Reliability of the thoracolumbar injury classification and severity score compared with the McAfee classification among young neurosurgeons

Je Hoon Lee¹, Deok-Ki Jeong², Hyun-Woo Lee³, Jae-Won Yu⁴, Dong-Joo Yun⁴, Sang-Soak Ahn⁵, Young-Min Kwon¹

¹Department of Neurosurgery, College of Medicine, Dong-A University, Busan, Korea
²Department of Neurosurgery, Cheomdan Medical Clinic, Gwangju, Korea
³Department of Neurosurgery, Gupo Sangshim Hospital, Busan, Korea
⁴Department of Neurosurgery, Spine Health Wooridul Hospital, Busan, Korea
⁵Barun Spine & Joint Neurosurgery, Busan, Korea

Objective: The reliability of the thoracolumbar injury classification and severity (TLICS) score is well established; however, its reliability among young neurosurgeons in particular has not been investigated. This study was designed to identify intra- and inter-observer differences between the TLICS system and the McAfee classification among young neurosurgeons, with the goal of facilitating communication between physicians and treatment decision-making for patients with thoracolumbar injuries.

Methods: Six young neurosurgeons reviewed thoracolumbar spinal fracture patients between January 2016 and October 2020 and analyzed thoracolumbar fractures according to the 2 classification systems. The intra- and inter-observer reliability of the TLICS and the McAfee scale was assessed with the Cohen and Fleiss kappa tests.

Results: The intra-observer kappa value for the TLICS exhibited excellent reliability (κ = 0.85) compared to the McAfee classification (κ = 0.79). The inter-observer kappa values for each category of the TLICS were 0.69 (morphology), 0.93 (neurologic status), 0.74 (posterior ligamentous complex), and 0.72 (total score). The kappa value of the McAfee classification was lower (κ = 0.52).

Conclusion: The TLICS system showed higher reliability than the McAfee classification. The TLICS score showed more consistent results for thoracolumbar spinal fractures and may thus serve as a guideline for young neurosurgeons in treating patients with thoracolumbar fractures.

Keywords: Spinal injuries; Reliability; Classification; Thoracolumbar fracture; McAfee

Introduction

Although spine fractures do not occur in all trauma patients, their impact on the patients’ life is more significant than that of other injuries. Therefore, early diagnosis and proper treatment are necessary to avoid complications and neurological deficits in patients with thoracolumbar injury. Currently, the following 4 thoracolumbar injury classification systems are widely used: the Denis classification [1], the AO classification [2], the load-sharing classification [3], and the thoracolumbar injury classification and severity score (TLICS) [4]. Despite extensive research on these systems, none is fully satisfying and spine surgeons therefore choose their own classification methods depending on their personal preferences [5,6]. We focused on the TLICS score for diagnosing thoracolumbar spine fractures and aimed to determine the reliability of this classification system in young neurosurgeons. Intra- and inter-observer differences between the TLICS system and the McAfee classification were studied to identify which classification method may facilitate communication between physicians and help young neurosurgeons establish a treatment plan for patients with thoracolumbar injury.
Material and Method

Six young Korean neurosurgeons, who had all obtained their specialist certification in neurosurgery in less than 4 years, reviewed thoracolumbar spinal fracture patients between January 2016 and October 2020. Patients with spinous process fractures, transverse process fractures alone, or pathologic fractures were excluded. The patients were categorized according to their level of injury: thoracic (T1–T10, n = 27), thoracolumbar (T11–L2, n = 131), and lumbar (L3–L5, n = 43) spine injury. The 6 observers analyzed patients’ thoracolumbar spinal fractures according to the 2 classification methods, the TLICS and the McAfee. They were well trained on both methods through a review of studies, and independently reviewed the patients’ history and neurologic, plain film, computed tomography (CT), and magnetic resonance imaging (MRI) findings. The patients’ data were re-evaluated after 4 weeks.

The TLICS is scored based on 3 categories: morphology of the injury, integrity of the posterior ligamentous complex (PLC), and the patient’s neurologic status (Table 1). Widening of the interspinous space, diastasis of the facet joint, and facet perch or subluxations on X-ray and/or CT indicate an injured PLC [7,8]. However, the most credible sign of PLC injury is discontinuation of the low-signal-intensity black strip on sagittal T1- or T2-weighted MR images [9].

