



Reliability of the AOSpine thoracolumbar spine injury classification system and thoracolumbar injury classification and severity score (TLICS)

Jetan H. Badhiwala, Michael G. Fehlings

Department of Surgery, Division of Neurosurgery, University of Toronto, Toronto Western Hospital, 399 Bathurst St., Toronto, Ontario M5T 2S8, Canada

Correspondence to: Michael G. Fehlings, MD, PhD, FRCSC. Department of Surgery, Division of Neurosurgery and Spinal Program, University of Toronto, Krembil Neuroscience Center, Toronto Western Hospital, 399 Bathurst St., Suite 4W-449, Toronto, Ontario M5T 2S8, Canada. Email: michael.fehlings@uhn.ca.

Comment on: 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;26:1470-6.

Received: 23 August 2017; Accepted: 26 September 2017; Published: 19 October 2017.

doi: 10.21037/amj.2017.09.16

View this article at: <http://dx.doi.org/10.21037/amj.2017.09.16>

Several classification schemes have been proposed for thoracolumbar spinal injuries over the years, with the goal of facilitating communication and streamlining selection of optimal treatment strategies. None of these have achieved universal acceptance. An ideal classification system is expected to categorize injuries in a way that permits identification of any injury, uses concise and descriptive terminology, reflects the mechanism of injury and biomechanical forces involved, guides choice of treatment, has easily recognizable clinical and radiological characteristics, describes and grades the severity of neurological injury, grades both osseous and ligamentous injury, and predicts the natural history and treatment end results (1).

The Denis and Magerl systems were among the earliest and most widely accepted classification schemes for thoracolumbar spinal injury. According to the Denis (2) three column model, instability is present and operative stabilization may be needed if two of the three columns are disrupted. This system drew its greatest strength from its simplicity and ease of use. As a result, it had greater inter-rater reliability than other systems in use around the same time (3,4). However, the Denis model is excessively simple and not comprehensive enough, not accounting for many fracture types, and lacking predictive value in aiding treatment decisions (5). By contrast, the

Magerl (6) system was arguably the most systematic and detailed classification scheme of fracture morphology. It employed a hierarchical system in which successive grades represent increasing fracture severity, with a comprehensive subdivision of variants within each injury grade. While comprehensive, the Magerl system was overly complex and had poor reproducibility, and hence found limited clinical use (3,4,7). The Denis and Magerl systems hence traded off simplicity in clinical use and reproducibility with comprehensiveness and all-inclusivity. Neither system accounted for the neurological status of the patient, which is actually a principal driver of treatment decisions. This ultimately led to the development of the Thoracolumbar Injury Classification and Severity Score (TLICS) (8,9) by the Spine Trauma Study Group (STSG). TLICS grades injury severity based on three characteristics: (I) injury morphology; (II) integrity of the posterior ligamentous complex (PLC); and (III) neurological status of the patient. Points are assigned for each category, with the total score suggesting a possible treatment option, either non-operative, operative, or indeterminate. TLICS has the advantages of being user-friendly, incorporating a patient's neurological status into the classification scheme, and guiding treatment. However, TLICS was also met with several criticisms, including the poor reproducibility of assessment of PLC integrity and the use of a scoring

scheme which may not be globally accepted across cultures and geographic regions (10,11). These inadequacies of TLICS prompted the introduction of the AOSpine Thoracolumbar Spine Injury Classification System (12) by the AOSpine Trauma Knowledge Forum, which in a way married the strengths of the Magerl and TLICS systems. This system consists of a morphological classification of the fracture based on a revision of the original Magerl system. In addition, two key improvements of the AOSpine system compared to prior classification schemes are: (I) inclusion of an assessment of the neurological status of the patient; and (II) inclusion of a description of clinically relevant patient-specific modifiers, such as osteoporosis and rheumatologic disease. The hope is that the AOSpine system will be accepted by the global spine community, but further validation studies are needed.

