

Reliability and Validity of Thoracolumbar Injury Classification and Severity Score (TLICS)

Young Do Koh, Dong Jun Kim, Young Won Koh

Department of Orthopedic Surgery, Ehwa Womans University School of Medicine, Seoul, Korea

Study Design: A new classification system for throacolumbar spine injury, Thoracolumbar Injury Classification and Severity Score (TLICS) was evaluated retrospectively.

Purpose: To evaluate intrarater and interrater reliability of newly proposed TLICS schemes and to estimate validity of TLICS's final treatment recommendation.

Overview of Literature: Despite numerous literature about thoracolumbar spine injury classifications, there is no consensus regarding the optimal system.

Methods: Using plain radiographs, computed tomography scanning, magnetic resonance imaging, and medical records, 3 classifiers, consisting of 2 spine surgeons and 1 senior orthopaedic surgery resident, reviewed 114 clinical thoracolumbar spine injury cases retrospectively to classify and calculate injury severity score according to TLICS. This process were repeated on 4 weeks intervals and the scores were then compared with type of treatment that patient ultimately received.

Results: The intrarater reliability of TLICS was substantial agreement on total score and injury morphology, almost perfect agreement on integrity of the posterior ligament complex (PLC) and neurologic status. The interrater reliability was substantial agreement on injury morphology and integrity of the PLC, moderate agreement on total score, almost perfect agreement on neurologic status. The TLICS schemes exhibited satisfactory overall validity in terms of clinical decision making.

Conclusions: The TLICS was demonstrated acceptable intrarater and interrater reliability and satisfactory validity in terms of treatment recommendation.

Key Words: Thoracolumbar spine injury, Classification, TLICS, Reliability, Validity

Introduction

The classification of thoracolumbar injury still remains controversial despite of the studies that have been conducted on this in the past several decades [1-6]. The classification of thoracolumbar fracture was first suggested by Böhler [7] in 1929. Although numerous authors have reported new classification systems afterwards, there is still no officially accepted classification system for thoracolumbar fracture.

In regard to the recent classification of thoracolumbar injury such as the Denis classification system and the Arbeitsgemeinschaft für osteosynthesefragen (AO) classification system, which have been commonly applied in clinical practice, problems with their validity, reliability and reproducibility have been continuously pointed out [8-10]. These problems include shortcomings in traumatic force assessment, accurate morphologic evaluation and effective prognostic assessment of traumatic injuries [11].

However, with the recent remarkable development in

Received Dec 16, 2009; 1st Revised Jan 21, 2010; 2nd Revised Feb 2, 2010; Accepted Feb 2, 2010

Corresponding author: **Young Do Koh, MD**

Department of Orthopedic Surgery, Ehwa Womans University School of Medicine,

911-1 Mok 5-dong, Yangcheon-gu, Seoul 158-710, Korea

Tel: +82-2-2650-5664, Fax: +82-2-2642-0349, E-mail: ydkoh@ewha.ac.kr

radiology and with the consequent widening knowledge of the anatomy as well the biomechanics of thoracolumbar injury, continuous efforts have been made to establish a more ideal classification system that is applicable to clinical practice, and as one of these efforts, Vaccaro et al. [12] in 2005 suggested the Thoracolumbar Injury Classification and Severity Score (TLICS).

Therefore, we evaluated the intraobserver and interobserver reliability of the TLICS, and we assessed the validity of selecting the treatment based on this classification system.

Materials and Methods

1. Materials

From January 2004 to June 2009, among the 168 patients who received operative or conservative treatment for traumatic thoracolumbar injury, 114 cases were recruited as the subjects who had accessible medical records, computed tomography (CT) and magnetic resonance imaging (MRI). In the process of selecting the patient group, to exclude osteoporotic compression fracture, the age of the patient group was arbitrary limited to younger than 56 years, and so the study group was composed of 74 male patients and 40 female patients with a mean age of 40.3 years (range, 15 to 56 years). As for the causative mechanism of injury, 58 cases (51%) were falls, 48 cases (42%) were traffic accidents and 8 cases (7%) were miscellaneous. Thirty two

cases (28%) were associated with musculoskeletal injury, 7 cases (6%) were associated with neurologic injury and 6 cases (5%) were associated with intraabdominal organ injury.