The McAfee classification is subdivided into 6 categories: wedge compression, stable burst fracture, unstable burst fracture, chance fracture, flexion-distraction injury, and translational injury. An unstable burst fracture is defined as a burst fracture with a neurologic deficit, three-column injury, kyphosis over 30 degrees, a decrease of anterior body height over 40%, and more than 50% canal compromise [10].

The intra- and inter-observer reliability was assessed by Cohen’s and Fleiss’ kappa values. Statistical analyzes were performed using IBM SPSS ver. 20.0 (IBM Corp., Armonk, NY, USA).

The kappa value was interpreted using the Landis and Koch [11] grading system, which defines k < 0.2 as slight agreement, 0.21 to 0.40 as fair agreement, 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as substantial reliability, and > 0.81 as excellent reliability.

Patients were informed of the use of medical information in the study and informed consent was obtained.

Results

During the study period, 201 patients (109 males and 92 females) visited our hospital. The mean age was 60.5 years (range, 18–88), with the age distribution as follows: 3 (10–19 years), 13 (20–29 years), 11 (30–39 years), 26 (40–49 years), 30 (50–59 years), 35 (60–69 years), 53 (70–79 years), and 30 (80–89 years). Thoracic spine fractures were noted in 27 (13.4%) patients, lumbar spine fractures in 43 (21.4%), and thoracolumbar spine fractures in 131 (65.2%). Spinal fusion operation was performed in 115 (57.2%) patients, percutaneous vertebroplasty or kyphoplasty in 62 (30.8%), and conservative treatment with a thoracolumbosacral orthosis for 8 to 12 weeks in 24 (11.9%) (Table 2).

The intra-observer kappa value showed almost perfect agreement according to the Landis and Koch [11] grading system in each category of the TLICS, with κ = 0.85 (morphology), κ = 0.95 (neurologic status), κ = 0.87 (integrity of the PLC), and κ = 0.85 (total score of the TLICS), while the kappa value for the McAfee classification was κ = 0.79, representing substantial agreement (Table 3).

Table 1. Thoracolumbar injury classification and severity scoring

| Category              | Qualifiers       | Score |
|-----------------------|------------------|-------|
| Injury morphology     |                  |       |
| Compression           |                  | 1     |
| Burst                 | +1               |       |
| Translational/rotational |              | 3     |
| Distraction           |                  | 4     |
| Neurological status   |                  |       |
| Intact                |                  | 0     |
| Nerve root            |                  | 2     |
| Cord, conus medullaris| Complete        | 2     |
|                       | Incomplete       | 3     |
| Cauda equina          |                  | 3     |
| Integrity of the PLC  |                  |       |
| Intact                |                  | 0     |
| Injury suspected or indeterminate | | 2     |
| Injured               |                  | 3     |

PLC, posterior ligamentous complex.

Table 2. Demographics of patients

| Demographic     | Value     |
|-----------------|-----------|
| No. of patients | 201       |
| Sex (male:female) | 109:92  |
| Mean age (range), yr | 60.5 (18–88) |
| Injury level (n, %) |         |
| Thoracic (T1–10) | 27 (13.4) |
| Thoracolumbar (T11–L2) | 131 (65.2) |
| Lumbar (L3–5) | 43 (21.4) |
| Treatment (n, %) |           |
| Spinal fusion | 115 (57.2) |
| Vertebroplasty/kyphoplasty | 62 (30.8) |
| Conservative care | 24 (11.9) |
The inter-observer kappa value in each category of the TLICS was $\kappa = 0.69$ (morphology; substantial agreement according to the Landis and Koch [11] grading system), $\kappa = 0.93$ (neurologic status; almost perfect agreement), $\kappa = 0.74$ (integrity of the PLC; substantial agreement), and $\kappa = 0.72$ (total score of the TLICS; substantial agreement), while the kappa value for the McAfee classification was lower, at $\kappa = 0.52$, representing moderate agreement (Table 4).

