Evaluation of the reliability and validity of the newer AOSpine subaxial cervical injury classification (C-3 to C-7)

Otávio Turolo da Silva, MD,1 Marcelo Ferreira Sabba, MD,1 Henrique Igor Gomes Lira, MD,1 Enrico Ghizoni, MD, PhD,1 Helder Tedeschi, MD, PhD,1 Alpesh A. Patel, MD,2 and Andrei Fernandes Joaquim, MD, PhD1

1Department of Neurology, University of Campinas, Campinas, São Paulo, Brazil; and 2Department of Orthopedics, Northwestern University, Chicago, Illinois

OBJECTIVE The authors evaluated a new classification for subaxial cervical spine trauma (SCST) recently proposed by the AOSpine group based on morphological criteria obtained using CT imaging.

METHODS Patients with SCST treated at the authors’ institution according to the Subaxial Cervical Spine Injury Classification system were included. Five different blinded researchers classified patients’ injuries according to the new AOSpine system using CT imaging at 2 different times (4-week interval between each assessment). Reliability was assessed using the kappa index (κ), while validity was inferred by comparing the classification obtained with the treatment performed.

RESULTS Fifty-one patients were included: 31 underwent surgical treatment, and 20 were managed nonsurgically. Intraobserver agreement for subgroups ranged from 0.61 to 0.93, and interobserver agreement was 0.51 (first assessment) and 0.6 (second assessment). Intraobserver agreement for groups ranged from 0.66 to 0.95, and interobserver agreement was 0.52 (first assessment) and 0.63 (second assessment). The kappa index in all evaluations was 0.67 for Type A, 0.08 for Type B, and 0.68 for Type C injuries, and for the facet modifier it was 0.33 (F1), 0.4 (F2), 0.56 (F3), and 0.75 (F4). Complete agreement for all components was attained in 25 cases (49%) (19 Type A and 6 Type C), and for subgroups it was attained in 22 cases (43.1%) (16 Type A0 and 6 Type C). Type A0 injuries were treated conservatively or surgically according to their neurological status and ligamentous status. Type C injuries were treated surgically in almost all cases, except one.

CONCLUSIONS While the general reliability of the newer AOSpine system for SCST was acceptable for group classification, significant limitations were identified for subgroups. Type B injuries were rarely diagnosed, and only mild (Type A0) and extreme severe (Type C) injuries had a high rate of interobserver agreement. Facet modifiers and intermediate injury patterns require better descriptions to improve their low agreement in cases of SCST.

http://thejns.org/doi/abs/10.3171/2016.2.SPINE151039

KEY WORDS cervical trauma; classification; evaluation; reliability; subaxial cervical spine; treatment

Subaxial cervical spine trauma (SCST) comprises injuries from C-3 to C-7. More than 50% of the fractures in this region affect the C5–7 segments, likely due to the greater mobility in this region. Spinal cord injury (SCI) in the subaxial cervical spine has a high rate of morbidity, as it is associated with a higher risk of tetraplegia and important functional disability, with a significant economic and social impact.

Historically, many SCST classifications have been proposed. None has achieved widespread acceptance due to different limitations and drawbacks. Allen et al. proposed a classification of cervical fractures that divided injuries into 8 groups based on the mechanism of trauma. The reliability and complexity of this system were key disadvantages. Another system proposed by Magerl et al. adapted the AOSpine thoracolumbar system to the cervical spine. Injuries were classified into 3 main groups, according to their mechanism: Type A (compression), Type B (distraction), and Type C (rotation) injuries. The criticism of this complex classification system was that it did not take into account the neurological status, and it did not guide treatment.
In this context, in 2007 the Spinal Trauma Study Group proposed an injury severity score known as the Subaxial Cervical Spine Injury Classification (SLIC). The system consisted of 3 main characteristics that are scored according to their degree of severity: 1) injury morphology, 2) neurological status, and 3) the integrity of the disc-ligamentous complex. Each item has a corresponding score assigned according to its degree of severity. Treatment is then proposed based on the final, summed score: a score of fewer than 4 points suggests conservative management and a score greater than 4 suggests surgical treatment; injuries with a score of 4 may be treated either conservatively or surgically, based on surgeon’s preference and patient characteristics (for example, comorbidities and expectations).

While many studies suggested that the SLIC is reliable and safe, the morphological characterization of injuries was limited and unspecific. In this context, in 2015, Vaccaro et al. proposed a new classification for SCST based on CT evaluation and morphology, adopted by AOSpine, in an attempt to improve the morphological classification (Table 1).

