

## Reliability and Clinical Usefulness of Current Classifications in Traumatic Thoracolumbar Fractures: A Systematic Review of the Literature

I. Curfs, M. Schotanus, W.L.W. Van Hemert, M. Heijmans, R.A. De Bie, L.W. Van Rhijn and P.C.P.H. Willems

*Int J Spine Surg* published online 29 December 2020  
<https://www.ijssurgery.com/content/early/2020/12/17/7145>

This information is current as of May 18, 2025.

---

**Email Alerts** Receive free email-alerts when new articles cite this article. Sign up at:  
<http://ijssurgery.com/alerts>

# Reliability and Clinical Usefulness of Current Classifications in Traumatic Thoracolumbar Fractures: A Systematic Review of the Literature

I. CURFS, MD,<sup>1</sup> M. SCHOTANUS, MSC,<sup>1</sup> W.L.W. VAN HEMERT, MD, PHD,<sup>1</sup> M. HEIJMANS, MSC,<sup>2</sup> R.A. DE BIE, PT, PHD,<sup>3</sup> L.W. VAN RHIJN, MD, PHD,<sup>4</sup> P.C.P.H. WILLEMS, MD, PHD<sup>4</sup>

<sup>1</sup>Zuyderland Medical Centre, Department of Orthopedic Surgery and Traumatology, Heerlen, Netherlands, <sup>2</sup>Zuyderland Medical Centre, Zuyderland Academy Heerlen, Netherlands, <sup>3</sup>University of Maastricht, Department of Epidemiology, Maastricht, Netherlands, <sup>4</sup>Maastricht University Medical Centre, Department of Orthopedic Surgery and Traumatology, Maastricht, Netherlands

## ABSTRACT

**Background:** A validated classification remains the key to an appropriate treatment algorithm of traumatic thoracolumbar fractures. Considering the development of many classifications, it is remarkable that consensus about treatment is still lacking. We conducted a systematic review to investigate which classification can be used best for treatment decision making in thoracolumbar fractures.

**Methods:** A comprehensive search was conducted using PubMed, Embase, CINAHL, and Cochrane using the following search terms: classification (mesh), spinal fractures (mesh), and corresponding synonyms. All hits were viewed by 2 independent researchers. Papers were included if analyzing the reliability (kappa values) and clinical usefulness (specificity or sensitivity of an algorithm) of currently most used classifications (Magerl/AO, thoracolumbar injury classification and severity score [TLICS] or thoracolumbar injury severity score, and the new AO spine).

**Results:** Twenty articles are included. The presented kappa values indicate moderate to substantial agreement for all 3 classifications. Regarding the clinical usefulness, > 90% agreement between actual treatment and classification recommendation is reported for most fractures. However, it appears that over 50% of the patients with a stable burst fracture (TLICS 2, AO-A3/A4) in daily practice are operated, so in these cases treatment decision is not primarily based on classification.

**Conclusion:** AO, TLICS, and new AO spine classifications have acceptable accuracy (kappa > 0.4), but are limited in clinical usefulness since the treatment recommendation is not always implemented in clinical practice. Differences in treatment decision making arise from several causes, such as surgeon and patient preferences and prognostic factors that are not included in classifications yet. The recently validated thoracolumbar AO spine injury score seems promising for use in clinical practice, because of inclusion of patient-specific modifiers. Future research should prove its definite value in treatment decision making.

**Level of evidence:** 2

**Clinical Relevance:** Without the appropriate treatment, the impact of traumatic thoracolumbar fractures can be devastating. Therefore it is important to achieve consensus in the treatment of thoracolumbar fractures.

Lumbar Spine

Keywords: thoracolumbar spine, fractures, classification, clinical usefulness, reliability, AO spine, TLICS

## INTRODUCTION

Thoracolumbar fractures are common injuries. Without appropriate treatment, their outcome can be devastating. Commonly, treatment decision is based upon accurate radiological diagnosis and concomitant use of a fracture classification system. Several classifications have been introduced during the past years. With the improvement of imaging (eg, computed tomography [CT], magnetic resonance imaging [MRI]), it has become possible to better understand the pathology of the thoracolumbar spine

fractures, and to recognize fracture patterns. These fracture patterns give insight into fracture morphology, trauma mechanism, and determination of stability, and have led to various classification systems. Classifications aim to create a common language with standardization and optimization of treatment. Currently the most used classifications are Magerl/AO and the thoraco-lumbar injury classification and severity score (TLICS).<sup>1–3</sup>

AO classification is primary based on the pathomorphological characteristics of the injury. It

**Table 1.** Description of the Magerl/AO, thoracolumbar injury classification and severity score (TLICS), and new AO spine classifications.

	Magerl/AO, 1994	TLICS, 2006	New AO Spine, 2013
Goal	More comprehensive classification, including all fracture types	Simplified classification and facilitating treatment decision making	Allows for a development of a globally accepted treatment algorithm
Concept	Primarily based on the pathomorphology of the injury pattern	Treatment algorithm with point allocation, based on <ul style="list-style-type: none"> <li>■ Fracture morphology</li> <li>■ Neurological status</li> <li>■ Posterior-ligament-complex integrity</li> </ul>	Combination of Magerl/AO and TLICS Classification based on the evaluation of 3 basic parameters: <ul style="list-style-type: none"> <li>■ Fracture morphology</li> <li>■ Neurological status</li> <li>■ Patient-specific modifiers</li> </ul>
Distinction	Three main categories of force: <ul style="list-style-type: none"> <li>■ A-type: compressive force which causes compression and burst injuries</li> <li>■ B-type: tensile force which causes injuries with transverse disruption</li> <li>■ C-type: axial torque which causes rotational injuries</li> </ul> Further division into 9 groups, 27 subtypes, and over 50 specifications	Fracture morphology: <ul style="list-style-type: none"> <li>■ Compression</li> <li>■ Translational/rotational</li> <li>■ Distraction</li> </ul> Neurological: <ul style="list-style-type: none"> <li>■ Intact</li> <li>■ Nerve root</li> <li>■ Cord/conus medullaris</li> <li>■ Cauda equina</li> </ul> Posterior-ligament complex: <ul style="list-style-type: none"> <li>■ Intact</li> <li>■ Injury suspected</li> <li>■ Indeterminate</li> <li>■ Injured</li> </ul>	Fracture morphology: <ul style="list-style-type: none"> <li>■ Type A: compression</li> <li>■ Type B: tension-band injuries</li> <li>■ Type C: translational injuries</li> </ul> Neurological status <ul style="list-style-type: none"> <li>■ N0: intact</li> <li>■ N1: transient deficit, no longer present</li> <li>■ N2: radiculopathy</li> <li>■ N3: incomplete spinal cord injury</li> <li>■ N4: complete spinal cord injury</li> <li>■ Nx: not able to evaluate</li> </ul> Clinical modifiers: <ul style="list-style-type: none"> <li>■ M1: PLC integrity</li> <li>■ M2: patient specific comorbidities (rheuma, ankylosing spondylitis etc)</li> </ul>
Treatment management	Classification contains no treatment recommendation Consensus in clinical application: conservative treatment for A-type (excluding burst), and surgical treatment for B- and C-type fractures.	Treatment algorithm: <ul style="list-style-type: none"> <li>■ Nonoperative 0–3points</li> <li>■ Nonoperative or operative 4–5points</li> <li>■ Operative &gt; 5points</li> </ul>	Treatment algorithm (validated): <ul style="list-style-type: none"> <li>■ Nonoperative 0–3points</li> <li>■ Nonoperative or operative 4–5points</li> <li>■ Operative &gt; 5points</li> </ul>