### Discussion

Spinal injury treatment remains a challenge, despite the developments in diagnostic and treatment systems for trauma patients. When accompanied by neurological deficits, thoracolumbar injuries have an emotional and economic effect on patients and their families, in addition to the physical disabilities they cause. Traumatic thoracolumbar injury is associated with a poor prognosis, even when aggressive rehabilitative treatments are provided. Numerous thoracolumbar spine injury classification systems have been introduced to aid clinical and surgical treatment [1–4,10,12–15]. The first classification of traumatic thoracolumbar injury was reported by Böhler and Böhler [12] in 1929. Watson-Jones [16] introduced a classification system based on the concept of “instability” in 1938, recognizing that posterior ligamentous integrity is a key element for spinal stability. In 1949, Nicoll [15] defined the concept of “stability” using an anatomic classification and identified 4 anatomical structures involved in mechanical stability—vertebral bodies, facet joints, posterior ligaments, and disks. Holdsworth [14] first recommended the modern classification of fractures based on a two-column theory, and this method employed simple x-rays. With the rapid development of radiology, Denis [1] proposed a three-column theory complementing the two-column theory and divided the spine into anterior, middle, and posterior columns. This method focused on the stability of the middle column measured by CT. Depending on the injury mechanism and the degree of injury, this system classified spinal injuries into 4 groups—compression fractures, burst fractures, seatbelt injuries, and fracture-dislocations. However, McAfee et al. [10] pointed out that this system overestimates the influence of middle column stability, thereby increasing the number of unnecessary surgical procedures. McAfee classified fractures into 6 types depending on compression, distraction, and direct shearing force on the middle column as determined on CT scans. Consequently, bursting fractures were subclassified into stable fractures and unstable fractures, and the latter were further subclassified based on the three-column theory, including posterior column injury. Magerl et al. [2] introduced the AO classification system and categorized thoracolumbar injuries into A type injuries by compression force, B type injuries by distraction force, and C type injuries by torsional force, with subcategories from group 1 to 3 based on the degrees of injuries.

Likewise, many classification methods of thoracolumbar injuries have been introduced, but these methods have been found to show diagnostic variability depending on personal point of view and low inter-observer reliability [7,17]. Moreover, the need for a new classification system reflecting the importance of soft tissue such as the PLC using MRI has emerged. Therefore, Vaccaro et al. [4] developed the TLICS classification to overcome these limitations. The TLICS is based on injury morphology, neurological status, and PLC integrity (Table 1). Treatment plans are decided according to scores for each category: $\leq 3$ points indicate conservative care, $\geq 5$ points indicate surgical treatment, and a score of 4 allows for either option. Existing classification systems emphasize mechanisms of injury assumed by observers, while the TLICS emphasizes objective analysis of the injury.

Consequently, many classification systems have been developed to offer better treatment to patients. Blauth et al. [5] have reported fair inter-observer reliability ($\kappa = 0.33$) for only the 3 main types (A, B, C) of the AO classification, and decreasing reliability with the inclusion of the AO subtypes. Oner et al. [17] and Wood et al. [7] have reported that the Denis classification system has higher

### Table 3. Intra-observer results of the McAfee classification and TLICS

| Category                      | Kappa value | P-value | Strength of agreement |
|-------------------------------|-------------|---------|-----------------------|
| McAfee classification         |             |         |                       |
| Injury morphology             | 0.79        | 0.020   | Substantial           |
| Neurologic status             | 0.85        | 0.032   | Almost perfect         |
| Intensity of the PLC          | 0.87        | $< 0.001$ | Almost perfect          |
| Total score                   | 0.85        | 0.021   | Almost perfect         |

### Table 4. Inter-observer results of the McAfee classification and TLICS

| Category                      | Kappa value | P-value | Strength of agreement |
|-------------------------------|-------------|---------|-----------------------|
| McAfee classification         |             |         |                       |
| Injury morphology             | 0.52        | 0.035   | Moderate              |
| Neurologic status             | 0.69        | 0.027   | Substantial           |
| Intensity of the PLC          | 0.93        | $< 0.001$ | Almost perfect          |
| Total score                   | 0.74        | 0.012   | Substantial           |

TLICS, thoracolumbar injury classification and severity; PLC, posterior ligamentous complex.
inter-observer reliability than the AO classification system (Oner, \( \kappa = 0.60, 0.35 \); Wood, \( \kappa = 0.606, 0.475 \)). However, both classification systems showed only fair to moderate inter-observer reliability. The AO classification system includes much information about the injury, and its complexity and the consequent low reproducibility limit its clinical and surgical application \[1,5,7,17\]. While the Denis classification is simple, it does not consider anatomical and pathophysiologic factors such as PLC or nerve injury \[18\].