Given the importance of an accurate classification system to standardize treatment and compare outcomes, evaluation of the reliability and safety of this new morphological classification system is necessary. In this paper, we evaluate the reliability of the new AOSpine classification system and the correlation of the new system with treatment for patients with subaxial cervical spine trauma.

Methods

After obtaining institutional review board approval, we retrospectively analyzed patients with SCST who were treated at our institution between 2011 and 2015. Patients were included if they had follow-up information; a complete CT scan reconstruction in the sagittal, coronal, and axial planes; and a documented traumatic cervical injury from C-3 to C-7. Images were analyzed using a DICOM System, part of the PACS (picture archiving and communication system) Aurora 3 (version 1.5.5, Pixeon Medical System). We excluded patients with incomplete clinical and radiological data.

The treatment of these injuries was performed by a single surgeon (A.F.J.) with expertise in the application of the SLIC system. Injuries were conservatively treated with a rigid cervical collar for 8 to 12 weeks in patients with 3 or fewer points, and surgical treatment was performed in patients with 4 or more points, based on fixation, realignment, and decompression, when necessary. Patients underwent routine follow-up after hospital discharge in our outpatient clinic performed by the same surgeon (A.F.J.) at 2 weeks, 1 month, 3 months, and then each 6 months after hospital discharge, with plain radiographs and CT scans obtained when necessary.

Clinical data analyzed included age, sex, level of injury (the most cranial bone level was considered when segmental injury occurred), mechanism of injury, neurological status evaluated according to the American Spinal Injury Association Impairment Scale (AIS), complications, and SLIC score.

Table 1. Morphological injury criteria of new AOSpine classification*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0</td>
<td>No/minor bony injury</td>
</tr>
<tr>
<td>A1</td>
<td>Compression involving single endplate</td>
</tr>
<tr>
<td>A2</td>
<td>Split vertebra</td>
</tr>
<tr>
<td>A3</td>
<td>Burst involving single endplate</td>
</tr>
<tr>
<td>A4</td>
<td>Burst involving both endplates</td>
</tr>
<tr>
<td>B1</td>
<td>Posterior bony injury</td>
</tr>
<tr>
<td>B2</td>
<td>Posterior bone-capsule-ligamentar injury</td>
</tr>
<tr>
<td>B3</td>
<td>Anterior tension band injury</td>
</tr>
<tr>
<td>C</td>
<td>Translation injuries</td>
</tr>
<tr>
<td>F1</td>
<td>Nondisplaced: &lt;1 mm, &lt;40%</td>
</tr>
<tr>
<td>F2</td>
<td>&gt;1 mm, &gt;40% or displaced</td>
</tr>
<tr>
<td>F3</td>
<td>Floating lateral mass</td>
</tr>
<tr>
<td>F4</td>
<td>Subluxation or perched</td>
</tr>
</tbody>
</table>

* From the description of Vaccaro et al.

After a complete presentation of the newly proposed AOSpine Classification System, discussed together with one of the senior authors (A.F.J.), 5 researchers blinded to the management of these injuries performed an individual classification of the included cases. The 5 researchers comprised a 2nd-year neurosurgery resident (M.F.S.), 2 final-year neurosurgery residents (O.T.S. and H.I.G.L.), 1 board-certified neurosurgeon (A.F.J.), and 1 board-certified orthopedic surgeon (A.A.P.). Classification was completed again 4 weeks after the first assessment. The data were organized according to the morphological groups (Types A, B, and C), subgroups (A0, A1, A2, A3, A4, B1, B2, B3, and C), and facet modifier (F1, F2, F3, and F4). The reliability between the intraobserver and interobserver classification was assessed by the kappa index, calculated using Stata MP for Windows (version 13.0, StatCorp). The kappa index was classified according to its numerical value: < 0.20, slight; 0.21–0.4, fair; 0.41–0.60 moderate; 0.61–0.80, substantial; and 0.81–1.00, excellent.

For validity, injuries with total morphological agreement between the 5 interobserver evaluations during all 10 assessments were correlated with the treatment performed based on the SLIC final score (conservative vs surgical treatment).

Results

A total of 51 patients treated between 2011 and 2015 were included; 45 patients (88.2%) were male and 6 were (11.7%) female. The mean age was 39.29 years, ranging from 18 to 82 years (mean 38, SD 15.37 years). With regard to injury mechanism, car accidents were the most prevalent etiology (33.3% of cases). The mean SLIC score among all patients was 4.52 points (ranging from 0 to 10 points, median 5, SD 3.39).
Conservatively Treated Group

Of the 20 patients (39.2%) treated conservatively, most were male (85%), with a mean age of 39.4 years. The mean follow-up was 86.8 days (median 30 days, SD 192.3 ranging from 20 to 829 days). The most commonly affected level of injury was C-7 with 7 cases (35%), followed by C-4 with 5 cases (25%), C-3 and C-6 with 3 cases (15%), and C-5 (10%) with 2 cases.

