



Original Article

Intraobserver and interobserver reproducibility of the old and new classifications of thoracolumbar fractures[☆]



Felipe Augusto Rozales Lopes*, Ana Paula Ribeiro Bonilauri Ferreira, Ricardo André Acácio dos Santos, Carlos Henrique Maçaneiro

Instituto de Ortopedia e Traumatologia de Joinville, Joinville, SC, Brazil

ARTICLE INFO

Article history:

Received 19 May 2017

Accepted 13 July 2017

Available online 2 August 2018

Keywords:

Spinal injuries

Magerl AO classification

AOSpine classification

Interobserver and intraobserver agreement

ABSTRACT

Objective: To evaluate the inter and intraobserver agreement of the Magerl AO and AOSpine thoracolumbar fracture classification systems.

Methods: The participants were divided into two groups, the first composed of six spinal surgeons and the other composed of 18 medical orthopedic residents. On two different occasions, separated by an interval of one month, the participants analyzed and classified 25 radiographs with thoracolumbar fractures using both thoracolumbar fracture classification systems, Magerl AO and AOSpine. The results were analyzed for classification reliability using the Kappa coefficient (k).

Results: The Magerl AO classification system showed a fair interobserver agreement ($k = 0.32$), considering the fractures type and subtype, whereas the AOSpine classification system showed a moderate interobserver agreement ($k = 0.59$). The Magerl AO classification showed a fair intraobserver agreement for both residents and specialists ($k = 0.21$ and 0.38 , respectively), while the AOSpine showed a substantial agreement between residents ($k = 0.62$) and moderate between specialists ($k = 0.53$).

Conclusions: When evaluating fracture morphology, the AOSpine thoracolumbar fracture classification system presented a better reliability and reproducibility compared to the Magerl AO classification system.

© 2018 Sociedade Brasileira de Ortopedia e Traumatologia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

[☆] Study conducted at Instituto de Ortopedia e Traumatologia de Joinville, Joinville, SC, Brazil.

* Corresponding author.

E-mail: felipearlobes@gmail.com (F.A. Lopes).

<https://doi.org/10.1016/j.rboe.2018.07.015>

2255-4971/© 2018 Sociedade Brasileira de Ortopedia e Traumatologia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Análise da reprodutibilidade intra e interobservadores das classificações antiga e atual da AO para fraturas toracolombares

R E S U M O

Palavras-chave:

Fraturas da coluna vertebral
Classificação Magerl AO
Classificação AOSpine
Concordância interobservadores
e intraobservadores

Objetivo: Avaliar a concordância inter e intraobservadores dos sistemas de classificação Magerl AO e AOSpine para fraturas toracolombares.

Métodos: Os participantes foram divididos em dois grupos, um com seis médicos ortopedistas especialistas em coluna e o outro com 18 médicos residentes em ortopedia. Os participantes analisaram 25 radiografias com fraturas toracolombares em duas oportunidades, com um mês de intervalo entre elas, e classificaram com o uso dos dois sistemas de classificação de fratura toracolombar, Magerl AO e AOSpine. Os dados de concordância foram analisados pelo método do coeficiente kappa.

Resultados: A classificação de Magerl AO apresentou uma concordância interobservadores leve ($k=0,32$), considerando o tipo e o subtipo das fraturas, enquanto a classificação AOSpine obteve uma concordância interobservadores moderada ($k=0,59$). A classificação de Magerl AO apresentou uma concordância intraobservadores leve entre médicos residentes e médicos especialistas ($k=0,21$ e $0,38$, respectivamente), enquanto a classificação AOSpine apresentou uma boa concordância intraobservadores entre médicos residentes ($k=0,62$) e moderada entre médicos especialistas ($k=0,53$).

Conclusão: O sistema de classificação da AOSpine para fraturas toracolombares apresentou uma melhor confiabilidade e reprodutibilidade comparado com o sistema de classificação Magerl AO, em relação à morfologia da fratura.

