deduced from the patterns of tissue failure as seen on CT. The authors thought the concept of dividing injuries into three anatomic columns did not take into account the biomechanical mechanism of failure of different anatomic regions and thus did not help to predict a treatment paradigm. The resulting classification strategy was an adaptation of the authors' popular classification for subaxial cervical trauma. Seven different mechanisms were described and included compressive flexion, distractive flexion, lateral flexion, translation, torsional flexion with torsion and compression of the anterior

elements and tension and torsion about the posterior elements, vertical compression, and the rare distractive extension injury. By understanding the mechanism of the fracture, the authors suggested that appropriate hardware and corrective forces could be applied to stabilize the spine and that no single mechanism of reduction could be applied to all injuries. For example, in vertical compressive injuries, distraction rods may be used to lengthen the shortened segments, but these devices should be avoided in a distractive flexion or torsional flexion injury. Another challenge with the Denis classification system noted by several authors is that not all injuries fit into one of the fracture pattern types, making the system less comprehensive than is preferred by clinicians and researchers.

The next major classification system was published by Magerl and the AO (Arbeitsgemeinschaft für Osteosynthesefragen) in 1994.¹⁴ The Comprehensive Classification system was derived from a retrospective review of 1445 thoracic and lumbar injuries. Three major injury patterns were identified with categorization along a scale of progressive severity. There were 3 main types of fracture with distinctly different mechanisms of occurrence: type A-axial compression, type B-distractive of anterior and/or posterior elements, and type C-axial torque leading to anterior and posterior element disruption with rotation. Each type was divided into 3 subtypes, and the subtypes were further divided into subdivisions. The result was a total of 53 distinct fracture patterns, leading to a comprehensive system with which to classify thoracolumbar trauma. The most common type of injury was the type A fracture, with a prevalence of 66%. Type B had a prevalence of 14.5%, and type C, 19.5%. The concept was a hierarchical classification system in which an A1 injury was less severe than a C3 injury,

and although neurological injury was not a part of the classification, the incidence of neurological injury did increase by type with an incidence of 14% in type A fractures, 32% in type B, and 55% in type C. Conceptually, the classification used both mechanism and morphology with the 3 major types differentiated by mechanism, but the groups and subgroups were based on specific morphology.

In attempts to determine the reliability, several studies provide level II evidence that externally validates the Comprehensive Classification System as a means to communicate fracture type when only the 3 main types are used in the classification. Oner²¹ examined the question by examining 53 consecutive patients with CT and MRI data who had thoracolumbar fractures. The intraobserver reliability for CT alone was 0.31 and for MRI alone was 0.28. When all studies were used, the intraobserver reliability was 0.47 for the complete classification system, which indicated fair to moderate reliability. When only the 3 major types were classified, the intraobserver κ was only 0.41. When interobserver reliability was examined, the results showed wide variation depending on who the observer was and ranged from poor to moderate when examining either the 3 major types or the complete classification system. The authors concluded that Comprehensive Classification was useful because it included all fracture patterns, but that MRI is important to better define the distinctive properties of the 3 different mechanisms of injury. There was difficulty in distinguishing between type A and type B fracture when only radiographs or CT were used without MRI. Leferink²² retrospectively reviewed 160 surgically treated patients with AO type A and type B fractures. Fracture classification was reviewed with the benefit of operative notes, and 17 fractures of the total population

were reclassified as type B. Based on this level III evidence, the authors concluded that up to 30% of type B fractures are misclassified as type A and suggested that preoperative MRI might be helpful to correctly use the Comprehensive Classification system.

Other researchers have also attempted to externally validate the Magerl (AO) classification system. Wood circulated 31 cases among 19 observers asking the participants to grade the 3 AO types and 9 AO subtypes.²⁰ The κ for interobserver reliability was 0.48 for AO type and 0.54 for AO subtype, with an intraobserver agreement of 82% for AO type and 67% for AO subtype. The authors concluded, using level II evidence, that well-trained spine surgeons demonstrated only high to moderate reliability when they used the AO classification at its simplest level and that intraobserver reliability at the subtype level or potentially beyond was of concern.

Kriek and Govender²³ examined the reliability of AO classification with radiographs and clinical information in 150 cases, with most cases being type C. The authors found interobserver reliability of 0.49 at the most basic level of classification by the second review of cases, but only fair ($\kappa = 0.33$) intraobserver reproducibility.²³ This study provided level III evidence that there is good interobserver reliability with the Comprehensive Classification system. Many observers believe that identification beyond the three basic types (A, B, or C) is confusing, and the AO system does not specifically include the degree of neurological injury,⁴⁻⁶ although the hierarchical grading scheme certainly confirms that type C injuries run a higher risk of neurological injury than type A injuries.

Although Magerl's Comprehensive Classification was inclusive of all injury patterns observed at the thoracolumbar junction, it did not help guide treatment. In an attempt to help guide treatment for the burst fracture (one of the more common types of thoracolumbar injury), McCormack and Gaines¹⁶ proposed the load sharing classification (LCS) scheme. Three characteristics were identified on CT: 1) comminution/involvement, 2) apposition of fragments, and 3) correction of kyphotic deformity in an attempt to determine if posterior short segment instrumentation would fail in the setting of a burst fracture. The CT patterns were assigned the following point scores: 1) involvement of <30% of the vertebral body received 1 point, 30-60% received two points, and >60% received 3 points; 2) apposition of fragments 0-1 mm received 1 point, >2 mm separation in ≤50% of the body received 2 points, and >2 mm of separation in >50% received 3 points, 3) kyphosis correction of <3° received 1 point, 4 to 9 degrees received 2 points, and >10 degrees was given 3 points. Using this scheme, a patient's CT pattern could be assigned a point total and a patient with a total of 7 to 9 points would be likely to benefit from both posterior and anterior fixation. In the original report, 5 out of the 10 patients who failed had 9 points, and no screw fractures occurred in patients with ≤6 points.