2. Methods

The score of the patient group, according to the TLICS, was retrospectively examined 2 times by 3 classifiers who were orthopedic surgeons, by viewing the thoracolumbar simple radiographs, the CT, the MRI and the medical records at 4 week intervals.

The TLICS consists of three variables: 1) the morphology of the fracture pattern as assessed by radiological tests, 2) whether injury in the posterior ligamentous complex (PLC) is involved or not, and 3) the neurologic status of the patient (Table 1). After scoring each variable, the treatment plans are determined according to the sum (Table 2). When scoring the variables of the TLICS, the morphology of injury, based on the radiology of the thoracolumbar injury, was classified as 3 types: 1) the compression type, 2) the translocation/rotation type and 3) the distraction type, and they were each scored as 1, 3 and 4 points, respectively. Additionally, concerning the presence of a burst fracture with the compression injuries, 1 point was added and thus it was scored as 2 points. For cases with multiple injuries of an identical type, no more than one injury was added to be scored. For cases with the combination of multiple different

Table 1. Summary of the revised scale, the TLICS

Component	Qualifiers	Score
Morphology type		
Compression		1
	Burst	1
Translational/Rotational		3
Distraction		4
Neurological involvement		
Intact		0
Nerve root		2
Cord, conus medullaris	Complete	2
	Incomplete	3
Cauda equina		3
PLC		
Intact		0
Injury suspected/Indeterminate		2
Injured		3

TLICS: Thoracolumbar Injury Classification and Severity Score, PLC: Posterior ligamentous complex.

Table 2. Summary of scores associated with various fracture management points

Management	Score
Nonoperative	≤ 3
Nonoperative or Operative	4
Operative	≥ 5

Table 3. Interpretation of Kappa statistics

k-value	Agreement
< 0	< Chance
0.01-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-0.99	Almost perfect

types of injuries, only one injury type with the highest points was scored. Regarding the score of the injury in the posterior ligamentous complex, the cases with no other injury were 0 points, the cases with unclear injury were 2 points and the cases with apparent injury were scored as 3 points. As for the score of the neurologic status, the cases without neurological symptoms on the neurologic examination were 0 points, the cases with injury at the nerve root were 2 points and the cases with complete or incomplete neurological injury in the spinal cord and the conus medullaris were 2 points and 3 points, respectively. The cases with injury in the cauda equine were scored as 3

points (Figs. 1 and 2).

Afterward, the intrarater reliability, as well as the interrater reliability of the score for each variable and sum, were evaluated by Cohen's unweighted k -value and Spearman's rank order correlation. For the statistical distinction of the Kappa values, the distinction according to the guideline of Viera and Garrett [13] was applied (Table 3). In addition, the sensitivity and specificity were evaluated by the percent of correct treatment protocols according to the sum of the TLICS and the treatment actually performed on the patients. The validity of selecting treatment protocols according to this classification was also assessed. SPSS ver. 13.0 (SPSS



Fig. 1. A neurologically intact patient (0 point) with compression fracture (1 point) and intact posterior ligamentous complex (0 point). Total score was 1, thus treatment recommendation is non-surgical treatment.

Inc., Chicago, IL, USA) was used for statistical analysis.