© 2018 Sociedade Brasileira de Ortopedia e Traumatologia. Publicado por Elsevier Editora Ltda. Este é um artigo Open Access sob uma licença CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Approximately 90% of spinal fractures affect the thoracic and lumbar regions.¹ The thoracolumbar junction is very susceptible to fractures since in this region there is a difference in the rigidity from a more rigid thoracic spine to a more flexible lumbar spine.¹ Over 50% of the injuries occur between T11 and L2, usually due to high-energy traumas, with other associated lesions,² such as intra-abdominal lesions (splenic and hepatic lesions), limb fractures, and brain trauma.¹

Several classification systems for thoracolumbar fractures had been described before Magerl et al.³ introduced the AO classification system for these fractures in 1994. A classification system that used vector forces as a classification criterion, included fractures that occur by compression, distraction, and torsional forces.^{1,3,4} This classification was established aiming at creating a standard classification system; however, it is not practical, which reduces its reliability.¹ The AO spine classification group proposed a new classification system, using the Magerl AO classification system as the main reference, in order to create a similar classification, but with a better and more immediate clinical application in medical practice. The AOSpine classification was then created, which has appeared to be fairly reliable and accurate. However, it needs further assessment.⁵

A bone fracture classification system should be reliable, valid, and accurate, as it will aid in the prognosis and indication of treatment. A classification system is considered to be reliable when a single evaluator obtains consistent results by classifying the same fracture at different times, or when

several evaluators produce the same result with the same classification. When this reliability is observed in medical practice, the classification is considered to be accurate.^{6,7} According to Vaccaro et al.,⁸ the current classification, AOSpine, allows a better understanding of the lesion, since not only it modified and simplified the old morphological classification, but also six neurological lesion topics and two patient modifiers were added, which help to guide treatment. However, there is still no global consensual classification for thoracolumbar fractures.⁹

This study was aimed at evaluating the reproducibility of the two AO classifications regarding the morphology of thoracolumbar fractures by assessing the intra- and inter-observer agreement.

Material and methods

Ethical considerations

The study was submitted to the Research Ethics Committee of the Hospital Municipal São José (HMSJ; Joinville, SC), and approved through the Brazil Platform, under No. 1.769.539.

Participants

The study included six spine specialists working in the same orthopedics department and 18 training physicians in their first, second, and third year of the orthopedic and traumatology medical residency at the Institute of Traumatology and Orthopedics (ITO) of Joinville (Santa Catarina State, Brazil).

The participants were divided into two groups, one with the spine specialists and the other with the resident physicians. Prior to their inclusion in the study, all participants signed an informed consent form.

Images

Twenty-five radiographs in the anteroposterior and lateral views showing different patterns of thoracolumbar fractures were selected from the HMSJ (Joinville, SC) and ITO (Joinville, SC) files. The images were selected by a resident physician of the second year of orthopedics and by an orthopedic surgeon specialized in the spine; both were familiar with the classification systems. All patient identification data were removed from the images. Radiographs with artifacts and those with poor image quality, poor patient positioning, or technical defects that could have compromised the evaluation were excluded.

Classification systems

Magerl AO classification

This classification system uses vector forces as a classification criterion. The fractures are divided into three types: A, B, and C. Type A consists of fractures caused by compression forces; type B, by distraction forces; and type C, by torsional, rotational forces. Each type is divided into three larger groups, in numerical order; each group is then subdivided into three subgroups, according to fracture morphology, allowing a more detailed description. Severity is defined by the classification and it increases from types A to C; the same occurs in the groups and subgroups.^{1,3,4}

AOSpine classification

This classification allows for a better understanding of the lesion. It was based on the Magerl AO morphological classification. Six topics of neurological injury and two patient modifiers were added. As for the morphological division, the types are the same as the old classification, from A to C, i.e., fractures caused by compression, distraction, and torsional forces, respectively. However, the main difference is in group modification; type A is subdivided into four groups, type B into three groups, and type C has no subdivision.^{5,8}

Procedures

The images to be classified were sent to the survey participants by E-mail. Images were assessed and classified into three stages. The first stage was the pre-test training, the purpose of which was to calibrate the classifications of participants. Participants received by E-mail a video tutorial explaining the two thoracolumbar fracture classification systems (the original Magerl and the current AOSpine). Participants also received illustrations of the two classification systems to consult during image classification. In this pre-test stage, five images were classified using the two classification systems. Participants then sent their responses by E-mail to the resident physician responsible for the study, who, together with the spine specialist, reviewed the answers and sent the participants a feedback

with explanations. This step was included so that, in the next two stages, the images could be classified in a more standardized manner. In the second stage, participants received by E-mail the 25 previously selected images, which were classified using the two classification systems for thoracolumbar fractures. In the third step, 30 days later, the same 25 images were sent by E-mail to the participants in a modified order, who were required to once again classify those images using the two classification systems for thoracolumbar fractures.