The mean SLIC score was 1 point (median 0 points, SD 1.82, ranging from 0 to 7 points). Seventeen patients were neurologically intact (AIS Grade E [85%]). Three patients were classified as AIS Grade C: 2 of them (with SLIC scores of 3 and 4 points) had spinal cord injury without evident instability or persistent compression, and one had a C7–T1 distraction injury with an SLIC of 7 points. However, this patient was treated for a subdural hematoma developed severe respiratory failure requiring tracheotomy and prolonged ventilatory support, all with upper cervical injuries (C-4) and paraplegia (1 AIS Grade A and 1 AIS Grade B). One patient (3%) required an additional surgical procedure for debridement of a wound infection, followed by intravenous antibiotics. No patient required additional surgery for hardware revision or instrumental removal.

Evaluation of Reliability

We present the 510 responses of the evaluations of 51 injuries for groups and subgroups, as well as 355 responses for facet morphology modifier. For group classification, there were 313 responses classifying injuries as Type A (61.3%), 36 as Type B (7%), and 161 as Type C (31.5%). For subgroup classification, there were 253 Type A0 (49.6%), 26 Type A1 (5%), 9 Type A2 (1%), 5 Type A3 (0.9%), 20 Type A4 (3%), 2 Type B1 (0.3%), 23 Type B2 (4.5%), 11 Type B3 (2.1%), and 161 Type C (31.5%) responses. Finally, classification of facet modifiers was as follows: 81 F1 (22.8%), 77 F2 (21.6%), 26 F3 (7%), and 170 F4 (47.8%) (Table 2).

Intraobserver Analysis

Groups. When assessing only major morphological groups (Types A, B, and C), intraobserver reliability reported ranged from 0.66 to 0.95 (substantial to excellent). Of note, intraobserver was excellent only for both senior spine surgeons (A.A.P and A.F.J), the κ ranged from 0.66 to 0.75 (substantial).

Subgroups. Reliability reported for subgroups (Types A0, A1, A2, A3, A4, B1, B2, B3, and C) ranged from κ = 0.61 (substantial) to κ = 0.93 (excellent). Again, for both senior spine surgeons (A.A.P and A.F.J), the κ was above 0.8 (0.8 and 0.93: excellent), whereas for the residents the κ ranged from 0.63 to 0.67 (moderate).

Facet Morphology. Reliability reported for facet morphology (F1, F2, F3, and F4) when assessed separately

Complications in the Surgical Group

Two patients (6%) died of infectious complications, one of septic shock and another with acute renal failure, at 4 and 22 days after surgery, respectively. Both were acutely tetraplegic patients (AIS Grade A). Three patients (9%) developed severe respiratory failure requiring tracheotomy and prolonged ventilatory support, all with upper cervical

---

**TABLE 2. Frequency of responses according to injury type in all evaluations**

<table>
<thead>
<tr>
<th>Morphology</th>
<th>No. of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertebral</td>
<td></td>
</tr>
<tr>
<td>A0</td>
<td>253 (49.6%)</td>
</tr>
<tr>
<td>A1</td>
<td>26 (5%)</td>
</tr>
<tr>
<td>A2</td>
<td>9 (1.7%)</td>
</tr>
<tr>
<td>A3</td>
<td>5 (0.9%)</td>
</tr>
<tr>
<td>A4</td>
<td>20 (3.9%)</td>
</tr>
<tr>
<td>B1</td>
<td>2 (0.3%)</td>
</tr>
<tr>
<td>B2</td>
<td>23 (4.5%)</td>
</tr>
<tr>
<td>B3</td>
<td>11 (2.1%)</td>
</tr>
<tr>
<td>C</td>
<td>161 (31.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>510</td>
</tr>
<tr>
<td>Facet</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>82 (23%)</td>
</tr>
<tr>
<td>F2</td>
<td>77 (21.6%)</td>
</tr>
<tr>
<td>F3</td>
<td>26 (7.3%)</td>
</tr>
<tr>
<td>F4</td>
<td>170 (47.8%)</td>
</tr>
<tr>
<td>Total</td>
<td>355</td>
</tr>
</tbody>
</table>
from the group morphology ranged from $\kappa = 0.66$ (substantial) to $\kappa = 0.95$ (excellent).

