

Research article

Evaluation of the safety and reliability of the newly-proposed AO spine injury classification system

Alexandre RD Yacoub¹, Andrei F. Joaquim¹, Enrico Ghizoni¹, Helder Tedeschi¹, Alpesh A. Patel²

¹Department of Neurology, Neurosurgery Division, University of Campinas (UNICAMP), Campinas-SP, Brazil,

²Department of Orthopedics, Northwestern University, Chicago, IL, USA

Objective: To evaluate the safety and reliability of the new AO Classification, a recent classification system for Thoraco-Lumbar Spine Trauma (TLST).

Design: Retrospective study.

Methods: We applied the new AO system in patients with TLST treated according to the TLICS. Two researchers classified injuries independently. Eight weeks later, the classification was repeated for intra and inter-observer agreement evaluation. To evaluate safety, we correlated the treatment performed based on the TLICS with the newer AO classification obtained.

Results: Fifty-four patients were included in this study, with a mean follow-up of 363.8 days. Twenty-three neurologically intact patients were initially treated conservatively. Their mean TLICS was 1.78 (1–4 points). Four patients underwent late surgery. Thirty-one patients were treated surgically. Their average TLICS was 7.22 points (4–10 points). Agreements in the four independent evaluations according to AO groups and subgroups were of 64.8% (35/54) and 55.5% (30/54) respectively. Kappa index for groups A, B and C was 0.75, 0.7 and 0.85 respectively. Kappa index for subgroups ranged from 0.16 to 0.85. Regarding safety, thirty (57.6%) patients with total subgroups agreement were analyzed. All patients with fracture in groups B and C underwent surgical treatment and patients in group A received surgery according to neurological status or failure of conservative treatment.

Conclusion: The newer AO spine classification demonstrated good reliability at the level of groups. Subgroups demonstrated worse and varying reliability. Although the safety analysis was limited due to the low level of total concordance among all evaluations, patients from group A can be treated conservatively or surgically, whereas those from groups B and C are treated surgically.

Keywords: Trauma, Classification, Thoracic, Lumbar, TLICS, AO spine

Introduction

Thoracolumbar spine trauma (TLST) is the most common site of spinal fractures, with a high morbidity rate and an important social and economic impact. Classification of TLST is important to compare treatment modalities as well as to evaluate patient outcomes.^{1–4} Throughout the years, many classification systems have been proposed to characterize different injuries as well as to guide the subsequent treatment to be chosen for each injury pattern.^{5–7} The most commonly described systems are “three-column” anatomical

view of Denis *et al.*, the mechanical classification of Magerl *et al.* (with more than 50 subtypes) and, more recently the Thoracolumbar Injury Classification System and Severity Score (TLICS) proposed by Vaccaro *et al.*^{1,7,8}

Today, the TLICS is the most widely-accepted system to guide treatment with evidence of its clinical safety in many clinical papers.^{3,4,9} However, the TLICS proposal is not based on a detailed morphological characterization of the spinal injuries. Based on this, the AO Spine Study Group recently published the new AO Spine Classification (new AO System) trying to solve these potential shortcomings.^{10,11} This new morphological classification divided injuries in three types:

Correspondence to: Andrei F. Joaquim, Rua Antônio Lapa 280, S 506, Cambuí, Campinas-SP, Brazil Zip 13025-240. Email: andjoaquim@yahoo.com

- A. anterior compression fractures;
- B. injuries within failure of posterior and/ or anterior tension band;
- C. displacement/dislocation.

Injuries in group A have an additional four subtypes, whereas in group B, three subtypes are included for a total of 8 different potential classifications (A1-4, B1-3, C).

Based on the potential benefits of the newer AO Spine classification system in the management of TLST, evaluation of its safety and reliability are necessary prior to its widespread adoption. The purpose of our study is to perform a retrospective evaluation of the reliability and the validity of the new AO system.

Materials and methods

We applied the new AO classification system in patients with acute TLST treated at our institution from January 2011 to February 2014 after Institutional Review Board approval. These patients were prospectively treated based on the TLICS system as part of our standard protocol, published previously.³ Patients with a score of three or fewer points generally received conservative treatment whereas patients with 4 or more points were surgically treated. Patients referred to surgical treatment were all operated by a posterior approach with fixation, decompression, reduction and realignment when necessary by the same surgeon. All the patients had a complete CT scan with reconstructions in our own hospital with the same multi-slice 64-channel device. Patients with CT scan performed in other facilities were not included. Patients with pathological fractures, such as osteoporosis, infection or tumor were excluded. Clinical data included age, gender, the TLICS score, and neurological status (in the admission and in the last hospital chart consultation). The neurological status was evaluated with the American Spine Injury Impairment System (AIS) and classified in AIS A (complete neurological deficits), B, C and D (incomplete neurological deficits) and E (neurologically intact). Ethical approval was obtained (protocol number 32917414.5.0000.5404).