All stages were performed individually by each participant. They had no access to the patients' medical history, treatment, or other complementary tests. In all stages, in addition to the 25 images to be classified, participants also received illustrations of the two classification systems to consult during image classification. The deadline for the participants to classify the images was one week. The responses were then sent by E-mail to the resident physician responsible for the study.

The AOSpine classification system characterizes fracture morphology, but also takes into account the neurological aspects of the patient for clinical decision. In turn, the Magerl AO system is predominantly morphological.¹⁰ Therefore, in this study, only the morphological part of the AOSpine classification, which is divided into types A, B, and C, with their respective subgroups, was considered for classification purposes.

Statistical analysis

Agreement was assessed using the weighted kappa coefficient method, which takes into account the fact that the variable is ordinal. The following interpretations of the kappa coefficient were used to assess the scores: <0.00, no agreement; 0.00–0.20, weak; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; 0.81–1.00, excellent agreement.¹¹

Inter- and intra-observer agreement tests were performed using SPSS 20.0 (IBM Statistics) and GraphPad software (Quick-Calcs).

Results

Inter-observer agreement

The Magerl AO classification presented a fair inter-observer agreement ($k=0.32$), considering the type and subtype of the fractures of all images, while the AOSpine classification obtained a moderate inter-observer agreement ($k=0.59$; Table 1). When considering only the type of fracture (A/B/C), without discriminating its subtype, the overall kappa value of the Magerl AO classification was 0.75, which is a good

Table 1 – Inter-observer agreement of each classification system between resident physicians and medical specialists, considering fracture type and subtype.

Classification	Residents vs. specialists (kappa value)	p
Magerl AO classification	0.32	0.138
AOSpine classification	0.59	0.089

Table 2 – Inter-observer agreement of each classification system, considering only the morphological fracture type.

Classification system	Kappa value ^a
<i>Magerl AO classification</i>	
A	0.78
B	0.65
C	0.74
General	0.75
<i>AOSpine classification</i>	
A	0.88
B	0.76
C	0.80
General	0.86

^a Magerl AO vs. AOSpine ($p=0.57$).

Table 3 – Intra-observer agreement of each classification system between resident physicians and medical specialists.

Classification	Residents Kappa value (p)	Specialists Kappa value (p)
Magerl AO classification	0.21 (0.074)	0.38 (0.084)
AOSpine Classification	0.62 (0.052)	0.53 (0.063)

agreement, and AOSpine presented $k=0.85$; nonetheless, no statistically significant difference was observed between the two classification systems ($p=0.57$; Table 2). The kappa values for the inter-observer agreement of the Magerl AO and AOSpine classifications for each morphological type of fractures are shown in Table 2.

Intra-observer agreement

Taking into account the type and subtype of the fractures, the Magerl AO classification showed a fair intra-observer agreement, both for resident physicians ($k=0.21$) and the specialist surgeons ($k=0.38$). In turn, the AOSpine classification showed a good intra-observer agreement among resident physicians ($k=0.62$) and a moderate agreement among specialists ($k=0.53$; Table 3).

Considering only the morphological type of fractures (A/B/C), the intra-observer reproducibility of the Magerl AO classification was good ($k=0.68$ for the resident physicians and $k=0.76$ for specialists). The AOSpine classification presented an excellent reproducibility for both resident physicians ($k=0.82$) and specialists ($k=0.96$). However, there was no statistically significant difference between the two classification systems, both for resident physicians ($p=0.67$) and specialists ($p=0.36$; Table 4). The kappa values that describe the intra-observer agreement for the Magerl AO and AOSpine classifications for each morphological type of fractures are shown in Table 4.