Given the numerous limitations of these classification systems, the TLICS system was found favorable in many studies. Whang et al. \[8\] reported satisfactory reliability of the TLICS with substantial agreement (\( \kappa = 0.626 \)) on injury morphology, moderate agreement (\( \kappa = 0.447 \)) on PLC integrity, and moderate agreement (\( \kappa = 0.455 \)) on the total score. However, the intra- and inter-observer reliability of the TLICS is not well-studied, especially among young neurosurgeons. Therefore, this study evaluated the intra- and inter-observer reliability between the TLICS and McAfee classification systems. The intra-observer reliability of the McAfee classification system shows substantial reliability, but that of the TLICS shows almost perfect reliability (Table 3). Fleiss’ kappa test on inter-observer reliability revealed high reliability in all categories (injury morphology = 0.69, neurologic status = 0.93, PLC integrity = 0.74, total score = 0.72) for the TLICS, while the kappa value for the McAfee classification was 0.52, representing only moderate reliability (Table 4, Fig. 1). Among the subcategories of the TLICS classification, injury morphology showed the lowest value, assuming that the ratio of flexion-distraction and bursting fracture were more significant than others among the 201 cases included in this study. Accordingly, the TLICS system has higher statistical significance than the McAfee classification regarding conformity and consistency. The young neurosurgeons in this study found the TLICS system to have higher reliability than the McAfee classification in suggesting treatment plans for patients with thoracolumbar injury. Moreover, it can facilitate communication among young neurosurgeons.

This study has, however, some limitations. First, this is a retrospective analysis based on clinical records of patient information. Therefore, records with insufficient information were excluded to improve the accuracy of our results. Second, all researchers in this study received their education from the same hospital; therefore, the neurosurgeons included in this study do not represent the general neurosurgeon population.

**Conclusion**

This study conducted in young neurosurgeons shows that the TLICS system has higher reliability than the McAfee classification. Hence, the TLICS system can assist young neurosurgeons better in establishing treatment plans for patients with thoracolumbar spinal fractures and eventually help establish a reliable guideline for all neurosurgeons.

**Conflicts of interest**

No potential conflict of interest relevant to this article was reported.

**ORCID**

Je Hoon Lee, https://orcid.org/0000-0002-8199-2536

https://doi.org/10.51638/jksgn.2022.00080
REFERENCES

1. 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.

2. Magerl F, Aebi M, Gertzbein SD, Harms J, Nazarian S. A comprehensive classification of thoracic and lumbar injuries. Eur Spine J 1994;3:184–201.

3. McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. Spine (Phila Pa 1976) 1994;19:1741–4.

4. Vaccaro AR, Lehman RA Jr, Hurlbert RJ, 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 1976) 2005;30:2325–33.

5. Blauth M, Bastian L, Knop C, Lange U, Tusch G. Inter-observer reliability in the classification of thoraco-lumbar spinal injuries. Orthopade 1999;28:662–81.

6. Lewkonia P, Paolucci EO, Thomas K. Reliability of the thoracolumbar injury classification and severity score and comparison with the Denis classification for injury to the thoracic and lumbar spine. Spine (Phila Pa 1976) 2012;37:2161–7.

7. Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB, Mehbod AA. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. J Bone Joint Surg Am 2005;87:1423–9.

8. 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.

9. Pizones J, Zúñiga L, Sánchez-Mariscal F, Alvarez P, Gómez-Rice A, Izquierdo E. MRI study of post-traumatic incompetence of posterior ligamentous complex: importance of the supraspinous ligament: prospective study of 74 traumatic fractures. Eur Spine J 2012;21:2222–31.

10. McAfee PC, Yuan HA, Fredrickson BE, Lubicky JP. The value of computed tomography in thoracolumbar fractures. An analysis of one hundred consecutive cases and a new classification. J Bone Joint Surg Am 1983;65:461–73.

11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

12. Böhler L, Böhler Jr. The treatment of fractures. 5nd ed. New York: Grune & Stratton; 1956.

13. Ferguson RL, Allen BL Jr. A mechanistic classification of thoracolumbar spine fractures. Clin Orthop Relat Res 1984;(189):77–88.

14. Holdsworth F. Fractures, dislocations, and fracture-dislocations of the spine. J Bone Joint Surg Am 1970;52:1534–51.

15. Nicoll EA. Fractures of the dorso-lumbar spine. J Bone Joint Surg Br 1949;31B:376–94.

16. Watson-Jones R. The results of postural reduction of fractures of the spine. J Bone Joint Surg 1938;20:567–86.

17. 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.

18. Koh YD, Kim DJ, Koh YW. Reliability and validity of Thoraco-lumbar Injury Classification and Severity Score (TLICS). Asian Spine J 2010;4:109–17.