Reliability

To evaluate reliability, injuries were classified according to the new AO system independently by two researchers—a board certified spine surgeon (AFJ) and a senior level neurosurgery resident (ARDY). The classification was performed again blinded after 8 weeks by both surgeons to assess intra and inter-observer agreement evaluation. The agreement rate was assessed using the Kappa coefficient and classified as poor, weak, moderate, good or almost perfect (Table 1).^{12,13}

Table 1 Kappa index and agreement characteristic

Agreement rate	Classification
<0.2	Poor
0.21–0.40	Weak
0.41–0.60	Moderate
0.61–0.80	Good
0.81–1.0	Almost Perfect

Safety

To evaluate clinical validity, we compared the treatment performed (conservative versus surgical) based on the TLICS score with that obtained from the new AO classification. Of note, only patients with total agreement in the new AO system classification between both surgeons were considered to assess safety.

Results

Clinical results

A total of 54 patients were included in this study. There were 40 (74%) men and 14 (26%) women. Mean age was 34.9 (range from 16 to 77 years-old). The mean follow-up of all the patients was 363.8 days (range from 40 to 991).

Twenty-three neurologically intact (AIS E) patients were initially treated conservatively. The mean TLICS in this group was 1.78 (range from 1 to 4 points). There were four patients who underwent late surgery. Two patients with burst fractures without neurological deficits underwent surgery for back pain. Another patient with a Chance fracture (bipedicular fracture) (TLICS of 4 points) had surgery as he also developed back pain and local kyphosis during the follow-up. Although our protocol includes surgical treatment for patients with TLICS of 4 points, this patient preferred an attempt of conservative treatment prior to early surgery. Lastly, the fourth patient had a severe systemic trauma and visceral injuries, with a mild compression fracture at L3. She was neurologically intact and initially classified as a TLICS score of 1 point. Two months later her acute event, after ambulation, she developed bilateral radicular pain and a listhesis between L3–4. This patient underwent an instrumented lumbar fusion with complete symptomatic relief. Her injury should have been correctly classified as a distractive one with PLC injury and total new TLICS score of 9 points (four points for distraction, three points for PLC injury and two additional points for radiculopathy).

Thirty-one patients were referred for initial surgical treatment. At the hospital admission, there were 13 patients with AIS E, 2 with AIS D, 6 with AIS C, and 10 with AIS A. In the last follow-up assessment, there

Table 2 Inter rater agreement rate and kappa index for major groups

Category	Agreement	Kappa
A	0.87	0.75
B	0.89	0.7
C	0.94	0.85

were 17 patients with AIS E, 3 with AIS D, 1 with AIS C, 1 with AIS B and 9 with AIS A. The average TLICS was 7.22 points, ranging from 4 to 10 points.

There was no neurological worsening in this series. One patient died after surgery due to a cerebral edema for traumatic brain injury. Minor complications included a revision surgery for a cerebral spinal fluid (CSF) leak, two wounds requiring debridement for infection, two surgery revisions for screw relocation, one lumbar external drain catheter for a thoracic CSF leak and two deep venous thromboses.

Reliability evaluation

Intra rater reliability according to major groups—A, B and C

Both researches achieved exactly 88% (48 of 54) agreement in overall group classification. Table 2 details the kappa index and the agreement for each major group (A, B or C). The kappa indexes for these groups were respectively, 0.75, 0.7 and 0.85.

Intra rater reliability according to sub-group classification (A1-4, B1-3, C)

The intra-rater agreement was of 85% (46 of 54) by one of the surgeons (ARDY) and 75% (41 of 54) by the other (AFJ).

Inter rater reliability according to major groups—A, B and C

Total agreement in the four independently evaluations according to groups were of 64.8% (35 of 54). Table 2 shows the inter rater agreement rate and kappa index for the major groups/categories.

Inter rater reliability according to sub-groups—A, B and C

Total agreement in the four independently evaluations according to sub-groups were 55.5% (30 of 54). Table 3 shows the inter rater agreement rate and kappa index for the sub-groups/subcategories.