Using the LSC, the same researchers reported on a group of patients in which this treatment paradigm was used.²⁴ Of the 51 patients reported, 39 had burst or Chance fractures, and 23 had point scores ≤6 and underwent posterior short segment fixation. All patients healed in near anatomic position. The remaining 16 patients had point scores of

7, 8, or 9, and all had anterior Kaneda type fixation, with 15 of 16 patients healing in anatomic position. The remaining patients had fracture dislocations and underwent initial reduction with posterior instrumentation. If their point score was 7, 8, or 9, that had supplemental anterior strut grafting of the fractured vertebrae. All of these patients healed in a near anatomic position.

Dai examined the LSC in a series of 45 consecutive burst fractures with 5 different observers and found a high degree of interobserver reliability ($\kappa = 0.82$) and intraobserver reliability ($\kappa = 0.89$).²⁵ Elzinga used the LSC in 40 consecutive fractures and found only fair intraobserver reliability on two observations that were 6 months apart ($\kappa = 0.29$). The interobserver reliability on the second assessment was moderate with a κ for involvement of 0.58, for apposition of 0.46, and for correction of 0.31, but the analysis of total score showed moderate to good agreement, with a κ of 0.67. These studies provide level II evidence that the LSC can be used reliably outside the original group to describe fracture patterns. However, there is insufficient evidence that LSC successfully predicts the ability of short segment instrumentation to treat thoracolumbar burst fractures.

Before 2005, no classification systems included the neurological status, which is one of the most important determinants for surgical intervention in a thoracolumbar fracture, and there were essentially no systems, except the McCormack and Gaines LSC, that guided operative intervention. As a result, Vaccaro led an effort by the Spine Trauma Study Group (STSG) to introduce a system that provided an injury severity score for thoracolumbar trauma that could potentially guide a clinician through the description and

management of these injuries. Vaccaro's original classification system, the Thoracolumbar Injury Severity Score (TLISS),¹⁸ relied on 3 variables that could be determined from radiographic data and the clinical examination. Injury mechanism could be recreated from the pattern of injury on the radiographs, and a point value was assigned for each mechanism. Simple compression received 1 point, with a mechanism severe enough to recreate a burst receiving an additional point, translation receiving 3 points, and a distraction mechanism gathering 4 points. To this, Vaccaro added the integrity of the posterior ligamentous complex, a concept initially introduced by Holdsworth, but quantified in TLISS. No points are assigned if the complex is intact, 3 points if it is ruptured and 2 points if the integrity of the ligaments is indeterminate. Finally, additional criteria (beyond radiographic) were added to TLISS. This was the first classification system to quantify the neurological status of the patient. Zero points were assigned to the intact patient, while those who were complete or have a nerve root injury receive two points, and patients with incomplete or cauda equine injuries were deemed the most urgent and received 3 points. If the point total was ≥ 5 , the injury was deemed operable and those injury patterns with only three points were thought capable of being treated nonsurgically. When patients received 4 points, surgery was left to the discretion of the treating physician, although physicians in North America would often proceed with surgical intervention. For example, a treatment option for a burst fracture with a complete neurological deficit would be decompression and stabilization to prevent late deformity and increase any potential for neurological improvement.

The internal reliability of the TLISS was determined by core members of the STSG from within the same institution and provided level III evidence that TLISS is a reproducible classification system. Vaccaro et al²⁶ circulated 71 cases with history, presumed mechanism of injury, neurological exam, and radiographs, including plain films, CT and MRI. Inter- and intrarater reliability was reported using a Cohen's κ statistic, with a κ of 0.33 for mechanism, 0.91 for the neurological score, 0.35 for integrity of the PLC, and 0.29 for total TLISS score, with an interrater reliability on repeated evaluation of 0.46.²⁶ Although the interrater reliability was only fair, surgeons agreed with the recommendation of the TLISS score 96% of time. Patel also determined the reliability of TLISS and assessed whether interobserver reliability of TLISS could improve with time.²⁷ The κ for mechanism improved from 0.26 to 0.64, and the κ for total TLISS improved from 0.19 to 0.51. The improvement from slight to moderate agreement for TLISS provided level II evidence that the classification system can be taught and learned with relative ease and that interobserver reliability improves with time and education.

The initial studies on reliability found only fair reliability of TLISS, so the researchers suggested that injury mechanism was often hard to recreate from the original radiographic studies and replaced the concept of mechanism (TLISS) with that of morphological pattern of the fracture.^{5,28,29} The resultant Thoracolumbar Injury Classification and Severity Score (TLICS) was used with the following modification: a compression fracture received 1 point, a burst received 2, a translation/rotation injury received 3 points, and a distraction injury contributed 4 points to the score.⁵ The hypothesis that morphology would lead to higher reliability when using the TLICS as

compared to the TLISS was tested by Whang.²⁹ Twenty-five consecutive cases of thoracolumbar trauma were presented to surgeons, ranging from orthopedic attendings to junior residents. The cases were scored based on TLISS and then 3 months later, the TLICS was scored. The κ statistic for the mechanism/morphology component of the injury score showed substantial agreement (0.64 - TLISS/0.63 - TLICS) and the overall agreement was moderate for both classification schemes (0.51 vs 0.46.). These authors provided level II evidence that TLICS was not necessarily more reliable than TLISS due to a moderate to substantial level of agreement for subcategories in both systems.