To that end, Kaul *et al.* (13) sought to assess and compare the reliability of the AOSpine classification system and TLICS. Clinical and radiological data of 50 consecutive patients admitted at a single center (the Indian Spinal Injuries Centre) with an acute traumatic thoracolumbar spine injury were circulated to 11 attending spine surgeons from six institutions in four different countries—the United States, Germany, India, and Bangladesh. Cases of osteoporotic fractures and pathological fractures secondary to infection or malignancy were excluded. Clinical data consisted of patient demographics, mechanism of injury, spinal level of injury, associated injuries, and neurological examination as per ASIA (ISNCSCI) standards. Radiological data took the form of representative stills of MRI, CT, and/or plain films. Participating surgeons were asked to classify each case according to the AOSpine system and TLICS. After a period of 6 weeks, the cases were rearranged randomly and sent back to participating surgeons for a second evaluation. For each system, the authors then calculated the inter-rater reliability using data from the first round, and the intra-rater reliability using data from the second round. Kappa values were interpreted by the scale described by Landis and Koch (14). Overall, the AOSpine Thoracolumbar Spine Injury Classification System demonstrated better reliability than TLICS. With regard to inter-rater reliability, TLICS was found to have moderate agreement ($\kappa=0.43$) for fracture morphology and integrity of the PLC ($\kappa=0.47$) and near perfect agreement ($\kappa=0.85$) for neurological function, but only fair agreement ($\kappa=0.29$) for total score. By contrast, the AOSpine system showed moderate reliability ($\kappa=0.59$) for fracture type and near perfect reliability ($\kappa=0.85$) for neurological

involvement. With regard to intra-rater reliability, TLICS had moderate agreement for fracture morphology ($\kappa=0.59$) and PLC ($\kappa=0.55$), near perfect reliability for neurological status ($\kappa=0.90$), and moderate agreement for total score ($\kappa=0.44$). On the other hand, the AOSpine classification system demonstrated substantial reliability for fracture type ($\kappa=0.68$) and near perfect agreement for neurological function ($\kappa=0.91$).

This paper represents an important contribution to the literature. The AOSpine Thoracolumbar Spine Injury Classification System was first developed and described in 2013 (12). While the system has many merits, broad validation studies are needed before widespread adoption is possible. The present study provides an international and cross-cultural external validation of the AOSpine system, and moreover, provides a direct comparison to a popular pre-existing classification system, TLICS. None of the participating surgeons, with one notable exception, were involved in development of either system. The results of this study suggest the AOSpine classification system performs well, in fact better than TLICS, in real world clinical application, as applied by practicing surgeons in many different countries. This is an important finding.

Strengths of this paper are many. Firstly, as discussed above, the study included surgeons from multiple centers in many different countries. The methodology of the study was robust. Participating surgeons were provided with standard training, in the form of published study materials, pertaining to both classification systems. Each surgeon used a standard sheet to score cases. Appropriate statistical techniques were used. Kappa values were interpreted according to standard criteria. A limitation is the use of representative stills, rather than full image sets, which may have influenced interpretation. Moreover, only 50 consecutive cases were used, meaning more severe injuries may have been underrepresented owing to lower incidence.

With regard to TLICS, disagreement between observers arose primarily in assessment of fracture morphology and the integrity of the PLC. The reliability and reproducibility of evaluating the integrity of the PLC on MRI is poor, and disagreement and controversy surrounding the use of MRI as an adjunctive diagnostic tool persists (10). The development of more accurate techniques to assess the PLC has been one of the key challenges in this area. Prior studies have reported greater agreement for fracture morphology than the authors report herein (11,15). We would agree with the authors that this is likely because this study involved, for the most part, surgeons who were not involved in the

Table 1 Inter- and intra-rater reliability of the AOSpine Thoracolumbar Spine Injury Classification System in published studies

Author & year	Reliability (κ)*	
	Inter-rater	Intra-rater
Vaccaro <i>et al.</i> (2013) (12)	0.64	0.77
Urrutia <i>et al.</i> (2014) (18)	0.55	0.71
Sadiqi <i>et al.</i> (2015) (19)	NR	0.67–0.69
Kepler <i>et al.</i> (2016) (20)	0.56	0.68
Cheng <i>et al.</i> (2017) (17)	0.36	0.41–0.48†
Kaul <i>et al.</i> (2017) (13) (present study)	0.45	0.61

*, including fracture subtype; †, 0.44 for type A injuries, 0.48 for type B injuries, 0.41 for type C injuries.

original development of the system. Hence, the current study may provide a more accurate estimate of the reliability of TLICS in a pragmatic setting. The greatest discrepancy arose for translation or rotation ($\kappa=0.36$) and distraction injuries ($\kappa=0.28$). This may be because classification into either of these categories requires some subjective inference about the mechanism of injury. Although this limitation underpinned revision of the original Thoracolumbar Injury Severity Score (TLISS) (8) to the TLICS (16), which included a description of fracture morphology, TLICS still involves some degree of retrospective reconstruction of the forces applied to the spinal column in classifying injuries. The definitions of compression and burst fractures, on the other hand, are more stringent and largely based upon observation of the morphological characteristics of a fracture. Compression fractures involve ‘wedging’ of the anterior column, whereas burst fractures are defined by concomitant injury to the middle column with breach of the cortex of the posterior vertebral body. Indeed, in the present study, the highest inter-rater reliability was seen for compression ($\kappa=0.55$) and burst fractures ($\kappa=0.60$).