is often used for fracture classification, but does not include a reliable estimation of prognosis for the determination of the best treatment.<sup>1,2</sup> In 2005, Vaccaro et al initially developed the thoraco-lumbar injury severity score (TLISS),<sup>3</sup> which was slightly modified to the TLICS in 2007.<sup>4</sup> As the name already states, this classification system includes a scoring system based on 3 variables, with subsequent treatment algorithms. Recently, the new AO spine classification has been published, which tries to simplify the comprehensive Magerl/AO classification and incorporates features of both TLICS and Magerl/AO classifications.<sup>5,6</sup> Table 1 shows a description of the Magerl/AO, TLICS, and new AO spine classification systems.

Considering the existence of various classification systems, and the quantity of research that has been done to classify thoracolumbar fractures, it is remarkable that consensus about treatment is still lacking. A validated classification of fractures remains the key to an appropriate treatment algorithm. In an attempt to achieve worldwide consensus in the treatment decision of traumatic thoracolumbar fractures, there is need for a classification that should have 2 important characteristics: (1) it needs to create a worldwide common

language concerning the recognition of injury types (accuracy) and (2) the treatment recommendation by the classification should be highly correlated to the actual treatment (clinical usefulness).

For this reason, we have performed a systemic review with the following research question: “In traumatic thoracolumbar spine fractures, which classification can be used best for treatment decision making?” This question considered participants, intervention, comparisons, outcomes, and study design (PICOS). We looked for participants who were treated for traumatic thoracolumbar spine fractures. We compared the Magerl/AO, new AO spine, and TLIC classifications with respect to treatment decision-making. We defined the outcome parameters, looking for accuracy (expressed in interobserver and intraobserver validity and kappa values) and clinical usefulness (expressed in the sensitivity or specificity of an algorithm); these parameters were the result of a systematic review.

## MATERIALS AND METHODS

A systematic literature review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement

**Table 2.** Quality assessment.

Study	Classification			Number		Study Design	Quality		
	TLICS/ TLISS	AO	New AO	Fractures or Patients	Observer		Patient Selection	Blinding	Level of Evidence
Reliability									
Oner et al <sup>8</sup>		x		60	5	Retrospective	+	+	4
Pishnamaz et al <sup>9</sup>		x		91	12	Case series	+	+	4
Kaul et al <sup>10</sup>	x		x	50	11	Retrospective	+	+	4
Urratia et al <sup>11</sup>			x	70	6	Retrospective	+	+	4
Lenarz et al <sup>12</sup>	x	x		97	3	Retrospective	+	+	4
Cheng et al <sup>13</sup>			x	109	6	Retrospective	+	?	4
Patel et al <sup>14</sup>	x			25	21	Prospective	+	+	3
Wood et al <sup>2</sup>		x		31	19	Retrospective	?	+	4
Whang et al <sup>4</sup>	x			25	5	Prospective	+	+	3
Park et al <sup>15</sup>	x			134	2	Retrospective	+	+	4
Kriek and Govender <sup>16</sup>		x		150	6	Retrospective	+	+	4
Kepler et al <sup>17</sup>			x	25	100	Retrospective, validity	+	+	4
Vaccaro et al <sup>18</sup>	x			71	5	Validity	?	+	4
Azimi et al <sup>19</sup>			x	56	2	Retrospective	?	-	4
Reinhold et al <sup>5</sup>			x	110	5	Validity	+	-	4
Harrop et al <sup>3</sup>	x			56	30	Validity	?	?	4
Vaccaro et al <sup>20</sup>			x	40	9	Validity	+	?	4
Clinical usefulness									
Pishnamaz et al <sup>9</sup>		x		91	12	Case series	+	+	4
Vaccaro et al <sup>18</sup>	x			71	5	Validity	?	+	4
Andrei et al <sup>21</sup>	x			49		Retrospective	+	-	4
Joaquim et al <sup>22</sup>	x			458		Retrospective	+	-	4
Park et al <sup>15</sup>	x			134	2	Retrospective	+	+	4
Whang et al <sup>4</sup>	x			25	5	Prospective	+	+	3
Rajasekaran et al <sup>23</sup>			x	30	41	Retrospective	?	+	4
Harrop et al <sup>3</sup>	x			56	30	Validity	?	?	4

Abbreviations: TLICS indicates thoracolumbar injury classification and severity score; TLISS, thoracolumbar injury severity score.

(PRISMA). The checklist of the PRISMA guidelines are attached in Appendix 1.

### Search

With help from the medical research librarian (M.H.) a comprehensive search of the English literature was conducted using PubMed, Embase, CINAHL, and the Cochrane Database. The literature was searched without any data limitations. Search terms included “classification (mesh),” with subsequent corresponding synonyms (ao spine, ao classification, tlics, tliiss, classification\*, systematics, taxonom\*); and “spinal fractures (mesh)” with subsequent synonyms (spinal fracture\*, spine fracture\*, thoracolumbar fracture\*, thoracic fracture\*, lumbar fracture\*, vertebra fracture\*, vertebral fracture\*). The full search process is shown in Appendix 2.

### Study Selection

All hits (PubMed: 1128, Embase: 2775, CINAHL: 279, Cochrane: 134; in total 4312 hits) were imported to Refworks. Two independent researchers (ICand MS) viewed all references and included

full text papers. In case of a difference of opinion, a third author (PW) was consulted.

Literature was included or excluded based on the following criteria. Inclusion: thoracolumbar fractures; English language; analysis of AO, new AO spine, TLICS or TLISS classification; measurements: intra- and/or interobserver validity (kappa values,) or clinical usefulness expressed in specificity/sensitivity of an algorithm, or any other way the applicability was scored. Exclusion: congress papers, instructional course lectures, reviews, cadaver studies, cervical spine fractures, all other classification methods except those mentioned in the inclusion, children, osteoporotic or other pathological fractures, the expression of the clinical usefulness associated with treatment-related outcome.

The Prisma evidence based medicine checklist for diagnostic articles was used for the qualitative analysis. See Table 2 for the qualitative analysis of the included literature.

### Outcome Parameters

Accuracy is defined as the interobserver and intraobserver reliability of the classification systems. The reliability is expressed in kappa values, which

are commonly used and accepted for the measurement of data collection accuracy.<sup>24</sup> As a summery measure for the kappa coefficients, mean kappa values were used. Some studies showed different kappa values depending on level of expertise and function of the observer. In that case we chose the kappa values represented by the attending spine surgeons, as these were most representative for clinical decision making.

For clinical usefulness, we decide to focus on the applicability of the current classifications. This is quantified by the percentage of agreement between classification recommendation and actual treatment, and shows the correlation between classification recommendations and decision making.