A number of researchers have demonstrated external validity to the TLISS/TLICS classification schema. Lenarz et al³⁰ looked at thoracolumbar trauma patients and graded fracture patterns based on three classification systems: AO comprehensive classification, Denis, and TLISS. Ninety-seven consecutive fractures were examined by four groups: 1) spine attendings, 2) spine fellows, 3) nonspine attendings, and 4) orthopedic residents. The TLISS classification showed substantial agreement for mechanism, neurological status, and PLC integrity for spine attendings and fellows, and moderate agreement for nonspine attendings and junior residents. Moderate agreement was found when examining the AO and Denis classifications by basic fracture type, providing level II evidence that the interobserver reliability can be substantial for TLISS, but varies by level of involvement and training. The same group then conducted a retrospective analysis of the same 97 consecutive trauma patients to determine if actual management agreed with the TLISS score and found that in those patients with a score ≤ 3 , 48 of 51 patients were successfully treated nonoperatively, while in 33 of 37 with a TLISS score ≥ 5 , surgery

was indeed chosen as the treatment option, supporting with level III evidence the use of TLISS in initial fracture management.³¹

Other authors have also examined whether TLICS/TLISS score could predict treatment (surgery vs. conservative treatment) outside the initial group of surgeons that developed the classification, thus validating its utility as a predictor for treatment. In 2010, Joaquin retrospectively reviewed a series of trauma patients from 2 centers in Brazil. A total of 49 patients who underwent surgery for thoracolumbar trauma had adequate records to determine a TLICS classification score.³² Forty-seven of the 49 patients (95.9%) had a TLICS score of 4 or greater, with 87.5% of those with a score > 6 exhibiting some degree of neurological injury. The authors provided level III evidence that TLICS was useful as an injury severity score in their trauma population. However, the same author reviewed a series of 458 consecutive patients from a North American center. A total of 310 patients were treated conservatively, and 148 patients were treated surgically.^{33,34} Of the 310 patients conservatively treated patients, 307 had a TLICS score of 4 or less (98%), but only 69 of the 148 patients (47%) in operative arm matched TLICS recommendations providing contradictory level II evidence that in North America, TLICS scoring does not accurately predict treatment for thoracolumbar trauma. The authors found inconsistencies with TLICS and treatment of thoracolumbar trauma, and it is likely that the population had a high number of “stable” burst fractures treated with early surgical stabilization in an effort to promote early mobilization. The authors noted that early in the study period, before the introduction of TLICS, several distractive injuries were missed, leading to delayed surgery. They surmised that integration of PLC injury into a classification

scheme would help reduce the number of missed distraction injuries, and indeed, none of these injuries were treated conservatively after the introduction of the TLICS system.

Finally, Choi et al³⁵ assessed the applicability of TLICS to a group of spine trauma patients previously treated between 2010 and 2013 in Korea. Decisions for operative intervention on thoracolumbar trauma are based on strict criteria from the Worker's Compensation board, HIRA, with 3-column injuries, burst fractures with 30° of kyphosis/40% loss of height/50% canal encroachment, injury of PLC, neurological deficit, and pain with conservative treatment all considered criteria for surgery. A total of 100 patients were retrospectively reviewed with 45 treated surgically and 55 treated nonsurgically. In the nonsurgically treated group, TLICS scores ranged 1 to 4 with no patients over 4. In the surgically treated group, all had TLICS scores ≥ 4 (mean, 5.62), except 1 patient who had an initial score of 2. This study provided level III evidence that the TLICS system has clinical applicability compared to real life experience in a select patient population.

Due to regional differences in the threshold for surgical intervention, and because of the often low reliability of discerning PLC injury and the wide variation in the availability of MRI to help determine PLC injury,³⁶⁻³⁸ the AO Spine Classification Group was tasked with the development of a morphologically based classification scheme that also paid attention to the critical determinant of neurological examination.^{6,17} The resultant AO Spine Thoracolumbar Injury Classification System is a comprehensive yet simple scheme which appears on initial evaluation to have greater reproducibility and reliability than

prior schemes. The wide availability and use of CT for evaluation of trauma patients is the basis for this scheme and uses the Magerl hierarchy of injury types with each successive type indicating ascending severity. Type A injuries are compression injuries with injury of the anterior elements and preservation of the posterior ligamentous complex: A0 fractures represent transverse or spinous process fractures; A1 are wedge compression fractures of 1 endplate without involvement of the posterior wall of the vertebral body; A2 are split or pincer fractures with involvement of both endplates; A3 are incomplete burst fractures which involve the posterior wall of the vertebral body but only 1 endplate; and A4 fractures are complete bursts, which involve both endplates and the posterior wall.

Type B injuries are failure of the posterior or anterior tension band in distraction: B1 injuries are transosseous monosegmental failure of the posterior tension band; B2 are bony and/or ligamentous failure of the posterior tension band in conjunction with an A fracture of the vertebral body; B3 injuries are hyperextension injuries through the disc space or bone as commonly seen in ankylosing spondylitis. There is some confusion because the first iteration of this new AO Classification System included these injuries under type C. However, for the purposes of this guideline, the authors will include them as type B as this is the classification which has been investigated for internal and external reliability.

Finally, type C injuries suffer disruption of all elements with displacement or dislocation of the cranial spinal elements relative to the caudal elements. There are no subtypes any

longer for this injury pattern. In addition to the morphological classification, there is also a neurological grading component (N0 = intact, N1 = transient symptoms, N2 = radiculopathy, N3 = incomplete or cauda injury, and N4 = complete) and case-specific modifiers. The goal will be to develop a spine injury score, though this is a work in progress and beyond the scope of this review.

The initial evaluation of the AO Trauma Knowledge Forum working group consisted of 7 face-to-face meetings with 9 experienced spine surgeons. The final evaluation to determine interobserver reliability was performed on 40 cases culled from the lead author's practice. There was agreement in 60% of cases when looking at basic type ($\kappa = 0.72$) and by complete classification or subtype there was agreement 35% of the time for a κ of 0.64, showing good agreement when looking at the 3 basic types or even the complete classification system. The intraobserver reliability showed a κ of 0.77 for the whole classification and 0.85 when only looking at subtype, which suggests excellent reproducibility when classifying fractures with this system.