In contrast to TLICS, the AOSpine system classifies fracture type essentially entirely based on morphology; that is, based on an observation of the fracture pattern, rather than any inference of mechanism. This is likely a large part of the reason why the AOSpine classification system achieved greater inter- and intra-rater reliability than assessment of fracture morphology according to TLICS. Like in other independent validations, the kappa values for both inter- and intra-rater reliability of the AOSpine system reported here were lower than those reported by the original group (17,18). *Table 1* presents kappa coefficients

from published studies.

One of the key elements of TLICS is the treatment recommendation. Non-operative management is recommended for patients with a score of 0 to 3 and surgery for patients scoring ≥ 5 points. In patients with a score of 4, either non-operative or operative treatment may be considered (9). However, a criticism levied against TLICS is that the scoring system guiding treatment decisions may not reflect global surgical preferences, but rather be region- or culture-dependent, reflecting the value placed on immediate surgical stabilization and accelerated rehabilitation (12). This has likely hindered global acceptance of TLICS. For example, in North America, thoracolumbar burst fractures are increasingly being treated conservatively, with or without brace (21) By contrast, there are several European reports of similar fractures being treated with 360° fusion (22,23). In fact, one of the key knowledge gaps and challenges lies in developing internationally accepted algorithms for the management of A3 and A4 (AOSpine) burst fractures in the neurologically intact individual. Recognizing this, the AOSpine group set out to develop a global injury severity scoring system. A survey of 100 AOSpine members from all six AO regions of the world (North America, South America, Europe, Africa, Asia, and the Middle East) was undertaken (24). Surgeons were asked to numerically rate, from 0 to 100, the severity of each variable of the AOSpine Thoracolumbar Spine Injury Classification System, including each category of morphology, neurological grade, and patient specific modifiers. The authors observed an increased perceived severity as the subtypes of fracture type A and B increased. Importantly, no difference in severity rating was observed by region or level of experience. In subsequent studies, the authors found no regional variability in ability to identify type A injuries or an injury to the PLC (25,26). Together, these results suggest that the development of a global algorithm for the treatment of thoracolumbar trauma is possible. This ultimately informed the development of the Thoracolumbar AOSpine Injury Score (TL AOSIS) (27). A worldwide survey of AOSpine members was then used to delineate the surgical threshold based on the TL AOSIS (28). We think it would be interesting for the authors of the present study, as a next step, to survey the same 11 participating surgeons on their proposed management strategy for each of the 50 cases (namely, operative or non-operative), and then compare this to the treatment recommendations provided by the TL AOSIS. This would provide a direct external evaluation of the performance TL

AOSIS in a global setting.

In summary, this is an important paper that provides a global, cross-cultural external validation of both the AOSpine Thoracolumbar Spine Injury Classification System and TLICS. Overall, the AOSpine system performed better than TLICS. The evidence so far would suggest the AOSpine system is perhaps the closest of any thoracolumbar spine injury classification system to achieving worldwide adoption, likely owing to the involvement of surgeons across the globe in its design and evaluation. Future studies are needed to validate the TL AOSIS.

Acknowledgements

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned and reviewed by the Section Editor Ai-Min Wu (Department of Spine Surgery, Zhejiang Spine Surgery Centre, Orthopaedic Hospital, The Second Hospital and Yuying Children's Hospital of Wenzhou Medical University, The Key Orthopaedic Laboratory in Zhejiang Province, Wenzhou, China).

