

The thoracolumbar injury severity scale and score (TLISS): inter-physician and inter-disciplinary validation of a new paradigm for the treatment of thoracolumbar spine trauma

Escala e pontuação da gravidade dos traumatismos da coluna toracolombar

Moe R. Lim¹
Alexander R. Vaccaro²
Joon Y. Lee,³
Steven C. Zeiller²
Sidney Jacoby²
James SanFilippo²
F. C. Öner MD⁴
R. John Hulbert⁵

Michael G. Fehlings⁶
Paul Arnold⁷
James S. Harrop²
Christopher M. Bono⁸
Paul A. Anderson⁹
D. Greg Anderson²
Eli M. Baron¹⁰

ABSTRACT

Objective: to analyze the application of a new classification system for thoracolumbar fractures and a scale of severity of the injury among physicians at various levels of training and across multiple medical specialties in regard to predicting the method of treatment. **Methods:** the classification is based on three major injury characteristics: 1) the mechanism of injury, 2) the integrity of the posterior ligamentous complex, and 3) the neurologic status of the patient. Based on the severity of injury of these three categories, specific points are allocated, and the sum of the points defines possible treatment alternatives. Higher total points indicate a more severe injury, and those injuries are more likely to benefit from surgical intervention based on the opinions of the Spine Trauma Study Group. Seventy-

RESUMO

Objetivos: analisar a aplicação de um novo sistema de classificação para as fraturas da coluna toracolombar e da escala de gravidade da lesão com médicos com vários níveis de treinamento e entre diferentes especialidades e sua relação com o método de tratamento proposto. **Métodos:** a classificação está baseada em três principais características de lesão: 1) no mecanismo da lesão; 2) na integridade do complexo ligamentar posterior; 3) no grau de lesão neurológica. Com base na gravidade das lesões correspondentes às três características é atribuída uma pontuação específica, cuja soma define as possibilidades alternativas de tratamento. Foram distribuídos em CD ROM 71 casos de trauma toracolombar para cinco médicos especialistas em coluna vertebral (três ortopedistas e dois neurocirurgiões); seis

Traumacentre and Neurocentre (*) of Liberec Regional Hospital

¹University of North Carolina - Chapel Hill, Chapel Hill, NC, EUA.

²Thomas Jefferson University and The Rothman Institute, Philadelphia, PA, EUA.

³University of Pittsburgh Medical Center, Pittsburgh, PA, EUA.

⁴University Hospital Utrecht, Utrecht, The Netherlands.

⁵University of Calgary, Foothills Hospital and Medical Centre, Calgary, Alberta, Canada.

⁶University of Toronto, Toronto, Ontario, Canada.

⁷Kansas University, Kansas City, KS.

⁸Boston University Medical Center, Boston, MA, EUA.

⁹University of Wisconsin, Madison, WI, EUA.

¹⁰Temple University, Philadelphia, PA, EUA.

Recebido: 17/02/2005 - Aprovado: 11/08/2006

one cases of thoracolumbar trauma were distributed on CD-ROM to five attending spine surgeons (three orthopedics' surgeons and two neurosurgeons), six orthopedics'/neurosurgery residents, four orthopedic spine fellows, and five non-surgeons (three radiologists, and two physiatrists). The participants were provided a detailed description of the novel Thoracolumbar Injury Severity Scale and Score and a scoring sheet in which points were to be assigned to various attributes of a particular injury. One month was allowed for completion of a work sheet, which included individual scores for various components of an injury, a total Thoracolumbar Injury Severity Score (TLISS), the TLISS management recommendations. One month later, the physicians were then re-assigned the same task with a second CD-ROM with the same cases randomly reordered. All final data sheets were then collected and submitted to an independent statistician for statistical analysis. Intra- and inter-observer reliability for physician's management decisions was calculated using Kappa coefficients and Spearman's correlation. For each management category, the proportions of cases in that category were compared across the different physician types by a 1-Way ANOVA, with physician type as the factor. **Results:** using TLISS to determine treatment, there was no statistically significant rate in recommending surgery between the four groups of physicians. The attending surgeons, fellows, non-surgeons, and residents recommend operative treatment in 34%, 35%, 35%, and 38% of the cases, respectively. The resident group tended to have a lower rate of recommending non-operative care than the other 3 groups, but the difference among groups did not reach statistical significance. Residents recommended non-operative treatment in 40% of cases, in contrast to 53%, 56%, and 49% for attending surgeons, fellows, and non-surgeons, respectively. Regarding the predicted method of treatment, there was moderate agreement between the various physician groups with Kappa values ranging from 0.38 to 0.53 and Spearman's correlation values ranging from 0.40 to 0.50. Between the various groups of physicians, the highest agreement occurred among the attending spine surgeons and spine fellows (Kappa value 0.53, Spearman's correlation value 0.50). The lowest agreement occurred between the spine fellows and the residents (Kappa value 0.38, Spearman's correlation value 0.40). There was significantly better agreement among the attendings and fellows, attendings and non-surgeons, and fellows and non-surgeons as compared to the various other physician combinations (p value <0.05). **Conclusion:** the results of our study demonstrate that the use of TLISS to direct treatment of thoracolumbar spine injuries by physicians of various training levels and specialties has moderate agreement in regard to whether these injuries should be treated surgically or non-surgically.