Safety analyses

From the 54 patients' eligible for this study, we analyzed only the 30 cases with total subgroups agreement in all evaluations. Table 4 demonstrated the treatment according to fracture's subgroups and Table 5 demonstrated

Table 3 Inter rater agreement rate and kappa index for the sub-groups/subcategories

Sub-category	Agreement	Kappa
A1	0.9	0.6
A2	0	0
A3	0.9	0.33
A4	0.87	0.59
B1	0.9	0.16
B2	0.85	0.48
B3	0.98	0.77
C	0.94	0.85

Table 4 Treatment performed according to fracture's subgroups

	Total	Initial Conservative Treatment	Initial Surgical Treatment
A1	4	4*	—
A2	—	—	—
A3	1	1	—
A4	7	3	4
B1	1	1 ⁺	—
B2	4	—	4
B3	—	—	—
C	13	—	13

*This patient with an initial A1 fractures was latter recognizing as having a PLC injury misdiagnosed with segmental listhesis and underwent late surgery.

⁺This patient with a B1 injury asked for trying for conservative treatment. Without success, he underwent late surgical treatment of low back pain and segmental kyphosis.

Patients with A1 and A3 classification underwent conservative treatment. Some of the patients with A4 classification had surgery (4 of 7 cases and all the patients with B or C morphology had surgery at some point during the follow-up).

the clinical data and TLICS of the 30 patients with total subgroup concordance in all evaluations.

From the 30 cases with perfect sub-group classification, four (13%) had A1 injuries, one (3.3%) A3 and seven A4 (23.3%).

One patient had a B1 (3.3%) and four (13%) had a B2. Thirteen cases (43.3%) of the 30 who had total reproducibility regarding AO subgroup were from group C.

With regards to the treatment received, all the four patients with A1 fractures where initially treated conservatively. However, as explained previously, one of these patients developed a L3-4 listhesis on follow up, after starting ambulation, as was reclassified as a B2 injury. This patient received late surgical treatment and was totally free of symptoms in the last follow-up. There were one A3 and three A4 fractures treated conservatively. All of them were neurologically intact.

One patient with a Chance fracture was in this group with total concordance between the evaluators: a B1

Table 5 Clinical data and TLICS score of the 30 patients with total subgroup concordance in all evaluations

	Sex	AIS Before	TLICS Score	Treatment performed	AIS last follow up	Spinal Level (fractures of group A) and Upper Level (fractures of groups B and C)	AO Classification
1	M	E	2	C	E	L3	A4
2	M	E	2	C	E	T6	A1
3	M	E	4	S	E	T9	B1
4	M	E	2	C	E	L1	A4
5	F	E	9	S	E	L3	B2
6	M	E	1	C	E	L3	A1
7	M	E	1	C	E	L1	A3
8	M	E	1	C	E	L5	A1
9	F	E	1	C	E	L1	A1
10	M	E	2	C	E	L1	A4
11	M	C	10	S	E	T9	B2
12	M	A	8	S	A	T9	C
13	M	E	7	S	E	T6	B2
14	M	A	4	S	A	T10	C
15	F	A	8	S	A	T11	C
16	F	A	9	S	A	T12	C
17	M	C	10	S	C	T11	C
18	M	A	10	S	A	T5	C
19	F	E	4	S	D	T5	A4
20	M	E	6	S	E	L3	C
21	M	E	7	S	E	T8	B2
22	M	C	10	S	D	T12	C
23	M	C	5	S	E	L3	A4
24	M	E	4	S	E	T2	A4
25	F	A	9	S	A	T6	C
26	M	A	9	S	B	T4	C
27	M	D	5	S	E	L1	A4
28	M	D	7	S	D	T11	C
29	M	A	9	S	A	T6	C
30	M	A	8	S	A	T2	C

injury. This patient also had failure in conservative treatment with a TLSO orthosis. His total TLICS score was of four points—four for morphology, zero for PLC injury (bone injury) and zero for neurological status (AIS E). This patient developed back pain and mild kyphosis underwent a posterior instrumented fusion, with complete relief of symptoms. Four patients with B2 injuries underwent early surgical treatment.