### Statistical Methods

Pooling of data could not be performed, because case cohorts and number of observers were too variable in the included studies. As raw data were not available, it was not possible to perform a meta-analysis of kappa values.

## RESULTS

Twenty-eight articles were selected for full text rating. Four more were selected by cross reference. After full text screening, the following 12 articles were excluded. The full text article of Yacoub et al<sup>25</sup> was not available. One article by Mirza et al<sup>26</sup> was a review of previous literature. Five articles were not applicable to the research question of this review: Salgado et al,<sup>27</sup> Pizonas et al,<sup>28</sup> and Winklhofer et al<sup>29</sup> studied the influence of the MRI on the classification system, instead of analyzing the reliability of the classifications itself; Shen et al<sup>30</sup> investigated the prognostic factors of failure of conservatively treated burst fractures; and Pneumoticos et al<sup>31</sup> compared TLICS 1-3 and TLICS 4 conservatively treated thoracolumbar spine fractures. Five articles contained subgroup analysis of the same cohort published earlier (Joaquim et al 2014<sup>32</sup> presented the same cohort as Joaquim et al 2013<sup>22</sup>; Ratliff et al<sup>33</sup> and Raja Rampersaud et al<sup>34</sup> performed subgroup analysis of the same cohort published by Harrop et al<sup>3</sup>; and Sadiqi et al<sup>35</sup> and Schroeder et al<sup>36</sup> had subgroup analysis of the cohort studied by Kepler et al<sup>17</sup>).

Seventeen papers were eligible for the first subquestion regarding the interobserver and intraobserver reliability. Eight articles could be used for

answering the second question, considering the clinical usefulness of the classification methods. The inclusion and exclusion processes are summarized in the PRISMA flow diagram (Figure 1).

### Reliability

Seventeen articles described the interobserver and/or intraobserver validity of at least one of the classification systems. The data for the interobserver and intraobserver kappa values are shown in Tables 3 and 4 respectively. Regarding the proposed guidelines of Landis and Koch,<sup>37</sup> the kappa values indicated moderate to substantial agreement (kappa > 0.4) for all 3 classification methods.

A wide range of kappa values have been described in literature. These values were influenced by various factors, including the number of observers, cases and options, prevalence, blinding and work-up bias. When taking that into account, all current classifications had acceptable reliability. Kappa values of the new AO spine classification seem slightly better than the Magerl/AO classification. In Figure 2 the mean interobserver kappa values of the total TLICS score were expressed against the number of observers. The larger the number of observers (and cases), the lower the kappa value. The study of Patel et al<sup>14</sup> shows an outlier with a mean interobserver kappa value of 0.51 in 21 observers. This is higher than expected, but could be explained by the fact that the attending observers had been involved in the development of the TLISS system and that they had trained the remaining observers on the use of the system. The kappa values of the validity studies were slightly higher than the kappa values of the independent prospective cohort studies.

### Clinical Usefulness

Eight articles were included with data concerning the clinical usefulness of the classification methods. Pishnamaz et al<sup>9</sup> used the AO classification and illustrated the treatment strategy depending on AO fracture type. There was a large difference between German and Dutch spine surgeons regarding treatment of burst fractures (AO type A3). Whereas in Germany 96.2% of the A3 fractures were treated surgically, in the Netherlands only 41.2% of these burst fractures were operated. They stated that despite the internationally used classification systems, there is insufficient evidence to install a

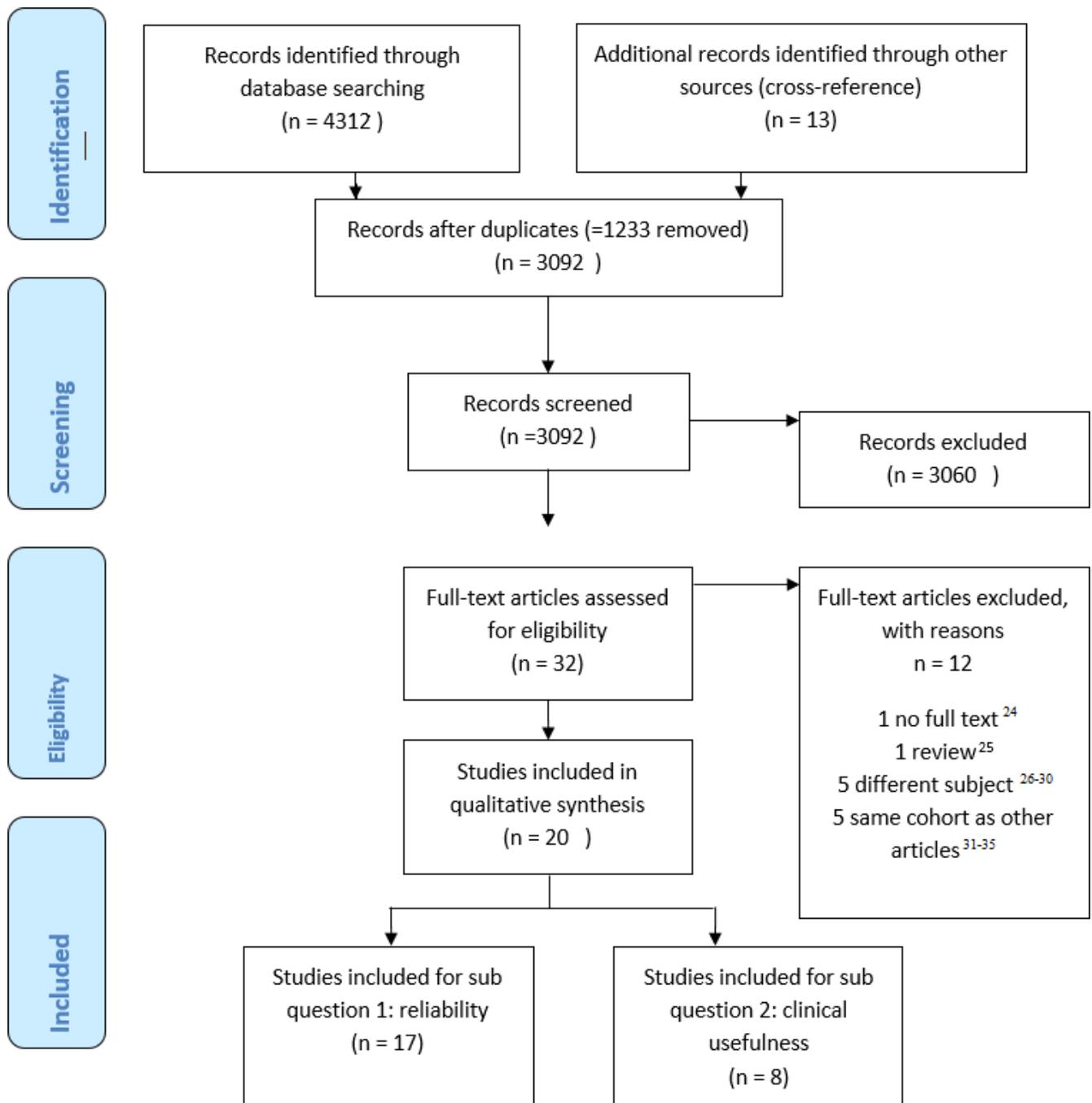


Figure 1.

standard treatment algorithm for fractures of the thoracolumbar spine.

