CrossMark

ORIGINAL ARTICLE

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

Rahul Kaul[1] · Harvinder Singh Chhabra[2] · Alexander R. Vaccaro[3] · Rainer Abel[4] ·
Sagun Tuli[5] · Ajoy Prasad Shetty[6] · Kali Dutta Das[2] · Bibhudendu Mohapatra[2] ·
Ankur Nanda[2] · Gururaj M. Sangondimath[2] · Murari Lal Bansal[2] · Nishit Patel[2]

Abstract

Purpose The aim of this multicentre study was to determine whether the recently introduced AOSpine Classification and Injury Severity System has better interrater and intrarater reliability than the already existing Thoracolumbar Injury Classification and Severity Score (TLICS) for thoracolumbar spine injuries.

Methods Clinical and radiological data of 50 consecutive patients admitted at a single centre with a diagnosis of an acute traumatic thoracolumbar spine injury were distributed to eleven attending spine surgeons from six different institutions in the form of PowerPoint presentation, who classified them according to both classifications. After time span of 6 weeks, cases were randomly rearranged and sent again to same surgeons for re-classification. Interobserver and intraobserver reliability for each component of TLICS and new AOSpine classification were evaluated using Fleiss Kappa coefficient (k value) and Spearman rank order correlation.

Results Moderate interrater and intrarater reliability was seen for grading fracture type and integrity of posterior ligamentous complex (Fracture type: $k = 0.43 \pm 0.01$ and 0.59 ± 0.16, respectively, PLC: $k = 0.47 \pm 0.01$ and

✉ Harvinder Singh Chhabra
drhschhabra@isiconline.org

Rahul Kaul
rahulkaul1977@gmail.com

Alexander R. Vaccaro
alexvaccaro3@aol.com

Rainer Abel
abel_rainer@hotmail.com

Sagun Tuli
tuli@flspinal.com; saguntuli@gmail.com

Ajoy Prasad Shetty
ajoyshetty@gmail.com

Kali Dutta Das
kalidutta@yahoo.com

Bibhudendu Mohapatra
bibhudendu1973@gmail.com

Ankur Nanda
nandaankur@yahoo.co.in

Gururaj M. Sangondimath
drgururajm@rediffmail.com

Murari Lal Bansal
drmlbansal@yahoo.com

Nishit Patel
doctor.ortho@gmail.com

[1] Department of Orthopedics, Flt. Lt. Rajan Dhal, Fortis Hospital, Vasant Kunj, New Delhi, India

[2] Indian Spinal Injuries Center, Sector C, Vasant Kunj, New Delhi, India

[3] Department of Orthopaedic Surgery, Delaware Valley Spinal Cord Injury Center, Rothman Institute, Sidney Kimmel Medical Center at Thomas Jefferson University, Philadelphia, PA, USA

[4] Klinik für Querschnittgelähmte, Orthopädie und Rheumatologie, Klinik Hohe Warte, Klinikum Bayreuth GmbH, Bayreuth, Germany

[5] Florida Spinal Surgery Center, Miami, FL, USA

[6] Division of Orthopaedics, Trauma and Spine Surgery, Ganga Hospital, Coimbatore, India

0.55 ± 0.15, respectively), and fair to moderate reliability ($k = 0.29 \pm 0.01$ interobserver and $0.44+/0.10$ intraobserver, respectively) for total score according to TLICS. Moderate interrater ($k = 0.59 \pm 0.01$) and substantial intrarater reliability ($k = 0.68 \pm 0.13$) was seen for grading fracture type regardless of subtype according to AOSpine classification. Near perfect interrater and intrarater agreement was seen concerning neurological status for both the classification systems.

Conclusions Recently proposed AOSpine classification has better reliability for identifying fracture morphology than the existing TLICS. Additional studies are clearly necessary concerning the application of these classification systems across multiple physicians at different level of training and trauma centers to evaluate not only their reliability and reproducibility, but also the other attributes, especially the clinical significance of a good classification system.