Urrutia et al³⁹ independently examined the inter- and intraobserver reliability of the modified AO classification scheme using 70 cases, evaluated by 6 surgeons, 6 weeks apart. The interobserver reliability was good for fracture type, $\kappa = 0.62$, and similar for subtype, $\kappa = 0.55$. The intraobserver reliability yielded a κ of 0.77 for type and 0.71 for subtype. The substantial agreement between observers with a wide variety of experience provides level II evidence that the modified AO classification may be a more reproducible classification system than previous systems.

Future Research

These studies show that TLICS/TLISS cannot yet be adapted to predict management in all thoracolumbar trauma populations because there is still wide variation in treatment recommendations for physicians who treat these types of injuries. Further prospective studies are necessary to validate the best treatment options for burst fractures that may be considered stable and have a TLICS score of 2 to 4. Prospective research is also lacking to demonstrate that the utilization of any classification system (compared to not using any system) in making treatment decisions results in superior clinical outcomes for patients with thoracolumbar spine injuries.

CONCLUSIONS

In summary, several classification systems for thoracolumbar trauma have been proposed over the last 100 years. Some systems follow mechanistic descriptions of the fracture patterns, while others are considered morphological classification systems. However, all systems had limitations with some being overly comprehensive or inclusive, and therefore, difficult to learn and use, while other systems had fewer fracture types and subtypes, which left gaps that did not allow for descriptions of all fracture types. In addition, none of the classification systems went through a rigorous validation process, and therefore were often difficult to reproduce outside of the original working group that proposed the system.

In the last 10 years, two classification systems have been proposed, TLICS and the AO Thoracolumbar Spine Injury Classification System. These have both undergone studies to measure internal and external reliability and were found to be inclusive and descriptive of

most thoracolumbar fractures. Hopefully, more studies using these systems will become available to determine if these systems can accurately predict fracture treatment through specific treatment protocols. Rigorous adoption and utilization of a specific classification description is needed for future researchers to perform studies to determine if a specific treatment algorithm is beneficial for a specific fracture pattern.

Potential Conflicts of Interest

The task force members were required to report all possible conflicts of interest (COIs) prior to beginning work on the guideline, using the COI disclosure form of the AANS/CNS Joint Guidelines Committee, including potential COIs that are unrelated to the topic of the guideline. The CNS Guidelines Committee and Guideline Task Force Chairs reviewed the disclosures and either approved or disapproved the nomination. The CNS Guidelines Committee and Guideline Task Force Chairs are given latitude to approve nominations of Task Force members with possible conflicts and address this by restricting the writing and reviewing privileges of that person to topics unrelated to the possible COIs. The conflict of interest findings are provided in detail in the companion [introduction and methods manuscript](#).

Disclaimer of Liability

This clinical systematic review and evidence-based guideline was developed by a multidisciplinary physician volunteer task force and serves as an educational tool designed to provide an accurate review of the subject matter covered. These guidelines are disseminated with the understanding that the recommendations by the authors and consultants who have collaborated in their development are not meant to replace the individualized care and treatment advice from a patient's physician(s). If medical advice

or assistance is required, the services of a competent physician should be sought. The proposals contained in these guidelines may not be suitable for use in all circumstances. The choice to implement any particular recommendation contained in these guidelines must be made by a managing physician in light of the situation in each particular patient and on the basis of existing resources.

Disclosures

These evidence-based clinical practice guidelines were funded exclusively by the Congress of Neurological Surgeons and the Section on Disorders of the Spine and Peripheral Nerves in collaboration with the Section on Neurotrauma and Critical Care, which received no funding from outside commercial sources to support the development of this document.

Acknowledgments

The guidelines task force would like to acknowledge the CNS Guidelines Committee for their contributions throughout the development of the guideline and the AANS/CNS Joint Guidelines Committee for their review, comments, and suggestions throughout peer review, as well as the contributions of Trish Rehring, MPH, CHES, Senior Manager of Clinical Practice Guidelines for the CNS, and Mary Bodach, MLIS, Guidelines Specialist and Medical Librarian, for assistance with the literature searches. Throughout the review process the reviewers and authors were blinded from one another. At this time, the guidelines task force would like to acknowledge the following individual peer reviewers for their contributions: Maya Babu, MD, MBA, Greg Hawryluk, MD, PhD, Steven Kalkanis, MD, Yi Lu, MD, PhD, Jeffrey J. Olson, MD, Martina Stippler, MD, Cheerag Upadhyaya, MD, MSc, and Robert Whitmore, MD.

REFERENCES

1. Audige L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma* 2005;19:401-406.
2. Bono CM, Vaccaro AR, Hurlbert RJ, et al. Validating a newly proposed classification system for thoracolumbar spine trauma: looking to the future of the thoracolumbar injury classification and severity score. *J Orthop Trauma* 2006;20:567-572.
3. Mirza SK, Mirza AJ, Chapman JR, Anderson PA. Classifications of thoracic and lumbar fractures: rationale and supporting data. *J Am Acad Orthop Surg* 2002;10:364-377.
4. Sethi MK, Schoenfeld AJ, Bono CM, Harris MB. The evolution of thoracolumbar injury classification systems. *Spine J* 2009;9:780-788.
5. Vaccaro AR, Lehman RA, Jr., Hurlbert RJ, et al. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. *Spine (Phila Pa 1976)* 2005;30:2325-2333.
6. Vaccaro AR, Oner C, Kepler CK, et al. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. *Spine (Phila Pa 1976)* 2013;38:2028-2037.
7. Bohler L. *Die Technik der Knochenbruchbehandlung im Frieden und im Kriegeed.* Wien, Austria: Maudrich; 1930.