Rajasekaran et al<sup>23</sup> published their results concerning the usefulness of the new AO spine classification in 2016. Forty-one AO Spine members classified 30 sets of images of patients with thoracolumbar spine trauma of varying severity. Cases were assessed independently and the reviewers were asked to answer questions regarding fracture

classification, type of treatment, and need for further investigations. The presented kappa values were not correlated with the observers, but with the diagnostics, as they were measured in plain radiographs, CT, and MRI. Hence, these data could not be included in the aforementioned reliability section. However, they also looked for the decision on fracture management. After the first assessment with plain radiographs, 72% of the patients were

**Table 3.** Overview interobserver kappa values.

Study	Cases (N)	Observers (N)	Diagnostics	Kappa Values			
				AO-ABC	AO-A	AO-B	AO-C
Oner et al <sup>8</sup>	60	5	X, CT, MRI	.31 (.16–.50)	.61 (.47–.86)		
Pishnamaz et al <sup>9</sup>	91	12	X,CT	<b>.45</b>			
Lenarz et al <sup>12</sup>	97	3	X, CT	<b>.71</b>			
Wood et al <sup>2</sup>	31	19	X, CT	<b>.475</b> (.39–.60)			
Kriek and Govender <sup>16</sup>	150	6	X	<b>.403</b>			
New AO spine classification							
Kaul et al <sup>10</sup>	50	11	X, CT, MRI	<b>.59</b> (±.01)			
Vaccaro et al <sup>20</sup>	40	9	X, CT, MRI	<b>.64</b>	.72	.58	.70
Reinhold et al <sup>5</sup>	110	5	X, CT	<b>.77</b>	.81	.71	.81
Cheng et al <sup>13</sup>	109	6	X, CT	<b>.362</b>	.385	.292	.552
Kepler et al <sup>17</sup>	25	100	CT	<b>.74</b>	.80	.68	.72
Urrutia et al <sup>11</sup>	70	6	X, CT	<b>.62</b>	.61	.57	.69
Azimi et al <sup>15</sup>	74	2	X, CT, MRI		.88 (.8–.94)	.86 (.84–.93)	.89 (.84–.94)
<b>TLICS/TLISS Classification</b>				<b>TLICS total</b>	<b>TLICS Mechanism</b>	<b>TLICS PLC</b>	<b>TLICS Neurological State</b>
Kaul et al <sup>10</sup>	50	11	X, CT, MRI	.29 (± .01)	.43 (± .01)	.47 (± .01)	.85 (± .01)
Lenarz et al <sup>12</sup>	97	3	X, CT		.65	.66	.73
Patel et al <sup>14</sup>	25	21	X, CT, MRI	<b>.509</b> (± .006)	.636 (± .04)	.534 (± .049)	
Whang et al <sup>4</sup>	25	5	X, CT, MRI	<b>.455</b>	.626	.447	
Park et al <sup>15</sup>	134	2	X, CT, MRI	<b>.880</b> (± .033)	.966 (± .024)	.858 (± .042)	
Vaccaro et al <sup>18</sup>	71	5	X, CT, MRI	<b>.46</b> (± .03)	.57 (± .04)	.48 (± .04)	.93 (± .02)
Harrop et al <sup>3</sup>	56	30	X, CT, MRI	.2403	.2951	.3359	.935

Abbreviations: X indicates x-ray; CT, computed tomography; MRI, magnetic resonance imaging; TLICS, thoracolumbar injury classification and severity score; TLISS, thoracolumbar injury severity score. Overall kappa values >0.4 are bolded, as these indicate acceptable accuracy.

indicated for surgical treatment. This percentage increased significantly to 81.7% with CT images. Additional MRI, however, did not alter treatment strategy.

Six papers debated the applicability of the treatment algorithm of the TLICS/TLISS system. Vaccaro et al,<sup>18</sup> who proposed the TLISS, showed in their study in 2006 a > 96% agreement on the treatment recommendation of the TLISS within a group of 5 observers scoring 71 clinical cases. Harrop et al,<sup>3</sup> who were also part of the development team for the TLISS classification, published the results of 48 observers who assessed 56 cases. They reported an agreement of > 90% among the surgeons on the preferred management of the fracture and the TLISS-graded management. In 2007, Whang et al<sup>4</sup> presented validity data of the assessment of 25 cases by 5 observers. They distinguished between TLICS and TLISS, but did not report any significant differences between these 2 (almost) equal classifications. A correct prediction was achieved in > 90% of the cases, with a sensitivity of 89% and a specificity of 95%. In 2010, Andrei et al<sup>21</sup> collected the data of a retrospective surgical cohort and presented the safety and applicability of the TLICS system. Forty-nine patients were included. In 47 of the 49 patients (95.9%), the TLICS accurately matched

surgical decision making. There were 2 patients with a TLICS score of 2 points who underwent surgical treatment. Both patients were diagnosed with an L1 burst fracture without neurological injury. Operative treatment was recommended by the surgeon because of concerns about the comminution and the possibility of progressive deformity. In addition to this study, in 2013 Joaquim et al<sup>22</sup> performed an analysis of a large retrospective cohort (N = 458), in which 310 patients were treated conservatively; 99% of these patients had a TLICS < 4. There were 9 failures, defined as patients that received surgical treatment in a second stage. Three missed B-type fractures required surgery because of progressive deformity and severe pain. One patient needed surgery after 6 months because of severe L5 radiculopathy (unknown if this was related to the fracture). Five patients with burst fractures underwent surgery because of persistent pain or progressive kyphosis. Only 2 of these had pain improvement postoperatively. Furthermore, the author stated that of the 125 patients with burst fractures without neurological deficit (TLICS 2), 96% were successfully treated without surgery. The second group consisted of 148 patients who all received surgical treatment. Twenty-four complications (16.2%) were reported, varying from instrumental removal and urinary infection to death (N =