KEYWORDS: Lumbar vertebrae; Spinal injuries; Trauma severity indices; Validation studies
[Publication Type]

residentes em ortopedia e em neurocirurgia; quatro "fellows" e cinco médicos de outras especialidades (três radiologistas e dois fisiatras). Os participantes eram providos de uma descrição detalhada da nova escala e pontuação da gravidade das lesões toracolombares, e de uma folha de escore, na qual cada ponto era assinalado com os vários atributos das lesões. Deu-se um mês para o preenchimento dessa folha, que incluiu um escore individual dos vários componentes da lesão, a escala total de gravidade das lesões (TLISS) e respectivas recomendações. Após um mês, os médicos tinham, então, re-assinaladas as mesmas questões com um segundo CD-ROM, com os mesmos casos reordenados aleatoriamente. Os dados coletados foram submetidos a uma análise estatística independente. A decisão intra e inter-observador dos médicos foi calculada usando o coeficiente de Kappa e o valor de correlação de Spearman. Para cada categoria de tratamento, as proporções dos casos foram comparadas com o cruzamento dos dados dos diferentes tipos de médicos por análise de variância (ANOVA), sendo o tipo de médico a variável estudada. **Resultados:** o uso da "TLISS" não apresentava um índice estatisticamente significativo de recomendação cirúrgica entre os quatro grupos de médicos. Os cirurgiões, os "fellows", os médicos de outras especialidades e os residentes recomendavam tratamento cirúrgico em: 34%, 35%, 35% e 38% dos casos, respectivamente. O grupo dos residentes tendeu a ter um mais baixo índice de recomendação de tratamento conservador (não-cirúrgico), do que os outros três grupos, mas a diferença entre os grupos não foi estatisticamente significativa. Os residentes recomendaram tratamento não-cirúrgico em 40% dos casos, em contraste com 53%, 56% e 49% dos cirurgiões, "fellows", e outras especialidades, respectivamente. Considerando-se a antecipação do método de tratamento escolhido houve uma concordância moderada entre os vários grupos médicos com o coeficiente de Kappa de 0,38 a 0,53 e o valor de correlação de Spearman de 0,40 a 0,50. Entre os vários grupos de médicos, a maior concordância ocorreu entre os médicos cirurgiões de coluna vertebral e os "fellows" (coeficiente de Kapa 0,53, e correlação de Spearman 0,50). A menor concordância ocorreu entre os "fellows" e os residentes (coeficiente de Kappa 0,38 e correlação de Spearman 0,40). Houve uma maior concordância entre os "fellows", de outra especialidade e não-cirurgiões, e "fellows" e não cirurgiões em comparação às demais combinações de tipos de médicos (valor $p <0,05$). **Conclusão:** os resultados do nosso estudo demonstrou que o uso do TLISS para direcionar o tratamento direto das lesões da coluna toracolombar por médicos de vários níveis e especialidades diversas tem uma moderada concordância na recomendação se essas lesões poderiam ser ou não tratadas cirurgicamente.