8. Watson-Jones R. The results of postural reduction of fractures of the spine. *J Bone Joint Surg Am* 1938;20:567-586.
9. Nicoll EA. Fractures of the dorso-lumbar spine. *J Bone Joint Surg Br* 1949;31B:376-394.
10. Holdsworth F. Fractures, dislocations, and fracture-dislocations of the spine. *J Bone Joint Surg Am* 1970;52:1534-1551.
11. Kelly RP, Whitesides TE, Jr. Treatment of lumbodorsal fracture-dislocations. *Ann Surg* 1968;167:705-717.
12. Denis F. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. *Spine (Phila Pa 1976)* 1983;8:817-831.
13. Ferguson RL, Allen BL, Jr. A mechanistic classification of thoracolumbar spine fractures. *Clin Orthop Relat Res* 1984;189:77-88.
14. Magerl F, Aebi M, Gertzbein SD, Harms J, Nazarian S. A comprehensive classification of thoracic and lumbar injuries. *Eur Spine J* 1994;3:184-201.
15. McAfee PC, Yuan HA, Fredrickson BE, Lubicky JP. The value of computed tomography in thoracolumbar fractures. An analysis of one hundred consecutive cases and a new classification. *J Bone Joint Surg Am* 1983;65:461-473.
16. McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. *Spine (Phila Pa 1976)* 1994;19:1741-1744.

17. Reinhold M, Audige L, Schnake KJ, Bellabarba C, Dai LY, Oner FC. AO spine injury classification system: a revision proposal for the thoracic and lumbar spine. *Eur Spine J* 2013;22:2184-2201.
18. Vaccaro AR, Zeiller SC, Hulbert RJ, et al. The thoracolumbar injury severity score: a proposed treatment algorithm. *J Spinal Disord Tech* 2005;18:209-215.
19. Ransohoff DF, Pignone M, Sox HC. How to decide whether a clinical practice guideline is trustworthy. *JAMA* 2013;309:139-140.
20. Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB, Mehbod AA. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. *J Bone Joint Surg Am* 2005;87:1423-1429.
21. Oner FC, Ramos LM, Simmermacher RK, et al. Classification of thoracic and lumbar spine fractures: problems of reproducibility. A study of 53 patients using CT and MRI. *Eur Spine J* 2002;11:235-245.
22. Leferink VJ, Veldhuis EF, Zimmerman KW, ten Vergert EM, ten Duis HJ. Classificational problems in ligamentary distraction type vertebral fractures: 30% of all B-type fractures are initially unrecognized. *Eur Spine J* 2002;11:246-250.
23. Kriek JJ, Govender S. AO-classification of thoracic and lumbar fractures--reproducibility utilizing radiographs and clinical information. *Eur Spine J* 2006;15:1239-1246.

24. Parker JW, Lane JR, Karaikovic EE, Gaines RW. Successful short-segment instrumentation and fusion for thoracolumbar spine fractures: a consecutive 41/2-year series. *Spine (Phila Pa 1976)* 2000;25:1157-1170.
25. Dai LY, Jin WJ. Interobserver and intraobserver reliability in the load sharing classification of the assessment of thoracolumbar burst fractures. *Spine (Phila Pa 1976)* 2005;30:354-358.
26. Vaccaro AR, Baron EM, Sanfilippo J, et al. Reliability of a novel classification system for thoracolumbar injuries: the Thoracolumbar Injury Severity Score. *Spine (Phila Pa 1976)* 2006;31:S62-69.
27. Patel AA, Vaccaro AR, Albert TJ, et al. The adoption of a new classification system: time-dependent variation in interobserver reliability of the thoracolumbar injury severity score classification system. *Spine (Phila Pa 1976)* 2007;32:E105-110.
28. Schweitzer KM, Jr., Vaccaro AR, Lee JY, Grauer JN. Confusion regarding mechanisms of injury in the setting of thoracolumbar spinal trauma: a survey of The Spine Trauma Study Group (STSG). *J Spinal Disord Tech* 2006;19:528-530.
29. Whang PG, Vaccaro AR, Poelstra KA, et al. The influence of fracture mechanism and morphology on the reliability and validity of two novel thoracolumbar injury classification systems. *Spine (Phila Pa 1976)* 2007;32:791-795.
30. Lenarz CJ, Place HM, Lenke LG, Alander DH, Oliver D. Comparative reliability of 3 thoracolumbar fracture classification systems. *J Spinal Disord Tech* 2009;22:422-427.

31. Lenarz CJ, Place HM. Evaluation of a new spine classification system, does it accurately predict treatment? *J Spinal Disord Tech* 2010;23:192-196.
32. Joaquim AF, Fernandes YB, Cavalcante RA, Fragoso RM, Honorato DC, Patel AA. Evaluation of the thoracolumbar injury classification system in thoracic and lumbar spinal trauma. *Spine (Phila Pa 1976)* 2011;36:33-36.
33. Joaquim AF, Lawrence B, Daubs M, et al. Measuring the impact of the Thoracolumbar Injury Classification and Severity Score among 458 consecutively treated patients. *J Spinal Cord Med* 2014;37:101-106.
34. Joaquim AF, Daubs MD, Lawrence BD, et al. Retrospective evaluation of the validity of the Thoracolumbar Injury Classification System in 458 consecutively treated patients. *Spine J* 2013;13:1760-1765.
35. Choi HJ, Kim HS, Nam KH, Cho WH, Choi BK, Han IH. Applicability of thoracolumbar injury classification and severity score to criteria of korean health insurance review and assessment service in treatment decision of thoracolumbar injury. *J Korean Neurosurg Soc* 2015;57:174-177.
36. Chhabra HS, Kaul R, Kanagaraju V. Do we have an ideal classification system for thoracolumbar and subaxial cervical spine injuries: what is the expert's perspective? *Spinal Cord* 2015;53:42-48.

37. Radcliff K, Kepler CK, Rubin TA, et al. Does the load-sharing classification predict ligamentous injury, neurological injury, and the need for surgery in patients with thoracolumbar burst fractures? Clinical article. *J Neurosurg Spine* 2012;16:534-538.
38. Vaccaro AR, Rihn JA, Saravanja D, et al. Injury of the posterior ligamentous complex of the thoracolumbar spine: a prospective evaluation of the diagnostic accuracy of magnetic resonance imaging. *Spine (Phila Pa 1976)* 2009;34:E841-847.
39. Urrutia J, Zamora T, Yurac R, et al. An independent interobserver reliability and intraobserver reproducibility evaluation of the new AOSpine Thoracolumbar Spine Injury Classification System. *Spine (Phila Pa 1976)* 2015;40:E54-58.