Results

The intrarater reliability of each variable of the TLICS showed substantial agreement ($k = 0.753$) for determining the injury pattern, almost perfect agreement ($k = 0.81$) for

whether there was injury in the posterior ligamentous complex, almost perfect agreement ($k = 0.96$) for the neurological status assessment and substantial agreement ($k = 0.724$) for the sum of the total score (Table 4). In regard to the interrater reliability of each variable, determination of the injury pattern showed substantial agreement ($k = 0.608$), there was substantial agreement ($k = 0.641$) for whether

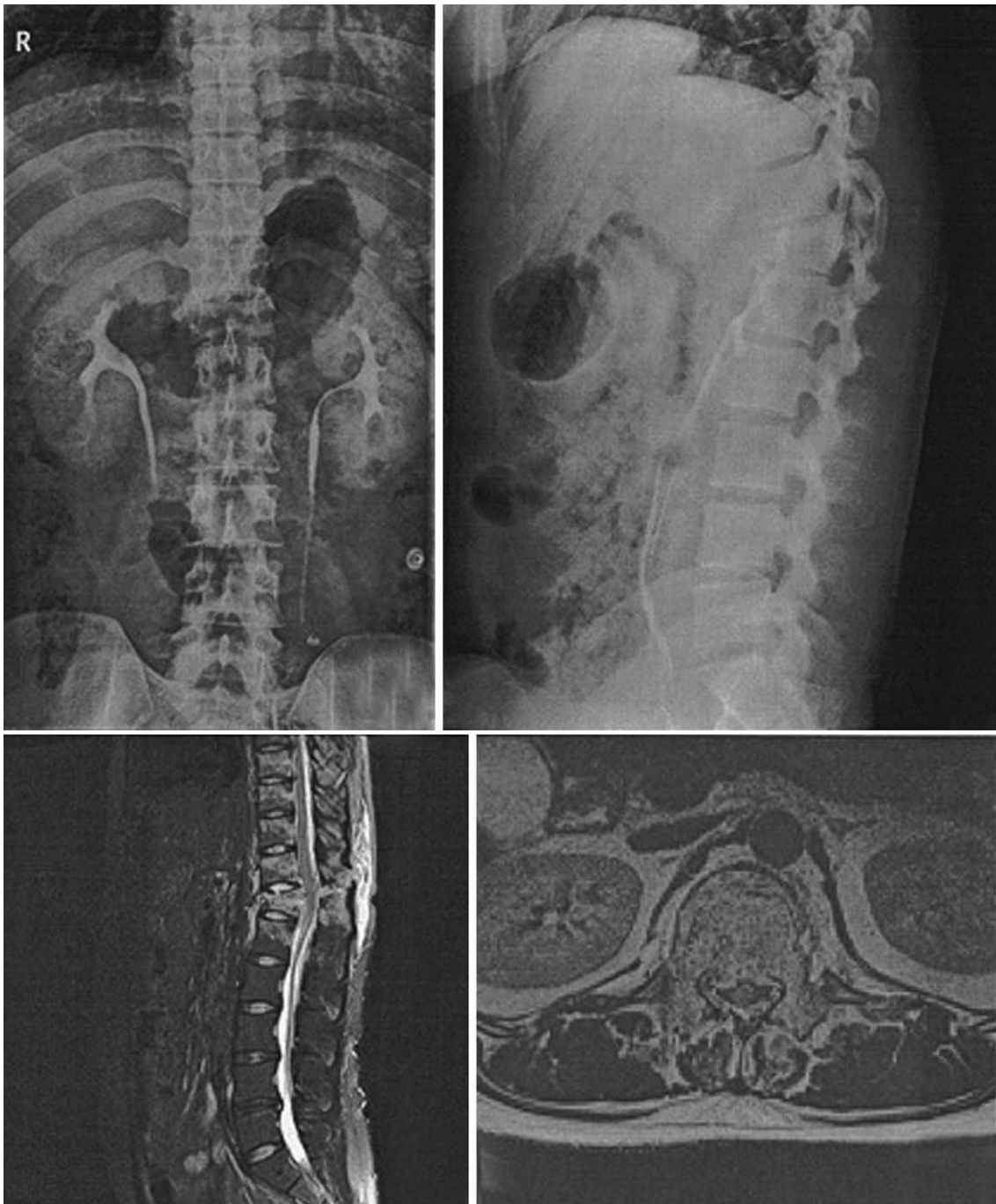


Fig. 2. T12 flexion burst fracture in patient with incomplete spinal cord injury. Compression injury (1 point) with burst component (1 point), injury to posterior ligament complex (3 point) and incomplete spinal cord injury (3 point). Total score was 8, thus treatment recommendation is surgery.