DESCRITORES: Vértebras lombares/lesões; Traumatismos da coluna vertebral; Índices de gravidade do trauma; Estudos de validação [Tipo de Publicação]

INTRODUCTION

Upon presentation to a tertiary care center, a patient with a suspected injury to the thoracolumbar spine is evaluated by a multitude of health care providers. The first physician to evaluate the patient is usually the emergency room (ER) physician and/or a general trauma surgeon. ER physicians' medical training is diverse and may be either through emergency medicine, internal medicine, and/or general surgery. General surgeons often spend several months rotating on an orthopaedic or neurosurgical service during training, and have some degree of experience in identifying spinal injuries. The radiographs ordered by the evaluating physician are then interpreted by the radiologist on-call, who may be a general radiologist, a fellowship-trained neuroradiologist, or a musculoskeletal radiologist. If the evaluating physician is concerned for a traumatic spinal injury, the orthopaedic or neurological spine surgeons are consulted for further treatment evaluation. Upon consultation, the first orthopaedic surgeon or neurosurgeon to see the patient is often the junior resident on-call. With a significant injury, the patient is subsequently evaluated by a chief resident or fellow, and then the attending orthopaedic or neurological spine surgeon. Lastly, after the injury has been stabilized, a physiatrist evaluates the patient to direct the rehabilitation regimen.

In an ideal situation, every physician evaluating the patient during the triage process should have a common "language" to describe and characterize the spinal injury. The injury characteristics are then used to classify the injury, assess severity, and determine management. With an effective classification system, the multi-disciplinary team of treating physicians should then ideally have general agreement on how the injury should be treated (operative vs. non-operative). Currently, no widely-accepted classification system exists which allows such inter-physician and inter-disciplinary agreement. We describe a novel classification and injury severity scoring system introduced by Vaccaro et al called the TLISS (Thoracolumbar Injury Severity Scale and Score).¹ Our study analyzes the application of this system between physicians at various levels of training and across multiple medical specialties in regard to predicting the method of treatment.

History of thoracolumbar classification systems

A widely-accepted classification system for thoracolumbar spine injuries has not been available for many reasons. First, spinal anatomy is complex and spinal injuries consist of a wide range of mechanisms and varying morphologies. Second, there are widely differing philosophies in treatment strategies, depending on specialty of training and geographic location^{2,3}. Third, the current systems have not been able to strike an adequate balance between being comprehensive while being user-friendly. For example, the AO classification system is comprehensive but too cumbersome for daily clinical use. In contrast, the Denis classification system is simplistic to use but does not account for many fracture patterns⁴. In 1930, Böhler⁵ initially attempted to classify thoracolumbar spine

fractures by defining five injury categories. He combined both anatomic appearance (morphology) and mechanisms of injury to describe 1) compression fractures, 2) flexion-distraction injuries, 3) extension fractures, 4) shear fractures, and 5) rotational injuries. Böhler, however, did not attempt to define spinal stability after trauma. In 1943, Watson-Jones⁶ recognized that the concept of "instability" and its importance in any effective treatment algorithm on thoracolumbar injuries. He further proposed ligamentous injury as one of the key determinants of instability. In 1949, Nicholl⁷ subsequently define the vertebral body, facet joints, posterior ligaments, and the intervertebral disks as the primary anatomic structures involved in determining an injury pattern. Nicholl also emphasized the potential for progressive neurological injury and deformity in an unrecognized instability pattern. Over 25 years later, Holdsworth⁸ introduced the principle of a "column" concept for spinal stability. He described the spine in two major columns: the anterior column (consisting of the vertebral body and intervertebral disk) and the posterior column (consisting of facet joints and the posterior ligamentous complex). He believed that the integrity of the posterior column was necessary for stability of the thoracolumbar spine. His classification scheme included anterior compression fractures, fracture dislocations, rotation fracture dislocations, extension injuries, burst fractures, and shear injuries. Although Holdsworth's descriptions remain the most influential to the modern classification schemes, the biomechanics principles in his injury classification were oversimplified. For example, all burst fractures were falsely categorized as "stable", when many of these fractures were proven to be "unstable" and developed kyphosis with progressive neurological deficits⁹⁻¹⁰.