Keywords AOSpine Thoracolumbar Spine Injury Classification System · TLICS · Reliability · Thoracolumbar spine injuries

Introduction

Classification systems are helpful to identify common attributes within a group to predict the behavior or outcome without sacrificing too much detail, being clinically relevant, reliable, and accurate [1]. An ideal spine injury classification is expected to provide details regarding injury severity, its pathogenesis, and causative biomechanical forces involved, in addition to clinical, neurological, and radiographical characteristics of the injury [2]. Following Böhler, who was the first to introduce a classification for thoracolumbar fractures, numerous other classifications have been proposed but there is still a lack of universal acceptance with regard to any of them [3, 4]. Among various existing classification systems for thoracolumbar spine injuries, only few have been assessed for reliability, reproducibility or clinical validity [5].

The three column model based Denis classification did not account for all fracture types or the neurological status of the patient and lacked predictive value to aid treatment decisions [4]. According to Denis, instability is present if two of the three columns are disrupted for which operative stabilization may be necessary [6]. However, studies have shown that nonoperative treatment of two column injuries may achieve a satisfactory outcome [7, 8]. The AO-Magerl classification was more inclusive by identifying a wide array of fractures including more than 50 subtypes using the 3-3-3 AO principle leading to its complexity, limiting its incorporation into routine clinical practice [4, 9]. In

addition it also does not account for neurological status, a critical determinant for surgical decision making [4, 10, 11].

In 2005 Spine Trauma Study Group (STSG) introduced Thoracolumbar Injury Severity Score (TLISS) which was based on three major injury characteristics: (1) the mechanism of injury, (2) the integrity of the posterior ligamentous complex, and (3) the patient's neurological status [12]. Although as a whole it had an excellent construct validity, interobserver agreement for injury mechanism was only fair ($\kappa = 0.33$) [13–15]. This led to the introduction of TLICS in which fracture mechanism was replaced by description of morphological injury [16]. Various authors have pointed out some inherent limitations even in the TLICS classification system [4, 11, 17, 18, 19]. AOSpine Knowledge Forum SCI and Trauma introduced the AOSpine Classification for traumatic fractures of the Thoracolumbar Spine in year 2013 and demonstrated substantial reliability for this newly proposed classification system [11]. Recent studies by authors other than those who have proposed it have also shown adequate agreement on classifying thoracolumbar spine injuries using this recently proposed classification system [20, 21].

To clarify the efficacy of a proposed classification system, a direct comparison with other already existing classification systems is required [4]. This current study evaluated the new AOSpine Classification and Injury Severity System introduced by the AO Spine Knowledge Forum, comparing its reliability to the already existing TLICS for thoracolumbar spine injuries. To the best of our knowledge no study has yet been published in the literature, which has assessed the reliability of the two together.

Materials and methods

The Institutional Research Review and Ethics committees of Indian Spinal Injuries Centre (ISIC) approved the protocol for the study and gave permission to review clinical charts as well as obtain images. For designing the study, literature survey was undertaken and advice from a statistical expert was sought to determine appropriate number of assessors and sample size. All the 11 participating Spine surgeons, including ten orthopedic surgeons and one neurosurgeon from six different institutions from four different countries (USA, Germany, India and Bangladesh), were sent appropriate published study materials with regard to both the classifications being evaluated in the study (Suppl. 1). Clinical and radiological data of 50 consecutive patients admitted at ISIC with a diagnosis of a traumatic thoracolumbar spine injury were distributed to the eleven spine surgeons in form of PowerPoint presentation. Pathological fractures due to infection, neoplasia and osteoporotic