**Table 4.** Overview intraobserver kappa values.

Study	Cases (N)	Observers (N)	Diagnostics	Kappa Values			
				AO-ABC	AO-A	AO-B	AO-C
Oner et al <sup>8</sup>	60	5	X, CT, MRI	.35			
Pishnamaz et al <sup>9</sup>	91	12	X, CT	—			
Lenarz et al <sup>12</sup>	97	3	X, CT	<b>.70</b>			
Wood et al <sup>2</sup>	31	19	X, CT	<b>.63</b>			
Kriek and Govender <sup>16</sup>	150	6	X	.334 (.18–.49)			
<b>New AO Spine Classification</b>							
				<i>AO-ABC</i>	<i>AO-A</i>	<i>AO-B</i>	<i>AO-C</i>
Kaul et al <sup>10</sup>	50	11	X, CT, MRI	<b>.61</b> (± .13)			
Vaccaro et al <sup>20</sup>	40	9	X, CT, MRI	<b>.85</b> (.75–.96)	.72	.43	
Reinhold et al <sup>5</sup>	110	5	X, CT	—			
Cheng et al <sup>13</sup>	109	6	X, CT		.442	.485	.412
Kepler et al <sup>17</sup>	25	100	CT	<b>.81</b> (.32–1.0)	.57	.43	
Urrutia et al <sup>11</sup>	70	6	X, CT	<b>.77</b> (.72–.83)			
Azimi et al <sup>15</sup>	74	2	X, CT, MRI		.84 (.82–.91)	.83 (.81–.88)	.86 (.83–.92)
<b>TLICS/TLISS Classification</b>				<b>TLICS total</b>	<b>TLICS Mech</b>	<b>TLICS PLC</b>	<b>TLICS neu</b>
Kaul et al <sup>10</sup>	50	11	X, CT, MRI	<b>.44</b> (± .10)			
Lenarz et al <sup>12</sup>	97	3	X, CT		.65	.65	.72
Patel et al <sup>14</sup>	25	21	X, CT, MRI	—			
Whang et al <sup>4</sup>	25	5	X, CT, MRI	—			
Park et al <sup>15</sup>	134	2	X, CT, MRI	—			
Vaccaro et al <sup>18</sup>	71	5	X, CT, MRI	.29 (± .02)	.33 (± .03)	.35 (± .03)	.91 (± .02)
Harrop et al <sup>3</sup>	56	30	X, CT, MRI		.429	.478	

Abbreviations: X indicates x-ray; CT, computed tomography; MRI, magnetic resonance imaging; TLICS, thoracolumbar injury classification and severity score; TLISS, thoracolumbar injury severity score; PLC, posterior-ligament complex. Overall kappa values >0.4 are bolded, as these indicate acceptable accuracy.

1). Surgical treatment matched the TLICS recommendation only in 46.6% of the cases. The 53.4% mismatches were all stable burst fractures (TLICS 2). No details about complications or other clinical implications in the subgroup of surgically treated patients with stable burst fractures were described.

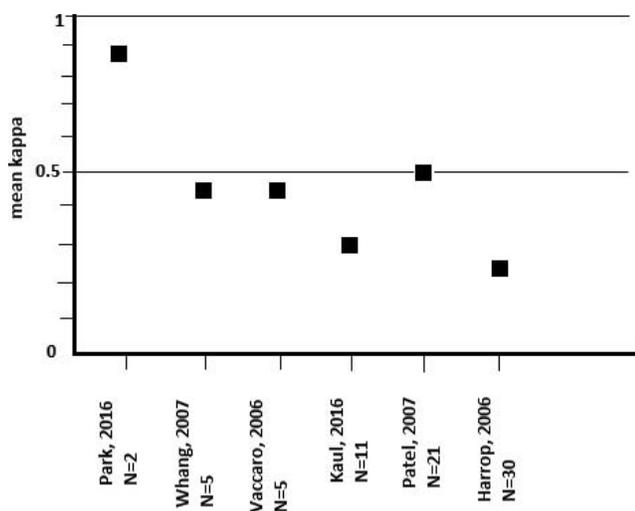
Recently, Park et al<sup>15</sup> described a modified TLICS score, and measured the clinical usefulness of this modified TLICS and the original TLICS

classification. The analysis was performed on 134 fractures, and images were independently interpreted by 2 observers. Thirty-one patients were treated surgically. Two of these patients had a TLICS < 4 (6%) and 58% (n = 18) of the surgically treated patients scored a TLICS 4. Of the 103 conservatively treated patients, only one scored TLICS 5 by both observers.

In summary, literature concerning the clinical usefulness of the classification methods is sparse. Joaquim et al<sup>22</sup> reported the largest series, in which it appeared that over 50% of patients suffering a stable burst fracture (TLICS 2, AO type A3/A4) were surgically treated, and treatment decision was based on other patient- and fracture-related factors, mainly persistent pain and progressive kyphosis.

## DISCUSSION

With the available literature, we would postulate that the accuracy of all 3 reviewed classifications is sufficient for use in clinical practice.<sup>2–5,8–20</sup> Although kappa values are in favor of the TLICS, we also believe that the accuracy of the AO spine classification is sufficient for use in clinical practice. Since these classifications have a different design, and the kappa values are calculated from different



**Figure 2.**

numbers of variables (more fracture morphology options in the AO spine classification), it is difficult to directly compare the kappa values of the TLICS versus the AO. With interobserver kappa values of 0.36–0.77 for the new AO spine, and 0.24–0.88 for TLICS management, they do not all reach the kappa value of  $> 0.55$ , which is necessary for a classification system to be clinically reliable, according to Sanders et al.<sup>38</sup> But in answer to that criterion, Oner et al.<sup>39</sup> stated that this is too stringent for assessing the reliability of a spinal fracture classification system. As kappa values depend on the number of options (fracture types), observers, and cases, one could state that in studies with many observers and many fracture types a kappa value of  $< 0.55$  may be deemed acceptable. Blinding of observers regarding treatment decision and outcome is important to have the lowest risk of bias in kappa values. Therefore, study design is paramount. The absolute kappa values found in the literature should therefore always be seen in relation to the quality of the studies, and numbers of cases and observers.

Although the research question regarding clinical usefulness seems very important to understand which classification can best be used in clinical practice, it is by far the hardest one. Regarding the clinical usefulness, scientific evidence remains poor. TLICS is the only current classification system that contains a point allocation with treatment recommendation in practice.

Very often, treatment decisions are not based on the classification alone. Current literature shows  $> 90\%$  agreement for the quite obvious treatment decision in simple compression fractures (conservative) and clearly unstable B and C type fractures (surgical).<sup>3,4,18,21,22</sup> However, in only 50% of the cases regarding stable burst fractures (AO A3/A4 and TLICS 2), treatment recommendation of the TLICS classification is followed by the surgeons, as shown by Joaquim et al.<sup>22</sup> Despite the evidence considering the safety of conservative treatment<sup>40,41</sup> and well-known negative clinical implications of surgery (eg, complications, limitation in spinal movement), in most patients with stable burst fractures an operation was performed.

In 2016, Bakhsheshian et al.<sup>42</sup> published a review of evidence-based management of stable thoracolumbar burst fractures. They concluded that a high level of evidence demonstrated similar functional outcomes with conservative management when compared with open surgical operative manage-

ment. However, some burst fractures treated conservatively had a poor outcome with progressive kyphosis and persistent pain, which could be the reason for uncertainty in the clinical management of burst fractures.

For an appropriate treatment algorithm, the prognostic factors responsible for worse outcome of these burst fractures should be elucidated, which we will discuss in the following sections.

However, next to prognostic factors, we should be aware that differences in treatment strategies also arise from other causes. These differences probably start with a lack of a worldwide uniform accepted definition of “right” or “optimal” care, as these definitions are mainly opinion based. These opinions are often formed by the surgeon’s culture and skills and institutional possibilities. Additionally, patient preferences and individual risk factors may play an important role.

### The Uncertainties Concerning Posterior-Ligament-Complex Integrity and Burst Fractures

Burst fractures and the posterior-ligament-complex (PLC) integrity remain the most important uncertainties. Clarification regarding these parameters would improve uniform decision. Schnake et al.<sup>43</sup> and Leferink et al.<sup>44</sup> showed in their studies that worse outcomes may be due to the fact that burst fractures are actually missed B-type fractures. The difference between burst and B-type fractures relies on the integrity of the PLC. So, reliable assessment of the PLC integrity is crucial in these cases.