In the early 1980's, the advent of computed tomography (CT) provided a new opportunity to improve upon the existing thoracolumbar classification schemes. CT imaging provided further visualization and definition of spinal injuries, including the osseous anatomy surrounding the spinal canal. Using this technology, Denis introduced a classification scheme based on a three-column concept. In contrast to Holdsworth's two-column system, Denis¹¹ defined the anterior column as the anterior longitudinal ligament to the anterior two thirds of the vertebral body, the middle column as the posterior one third of the vertebral body including the anulus fibrosus and posterior longitudinal ligament, and the posterior column as including all structures posterior to the posterior longitudinal ligament. Denis defined four distinct fractures types including compression fractures, burst fractures, fracture dislocations, and seatbelt injuries. Denis also recognized that mechanical instability and progressive neurologic deterioration could either occur separately or concurrently. Mechanical instability may lead to progressive kyphosis without neurologic instability. Neurologic deterioration may occur without radiographic signs of instability as in the case burst fractures. More unstable fractures, such as fractures associated with dislocations of the facets or the disk interspace were usually associated with neurologic deterioration. Denis described

isolated mechanical instability as “first degree” injuries, neurologic deterioration as “second degree” injuries, and combined mechanical and neurologic deterioration as “third degree” injuries. Presently, this three-column classification scheme remains the most popular classification system presumably due to its inherent simplicity.

Denis’ classification, however, does not provide guidance for treatment decision-making. With the over-simplification of his original scheme over time, it has become widely accepted that when two of three columns are injured, operative stabilization may be required. This oversimplification, however, has led to the loss of Denis’ original emphasis on the distinction between mechanical and neurologic instability. Several studies have shown that non-operative treatment of two column injuries achieves satisfactory outcome¹²⁻¹⁴. Also, Denis’ classification does not address ligamentous injuries which may lead to progressive instability. With the advent of magnetic resonance imaging (MRI), previously unidentified ligamentous injuries can now be readily recognized. Yet, no current classification system incorporates MRI findings in its scheme.

A modern classification system, thus, should incorporate the current understanding of the biomechanics of thoracolumbar injuries and modern imaging modalities such as MRI. Also, the classification system should strike a balance between being comprehensive while being simple and intuitive. In addition, it should direct general guidelines of treatment, based on current understanding of the natural history of thoracolumbar spine injuries. Lastly, the classification should have good intra-observer and inter-observer reliability amongst physicians of diverse specialties and varying training levels involved in the care of the patient with thoracolumbar spine trauma.

The thoracolumbar injury severity scale and score: a novel classification system and treatment algorithm

The thoracolumbar injury severity scale and score (TLISS) was conceptualized based on a survey of the Spine Trauma Study Group which consists of an international panel of experts in the field of spine trauma. The goal of the survey was to identify similarities in treatment algorithms for common thoracolumbar injuries and to identify characteristics of injuries that played a key role in the decision-making process. Using these important characteristics, a new classification system and algorithm of treatment was developed by Vaccaro et al.¹⁵

TLISS is based on three major injury characteristics: 1) the mechanism of injury, 2) the integrity of the posterior ligamentous complex, and 3) the neurologic status of the patient¹⁵. The mechanism of injury and integrity of the posterior ligamentous complex are both inferred by the appearance of the injury on imaging studies. Neurologic status is determined by physical

exam. Based on the severity of injury of these three categories, specific points are allocated, and the sum of the points defines possible treatment alternatives (Table 1). Higher total points indicate a more severe injury, and those injuries are more likely to benefit from surgical intervention based on the opinions of the Spine Trauma Study Group.

Mechanism of injury

The initial step in obtaining a TLISS score is to scrutinize the available imaging studies to identify the mechanism of injury. The three major injury mechanism sub-categories include: 1) Compression injuries, 2) Translational/Rotational injuries, and 3) Distraction injuries (Figure 1). One to four points are assigned to these mechanisms, with more severe bony and ligamentous injuries receiving higher scores. If there are injuries in multiple levels, only the level with the most severe mechanism is counted. If multiple primary mechanisms are involved at a single level, the score is additive. For example, a patient with both a compression and distraction injury at the same level would receive scores for each, whereas, a patient with a compression and a distraction injury at separate levels would only be scored at the level with the higher total point value. Compression injuries are the most common form of thoracolumbar fractures and typically arise due to axial load to the spine. (Figure 1). A vertebral compression fracture, where an axial load is transferred to the anterior vertebral body is the most common thoracolumbar traumatic fracture. The anterior vertebral body deforms into a wedge, causing varying degrees of kyphosis, while the posterior vertebral body remains intact. Another form of a compression injury pattern is the burst fracture, where an axial load is transferred to both the anterior and posterior vertebral walls. In this case, both the anterior and posterior cortex of the vertebral body is disrupted in varying degrees, with retropulsion of the posterior vertebral wall into the spinal canal. The TLISS score assigns 1 point to the compression fracture mechanism, with an additional point added if a burst fracture is present. The score is further increased if there is evidence of a lateral flexion component ($>15^\circ$ in coronal deformity) by an additional point.