fractures of spine without associated trauma were excluded from study. Patient related data included age, gender, mode of trauma, fracture level, associated injuries and neurological status. The neurological status included motor power, sensory function, perianal sensation and voluntary anal contraction assessed as per International standards for neurological classification of spinal cord injury [22]. Radiographic images consisting of plain films, computerized tomography (CT), and magnetic resonance imaging (MRI) were also provided. In case of multiple level spine injuries, more severe injury was considered for assessment by the author who chose the cases, to ascertain that all the evaluators involved in the study were grading the same lesion (Suppl. 2). With regard to Type B and Type C injuries according to new AOSpine classification, concurrent Type A or Type B at the same level were not included in the analysis. The additional modifiers M1 for fractures with an indeterminate injury to posterior tension band and M2 to designate a patient specific co-morbidity having an impact on surgical decision making, according to new AOSpine classification were not evaluated. A scannable sheet was provided to each participant, in which data regarding various components of AOSpine classification and TLICS were to be noted (Suppl. 3). After a time lapse of 6 weeks, cases were randomly rearranged to prevent recall bias and sent back to the same spine surgeons for reevaluation in form of PowerPoint presentation. All the final scoring sheets were then collected and submitted to an independent statistician analysis. Data obtained from the first round were analyzed to determine the interobserver agreement, whereas second round data analysis determined the intraobserver agreement. Interobserver and intraobserver reliability for each component of TLICS and new AOSpine classification were evaluated using Fleiss Kappa coefficient (k value) and Spearman rank order correlation. Statistical distinction of the Kappa values was done according to the Landis and Koch criteria [23] (Table 1).

Results

For TLICS, 44 % (242 of the 550) classifications were described as compression injuries (included both compression and burst fractures), 30 % (165) as distraction injuries and 26 % (143) as translational/rotation injuries. In 11 of 50 cases (22 %), evaluators classified fracture morphology unanimously according to TLICS and all of these were compression injuries. The interrater reliability for TLICS showed moderate agreement for determining fracture morphology and posterior ligamentous complex status (PLC) ($k = 0.43 \pm 0.01$ and 0.47 ± 0.01 respectively), near perfect agreement for the neurological status ($k = 0.85 \pm 0.01$) and fair agreement for the sum of the total score ($k = 0.29 \pm 0.01$). For the fracture morphology highest interobserver reliability was seen for compression injuries including both compression fractures and burst fractures ($k = 0.55$ and 0.60, respectively), then for translation or rotation injuries ($k = 0.36$) and least for distraction injuries ($k = 0.28$). Interrater reliability using the Spearman correlation was 0.57 ± 0.13 for fracture morphology, 0.68 ± 0.11 for PLC status, 0.86 ± 0.09 for neurological status and 0.69 ± 0.07 for total score (Table 2). Moderate intrarater reliability was noted for fracture morphology and PLC status ($k = 0.59 \pm 0.16$ and 0.55 ± 0.15, respectively), near perfect agreement for the neurological status ($k = 0.90 \pm 0.12$) and moderate agreement for the sum of the total score ($k = 0.44 +/0.10$). Intrarater reliability using the Spearman correlation was 0.69 ± 0.12 for fracture morphology, 0.77 ± 0.07 for PLC status, 0.88 ± 0.13 for neurological status and 0.79 ± 0.08 for total score (Table 3).

About 39.45 % (217 of the 550) AOSpine classifications were noted as type A, 36.55 % (201) as type C and 24 % (132) as type B injuries. In 16 of 50 cases (32 %) surgeons classified fracture morphology unanimously which included 8 type A, 1 type B and 7 type C injuries. The overall interrater agreement on grading fracture type and subtypes was moderate ($k = 0.59 \pm 0.01$ and 0.45 ± 0.01, respectively), and near perfect for neurological injury ($k = 0.85 \pm 0.01$). Among the main fracture types substantial interobserver reliability was seen for type A and type C ($k = 0.64$ and 0.71, respectively), and moderate for type B injuries ($k = 0.40$). Interrater reliability using the