Unfortunately, evidence about the role of standard MRI in addition to plain radiographs and CT is contradictory. Oner et al.,<sup>8</sup> Salgado et al.<sup>27</sup>, Pizones et al.,<sup>28</sup> and Winklhofer et al.<sup>29</sup> presented results in which MRI seemed to improve reliability and influence treatment strategies compared to CT. But Rajasekaran et al.<sup>23</sup> performed a study with similar sensitivity for CT and MRI, and reported no change in treatment decision after additional MRI. This would indicate that plain radiographs and CT suffice for classification and treatment decision. Literature regarding the reliability of MRI in PLC status resulted in fair to moderate kappa values ( $\text{kappa} \pm .4$ ),<sup>45</sup> and demonstrated relatively high negative predictive values and relatively low positive predictive values for PLC injuries.<sup>46</sup>

Despite the controversial evidence regarding MRI and PLC, agreement on the PLC status is important, especially in burst fractures. We suggest

in burst fractures routine MRI could be of additional value. Without any edema in the PLC on MRI, the integrity is established. But without an additional MRI, it is probably safer to value PLC as undetermined in most burst fractures, leading to recommendation of surgical treatment.

### Prognostic Patient- and Fracture-Related Parameters

In addition to the PLC status, anterior comminution remains an important risk factor for worse outcome in burst fractures, although definite evidence concerning its role is still lacking. Recently, Spiegl et al<sup>47</sup> discussed a key role for the intervertebral disc in determining the long-term clinical and radiological outcome of burst fractures. Incorporation of the intervertebral disc pathology into the existing classification systems might be a valuable prognostic factor. In addition to these factors, previous studies also stated that several other parameters might influence outcome in thoracolumbar fractures. Shen et al<sup>30</sup> published results of a radiological and binary logistic regression analysis. They showed that visual analog scale pain scores and interpedicular distance could be significant risk factors for failure of nonoperative treatment of burst fractures. Furthermore, lower bone quality and bone regeneration (eg, osteoporosis), higher age, and fracture localization at the thoracolumbar junction seem to be responsible for worse radiological outcome.<sup>30,48</sup>

Clinical usefulness of the current classifications is still limited as outcome is influenced by the above-mentioned patient- and fracture-related parameters. Including these parameters in future classification systems may enhance prognostic value and thus clinical usefulness of such classifications.

As an extension of the new AO spine classification, Vaccaro et al<sup>20</sup> recently introduced the thoracolumbar AO spine injury score. This score contains a treatment algorithm, not only based on the classification of the fracture morphology, but including a point allocation for neurological status and patient-specific modifiers (eg, PLC status and ankylosing spondylitis). In 2016, Kepler et al<sup>49</sup> and Vaccaro et al<sup>50</sup> presented a validation study of this AO spine injury score. In these studies, input from surgeons worldwide was used to determine the initial treatment recommendation.

In this respect, with the addition of patient-specific modifiers, the thoracolumbar AO spine

injury score shows insight that other patient- and fracture-related parameters are important in the search for a worldwide applicable and accepted classification and a treatment algorithm for thoracolumbar spine fractures.

## CONCLUSION

Current TLICS and new AO spine classifications have acceptable accuracy regarding their reproducibility, but are limited in clinical usefulness since the treatment recommendation is not always implemented in clinical practice, mainly in burst fractures. Differences in treatment decision making arise from several causes, such as surgeon and patient preferences, culture, and prognostic factors that are not included in classifications yet.

The recently validated thoracolumbar AO spine injury score including patient-specific modifiers seems promising for use in clinical practice. However, we would suggest further evaluation of the clinical usefulness of this score, and consider adding more relevant parameters associated with a worse outcome.