Interobserver Analysis

**Groups.** Interobserver reliability for Types A, B, and C was $\kappa = 0.52$ (moderate) and $\kappa = 0.63$ (substantial) in the first and in the second evaluation, respectively.

**Subgroups.** Interobserver reliability for subgroups (Types A0–4, B1–3, and C) was $\kappa = 0.51$ (moderate) and $\kappa = 0.60$ (moderate) in the first and second evaluations, respectively.

**Facet Morphology.** Interobserver reliability for facet morphology (F1, F2, F3, and F4) was $\kappa = 0.53$ (moderate) in the first evaluation and $\kappa = 0.62$ (substantial) in the second evaluation.

**Kappa Index for All 10 Assessments**

The kappa index for all 10 evaluations performed in the study was also analyzed.

**Group Reliability.** The final kappa index was 0.67 (substantial) for Type A, 0.08 for Type B (slight), and 0.68 for Type C (substantial).

**Subgroup Reliability.** The final kappa index was 0.76 (Type A0), 0.26 (Type A1), 0.1 (Type A2), 0.03 (Type A3), 0.29 (Type A4), 0.08 (Type B2), 0.06 (Type B3), and 0.7 (Type C).

**Facet Modifiers Reliability.** The final kappa index was 0.33 (F1), 0.4 (F2), 0.56 (F3), and 0.75 (F4). In combination, the greatest degree of interobserver reliability ($\kappa = 0.7$) was attained for Type C injuries with a facet modifier of F4 (149 answers).

**Validity**

Total agreement between the observers was attained in 25 cases (49%) when considering only groups and in 22 cases (43.1%) when analyzing subgroups in the 10 evaluations.

For group classification, the majority of the cases with total agreement were from Type A, with 19 cases, followed by 6 cases of Type C.

Considering subgroups, total agreement was attained in 16 cases of AO Spine Type A0 and in the same 6 cases of Type C injuries (no subgroup for Type C). Of the 6 patients with Type C injuries and total agreement, 5 were treated surgically, with the exception of the patient who had a severe traumatic brain injury and poor clinical status, as explained previously. There were no patients with Type B injuries for whom there was total agreement in our series. Additionally, no patient with morphology of Subtype A1–4 had total agreement. Patients with Type A0 (with or without facet fractures) had variable treatment, according to patients’ neurological status: 5 were treated surgically, whereas 11 were managed conservatively. The explanation for surgery in the 5 patients with Type A0 injuries is that some had traumatic disc herniation or spondylotic changes and had neurological worsening after acute trauma, without severe fracture or dislocation.

When considering morphology subgroups and facet modifiers, only 8 cases attained total agreement: 2 cases of Type A0 without any facet modifier and 6 cases of Type C injury with a modifier F4. These cases are summarized in Table 3.

In Figs. 1 and 2 we report 2 illustrative cases.

**Discussion**

Classification of SCST is necessary to facilitate clinical management, such as in decision making for conservative versus surgical management, for standardizing language among health care providers, and also for research purposes.2,4,7,12,16,20,23 However, a classification must be practical, reliable, and valid prior to its adoption in clinical practice.

In this new proposal, injuries were also divided into 3 major groups: Type A or compression (with 6 subgroups: A0 to A5), Type B with fractures with lesions on the anterior or posterior band tension (with 3 subgroups: B1–B3) and, finally, Type C or fractures with translation and severe dislocation (Table 1). Additionally, the proposal included a detailed classification of facet lesions in 4 degrees (F), a classification of the severity of the neurological deficit (N), and characterization of special subtypes of injuries (M).19

The reliability of this newly proposed system was tested by 10 spine experts selected by the classification’s authors, using 30 SCST cases.19 The agreement rate was assessed using the kappa score and attained 0.75 for intraobserver agreement and 0.64 for interobserver agreement for all subgroups of injuries analyzed together. The agreement rate was also evaluated separately for each injury morphology, with a higher agreement for B3 ($\kappa = 0.87$; excellent). The agreement rate for A3 morphology was $\kappa = 0.25$, for B2 it was $\kappa = 0.29$, for F1 it was $\kappa = 0.0$, and for F4 injury it was $\kappa = 0.06$. The lower agreement for these morphologies was associated with a higher prevalence of injury subtypes in the series of patients studied.