Appendix I. Literature Searches

Search Strategies

PubMed

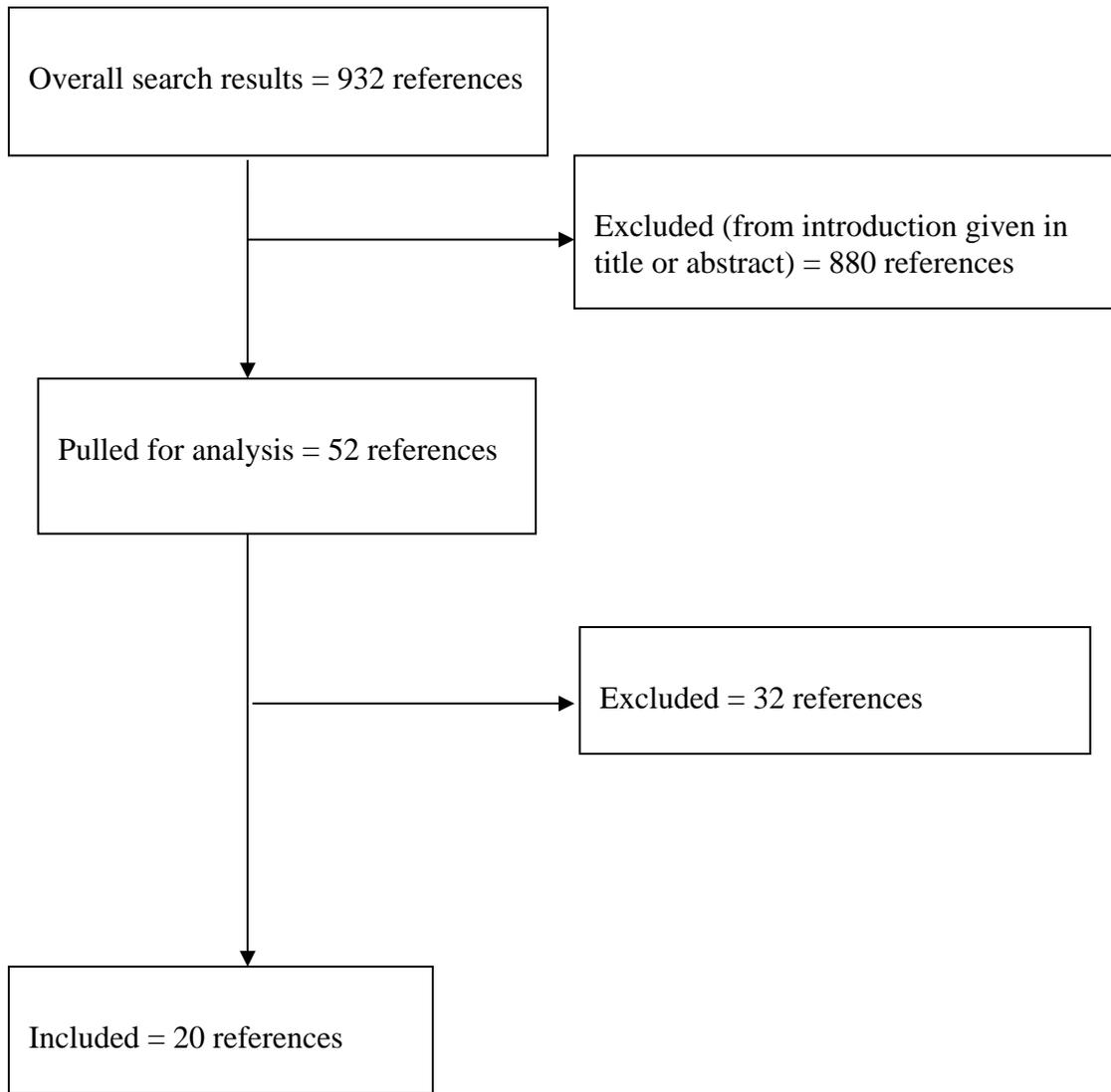
1. Lumbar vertebrae [MeSH] OR Thoracic vertebrae [MeSH]
2. Spinal Injuries [MeSH] OR Spinal Cord Injuries [MeSH]
3. #1 AND #2
4. Thoracolumbar [TIAB] OR thoraco-lumbar [TIAB] OR thoraco lumbar [TIAB] OR burst [Title]
5. Injur* [TIAB] OR trauma* [TIAB] OR fractur* [TIAB] OR dislocation* [TIAB]
6. #4 AND #5
7. Lumbar vertebrae/injuries [MeSH] OR Thoracic vertebrae/injuries [MeSH]
8. #3 OR #6 OR #7
9. Trauma Severity Indices [MeSH] OR (Wounds and Injuries/classification [MeSH:noexp] AND 1966:1989 [MHDA]) OR classification [SH]
10. Classif* [TIAB] OR categor* [TIAB]
11. #9 OR #10
12. #8 AND #11
13. #12 AND English [Lang]
14. (animal [MeSH] NOT human [MeSH]) OR cadaver [MeSH] OR cadaver* [Titl] OR comment [PT] OR letter [PT] OR editorial [PT] OR addresses [PT] OR news [PT] OR “newspaper article” [PT] OR case reports [PT]
15. #13 NOT #14
16. osteoporosis [MH] OR osteoporotic fractures [MH] OR osteoporo* [TITLE] OR spinal neoplasms [MH] OR tumor* [TITLE] OR tumour* [TITLE] OR malignan* [TITLE]
17. #15 NOT #16