## REFERENCES

1. Magerl F, Aebi M, Gertzbein SD, Harms J, Nazarian S. A comprehensive classification of thoracic and lumbar injuries. *Eur Spine J.* 1994;3(4):184–201.
2. 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 Jul;87(7):1423–1429.
3. Harrop JS, Vaccaro AR, Hurlbert RJ, et al. Intrarater and interrater reliability and validity in the assessment of the mechanism of injury and integrity of the posterior ligamentous complex: a novel injury severity scoring system for thoracolumbar injuries. *J Neurosurg Spine.* 2006;4(2):118–122.
4. 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(7):791–795.
5. Reinhold M, Audigé 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(10):2184–201. doi:10.1007/s00586-013-2738-0
6. Aebi M. AO spine classification system for thoracolumbar fractures. *Eur Spine J.* 2013;22(10):2147–2148.
7. Moher D, Liberati A, Tetzlaff J, et al. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7):e1000097. doi:10.1371/journal.pmed1000097
8. Oner FC, Ramos LM, Simmermacher R, 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(3):235–245.
9. Pishnamaz M, Curfs I, Balosu S, et al. Two-nation comparison of classification and treatment of thoracolumbar fractures: an internet-based multicenter study among spine surgeons. *Spine (Phila Pa 1976)*. 2015;40(22):1749–1756.
  10. Kaul R, Chhabra HS, Vaccaro AR, et al. Reliability assessment of AOSpine thoracolumbar spine injury classification system and Thoracolumbar Injury Classification and Severity Score (TLICS) for thoracolumbar spine injuries: results of a multicentre study. *Eur Spine J*. 2017;26(5):1470–1476.
  11. 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(1):E54–E58. doi:10.1097/BRS.0000000000000656
  12. Lenarz CJ, Place HM, Lenke LG, Alander DH, Oliver D. Comparative reliability of 3 thoracolumbar fracture classification systems. *J Spinal Disord Tech*. 2009;22(6):422–427.
  13. Cheng J, Liu P, Sun D, Qin T, Ma Z, Liu J. Reliability and reproducibility analysis of the AOSpine thoracolumbar spine injury classification system by Chinese spinal surgeons. *Eur Spine J*. 2017;26(5):1477–1482.
  14. Patel AA, Whang PG, Brodke DS, et al. Evaluation of two novel thoracolumbar trauma classification systems. *Indian J Orthop*. 2007;41(4):322–326.
  15. Park HJ, Lee SY, Park NH, et al. Modified thoracolumbar injury classification and severity score (TLICS) and its clinical usefulness. *Acta Radiol*. 2016;57(1):74–81.
  16. Kriek JJ, Govender S. AO-classification of thoracic and lumbar fractures—reproducibility utilizing radiographs and clinical information. *Eur Spine J*. 2006;15(8):1239–1246.
  17. Kepler CK, Vaccaro AR, Koerner JD, et al. Reliability analysis of the AOSpine thoracolumbar spine injury classification system by a worldwide group of naïve spinal surgeons. *Eur Spine J*. 2016;25(4):1082–1086.
  18. 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(suppl 11):S62–S69; discussion S104.
  19. Azimi P, Mohammadi HR, Azhari S, Alizadeh P, Montazeri A. The AOSpine thoracolumbar spine injury classification system: a reliability and agreement study. *Asian J Neurosurg*. 2015;10(4):282–285.
  20. 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(23):2028–2037.
  21. Andrei F, Joaquim AF, Yvens B, et al. Evaluation of the thoracolumbar injury classification system in thoracic and lumbar spinal trauma. *Spine*. 2011;36(1):33–36.
  22. 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;23(12):1760–1765.
  23. Rajasekaran S, Vaccaro AR, Kanna RM, et al. The value of CT and MRI in the classification and surgical decision-making among spine surgeons in thoracolumbar spinal injuries. *Eur Spine J*. 2017;26(5):1463–1469.
  24. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–282.
  25. Yacoub AR, Joaquim AF, Ghizoni E, Tedeschi H, Batista UC, Patel AA. Evaluation of the safety and reliability of the newly-proposed AO spine injury classification system. *J Spinal Cord Med*. 2017;40(1):70–75.
  26. 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(5):364–377.
  27. Salgado Á, Pizones J, Sánchez-Mariscal F, Álvarez P, Zúñiga L, Izquierdo E. MRI reliability in classifying thoracolumbar fractures according to AO classification. *Orthopedics*. 2013;36(1):e75–e78. doi:10.3928/01477447-20121217-22
  28. Pizones J, Izquierdo E, Alvarez P, et al. Impact of magnetic resonance imaging on decision making for thoracolumbar traumatic fracture diagnosis and treatment. *Eur Spine J*. 2011;20(suppl 3):390–396.
  29. Winklhofer S, Thekkumthala-Sommer M, Schmidt D, et al. Magnetic resonance imaging frequently changes classification of acute traumatic thoracolumbar spine injuries. *Skeletal Radiol*. 2013;42(6):779–786.
  30. Shen J, Xu L, Zhang B, Hu Z. Risk factors for the failure of spinal burst fractures treated conservatively according to the Thoracolumbar Injury Classification and Severity Score (TLICS): a retrospective cohort trial. *PLoS One*. 2015;10(8):e0135735. doi:10.1371/journal.pone.0135735
  31. Pneumoticos, Karampinas PK, Triantafilopoulos G, Koufos S, Polyzois V, Vlamis J. Evaluation of TLICS for thoracolumbar fractures. *Eur Spine J*. 2016;25(4):1123–1127.
  32. Joaquim AF, Lawrence B, Daubs M, Brodke D, Tedeschi H. Measuring the impact of the Thoracolumbar Injury Classification and Severity Score among 458 consecutively treated patients. *J Spinal Cord Med*. 2014;37(1):101–106.
  33. Ratliff J, Anand N, Vaccaro AR, et al. Regional variability in use of a novel assessment of thoracolumbar spine fractures: United States versus international surgeons. *World J Emerg Surg*. 2007;2:24. doi:10.1186/1749-7922-2-24
  34. Raja Rampersaud Y, Fisher C, Wilsey J, et al. Agreement between orthopedic surgeons and neurosurgeons regarding a new algorithm for the treatment of thoracolumbar injuries: a multicenter reliability study. *J Spinal Disord Tech*. 2006;19(7):477–482. doi:10.1097/01.bsd.0000211237.14211.21
  35. Sadiqi S, Oner FC, Dvorak MF, et al. The influence of spine surgeons; experience on the classification and intra-observer reliability of the novel AOSpine thoracolumbar spine injury classification system—an international study. *Spine (Phila Pa 1976)*. 2015;40(23):E1250–E1256. doi:10.1097/BRS.0000000000001042
  36. Schroeder GD, Kepler CK, Koerner JD, et al. Is there a regional difference in morphology interpretation of A3 and A4 fractures among different cultures? *J Neurosurg Spine*. 2016;24(2):332–339. doi:10.3171/2015.4.SPINE1584
  37. Landis JR, Koch GC. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
  38. Sanders RW. Editorial. The problem with apples and oranges. *J Orthop Trauma*. 1997;11(7):465–466.
  39. Oner FC, van Gils AP, Dhert WJ, Verbout AJ. MRI findings of thoracolumbar spine fractures: a categorization based on MRI examinations of 100 fractures. *Skeletal Radiol*. 1999;28(8):433–443.
  40. Weninger P, Schultz A, Hertz H. Conservative manage-

ment of thoracolumbar and lumbar spine compression and burst fractures: functional and radiographic outcomes in 136 cases treated by closed reduction and casting. *Arch Orthop Trauma Surg.* 2009;129(2):207–219.

41. Abudou M, Chen X, Kong X, et al. Surgical versus non-surgical treatment for thoracolumbar burst fractures without neurological deficit. *Cochrane Database Syst Rev.* 2013. doi:10.1002/14651858.CD005079.pub3

42. Bakhsheshian J, Dahdaleh NS, Fakurnejad S, Scheer JK, Smith ZA. Evidence-based management of traumatic thoracolumbar burst fractures: a systematic review of nonoperative management. *Neurosurg Focus.* 2014;37(1):E1. doi:10.3171/2014.4.FOCUS14159

43. Schnake KJ, von Scotti F, Haas NP, Kandziora F. Unfallchirurg. Type B injuries of the thoracolumbar spine: misinterpretations of the integrity of the posterior ligament complex using radiologic diagnostics. *Unfallchirurg.* 2008;111(12):977–984.

44. 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 unrecognised. *Eur Spine J.* 2002;11(3):246–250.

45. Lee GY, Lee JW, Choi SW, et al. MRI inter-reader and intra-reader reliabilities for assessing injury morphology and posterior ligamentous complex integrity of the spine according to the thoracolumbar injury classification system and severity score. *Korean J Radiol.* 2015;16(4):889–898.

46. van Middendorp J, Patel A, Schuetz M, Joaquim AF. 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.* 2013;22(3):461–474.

47. Spiegl UJ, Josten C, Devitt BM, Heyde CE. Incomplete

burst fractures of the thoracolumbar spine: a review of literature. *Eur Spine J.* 2017;26(12):3187–3198. doi:10.1007/s00586-017-5126-3

48. Curfs I, Grimm B, van der Linde M, Willems P, van Hemert W. Radiological prediction of posttraumatic kyphosis after thoracolumbar fracture. *Open Orthop J.* 2016;10:135–142. doi:10.2174/1874325001610010135

49. Kepler CK, Vaccaro AR, Schroeder GD, et al. The thoracolumbar AOSpine injury score. *Global Spine J.* 2016;6(4):329–334.

50. Vaccaro AR, Schroeder GD, Kepler CK, et al. The surgical algorithm for the AOSpine thoracolumbar spine injury classification system. *Eur Spine J.* 2016;25(4):1087–1094.

**Disclosures and COI:** The authors have no disclosures, and did not receive any funding. The manuscript submitted does not contain information about medical devices or drugs.

**Corresponding Author:** I. Curfs, MD, Department of Orthopedic Surgery and Traumatology, Zuyderland MC Heerlen, Netherlands. Phone: +31 (088) 459 8240. Email: i.curfs@zuyderland.nl.

Published 0 Month 2020

This manuscript is generously published free of charge by ISASS, the International Society for the Advancement of Spine Surgery. Copyright © 2020 ISASS. To see more or order reprints or permissions, see <http://ijssurgery.com>.