Table 1 Interpretation of Kappa

κ value	Agreement
>0.80	Near perfect
0.61–0.80	Substantial
0.41–0.60	Moderate
0.21–0.40	Fair
0.00–0.20	Slight

Table 2 Interrater statistics for TLICS

	k value	Spearman	p value
Fracture morphology	0.43±	0.57	<0.05
SE	0.01	0.13	
PLC status	0.47	0.68	<0.05
SE	0.01	0.11	
Neurological involvement	0.85	0.86	<0.05
SE	0.01	0.09	
Total score	0.29	0.69	<0.05
SE	0.01	0.07	

SE standard error

Table 3 Intrarater statistics for TLICS

	k value	Spearman	p value
Fracture morphology	0.59	0.69	<0.05
SE	0.16	0.12	
PLC status	0.55	0.77	<0.05
SE	0.15	0.07	
Neurological involvement	0.90	0.88	<0.05
SE	0.12	0.13	
Total score	0.44	0.79	<0.05
SE	0.10	0.08	

SE standard error

Table 4 Interrater statistics for AOSpine Classification and Injury Severity System

	k value	Spearman	p value
Fracture type (overall value irrespective of fracture subtype)	0.59±	0.75	<0.05
SE	0.01	0.07	
Neurological involvement	0.85	0.88	<0.05
SE	0.01	0.08	

SE standard error

Table 5 Intrarater statistics for AOSpine Classification and Injury Severity System

	k value	Spearman	p value
Fracture type (overall value irrespective of fracture subtype)	0.68	0.78	<0.05
SE	0.13	0.08	
Neurological involvement	0.91	0.93	<0.05
SE	0.08	0.07	

SE standard error

Spearman correlation was 0.75 ± 0.13 for fracture type and 0.88 ± 0.08 for neurological injury (Table 4). Substantial intrarater agreement was noted for fracture type without and with regard to subtype ($k = 0.68 \pm 0.13$ and 0.61 ± 0.13, respectively), and near perfect for neurological injury ($k = 0.91 \pm 0.08$). Intrarater reliability using the Spearman correlation was 0.78 ± 0.08 for fracture type and 0.93 ± 0.07 for neurological injury (Table 5).

Discussion

The desired objectives of a good spinal injury classification system have been well described in the literature [2]. A good classification system should be straightforward, easy to use for all concerned and practically implementable in day-to-day practice. In addition, it should be replicable, that is, have a good interobserver and intraobserver reliability, it should be able to predict natural history, provide tool for future studies, take into consideration patterns of neurological injury as well as grade its severity and appropriately guide choice of treatment [1]. A survey conducted amongst members of International Spinal Cord Society's Spine Trauma Study Group members and other spine surgeons brought out that none of the classification systems meet the desired objectives appropriately and hence there is a need for developing newer ones [24].

There are no clear cut published guidelines in the existing literature regarding the development of a spine injury classification, and most of the past efforts have been based on expert opinion [1]. There exists a delicate balance between simplicity and inclusiveness of a classification system. In general, if the classification system is made exceeding simple, it often leads to loss of information. If the system is all inclusive, it becomes cumbersome to use and its reproducibility is markedly affected [25].

TLICS was introduced to overcome the limitations of the previous existing classification systems [4]. It has also faced criticisms in the recent literature [4, 17]. One of the concerns is relatively poor reliability and reproducibility of evaluating the posterior ligament complex (PLC) status based on MRI [26]. There exists scant clinical evidence to demonstrate true prognostic value of detected PLC injuries in patients with thoracolumbar spine injuries [27]. It has also not yet been proven whether MRI has any additional role for decision making regarding treatment in spinal injury patients without neurological deficit [28]. Due to these existing controversies, recommended use of disco-ligamentous characteristics in spinal injury classifications has been recently questioned [1]. The total injury scores which are supposed to guide the treatment may be culture or region specific decisions and may not reflect global surgical preferences or the most rational approach to treatment, thus preventing its worldwide acceptance [11]. In the endeavor to strive towards a better classification system overcoming the limitations of pre existing ones, the AOSpine Thoracolumbar Spine Injury Classification System was introduced [11]. To overcome the lacunae, Vaccaro et al. have proposed a data driven surgical algorithm as a result of worldwide survey concerning the treatment of thoracolumbar spine injuries, based on the recently proposed thoracolumbar AOSpine injury score. The authors have expressed their optimism concerning this new algorithm for its potential to become the new standard for research, teaching, and clinical decision making for thoracolumbar spine injuries. However, they believe that further prospective clinical studies are necessary to validate this algorithm and to assess its outcome [29].