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/amj.2017.09.16>). Dr. Fehlings reports personal fees and other from Neuraxis, outside the submitted work. The other author has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Mirza SK, Mirza AJ, Chapman JR, et al. Classifications of thoracic and lumbar fractures: rationale and supporting data. *J Am Acad Orthop Surg* 2002;10:364-77.
2. Denis F. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. *Spine (Phila Pa 1976)* 1983;8:817-31.
3. Oner FC, Ramos LM, Simmermacher RK, et al. Classification of thoracic and lumbar spine fractures: problems of reproducibility. A study of 53 patients using CT and MRI. *Eur Spine J* 2002;11:235-45.
4. Wood KB, Khanna G, Vaccaro AR, et al. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. *J Bone Joint Surg Am* 2005;87:1423-9.
5. Patel AA, Vaccaro AR. Thoracolumbar spine trauma classification. *J Am Acad Orthop Surg* 2010;18:63-71.
6. Magerl F, Aebi M, Gertzbein SD, et al. A comprehensive classification of thoracic and lumbar injuries. *Eur Spine J* 1994;3:184-201.
7. Blauth M, Bastian L, Knop C, et al. [Inter-observer reliability in the classification of thoraco-lumbar spinal injuries]. *Orthopade* 1999;28:662-81.
8. Vaccaro AR, Zeiller SC, Hulbert RJ, et al. The thoracolumbar injury severity score: a proposed treatment algorithm. *J Spinal Disord Tech* 2005;18:209-15.
9. Lee JY, Vaccaro AR, Lim MR, et al. Thoracolumbar injury classification and severity score: a new paradigm for the treatment of thoracolumbar spine trauma. *J Orthop Sci* 2005;10:671-5.
10. Rihn JA, Yang N, Fisher C, 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;12:391-6.
11. Whang PG, Vaccaro AR, Poelstra KA, et al. The influence of fracture mechanism and morphology on the reliability and validity of two novel thoracolumbar injury classification systems. *Spine (Phila Pa 1976)* 2007;32:791-5.
12. Vaccaro AR, Oner C, Kepler CK, et al. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. *Spine (Phila Pa 1976)* 2013;38:2028-37.
13. 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;26:1470-6.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
 15. Koh YD, Kim DJ, Koh YW. Reliability and Validity of Thoracolumbar Injury Classification and Severity Score (TLICS). *Asian Spine J* 2010;4:109-17.
 16. Schweitzer KM, Jr., Vaccaro AR, Lee JY, et al. Confusion regarding mechanisms of injury in the setting of thoracolumbar spinal trauma: a survey of The Spine Trauma Study Group (STSG). *J Spinal Disord Tech* 2006;19:528-30.
 17. Cheng J, Liu P, Sun D, et al. Reliability and reproducibility analysis of the AOSpine thoracolumbar spine injury classification system by Chinese spinal surgeons. *Eur Spine J* 2017;26:1477-82.
 18. Urrutia J, Zamora T, Yurac R, et al. An independent interobserver reliability and intraobserver reproducibility evaluation of the new AOSpine Thoracolumbar Spine Injury Classification System. *Spine (Phila Pa 1976)* 2015;40:E54-8.
 19. Sadiqi S, Oner FC, Dvorak MF, et al. The Influence of Spine Surgeons' Experience on the Classification and Intraobserver Reliability of the Novel AOSpine Thoracolumbar Spine Injury Classification System- An International Study. *Spine (Phila Pa 1976)* 2015;40:E1250-6.
 20. Kepler CK, Vaccaro AR, Koerner JD, et al. Reliability analysis of the AOSpine thoracolumbar spine injury classification system by a worldwide group of naive spinal surgeons. *Eur Spine J* 2016;25:1082-6.
 21. Bailey CS, Urquhart JC, Dvorak MF, et al. Orthosis versus no orthosis for the treatment of thoracolumbar burst fractures without neurologic injury: a multicenter prospective randomized equivalence trial. *Spine J* 2014;14:2557-64.
 22. Reinhold M, Knop C, Beisse R, et al. Operative treatment of 733 patients with acute thoracolumbar spinal injuries: comprehensive results from the second, prospective, Internet-based multicenter study of the Spine Study Group of the German Association of Trauma Surgery. *Eur Spine J* 2010;19:1657-76.
 23. Schnake KJ, Stavridis SI, Kandziora F. Five-year clinical and radiological results of combined anteroposterior stabilization of thoracolumbar fractures. *J Neurosurg Spine* 2014;20:497-504.
 24. Schroeder GD, Vaccaro AR, Kepler CK, et al. Establishing the injury severity of thoracolumbar trauma: confirmation of the hierarchical structure of the AOSpine Thoracolumbar Spine Injury Classification System. *Spine (Phila Pa 1976)* 2015;40:E498-503.
 25. 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* 2015:1-8.
 26. Schroeder GD, Kepler CK, Koerner JD, et al. A Worldwide Analysis of the Reliability and Perceived Importance of an Injury to the Posterior Ligamentous Complex in AO Type A Fractures. *Global Spine J* 2015;5:378-82.
 27. Kepler CK, Vaccaro AR, Schroeder GD, et al. The Thoracolumbar AOSpine Injury Score. *Global Spine J* 2016;6:329-34.
 28. Vaccaro AR, Schroeder GD, Kepler CK, et al. The surgical algorithm for the AOSpine thoracolumbar spine injury classification system. *Eur Spine J* 2016;25:1087-94.

doi: 10.21037/amj.2017.09.16

Cite this article as: Badhiwala JH, Fehlings MG. Reliability of the AOSpine thoracolumbar spine injury classification system and thoracolumbar injury classification and severity score (TLICS). *AME Med J* 2017;2:158.