there was injury in the posterior ligament complex, almost perfect agreement ($k = 0.91$) for the neurological status assessment and moderate agreement ($k = 0.576$) for the sum of the total score (Table 5).

Regarding the validity of the selection of treatment protocols according to the TLICS classification, of the total 684 cases, 114 cases were double-measured by the 3 investigators. There were 362 cases for which the actual total TLICS score was higher than 5 points, and among them, surgery was performed for 355 cases. Among the 195 cases with the total score lower than 3 points, surgery was not performed for 176 cases. It was shown that the percent of correctly selecting treatment according to the TLICS was 95%, the sensitivity was 98%, and the specificity was 90% (Table 6).

Discussion

The classification of traumatic thoracolumbar injury was first reported in 1929 by Böhler [7], and afterwards, spinal fracture was classified to 5 types according to the mechanism of injury. In 1949, Nicoll [14] revised the concept of instability, and Nicoll [14] claimed that the fracture gap caused by the comminution of the vertebral body as well as injury of the PLC could induce instability. In 1970, Holdsworth [15] explained thoracolumbar injury by introducing the concept of the two-column theory. Based on this, Louis [16] in 1977 was able to explain the structure of the vertebral body by the three-column concept.

With the rapid development of radiology, Denis [17] in

1983 introduced the three-column concept based on the findings of thoracolumbar CT. According to the injury mechanism and the degree of injury, he classified spinal injury into 4 types of major spinal injury (compression fracture, burst fracture, seat-belt injury and fracture-dislocation) and 4 types of minor spinal injury (articular process fracture, transverse process fracture, spinous process fracture and isthmus fracture). With the widening knowledge of the mechanisms of spinal injury, McCormack et al. [18] in 1994 suggested the loading-sharing classification for determining the necessity of reinforcing the anterior column during surgical treatment for burst fracture. In addition, Magerl et al. [19] considered the major external forces placed on the vertebral body as compression, distraction and rotation and according to this, they reported the AO classification that divides thoracolumbar injury to a total of 53 fracture groups.

In such a manner, numerous classification systems for thoracolumbar injury have been reported, and studies on their validity have been continuously conducted, yet none of them is currently accepted as an optimal classification. The prerequisite of an ideal classification method for thoracolumbar injury includes the following factors: it should be comprehensive and it can be readily applied, the reliability and reproducibility should also be ensured and it should be of help to determine the degree of injury and to decide the treatment strategy, and furthermore, it should be of help to evaluate the prognosis. In addition, it should be able to be used as a common language for effective and reliable com-

Table 4. Intrarater statistics

	Value	Standard error	<i>p</i> -value
Injury morphology			
<i>k</i> -value	0.753	0.032	0
PLC			
<i>k</i> -value	0.81	0.027	0
Neurology change			
<i>k</i> -value	0.96	0	0
TLICS total			
<i>k</i> -value	0.724	0.027	0

PLC: Posterior ligamentous complex, TLICS: Thoracolumbar Injury Classification and Severity Score.

Table 5. Interrater statistics

	Value	Standard error	<i>p</i> -value
Injury morphology			
<i>k</i> -value	0.608	0.026	0
PLC			
<i>k</i> -value	0.641	0.024	0
Neurology change			
<i>k</i> -value	0.91	0	0
TLICS total			
<i>k</i> -value	0.576	0.021	0

PLC: Posterior ligamentous complex, TLICS: Thoracolumbar Injury Classification and Severity Score.