This study compared the reliability of this newly developed AOSpine Thoracolumbar Spine Injury Classification system with that of TLICS. There exists paucity of multicentre studies evaluating the reliability of TLICS across geographic boundaries [4]. Our study results revealed moderate interobserver and intraobserver reliability with regard to diagnosing posterior ligamentous complex injury according to TLICS. With regard to classifying fracture morphology according to TLICS, majority of studies mentioned in the existing literature have shown substantial agreement [30–32]. However, in our study only moderate interobserver and intraobserver reliability ($k = 0.43 \pm 0.01$ and 0.59 ± 0.16, respectively) was observed for classifying fracture type. The possible reason explaining this discrepancy may be because majority of surgeons involved with prior studies mentioned in the literature were those who were involved in the creation of these classification systems or most of them were being carried out in a single geographic region [30–32]. It may be the reflection of the disconnection between the "developer" and "evaluator" which prevents a classification system from adapting, expanding, and evolving beyond its original form [5].

The previous AO-Magerl classification though inclusive was complex and did not account for neurological status resulting in its limited use in routine clinical practice [4]. Blauth et al. reported fair interobserver reliability for the three main AO-Magerl categories ($k = 0.33$), which further decreased on inclusion of the subgroups [33]. Similarly Oner et al. also showed fair interobserver reliability ($k = 0.35$) and moderate intraobserver reliability ($k = 0.41$) in their study on assessment of reliability for AO-Magerl classification [34]. Kriek and Govender reported fair interobserver reliability in the first session ($k = 0.291$) and moderate during second session ($k = 0.403$) for the three main AO-Magerl categories. Intraobserver reliability values ranged from $k = 0.181$ to 0.488 in their study [35]. Wood et al. evaluating AO-Magerl classification system at its simplest level (type A, B or C), demonstrated moderate interobserver reliability ($k = 0.475$) and substantial intraobserver reliability ($k = 0.63$) [25]. In comparison to prior studies regarding reliability of AO-Magerl classification, our interrater and intrarater reliability ($k = 0.59 \pm 0.01$ and 0.68 ± 0.13 respectively) is better with regard to fracture classification using the recently introduced AOSpine classification. This new classification has been developed taking measures to overcome the lacunae which existed with prior AO-Magerl classification. Thus even though it is comprehensive, it is simple as shown by better reliability results in our study. It also includes clinical factors relevant for surgical decision making, like assessment of neurological status and patient specific co morbidities [11].

The authors of the newly proposed AOSpine classification demonstrated substantial interobserver and intraobserver reliability for evaluating fracture type (interobserver k value $= 0.64$ and 0.72, intraobserver k value $= 0.77$ and 0.85 with and without regard fracture subtype respectively) in their study [11]. Urrutia et al. in their independent reliability study with regard to AOSpine classification demonstrated moderate interobserver reliability (k value $= 0.57$) according to fracture subtype, substantial without fracture subtype (k value $= 0.62$), and substantial intraobserver reliability (k value $= 0.71$ and 0.77, with and without fracture subtype respectively) [20]. Kepler et al. assessed the reliability of the AOSpine classification among a worldwide group of spinal surgeons. Similar to results of Urrutia et al. study, they demonstrated moderate interobserver reliability (k value $= 0.56$) according to fracture subtype and substantial without fracture subtype ($k = 0.76$), and substantial intraobserver reliability (k value $= 0.68$ and 0.81, with and without fracture subtype respectively) [21]. In the present study our results show a moderate interobserver reliability with and without fracture subtype (k value $= 0.45 \pm 0.01$ and 0.59 ± 0.01, respectively) and substantial intraobserver reliability (k value $= 0.61 \pm 0.13$ and 0.68 ± 0.13, with and without fracture subtype respectively). In the existing literature it has been shown that reliability coefficients (kappa value) of independent studies have been lower than those performed by the original group who has proposed it, similar to results in the present study [20].