Our study, by comparison, had a slightly lower rate of agreement for group evaluation, with an intraobserver agreement of 0.52 (first evaluation) and 0.63 (second evaluation), which, according to the Landis and Koch4 system means moderate (first) and substantial (second) agreement. This improvement from moderate to substantial (0.52 to 0.63) between the 2 evaluations may reflect inexperience and unfamiliarity with the classification of the evaluators at the first time of assessment. Patel et al. had already reported that agreement improves with time, suggesting the benefits of learning and familiarity with training for application of a thoracolumbar classification.11 The same was emphasized in our results of intraobserver analysis, where a higher intraobserver reliability for both senior spine surgeons were reported compared with the residents scores.

Our study had only 36 responses of Type B injury mor-
Evaluation of the newer AO cervical spine classification

phyiology (2 Type B1, 23 Type B2, and 11 Type B3 in 510 responses). Contrarily, in the reported AO spine study, a high agreement rate was attained for Type B3 injuries ($\kappa = 0.87$). On the other hand, a high agreement rate was attained in this study for Type A and Type C injuries, potentially explained by the fact that extremely mild and severe injury morphologies are easier to identify and are less prone to confounding factors. This may also be explained by methodology differences between the 2 studies. In the AO spine study, the cases were preselected for evaluation while in the present study, we have used consecutively treated patients, which may decrease the prevalence of Type B injuries and influence the final agreement rate.

When assessing the reliability of the subgroups, we realize that the least common subtypes (Types A2, A3, B1, B2, and B3) did not have statistical relevance due to their low prevalence, with no agreement in any case. The small number also did not allow us to evaluate validity in these subtypes.

Reliability of facet morphology ranged from 0.59 to 0.84 in the intraobserver evaluation and 0.53 (first assessment) and 0.62 (second assessment) in the interobserver evaluation, with just moderate agreement. We believe that a more precise definition of facet injuries or simplification of the categories may improve reliability. As an example, distinguishing F1 from F2 injuries may be extremely subjective since the differences are small. Consistent with this, F1, F2, and F3 attained only fair to moderate agreement (final $\kappa$ of 0.33, 0.4, and 0.56, respectively), compared with a $\kappa$ of 0.75 (substantial) for severe facet injuries (F4).

When assessing the reliability of the subgroups, we realize that subtypes (A1–4 and B1–3) did not have total agreement in any case. This precluded evaluation of the treatment performed in these injuries compared with the SLIC system. For this reason, safety evaluation was only performed for the 2 extreme injury morphologies, Type A0 and Type C, thus limiting the analysis of validity in this study. We could infer that an isolated analysis of the morphology of Type A was, however, not sufficient to guide treatment. Some of these patients had neurological deficits and disc herniation/ligamentous injuries requiring surgery whereas others had only minor fractures. Of note, these instances demonstrated the importance of MRI in the decision making of some patients with SCST.

Our study is limited to a single center and is retrospective in nature. Despite these methodological limitations, many conclusions could be inferred, such as the low prevalence and also low agreement of intermediate morphology types (Types A1–4 and B), the need for clarity with
facet modifiers, and, finally, the redundancy of classifying Type C injuries with an F4 modifier, as the dislocation morphology necessarily has an associated facet dislocation or subluxation.

Conclusions

Our assessment has shown that the new classification of AOSpine demonstrates good statistical agreement for group (Types A, B, and C) classification. However, it is limited in guiding treatment in isolation, with a low agreement rate for some subtypes of injuries. Only extremely mild or severe injury morphologies had a higher agreement. Additionally, facet modifiers are imprecise. Validity was not evaluated completely in our study due to the low prevalence of moderate (Type B) injury morphologies. The use of this new classification system to guide treatment, therefore, requires additional, and preferably prospective, investigation.

References


Disclosures

Dr. Patel reports the following: ownership in Amedica, Vital 5, Nocimed, and Cytonics; consultancy with Amedica, Biomet/Zimmer, DePuy, Pacira, and Ulrich.

Author Contributions

Conception and design: Joaquim, da Silva, Ghizoni, Tedeschi. Acquisition of data: Joaquim, da Silva, Sabbia, Lira, Ghizoni. Analysis and interpretation of data: Joaquim, da Silva, Sabbia, Ghizoni, Tedeschi, Patel. Drafting the article: Joaquim, da Silva, Lira, Ghizoni. Critically revising the article: all authors. Reviewed submitted version of manuscript: Joaquim, Tedeschi, Patel. Approved the final version of the manuscript on behalf of all authors: Joaquim. Statistical analysis: da Silva, Lira. Administrative/technical/material support: Joaquim, Patel. Study supervision: Joaquim.

Correspondence

Andrei Fernandes Joaquim, Department of Neurology, Universidade Estadual de Campinas. Rua Tessália Vieira de Camargo, 126, Campinas, SP 13083-887, Brazil. email: andjoaquim@yahoo.com.