Table 6. Summary of validity data

	% Correct	Sensitivity	Specificity	PPV	NPV
TLICS	95	0.98	0.9	0.95	0.96

PPV: Positive predictive value, NPV: Negative predictive value, TLICS: Thoracolumbar Injury Classification and Severity Score.

munication among clinicians. For this, assessment of both the intraobserver and interobserver reliability is essential.

The classification systems that have been recently applied for thoracolumbar injury most frequently in clinical practice include the Denis [17] classification and the AO classification according to Magerl et al. [19]. Blauth et al. [20] have reported that the interobserver reliability of the basic type of AO classification (A, B and C type) was low (fair agreement, $k = 0.33$), and in regard to the AO classification, when the basic type was subclassified into subgroups, the interobserver reliability decreased even more. Oner et al. [8] and Wood et al. [9] have reported in their study that the Denis classification system showed higher interobserver reliability than did the AO classification system (Oner, $k = 0.60, 0.35$; Wood, $k = 0.606, 0.475$). Nonetheless, the two classification methods showed unsatisfactory interobserver reliability. The AO classification system includes the most abundant information on the classification of injury, and so it has the advantage that it could classify almost any types

of injury, yet it has many limitations to be applied to clinical practice due to the unavoidable complexity and the consequent low reproducibility. On the other hand, the Denis classification is too simple, and so it has the limitation that it does not contain important anatomical and pathophysiologic factors that play an important role in deciding the treatment protocols for injury in the PLC or for nerve injury. In 1986, Bucholz and Grill [21] pointed out that the limitation of the Denis [17] classification system results from the absence of considering the dynamic mechanism of spinal injury and the level of the associated nerve injury.

Vaccaro et al. [22] in 2005 suggested a new classification method, the Thoracolumbar Injury Severity Score (TLISS) and this would compensate for the problems of the previous classification systems and satisfy the requirements of an ideal classification system. This classification scores three variables that are considered to be directly associated with the stability of the vertebral body and the prognosis. These variables are composed of the mechanism of injury as deter-



Fig. 3. A neurologically intact patient (0 point) with compression injury of L2 vertebral body (1 point) with burst component (1 point) and intact posterior ligamentous complex (0 point). Total score was 2, thus treatment recommendation is non-surgical treatment. But, about 40% canal compromise with kyphotic deformity was seen and we treated this patient with surgery.

mined by radiologic tests, the present of an injury in the PLC and the neurologic status, and all of these are scored to help decide on the treatment plans. Afterward, the TLISS was modified to the TLICS that includes the radiological injury pattern (the morphology and the fracture pattern), which could be classified objectively instead of subjectively for determining the mechanisms of injuries. Vaccaro et al. [12] later reported the high reliability and reproducibility of the TLICS. In addition, Whang et al. [23] reported the results of a study conducted on the reliability of the TLICS, and this revealed the satisfactory reliability of the TLICS by showing that determining the injury pattern showed moderate agreement ($k = 0.626$), the presence of injury in the PLC showed moderate agreement ($k = 0.447$) and the total score showed moderate agreement ($k = 0.455$).

As for our study, the interobserver reliability of the TLICS was higher than substantial agreement for all the categories, and the interobserver reliability showed higher than substantial agreement for all the categories, except that the total score showed moderate agreement ($k = 0.576$), and our results were found to be better than the results reported

by Whang et al. [23]. Until now, the k -value required for a clinically reliable classification system has been determined to be higher than 0.55 [24], and when considering the opinion of Oner et al. [25] that the standard is too strict to apply to a classification system for traumatic vertebral fracture, our results could be considered to show excellent reliability. In addition, the results of our study was found to be excellent as compared with the intraobserver and interobserver reliability of the Denis classification system and the AO classification system reported by Oner et al. [8] and Wood et al. [9], respectively.