According to TLICS three types of fracture pattern have been described: (1) Compression injuries, (2) translation/rotation injuries, and (3) distraction injuries [14]. These are similar to three main fracture types as mentioned by previous AO-Magerl classification (type A, B and C) [9]. The compression injury pattern (type A) according to new AOSpine classification is similar to that mentioned according to TLICS or main type A of previous AO-Magerl. An exception is the "A0" type in the new AOSpine classification, representing no vertebral fracture or insignificant transverse process or spinous process fractures [9, 11]. The lack of involvement of posterior structures clearly differentiates them from other injury patterns. This was evident by reasonable kappa agreement values for compression injuries according to both TLICS and new AOSpine classification in our study. Least interobserver agreement was seen for distraction injuries according to TLICS ($k = 0.28$), similarly for type B (distraction) injuries according to new AOSpine classification ($k = 0.40$). Significant difference was seen between interobserver kappa values for grading type C injuries according to new AOSpine classification ($k = 0.71$) and translation/rotational injuries according to TLICS ($k = 0.36$), similar to main type C injuries according to previous AO-Magerl

classification. Similarly overall interobserver and intraobserver agreement for grading fracture type regardless of subtype using new AOSpine classification were better than that using TLICS. Both interobserver and intraobserver Spearman correlation statistics with regard to grading of fracture type were higher for the AOSpine classification than TLICS. Also on classifying fracture morphology unanimously according to both classification systems, better agreement was seen with AOSpine classification (32 vs. 22 %). These differences may reflect the changes brought about in new AOSpine classification which has made it simpler and comprehensive with better reproducibility in comparison to TLICS morphology component, which was similar to previous AO-Magerl classification. To further substantiate this view regarding AOSpine classification, Sadiqi et al. in their recent international validation study involving 100 spine surgeons demonstrated that the experience level of spine surgeons did not substantially influence the classification and intraobserver reliability concerning newly introduced AOSpine classification for thoracolumbar spine injuries [36]. With regard to evaluating agreement between surgeons for assessment of neurological status, near perfect agreement was seen using both the classification systems.

Reliability and reproducibility of fracture morphology forms the backbone of any classification system [11]. Our study data demonstrate that the newly proposed AOSpine classification has better reliability for identifying fracture morphology than the existing TLICS. The previous AO-Magerl and TLICS classifications had an element of bias with regard to their selective adoption by surgeons in the regions where they have been developed (Europe for AO-Magerl system and North America for TLICS), although neither of them gained a worldwide acceptance [20]. Highlights of our study were that majority of surgeons involved with the present study were not the members of STSG or AOSpine Knowledge forum, who have proposed the classification systems being reviewed in the present study thus minimizing any bias. Although quite a few of them belonged to the Spine Trauma Study Group of International Spinal Cord Society (ISCoS), it was a multicentre study involving a representation of surgeons from India, USA, Germany and Bangladesh (Table 6) ensuring a more authentic reliability assessment regardless of locale and minimizing any bias as discussed above.