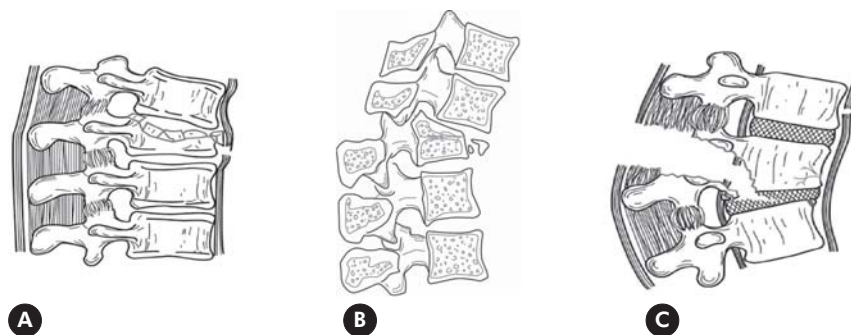


Figure 1
The three major injury mechanism sub-categories include compression injuries (A), translational/rotational injuries (B), and distraction injuries (C)

Translational/rotational injuries are significant injuries that result from violent torsional and/or shear forces (Figure 1). These injuries usually cause significant ligamentous or osseous damage that result in an unstable spine. Due to the severe instability created by this mechanism of injury 3 points are assigned.

Distraction injuries describe a tensile disruption of the spinal column (Figure 1). The distraction can cause osseous, ligamentous, or combined injuries of the spine, and usually results in circumferential instability. Due to the extreme instability, this mechanism is assigned 4 points in TLISS.

Neurologic injury

The TLISS provides for five subcategories of neurologic injury: 1) Intact, 2) Nerve root injury, 3) Complete spinal cord injury, 4) Incomplete spinal cord injury, and 5) Cauda equina syndrome (Table 1). Patients with an intact neurologic status are assigned 0 points and those with either nerve root injury or a complete spinal cord injury are assigned 2 points. Patients with an incomplete spinal cord injury or cauda equina syndrome are assigned 3 points, due to relative value of surgical decompression in this neurologically impaired subgroup.

Posterior ligamentous complex:

The TLISS includes three subcategories of description when evaluating the integrity of the posterior ligamentous complex (Table 1). A disrupted posterior ligamentous complex is determined by the presence of a palpable gap between spinous processes and/or interspinous widening on plain films or CT images. Disrupted posterior ligaments may also be determined by MRI. An intact posterior ligamentous complex is assigned 0 points while definite ligamentous disruption is assigned 3 points. When the status of the posterior ligamentous complex is indeterminate, 2 points is assigned.

Total points:

The total TLISS score is determined by adding the assigned points in all three major categories. Determination of the total score is designed to help surgeons and non-surgeons assess the severity of the injury, as well as to guide the decision between operative and non-operative management. Patients with 3 or less total points are considered non-operative candidates, whereas patients with 5 or greater points are operative candidates. Patients with a total score of 4 fall into an option category, where either non-operative or operative treatment may be considered based on physician preference and other factors not included in TLISS (Table 2).

TLISS caveats:

The TLISS algorithm represents an objective way to determine treatment for patients with thoracolumbar injuries. Although it is a systematic approach to determine injury severity, the TLISS score does not provide subjective criteria that may also be crucial in the decision-making process. These factors include medical co-morbidities, traumatic injuries other than those involving the spine (multiple limb fractures, closed head injury, internal organ injuries, etc), abrasions over potential operative sites, or excessive kyphosis^{3,15}. Other pre-existing osseous disorders, such as ankylosing spondylitis, diffuse idiopathic skeletal hyperostosis, or osteoporosis may also affect treatment decisions. These factors have not been incorporated in TLISS score in order to keep the algorithm purely objective and as simple as possible. The TLISS score is designed only as a guideline for physicians to help determine possible treatment options, and is not meant to be absolutely authoritative in decision-making. Therefore, incorporation of these subjective factors is critical in ultimately choosing non-operative or operative treatment.