Cochrane Library

1. Lumbar vertebrae: MeSH descriptor, explode all trees
2. Thoracic vertebrae: MeSH descriptor, explode all trees
3. #1 OR #2
4. Spinal Injuries: MeSH descriptor
5. Spinal Cord Injuries: MeSH descriptor
6. #4 OR #5
7. #3 AND #6
8. (Thoracolumbar OR thoraco-lumbar OR thoraco lumbar OR burst) NEAR/4 (Injur* OR trauma* OR fractur* OR dislocation*):ti,ab,kw
9. Lumbar vertebrae/injuries: MeSH descriptor, explode all trees
10. Thoracic vertebrae/injuries: MeSH descriptor, explode all trees
11. #9 OR #10
12. #7 OR #8 OR #11
13. mh osteoporosis or mh osteoporotic fractures or mh spinal neoplasms
14. osteoporo* or tumor* or malignan*:ti

15. #13 OR #14
16. #12 NOT #15

Appendix II. Article Inclusions and Exclusions



**Appendix III. Rating Evidence Quality
Levels of Evidence for Primary Research Question^a**

Types of studies				
	Therapeutic studies – Investigating the results of treatment	Prognostic studies – Investigating the effect of a patient characteristic on the outcome of disease	Diagnostic studies – Investigating a diagnostic test	Economic and decision analyses – Developing an economic or decision model
Level I	<ul style="list-style-type: none"> • High-quality randomized trial with statistically significant difference or no statistically significant difference but narrow confidence intervals • Systematic review^b of level I RCTs (and study results were homogenous^c) 	<ul style="list-style-type: none"> • High-quality prospective study^d (all patients were enrolled at the same point in their disease with ≥80% follow-up of enrolled patients) • Systematic review^b of level I studies 	<ul style="list-style-type: none"> • Testing of previously developed diagnostic criteria on consecutive patients (with universally applied reference “gold” standard) • Systematic review^b of level I studies 	<ul style="list-style-type: none"> • Sensible costs and alternatives; values obtained from many studies; with multiway sensitivity analyses • Systematic review^b of level I studies
Level II	<ul style="list-style-type: none"> • Lesser quality RCT (e.g., ≤80% follow-up, no blinding, or improper randomization) • Prospective^d comparative study^e • Systematic review^b of level II studies or level I studies with inconsistent results 	<ul style="list-style-type: none"> • Retrospective^f study • Untreated controls from an RCT • Lesser quality prospective study (e.g., patients enrolled at different points in their disease or ≤80% follow-up) • Systematic review^b of level II studies 	<ul style="list-style-type: none"> • Development of diagnostic criteria on consecutive patients (with universally applied reference “gold” standard) • Systematic review^b of level II studies 	<ul style="list-style-type: none"> • Sensible costs and alternatives; values obtained from limited studies; with multiway sensitivity analyses • Systematic review^b of level II studies

Level III	<ul style="list-style-type: none"> • Case control study^g • Retrospective comparative study^e • Systematic review^b of level III studies 	<ul style="list-style-type: none"> • Case control study^g 	<ul style="list-style-type: none"> • Study of non consecutive patients; without consistently applied reference “gold” standard • Systematic review^b of level III studies 	<ul style="list-style-type: none"> • Analyses based on limited alternatives and costs; and poor estimates • Systematic review^b of level III studies
Level IV	Case series ^h	Case series	<ul style="list-style-type: none"> • Case-control study • Poor reference standard 	<ul style="list-style-type: none"> • Analyses with no sensitivity analyses

RCT, Randomized controlled trial.

^aA complete assessment of quality of individual studies requires critical appraisal of all aspects of the study design.

^bA combination of results from ≥ 2 previous studies.

^cStudies provided consistent results.

^dStudy was started before the first patient enrolled.

^ePatients treated one way (e.g., instrumented arthrodesis) compared with a group of patients treated in another way (e.g., uninstrumented arthrodesis) at the same institution.

^fThe study was started after the first patient enrolled.

^gPatients identified for the study based on their outcome, called “cases” (e.g., pseudoarthrosis) are compared to those who did not have outcome, called “controls” (e.g., successful fusion).

^hPatients treated one way with no comparison group of patients treated in another way.

Appendix IV. Linking Levels of Evidence to Grades of Recommendation

Grade of recommendation	Standard language	Levels of evidence	
A	Recommended	Two or more consistent level I studies	
B	Suggested	One level I study with additional supporting level II or III studies	Two or more consistent level II or III studies
C	Is an option	One level I, II, or III study with supporting level IV studies	Two or more consistent level IV studies
Insufficient (insufficient or conflicting evidence)	Insufficient evidence to make recommendation for or against	A single level I, II, III, or IV study without other supporting evidence	>1 study with inconsistent findings ^a

^aNote that in the presence of multiple consistent studies, and a single outlying, inconsistent study, the Grade of Recommendation will be based on the level of the consistent studies.

Appendix V. Criteria Grading the Evidence

The task force used the criteria provided below to identify the strengths and weaknesses of the studies included in this guideline. Studies containing deficiencies were downgraded one level (no further downgrading allowed, unless so severe that study had to be excluded). Studies with no deficiencies based on study design and contained clinical information that dramatically altered current medical perceptions of topic were upgraded.