Table 6 Demographics of participant respondents

Region	Number
India	7
USA	2
Germany	1
Bangladesh	1

We do recognize limitations in the study. Fifty consecutive cases were selected retrospectively and their representative images provided to assessors in power point format, which may have limited the accuracy in image interpretation. Since case series selection was done as 50 consecutive cases rather than random selection less severe grades of injuries may have predominated on the basis of incidence, thus having an impact on the reliability analysis.

In conclusion the present multicentre study, being the first in the existing literature to the best of our knowledge comparing reliability of recently proposed AOSpine classification and existing TLICS for thoracolumbar injuries, showed better reliability of the new AOSpine classification for identifying the fracture morphology. Additional studies are clearly necessary concerning the application of these classification systems across multiple physicians at different level of training and trauma centers to evaluate not only their reliability and reproducibility, but also the other attributes of a good classification system. Even though newer classification systems have come out which supposedly have improvement over the previous ones, the endeavor to improve to strive for an ideal one is likely to continue.

Compliance with ethical standards

Conflict of interest None of the authors has any potential conflict of interest.

References

1. Middendorp JJV, Audige L, Hanson B et al (2010) What should an ideal spinal injury classification system consist of? A methodological review and conceptual proposal for future classifications. Eur Spine J 19:1238–1249
2. Mirza SK, Mirza AJ, Chapman JR et al (2002) Classifications of thoracic and lumbar fractures: rationale and supporting data. J Am AcadOrthop Surg 10:364–377
3. Bohler L (1930) Die techniek de knochenbruchbehandlungimgrieden und imkriege. Veralag von Wilhelm Maudrich. **(in German)**
4. Patel AA, Vaccaro AR (2010) Thoracolumbar spine trauma classification. J Am Acad Orthop Surg 18:63–71
5. Bono CM, Vaccaro AR, Hurlbert RJ et al (2006) 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 20:567–572
6. Denis F (1983) The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. Spine 8:817–831
7. Agus H, Kayali C, Arslantas M (2004) Non-operative treatment of burst type thoracolumbar vertebral fractures: clinical and radiological results of 29 patients. Eur Spine J 14:536–540
8. Wood K, Buttermann G, Mehbod A et al (2003) Operative compared with non-operative treatment of thoracolumbar burst fractures without neurological deficit: a prospective randomized study. JBJS Am 85:773–781

9. Magerl F, Aebi M, Gertzbein SD, Harms J et al (1994) A comprehensive classification of thoracic and lumbar injuries. Eur Spine J 3:184–201

10. Sethi MK, Schoenfeld AJ, Bono CM et al (2009) The evolution of thoracolumbar injury classification systems. Spine J 9:780–788

11. Vaccaro AR, Oner C, Kepler CK et al (2013) AOSpine thoracolumbar spine injury classification system: fracture description, neurological status and key modifiers. Spine (Phila Pa 1976) 38:2028–2037

12. Vaccaro AR, Zeiller SC, Hulbert RJ et al (2005) The thoracolumbar injury severity score: a proposed treatment algorithm. J Spinal Disord Tech 18:209–215

13. Harrop JS, Vaccaro AR, Hurlbert RJ et al (2006) 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. Invited submission from the joint section meeting on disorders of the spine and peripheral nerves, March 2005. J Neurosurg Spine 4:118–122

14. Vaccaro AR, Baron EM, Sanfilippo J et al (2006) Reliability of a novel classification system for thoracolumbar injuries: the thoracolumbar injury severity score. Spine (Phila Pa 1976) 231:S62–S69

15. Rampersaud YR, Fisher C, Wilsey J et al (2006) Agreement between orthopedic surgeons and neurosurgeons regarding a new algorithm for the treatment of thoracolumbar injuries. A multicenter reliability study. J Spinal Disord Tech 19:477–482

16. Vaccaro AR, Lehman RA Jr, Hurlbert RJ et al (2005) A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. Spine (Phila Pa 1976) 30:2325–2333

17. Patel AA, Dailey A, Brodke DS et al (2009) Thoracolumbar spine trauma classification: the Thoracolumbar Injury Classification and Severity Score system and case examples. J Neurosurg Spine 10:201–206