TABLE 1 - Injury characteristics, qualifier and specific points

MECHANISM TYPE	QUALIFIER	POINTS
Compression	Lateral angulation > 15°	1
	Burst	1
		3
Translational/rotational		3
Distraction		4
NEUROLOGIC INVOLVEMENT	QUALIFIER	POINTS
Intact		0
Nerve root		2
Cord, conus medullaris	Incomplete	3
	Complete	2
Cauda Equina		3
POSTERIOR LIGAMENOUS COMPLEX		POINTS
Intact		0
Injury suspected/indeterminate		2
Injured		3

TABLE 2 - Management of the fracture according to the total score

Management	Points
Nonoperative	0-3
Option: nonoperative or operative	4
Operative	5 and above

TABLE 3 - Distribution of management decisions for attending spine surgeons, spine fellows, residents, and non-surgeons

	Attending spine Surgeon(n=5)	Spine surgery fellows(n=4)	Non-surgeons (n=5)	Residents (n=6)
Operative	34%	35%	35%	38%
Option	12%	9%	15%	21%
Non-Operative	53%	56%	49%	40%

METHODS

Seventy-one clinical cases of thoracolumbar spinal trauma were distributed on CD-ROM to 5 attending spine surgeons (three orthopedics' surgeons and two neurosurgeons), six orthopedics'/neurosurgery residents, four orthopedic spine fellows, and five non-surgeons (three radiologists, and two physiatrists).

The case histories included a description of the injury (fall, motor vehicle collision, etc.), the clinical examination (ASIA score and/or neurologic exam), and radiographic images consisting of plain films, CT and MRI.

The participants were provided a detailed description of the novel Thoracolumbar Injury Severity Scale and Score and a scoring sheet in which points were to be assigned to various attributes of a particular injury. One month was allowed for completion of a work sheet (Tables 1 and 2), which included individual scores for various components of an injury, a total Thoracolumbar Injury Severity Score (TLISS), the TLISS management recommendations. One month later, the physicians were then re-assigned the same task with a second CD-ROM with the same cases randomly reordered.

All final data sheets were then collected and submitted to an independent statistician for statistical analysis. All statistics were calculated using SPSS software (SPSS Inc, Chicago, IL). Intra- and inter-observer reliability for physician's management decisions was calculated using Kappa coefficients and Spearman's correlation. For each management category, the proportions of cases in that category were compared across the different physician types by a 1-Way ANOVA, with physician type as the factor.

RESULTS

The distribution of management decisions directed by TLISS for attending spine surgeons, spine surgery fellows, non-surgeons, and residents is shown in Table 3. Using TLISS to determine treatment, there was no statistically significant rate in recommending surgery between the four groups ($p = 0.6278$). The attending surgeons, fellows, non-surgeons, and residents recommend operative treatment in 34%, 35%, 35%, and 38% of the cases, respectively. The resident group tended to have a lower rate of recommending non-operative care than the other 3 groups, but the difference among groups did not reach statistical significance ($p = 0.1589$). Residents recommended non-operative treatment in 40% of cases, in contrast to 53%, 56%, and 49% for attending surgeons, fellows, and non-surgeons, respectively.

The inter-rater agreement indices within and between the groups are shown in Table 4. In regard to the predicted method of treatment, there was moderate agreement between the various physician groups with Kappa values ranging from 0.38 to 0.53 and Spearman's correlation values ranging from 0.40 to 0.50. Between the various groups of physicians, the highest agreement occurred between the attending spine surgeons and spine fellows (Kappa value 0.53, Spearman's correlation value 0.50). The lowest agreement occurred between the spine fellows and the residents (Kappa value 0.38, Spearman's correlation value 0.40). There was significantly better agreement between the attendings and fellows, attendings and non-surgeons, and fellows and non-surgeons as compared to the various other physician combinations (p value <0.05).