In other words, when establishing a treatment plan through the TLICS, as compared with the treatment method that was actually applied to patients, the validity of the decision-making for surgery was satisfactory for the percent correct, the sensitivity and the specificity. Nevertheless, the specificity was relatively low (90%), which might have resulted from the fact that the choice of treatment method varied depending on the preference of clinicians in some cases. Indeed, in our study we administered operative treatments for the burst fracture with the invasion of a bony

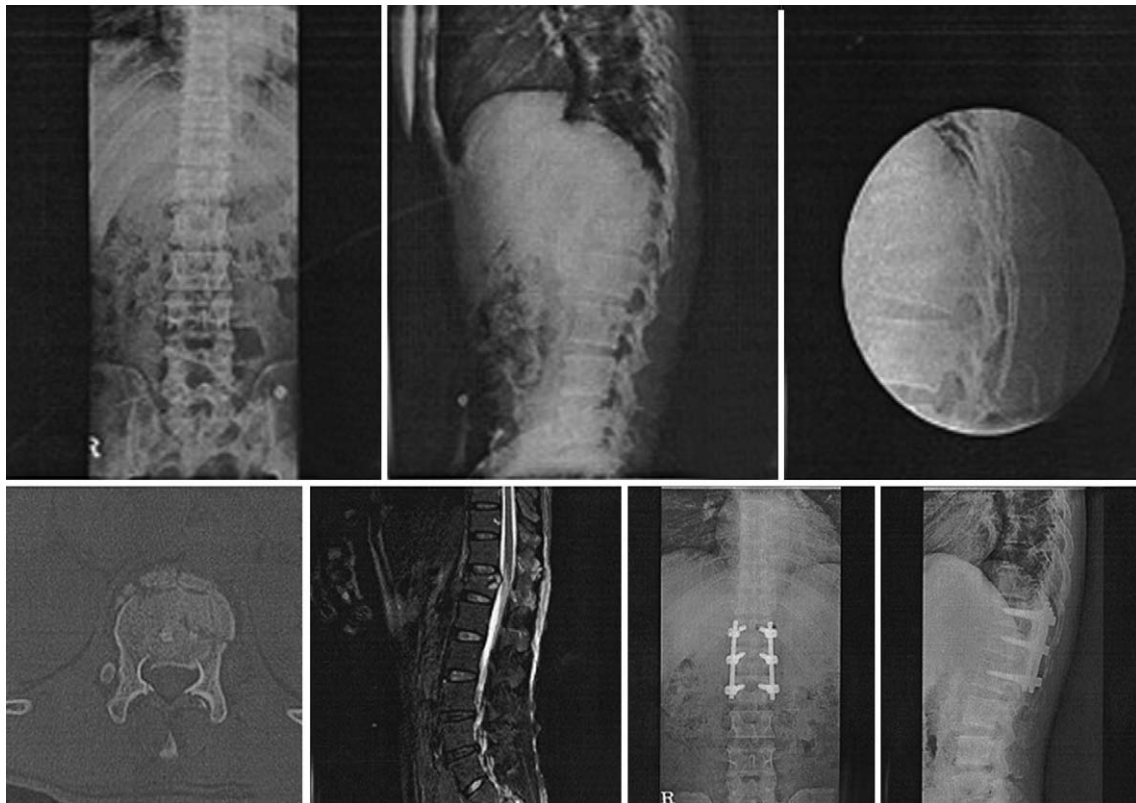


Fig. 4. L1 burst fracture in patient without neurologic symptom (0 point). Compression injury (1 point) with burst component (1 point) and intact posterior ligamentous complex (0 point). Total score was 2, thus treatment recommendation is non-surgical treatment. But, fracture of T12 spinous process was seen and our treatment choice was surgery.

fragment to the spinal canal reaching 50% that were without injury in the PLC and neurologic symptoms (TLICS total score = 2), and for the cases of burst fracture associated with fracture in the lamina or fracture in the adjacent spinous process and these case were without neurologic symptoms and injury in the PLC (TLICS total score = 2) (Figs. 3 and 4). In addition, for the cases that injury in the PLC is not apparent and it is only suspected, operative treatment may be preferred depending on the clinician's opinion. Further evaluation is required when considering that the previous classification systems were unrelated to the selection of treatment plans and the studies on the validity of selecting treatment plans were meager as compared with the studies conducted on the reliability of TLICS.