Discussion

Spinal trauma classification is essential to compare treatments and evaluate patient outcomes.¹⁴ However, the classification must have a good reliability and also be clinically relevant to be used. The new AO system is the newest attempt of the spinal community to improve the quality of thoracolumbar spine classification and consequently its treatment, addressing many potential shortcomings of the previous systems, which had been criticized for their limited reproducibility.^{15–17}

This classification system is derived from the Magerl *et al.* system and also influenced by the Thoracolumbar Injury Classification System (TLICS).^{6,7} The relationship with the TLICS is based on the inclusion of an osteo-ligamentous injury

grading. This was an attempt to address the shortcomings of the previous system. For example, the controversial evaluation of the status of the Posterior Ligament Complex (PLC), are instead integrated into the A, B and C classification and potentially minimize inter-observer disagreement.^{11,16,18} Moreover, it was designed primarily as a Computer Tomography (CT) classification, as these are widely available imaging modalities in trauma centers around the world. The authors suggested that Magnetic Resonance Imaging (MRI) could be used for further investigation of ligamentous injuries and was found to be helpful in some cases.¹⁸

In our study, 64.8% of the patients had total reproducibility between the newly AO spine groups (A, B and C), whereas only 55.2% had perfect concordance regards to sub-type classification. The Kappa index for major groups, A, B and C, was respectively 0.75 (good), 0.7 (good) and 0.85 (almost perfect). However, for subgroups, range from 0.16 to 0.85, suggesting some difficult in performing and accurate comparison of subgroups, which can influence conclusions. We observed a trend towards a higher reproducibility for the most severe injuries (dislocations and those with

misalignments), with 13 (43.3%) patients from 30 achieving perfect agreement

Safety

In evaluating safety, only 30 patients with perfect agreement were evaluated between the subgroups. This limited any further conclusion about safety of the newer AO system once 24 patients were excluded from the analysis (44.4%). We had to exclude them once misclassification would result in false conclusions about safety.

With regards to treatment rendered, according to subgroup classification, our data suggested that fractures from the newer proposed group B and C were treated surgically (their final TLICS score is generally higher than 4), identifying more severe and unstable injuries. These findings are similar to the suggested with the Magerl classification system published in 1994.⁶ However, in the Magerl classification, Group B included distractive injuries, even with some dislocations, that were included in group C in the new AO classification proposed by Vaccaro *et al.* Concerns should be raised however as only 57.6% of the patients could be perfectly evaluated regarding safety in our current study.

All the injuries of group B were surgically treated. It is suggested that the injuries to the anterior or posterior tension band can lead to late deformities and refractory pain, or can be associated with neurological deficits. The only patient with a type B injury who refused surgery initially ultimately underwent a late instrumented and fixation for refractory pain and segmental kyphosis.

Potential information retrieved from our series was that static radiological evaluation using only a supine CT scan can result in misclassification: the patient who had a systemic trauma and was initially classified as an A1 fracture by both evaluators in the two assessments had an important PLC injury that was misdiagnosed due severe clinical status at admission and longtime recumbence. It is our practice to perform standing upright plain lateral spine radiography prior to hospital discharge in patients that we intend to perform conservative treatment.¹⁹ The other three patients with type A1 fractures treated non-surgically did not have any further complications during follow-up. One case of an A3 fracture as well as three patients with A4 fractures treated conservatively also did not have any further complication.

In the paper defining the newer AO classification, Vaccaro *et al.* evaluated 40 cases selected from one author's practice and asked for 9 fellowship-trained spine surgeons to grade the cases.¹⁰ They graded the patients again one month after the first round,

scrambling the cases. Cases were graded by injury type, with a proportion of 54% of type A, 24% type B and 22% type C according to surgeons' classification. Statistical analysis was performed utilizing the Kappa coefficient. They obtained a full agreement when classifying the type of injury in 14 of 40 (35%) of the cases, with an overall Kappa coefficient was 0.64. They also compared grading regarding subtypes, they classified unanimously in 24 of 40 cases (60%). These cases included 16 types A, 3 types B and 5 type C, with a kappa for overall agreement of 0.72. For type A, the kappa was of 0.72, 0.58 for type B and 0.7 for type C. The lowest level of agreement was for fracture type B2 (kappa = 0.34) and B3 (kappa = 0.41).

Interestingly, we obtained better kappa indexes for the major groups, namely, 0.75 for type A; 0.7 for type B and 0.85 for type C. This may be due to the low number of some subtypes (for instance A2) of fractures in our sample and the smaller number of observers in our study (two).