18. Moore TA, Bransford RJ, France JC et al (2014) Low lumbar fractures. Does Thoracolumbar Injury Classification and Severity Score work? Spine (Phila Pa 1976) 39:E1021–E1025

19. Joaquim AF, Bastos DCDA, Torres HHJ (2015) Thoracolumbar Injury Classification and Injury Severity Score system: a literature review of its safety. Global Spine J. **(Epub ahead of print)**

20. Urrutia J, Zamora T, Yurac R et al (2014) An independent interobserver reliability and intraobserver reproducibility evaluation of the new AOSpine Thoracolumbar Spine Injury Classification System. Spine (Phila Pa 1976) 40:E54–E58

21. Kepler CK, Vaccaro AR, Koerner JD, et al (2015) Reliability analysis of the AOSpine Thoracolumbar Spine Injury Classification System by a worldwide group of naïve spinal surgeons. Eur Spine J. **(Epub ahead of print)**

22. Kirshblum SC, Burns SP, Biering-Sorenson F et al (2011) International standards for neurological classification of spinal cord injury (Revised 2011). J Spinal Cord Med 34:535–546

23. Landis JR, Koch GC (1977) The measurement of observer agreement for categorical data. Biometrics 36:207–216

24. Chhabra HS, Kaul R, Kanagaraju V (2015) Do we have an ideal classification system for thoracolumbar and subaxial cervical spine injuries: what is the expert's perspective? Spinal Cord 53:42–48

25. Wood KB, Khanna G, Vaccaro AR et al (2005) Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. J Bone Joint Surg Am 87:1423–1429

26. Rihn JA, Yang N, Fischer C et al (2010) 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 12:391–396

27. Middendorp JJV, Patel AA, Schuetz, Joaquim AF (2013) The precision, accuracy and validity of detecting posterior complex injuries of the thoracic and lumbar spine: a critical appraisal of the literature. Eur Spin J 22:461–474

28. Dai LYMDP, Ding WGMDM, Wang XYMDP et al (2009) Assessment of ligamentous injury in patients with thoracolumbar burst fractures using MRI. J Trauma Injury Infect Crit Care 66:1610–1615

29. Vaccaro AR, Schroeder GD, Kepler CK, et al (2015) The surgical algorithm for the AOSpine thoracolumbar spine injury classification system. Eur Spin J. **(Epub ahead of print)**

30. Whang PG, Vaccaro AR, Poelstra KA et al (2007) The influence of fracture mechanism and morphology on the reliability and validity of two novel Thoracolumbar Injury Classification Systems. Spine. 32:791–795

31. Koh YD, Kim DJ, Koh YW (2010) Reliability and validity of Thoracolumbar Injury Classification and Severity Score (TLICS). Asian Spine J 4(2):109–117

32. Kim WC, Lee KY, Kang JH et al (2012) Comparison of TLICS and McAfee classification in thoracolumbar injuries. J Korean Soc Spine Surg 19:8–15

33. Blauth M, Bastian L, Knop C et al (1999) Inter-observer reliability in the classification of thoracolumbar spinal injuries. Orthopade 28:662–681 **(in German)**

34. Oner FC, Ramos LM, Simmermacher RK et al (2002) Classification of thoracic and lumbar spine fractures: problems of reproducibility. a study of 53 patients using CT and MRI. Eur Spine J 11:235–245

35. Kriek JJ, Govender S (2006) AO-classification of thoracic and lumbar fractures—reproducibility utilizing radiographs and clinical information. Eur Spine J 15:1239–1246

36. Sadiqi S, Oner FC, Dvorak MF et al (2015) The influence of spine surgeon's experience on the classification and intraobserver reliability of the novel AOSpine thoracolumbar spine injury classification system—an international validity study. Spine (Phila Pa 1976) 40(23):E1250–E1256