Conclusions

The ideal classification system of thoracolumbar injury should be comprehensive, easy to apply, it ensures sufficient reliability and reproducibility, and it should be of help to determine the degree of injury and for deciding the treatment plan. Furthermore, the ideal classification system should be helpful for determining the prognosis. The TLICS showed excellent interobserver and intraobserver reliability as well as satisfactory validity for selecting a treatment plan. Therefore, it could be considered a superior classification system that is applicable in clinical practice, as compared with the previous classification systems.

REFERENCES

1. Holdsworth FW. Fractures, dislocations, and fracture-dislocations of the spine. *J Bone Joint Surg Br* 1963;45:6-20.
2. Kelly RP, Whitesides TE Jr. Treatment of lumbodorsal fracture-dislocations. *Ann Surg* 1968;167:705-17.
3. 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.
4. Ferguson RL, Allen BL Jr. A mechanistic classification of thoracolumbar spine fractures. *Clin Orthop Relat Res* 1984;(189):77-88.
5. Gertzbein SD. Spine update: classification of thoracic and lumbar fractures. *Spine (Phila Pa 1976)* 1994;19:626-8.
6. Mirza SK, Mirza AJ, Chapman JR, Anderson PA. Classifications of thoracic and lumbar fractures: rationale and supporting data. *J Am Acad Orthop Surg* 2002;10:364-77.
7. Böhler L. The treatment of fractures. 5th ed. New York: Grune & Stratton, 1956. p. 300-29.
8. 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.
9. Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB, Mehdod AA. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. *J Bone Joint Surg Am* 2005;87:1423-9.
10. Wood K, Buttermann G, Mehdod A, et al. Operative compared with nonoperative treatment of a thoracolumbar burst fracture without neurological deficit: a prospective, randomized study. *J Bone Joint Surg Am* 2003;85:773-81.
11. Dai LY, Jin WJ. Interobserver and intraobserver reliability in the load sharing classification of the assessment of thoracolumbar burst fractures. *Spine (Phila Pa 1976)* 2005;30:354-8.
12. 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.
13. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360-3.
14. Nicoll EA. Fractures of the dorso-lumbar spine. *J Bone Joint Surg Br* 1949;31:376-94.
15. Holdsworth F. Fractures, dislocations, and fracture-dislocations of the spine. *J Bone Joint Surg Am* 1970;52:1534-51.
16. Louis R. Unstable fractures of the spine: III. Instability. A. Theories concerning instability. *Rev Chir Orthop Reparatrice Appar Mot* 1977;63:423-5.
17. 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.
18. McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. *Spine (Phila Pa 1976)* 1994;19:1741-4.
19. 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.
20. Blauth M, Bastian L, Knop C, Lange U, Tusch G. Interobserver reliability in the classification of thoraco-lumbar spinal injuries. *Orthopade* 1999;28:662-81.
21. Bucholz RW, Gill K. Classification of injuries to the thoracolumbar spine. *Orthop Clin North Am* 1986;17:67-73.
22. Vaccaro AR, Zeiller SC, Hulbert RJ, et al. The thoracolum-

- bar injury severity score: a proposed treatment algorithm. *J Spinal Disord Tech* 2005;18:209-15.
23. 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.
24. Sanders RW. The problem with apples and oranges. *J Orthop Trauma* 1997;11:465-6.
25. Oner FC, van Gils AP, Dhert WJ, Verbout AJ. MRI findings of thoracolumbar spine fractures: a categorisation based on MRI examinations of 100 fractures. *Skeletal Radiol* 1999;28:433-43.