TABLE 4 - Inter-rater agreement indices within and between groups on management decision

		Attending spine surgeons (AS)	Spine surgery fellows (SF)	Non-surgeons (NS)	Residents (R)
Attending spine	Kappa	0.52	0.53 ^a	0.49 ^b	0.43
Surgeons	Spearman's	0.51	0.50	0.46	0.48
	% Agreement	72.0	73.0	69.3	63.9
Spine surgery	Kappa		0.52	0.47 ^c	0.38
Fellows	Spearman's		0.50	0.43	0.40
	% Agreement		73.2	69.0	61.3
Non-surgeons	Kappa			0.44	0.43
	Spearman's			0.47	0.46
	% Agreement			65.5	63.7
Residents	Kappa				0.46
	Spearman's				0.54
	% Agreement				65.2

^a Inter-rater kappa score for AS versus SF is significantly greater ($p < 0.05$) than that between AS and R, SF and R, and NS and R.

^b Inter-rater kappa score for AS versus NS is significantly greater ($p < 0.05$) than that between AS and R, SF and R, and NS and R.

^c Inter-rater kappa score for SF versus NS is significantly greater ($p < 0.05$) than that between SF and R

DISCUSSION

One of the most commonly used statistical instruments for measuring a system's reliability are Cohen's Kappa value and the Spearman correlation. The Kappa value ranges from -1.0 (complete disagreement beyond that related to chance) to 1.0 (complete agreement beyond that related to chance). The most widely accepted criteria for assessing the extent of agreement per the Kappa value are the criteria of Landis and Koch.¹⁶ Based on their criteria, $k > 0.80$ indicates near perfect agreement, $k = 0.61$ to 0.80 indicates substantial agreement, $k = 0.41$ to 0.60 indicates moderate agreement, $k = 0.21$ to 0.40 indicates fair agreement, and $k = 0.00$ to 0.20 indicates slight agreement. Another widely used method for assessing the relationship between sets of data is the Spearman's rank correlation. Spearman's correlation values can range between -1.0 and 1.0 , where correlation values are viewed along a continuum. Values of 0.00 to 0.30 are considered no positive correlation, 0.40 to 0.70 a mildly positive correlation, and 0.80 to $.99$ a strongly positive correlation¹⁷. Therefore, between physicians of varying training levels and varying specialties, the use of TLISS to predict treatment for thoracolumbar injuries leads to moderate agreement (range of Kappa values 0.38 to 0.53 , and range of Spearman's correlation values 0.40 to 0.50).

The highest agreement for treatment prediction was amongst the spine surgeon attendings and the fellows (Kappa value 0.53). This likely relates to the close working relationship between the fellows and the attending spine surgeons. Although the ideal classification system is purely objective without any observer bias, our results indicate the susceptibility of the TLISS to a moderate amount of bias.

A modern classification system for thoracolumbar spine injuries requires a balance between being user-friendly while also being comprehensive. Denis' classification system, which represents the most popular in current use, is user-friendly and has a good inter- and intra-observer reliability with a Kappa values of approximately 0.6 .¹⁸ The use of this simple classification system, however, may lead to incorrectly treating potentially unstable injuries non-operatively or performing unnecessary surgery on stable fractures. In contrast, the AO classification of thoracolumbar fractures is comprehensive with categorization of more than 50 types of fractures. However, the complicated categorization makes this system cumbersome and unwieldy for practical clinical use. The AO classification has been shown to have poor inter- and intra-observer reliability⁴. In designing TLISS, an attempt was made to achieve simplicity by emphasizing the few critical factors useful in guiding treatment. These key factors including injury mechanism, neurologic findings, and status of the posterior ligamentous complex were used to create a treatment algorithm called TLISS. The result was a simple yet comprehensive classification system which may effectively direct the management of the injuries. Our study has shown that the use of the TLISS leads to moderate agreement between physicians of various training levels and disciplines in terms of predicting treatment.