Our study has some limitations. Treating surgically patients with a TLICS of more than 3 points (including the TLICS of 4 points) may bias external validity of our correlation between the AO system and the TLICS. Additionally, we had a limited sample size, a relative short follow-up and also eight complications in 54 patients, including one death, which can be considered high. However, even considering these limitations, we could infer that there is a high inter rater agreement for the major categories of the newer AO system. Moreover, there seems to be some correlation between the initial treatment choice and the major group classification.

Conclusions

The new AO spine classification has a good inter rater reproducibility for the major categories, and moderate for the sub-categories.

The reproducibility obtained with the newer AO spine classification in this study seemed to correlate in some degree with the choice of treatment for the major groups. However the classification in subgroups raised some concerns regards to its use in surgical decision-making. Although safety analysis was done in only 30 of 54 patients (57.6% of the total sample), when total consensus of fracture subtypes was obtained, fractures of group A can be treated conservatively or surgically, depending on other factors, such as neurological status. For injuries of group B and C, all the patients would receive surgical treatment. Further multicenter prospective studies are necessary to clarify these issues and validate the newer AO classification system.

Disclaimer statements

Contributors All the authors participated in the manuscript elaboration, such as acquisition of the data, interpretation of the data, manuscript writing and final review for submission.

Funding None.

Conflicts of interest There is no interest conflict regarding this manuscript content.

Ethics approval IRB approval was obtained. (protocol number 32917414.5.0000.5404).

References

- 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)* 2011;36(1):33–6.
- Joaquim AF, Daubs MD, Lawrence BD, Brodke DS, Cendes F, Tedeschi H, *et al.* Retrospective evaluation of the validity of the Thoracolumbar Injury Classification System in 458 consecutively treated patients. *Spine J* 2013;13(12):1760–5.
- Joaquim AF, Ghizoni E, Tedeschi H, Batista UC, Patel AA. Clinical results of patients with thoracolumbar spine trauma treated according to the Thoracolumbar Injury Classification and Severity Score. *J Neurosurg Spine* 2014;20(5):562–7.
- Joaquim AF, Lawrence B, Daubs M, Brodke D, Tedeschi H, Vaccaro AR, *et al.* Measuring the impact of the Thoracolumbar Injury Classification and Severity Score among 458 consecutively treated patients. *J Spinal Cord Med* 2014;37(1):101–6.
- Denis F. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. *Spine* 1983;8(8):817–31.
- 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.
- Vaccaro AR, Lehman RA, Jr., Hurlbert RJ, Anderson PA, Harris M, Hedlund R, *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)* 2005;30(20):2325–33.
- Patel AA, Dailey A, Brodke DS, Daubs M, Harrop J, Whang PG, *et al.* Thoracolumbar spine trauma classification: the Thoracolumbar Injury Classification and Severity Score system and case examples. *J Neurosurg Spine* 2009;10(3):201–6.
- Bono CM, Vaccaro AR, Hurlbert RJ, Arnold P, Oner FC, Harrop J, *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(8):567–72.
- Vaccaro AR, Oner C, Kepler CK, Dvorak M, Schnake K, Bellabarba C, *et al.* AO spine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. *Spine (Phila Pa)* 2013;38(23):2028–37.
- van Middendorp JJ, Patel AA, 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):461–74.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–74.
- Aebi M. Classification of thoracolumbar fractures and dislocations. *Eur Spine J* 2010;19(1):S2–7.
- Oner FC, Ramos LM, Simmermacher RK, Kingma PT, Diekerhof CH, Dhert WJ, *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):235–45.
- Rihn JA, Yang N, Fisher C, Saravanja D, Smith H, Morrison WB, *et al.* Using magnetic resonance imaging to accurately assess injury to the posterior ligamentous complex of the spine: a prospective comparison of the surgeon and radiologist. *J Neurosurg Spine* 2010;12(4):391–6.
- Whang PG, Vaccaro AR, Poelstra KA, Patel AA, Anderson DG, Albert TJ, *et al.* The influence of fracture mechanism and morphology on the reliability and validity of two novel thoracolumbar injury classification systems. *Spine (Phila Pa)* 2007;32(7):791–5.
- Winklhofer S, Thekkumthala-Sommer M, Schmidt D, Rufibach K, Werner CM, Wanner GA, *et al.* Magnetic resonance imaging frequently changes classification of acute traumatic thoracolumbar spine injuries. *Skeletal Radiol* 2013;42(6):779–86.
- Joaquim AF, Patel AA. Letters to the editor: burst fractures. *J Neurosurg Spine* 2013;18(3):314.