1. Baseline study design (i.e., therapeutic, diagnostic, prognostic) determined to assign initial level of evidence.
2. Therapeutic studies reviewed for following deficiencies:
 - Failure to provide a power calculation for an RCT;
 - High degree of variance or heterogeneity in patient populations with respect to presenting diagnosis/demographics or treatments applied;
 - <80% of patient follow-up;
 - Failure to utilize validated outcomes instrument;
 - No statistical analysis of results;
 - Cross over rate between treatment groups of >20%;
 - Inadequate reporting of baseline demographic data;
 - Small patient cohorts (relative to observed effects);
 - Failure to describe method of randomization;
 - Failure to provide flowchart following patients through course of study (RCT);
 - Failure to account for patients lost to follow-up;

- Lack of independent post-treatment assessment (e.g., clinical, fusion status, etc.);
 - Utilization of inferior control group:
 - Historical controls;
 - Simultaneous application of intervention and control within same patient.
 - Failure to standardize surgical/intervention technique;
 - Inadequate radiographic technique to determine fusion status (e.g., static radiographs for instrumented fusion).
3. Methodology of diagnostic studies reviewed for following deficiencies:
- Failure to determine specificity and sensitivity;
 - Failure to determine inter- and intraobserver reliability;
 - Failure to provide correlation coefficient in the form of kappa values.
4. Methodology of prognostic studies reviewed for following deficiencies:
- High degree of variance or heterogeneity in patient populations with respect to presenting diagnosis/demographics or treatments applied;

Failure to appropriately define and assess independent and dependent variables (e.g., failure to use validated outcome measures when available).

Appendix VI. Evidence Tables

Table 1. Systems for Classification of Thoracolumbar Fractures

Author, Year	Level of Evidence	Task Force Conclusions Relative to Question and Rationale for Evidence Grading
Kelly and Whitesides, ¹¹ 1968	III	Further promoted the concept of 2 anatomic columns of the spine
Denis, ¹² 1984	III	Original classification reviewing 412 fractures with 4 distinct types and 16 subtypes. Introduced the concept of three different anatomic columns
McAfee et al, ¹⁵ 1986	III	Original classification based on CT appearance of fracture. Importance of mode of failure of middle column in distinguishing stable from unstable burst fractures
Ferguson and Allen, ¹³ 1984	III	Original classification with 7 different fracture patterns based on mechanism of failure
Magerl et al, ¹⁴ 1994	III	Original classification with 3 major types but 53 subtypes, which provided a comprehensive description of all fractures. Increasing risk of neurological injury with increased instability
McCormack et al, ¹⁶ 1994	III	Classification of burst fracture with anatomic pattern determining whether short segment posterior fixation is adequate
Vaccaro et al, ⁵ 2005	III	Original classification with a scoring system derive from injury morphology or mechanism, neurological deficit, and integrity of the PLC. Surgery should be offered with injury to PLC and/or neurologic deficit
Vaccaro et al, ⁶ 2013; Reinhold et al, ¹⁷ 2013	III	Original classification which simplifies the original AO classification system into many fewer subtypes.

AO, Arbeitsgemeinschaft fur Osteosynthesenfragen (Association for the Study of Internal Fixation); CT, computed tomography; PLC, posterolateral corner.

Table 2. Studies Examining the Reliability and Validity of the Major Classification Systems

Author, Year	Level of Evidence	Task Force Conclusions Relative to Question and Rationale for Evidence Grading
Choi et al, ³⁵ 2015	III	TLICS provided similar recommendations for surgical or nonsurgical treatment in a population of Korean workers compared to formal recommendations from the Korean Worker's Compensation Board
Dai and Jin, ²⁵ 2005	II	There is excellent inter- and intraobserver reliability using the load sharing classification to describe fractures. However, there is insufficient evidence to suggest that this classification scheme can be used to predict failure of short segment instrumentation
Joaquim et al, ³² 2010	III	This paper provides evidence that TLICS is useful as a predictor for surgical or nonsurgical treatment outside the original group of surgeons who developed the classification
Joaquim et al, ³⁴ 2013	II	Retrospective review of 458 consecutive patients to determine if TLICS classification scoring corresponded with actual management. 98% of patients in the conservative arm matched TLICS recommendations while only 47% in operative arm matched TLICS recommendations
Kriek and Govender, ²³ 2006	III	In this retrospective review of type A and B fractures using the comprehensive classification, up to 30% of type B fractures are misclassified as type A. The authors point to the potential importance of MRI
Leferink et al, ²² 2002	III	There is good interobserver reliability of the comprehensive classification at the basic level of the 3 major types
Lenarz et al, ³⁰ 2009	II	This is a reliability study comparing the AO, Denis, and TLISS classification systems which shows good interobserver reliability among more senior reviewers. Suggests the level of reliability varies by experience
Lenarz and Place, ³¹ 2010	III	Retrospective review of cases to determine if actual management of cases agreed with TLISS score. 48/51 with a score <3 were successfully treated nonoperatively and 33/37 with a score ≥5 were treated surgically. The authors believe this supports the utility of TLISS in initial fracture management
Oner et al, ²¹ 2002	II	There is good interobserver reliability when classifying fractures with the Denis classification by type, but it drops to fair to moderate when reviewing by subtype. The comprehensive classification system yielded on fair to moderate reliability, but it depended on the level of knowledge of the person interpreting the scans. It also points to the importance of MRI in using the comprehensive classification

Patel et al, ²⁷ 2007	II	Interobserver reliability improved with time for poor or fair to good with repeated use of the TLISS system. This paper suggests that TLISS can be taught and learned with relative ease
Urrutia et al, ³⁹ 2014	II	The modified AO classification system had good to excellent inter- and intraobserver reliability, suggesting this classification system may be more reproducible than earlier systems
Vaccaro et al, ²⁶ 2006	III	This is a retrospective review of 71 cases which showed fair inter- and intraobserver reliability for the TLISS when looking at the subcategories for posterior ligamentous disruption and mechanism. However, surgeons agreed with the TLISS recommendation for or against surgery 96% of the time
Wood et al., ²⁰ 2005	II	There is moderate inter- and intraobserver reliability when using either the AO-comprehensive or Denis classification systems at the most basic types of injury. However, the reliability is only fair to moderate when trying to classify subtypes of injury

AO, Arbeitsgemeinschaft für Osteosynthesefragen (Association for the Study of Internal Fixation); MRI, magnetic resonance imaging; TLICS, Thoracolumbar Injury Classification and Severity Scale; TLISS, Thoracolumbar Injury Severity Score.