The results of this survey and others has lead to a recent modification of the TLISS system¹⁵. Due to the difficulties in inferring the mechanism of injury from initial imaging modalities and disagreement among observers,

the category of mechanism of injury has been substituted with injury morphology. The new classification system is referred to as the Thoracolumbar Injury Classification and Severity Score (TLICS). In this system only the most severe (highest total points) injury morphology category is included in the scoring. One point is assigned for a compression morphology, and an additional point is assigned for a burst morphology. Three points are assigned for a translational/rotational morphology and 4 points are assigned for a distraction morphology. The descriptive morphological term distraction is only applied if there is absolute objective imaging evidence of distraction noted. The scores of the morphology subgroups are not additive if multiple injury morphologies

are present. The remaining details of the classification and severity scoring are identical to the original TLISS system. Further studies are currently underway to refine this classification system to improve observer agreement and improve its ability to direct treatment.

CONCLUSIONS

The results of our study demonstrate that the use of TLISS to direct treatment of thoracolumbar spine injuries by physicians of various training levels and specialties has moderate agreement in regard to whether these injuries should be treated surgically or non-surgically (Kappa values ranging from 0.38 to 0.53, and Spearman's correlation values ranging from 0.40 to 0.50).

REFERENCES

- Vaccaro AR, Zeiller SC, Hulbert RJ, Anderson PA, Harris M, Hedlund R, et al. The thoracolumbar injury severity score: a proposed treatment algorithm. *J Spinal Disord Tech.* 2005; 18(3):209-15.
- Oner FC. Thoracolumbar Spine Fractures: Diagnostic and Prognostic Parameters. Universiteit Utrecht; 1999. p.32-50.
- Grauer JN, Vaccaro AR, Beiner JM, Kwon BK, Hilibrand AS, Harrop JS, et al. Similarities and differences in the treatment of spine trauma between surgical specialties and location of practice. *Spine.* 2004; 29(6):685-96.
- Blauth M, Bastian L, Knop C, Lange U, Tusch G. [Inter-observer reliability in the classification of thoraco-lumbar spinal injuries]. *Orthopade.* 1999; 28(8):662-81. German.
- Bohler L. Die Technik der Knochenbruchbehandlung im Frieden und im Kriege: Verlag von Wilhelm Maudrich; 1930. p.9-11.
- Watson-Jones R. Fractures and joint injuries. Edinburgh: E & S Livingstone; 1943.
- Nicholl EA. Fractures of the dorso-lumbar spine. *J Bone Joint Surg Br.* 1949; 31:376-94.
- Holdsworth F. Fractures, dislocations, and fracture-dislocations of the spine. *J Bone Joint Surg Am.* 1970; 52(8): 1534-51.
- Roberts JB, Curtiss PH Jr. Stability of the thoracic and lumbar spine in traumatic paraplegia following fracture or fracture-dislocation. *J Bone Joint Surg Am.* 1970; 52(6):1115-30.
- Whitesides TE Jr. Traumatic kyphosis of the thoracolumbar spine. *Clin Orthop Relat Res.* 1977;(128):78-92.
- Denis F. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. *Spine.* 1983; 8(8):817-31.
- Agus H, Kayali C, Arslantasi M. Nonoperative treatment of burst-type thoracolumbar vertebra fractures: clinical and radiological results of 29 patients. *Eur Spine J.* 2004; 14(6): 536-40. Epub 2004 May 28.
- Wood K, Buttermann G, Mehdod A, Garvey T, Jhanjee R, Sechriest V, Butterman G. Operative compared with nonoperative treatment of a thoracolumbar burst fracture without neurological deficit. A prospective, randomized study. *J Bone Joint Surg Am.* 2003; 85-A(5):773-81.
- Kelly RP, Whitesides TE Jr. Treatment of lumbodorsal fracture-dislocations. *Ann Surg.* 1968; 167(5):705-17.
- Vaccaro AR, Lehman RA Jr, Hurlbert RJ, Anderson PA, Harris M, Hedlund R, et al. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. *Spine.* 2005; 30(20):2325-33.
- Landis JR, Koch GC. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33(1):159-74.
- MacFarland TW. Spearman's Rank-Difference Coefficient of Correlation. Boca Raton; 1998.
- Oner FC, Ramos LM, Simmermacher RK, Kingma PT, Diekerhof CH, Dhert WJ, Verbout AJ. Classification of thoracic and lumbar spine fractures: problems of reproducibility. A study of 53 patients using CT and MRI. *Eur Spine J.* 2002; 11(3):235-45.

Correspondence:

Alexander R. Vaccaro
925 Chestnut St. 5th Floor
Philadelphia, EUA
PA 19107