APPENDIX 1. PRISMA GUIDELINES CHECKLIST<sup>a,b</sup>

Section and Topic	No.	Checklist Item	Page No.
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Title page
Abstract			
Structured summary	2	Provide a structured summary including the following, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2–3
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (eg, web address), and, if available, provide registration information including registration number.	NA
Eligibility criteria	6	Specify study characteristics (eg, PICOS, length of follow-up) and report characteristics (eg, years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (eg, databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	6; Appendix 1
Study selection	9	State the process for selecting studies (ie, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (eg, piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (eg, PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	6; Table 2
Summary measures	13	State the principal summary measures (eg, risk ratio, difference in means).	7
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (eg, $I^2$ ) for each meta-analysis.	NA
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (eg, publication bias, selective reporting within studies).	6
Additional analyses	16	Describe methods of additional analyses (eg, sensitivity or subgroup analyses, meta-regression), if done, indicating which were prespecified.	7
Results			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8
Study characteristics	18	For each study, present characteristics for which data were extracted (eg, study size, PICOS, follow-up period) and provide the citations.	Tables 3, 4
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	9; Table 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Summary; 9, forest plot NA
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see item 15).	8
Additional analysis	23	Give results of additional analyses, if done (eg, sensitivity or subgroup analyses, meta-regression [see item 16]).	NA
Discussion			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (eg, healthcare providers, users, and policy makers).	13
Limitations	25	Discuss limitations at study and outcome level (eg, risk of bias), and at review-level (eg, incomplete retrieval of identified research, reporting bias).	14–16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	17
Funding	27	Describe sources of funding for the systematic review and other support (eg, supply of data); role of funders for the systematic review.	NA

Abbreviations: PICOS indicates participants, interventions, comparisons, outcomes, and study design; NA, not applicable.

<sup>a</sup>From: Moher D<sup>7</sup>

## APPENDIX 2. SEARCH PROCESS

Searches were conducted March 16, 2017, on PubMed, Embase, Cochrane Database, and CINAHL.  
PubMed: 1128 hits. See Table A1.

**Table A1.** . List of search strings and hits.

Search No.	Query	Items found
15	Search no. 10 AND no. 14	1128
14	Search no. 12 OR no. 13	17 811
13	Search spinal fracture*[tiab] OR spine fracture*[tiab] OR thoracolumbar fracture*[tiab] OR thoracic fracture*[tiab] OR lumbar fracture*[tiab] OR vertebra fracture*[tiab] OR vertebral fracture*[tiab]	10 664
12	Search "Spinal Fractures"[Mesh]	12 457
10	Search no. 7 OR no. 8 OR no. 9	818 927
9	Search ao spine[tiab] OR TLICS[tiab] OR ao classification [tiab] OR TLISS[tiab]	808
8	Search Classification*[tiab] OR Systematics[tiab] OR Taxonom*[tiab]	303 573
7	Search "Classification"[Mesh] OR "classification" [Subheading]	606 407

Embase: 2775 hits.

Search No.	Query	Items found
#9	#3 AND #8	2,775
#8	#4 OR #6 OR #7	1,690,247
#7	'ao spine'.ab,i OR 'ao classification'.ab,i OR 'tlcs'.ab,i OR 'tlss'.ab,i	808
#6	'classification'.ab,i OR 'classification tree'.ab,i OR 'systematics'.ab,i OR 'taxonomy'.ab,i	304,450
#4	'classification'/exp	1,515,199
#3	#1 OR #2	31,405
#2	'spine fracture'/exp	28,517
#1	'spine fracture'.ab,i OR 'spinal fracture'.ab,i OR 'vertebra fracture'.ab,i OR 'vertebral fracture'.ab,i OR 'lumbar fracture'.ab,i OR 'thoracic fracture'.ab,i OR 'thoracolumbar fracture'.ab,i	14,382

Cochrane: 134 hits.

The screenshot shows the Cochrane Library search interface. At the top, there's a navigation bar with 'Wiley Online Library' and 'Cochrane Library' logos. Below that, there are tabs for 'Search', 'Search Manager', 'Medical Terms (MeSH)', and 'Browse'. The main search area contains a search strategy with five numbered steps:

- #1: spinal or spine or vertebra or vertebral or thoracolumbar or thoracic or lumbar;ab,iw (Word variations have been searched) - 33959
- #2: fracture\* - 14769
- #3: classification\* or systematic\* or taxonom\* or ao spine\* or ao classification\* or TLICS or TLISS - 20855
- #4: #1 and #2 - 3302
- #5: #3 and #4 - 134

Below the strategy, there are options to 'Clear Strategy', 'Save Strategy', and 'Search Help'. The search results section shows 'All Results (134)' and a list of filters including 'Cochrane Reviews (66)', 'Review', 'Protocol', 'Other Reviews (4)', 'Trials (63)', 'Methods Studies (0)', 'Technology Assessments (0)', and 'Economic Evaluations (1)'. A specific result is highlighted: 'Multicenter retrospective evaluation of the validity of the Thoracolumbar Injury Classification and Severity Score system in children. Journal of Neurosurgery, Pediatrics. 18 (2) (pp 164-170), 2016. Date of Publication: August 2016, 2016. Publication Year: 2016'.

CINAHL: 279 hits.

a	Search ID#	Search Terms	Search Options	b	Actions
<input type="checkbox"/>	S7	S3 AND S6	Search modes - Boolean/Phrase	<input type="checkbox"/>	<a href="#">View Results (279)</a>
<input type="checkbox"/>	S6	S4 OR S5	Search modes - Boolean/Phrase	<input type="checkbox"/>	<a href="#">View Results (106,601)</a>
<input type="checkbox"/>	S5	TI ( (classification* OR systematic* OR taxonom* OR ao spine* OR ao classification* OR TLICS OR TLISS) OR AB ( (classification* OR systematic* OR taxonom* OR ao spine* OR ao classification* OR TLICS OR TLISS) )	Search modes - Boolean/Phrase	<input type="checkbox"/>	<a href="#">View Results (83,746)</a>
<input type="checkbox"/>	S4	MH "Classification+"	Search modes - Boolean/Phrase	<input type="checkbox"/>	<a href="#">View Results (26,980)</a>
<input type="checkbox"/>	S3	S1 OR S2	Search modes - Boolean/Phrase	<input type="checkbox"/>	<a href="#">View Results (4,798)</a>
<input type="checkbox"/>	S2	TI ( ("spinal fracture*" OR "spine fracture*" OR "vertebra fracture*" OR "vertebral fracture*" OR "thoracolumbar fracture*" OR "thoracic fracture*" OR "lumbar fracture*") ) OR AB ( ("spinal fracture*" OR "spine fracture*" OR "vertebra fracture*" OR "vertebral fracture*" OR "thoracolumbar fracture*" OR "thoracic fracture*" OR "lumbar fracture*") )	Search modes - Boolean/Phrase	<input type="checkbox"/>	<a href="#">View Results (2,194)</a>
<input type="checkbox"/>	S1	MH spinal fractures+	Search modes - Boolean/Phrase	<input type="checkbox"/>	<a href="#">View Results (3,721)</a>