Applicability of the classification systems

Of the participants, 68.2% believe that the AOSpine classification has better applicability when compared with the Magerl AO classification.

Table 4 – Intra-observer agreement of each classification system, considering only the morphological fracture type.

Classification	Residents (kappa value)	Specialists ^{a,b} (kappa value)
<i>Magerl AO classification</i>		
A	0.75	0.79
B	0.62	0.69
C	0.72	0.72
General	0.68	0.76
<i>AOSpine classification</i>		
A	0.86	0.96
B	0.84	0.94
C	0.78	0.89
General	0.82	0.95

^a Magerl AO vs. AOSpine for residents ($p=0.67$).

^b Magerl AO vs. AOSpine for specialists ($p=0.36$).

Discussion

To the best of the authors' knowledge, the literature presents very few studies that have evaluated the inter- and intra-observer agreement between the two AO classification systems for thoracolumbar fractures, the former Magerl AO and the current AOSpine.

The present study demonstrated that the AOSpine classification presented better inter- and intra-observer agreement when compared with the Magerl AO classification, when the type and subtype of the fractures were included. When considering only the type of fractures (A/B/C), the inter- and intra-observer agreement increased considerably in the two classification systems evaluated, and no statistically significant difference was observed between them; this result is justified by the fact that the criteria used to classify the type of fractures are the same for both classification systems. Kepler et al.¹⁰ also used the AOSpine classification system to analyze 25 cases, which were classified by 100 spine surgeons. Considering only the type of fracture (A/B/C), the inter-observer agreement was good ($k=0.74$) and the intra-observer agreement was excellent ($k=0.81$).

Considering the type and subtype of the fractures, the present findings corroborate those of previous studies that also found low to moderate inter- and intra-observer agreement when using the Magerl AO classification. Oner et al.¹² assessed the Magerl AO classification system and indicated a low inter-observer agreement ($k=0.35$) and a moderate intra-observer agreement ($k=0.41$). Wood et al.¹³ observed that considering the type and the group of the Magerl AO classification (A1, A2, A3, B1, B2, etc.) to assess interobserver agreement, the k -value was 0.53 (from 0.33 to 0.68), which indicates a moderate agreement. That is, the more complex that classification becomes, including types, groups, and subgroups, the lower the agreement, which affects its reproducibility. Maçaneiro et al.¹⁴ assessed the inter-observer agreement of 40 cases of thoracolumbar spine fractures using the Magerl AO classification and observed a fair agreement, both for fracture type ($k=0.39$) and group ($k=0.32$).

There is no consensus on the k values that are considered acceptable for fracture classification systems.¹⁵ However, it is suggested that the inter-observer agreement should have a k -value >0.55 .¹⁶

In the present study, the low inter- and intra-observer agreement of the Magerl AO classification system can be justified by its complexity, as it is a very inclusive system in which the observer needs to evaluate many variables, hindering its use in clinical practice. Another possible reason would be the high number of training physicians included in the present study; their inexperience in using such a classification may have statistically contributed to the low agreement. Furthermore, it could be suggested that the lack of complementary exams in the evaluation of the images, such as computed tomography or magnetic resonance imaging, may have interfered in the classification results.

The AOSpine Trauma Knowledge Forum developed the new classification system, AOSpine, which combined the characteristics of the Magerl AO system and the Thoracolumbar Injury Classification System (TLICS), aiming to create a globally accepted system. After this classification system was developed, that group also assessed the morphological aspects of the classification, measuring the inter- and intra-observer reliability. Considering the type and group of fractures, the inter-observer agreement was good ($k=0.64$), and so was the intra-observer agreement ($k=0.77$).⁸ In a recent study, the inter-observer reliability of three classification systems (Magerl AO, TLICS, and AOSpine) was assessed, considering the type and group of fractures. Similarly to the present study, the Magerl AO classification presented a slight agreement ($k=0.38$). In turn, contrary to the present results ($k=0.59$), the AOSpine classification presented a good inter-observer agreement ($k=0.62$).¹⁷ Azimi et al.¹⁸ also observed excellent k -values (from 0.83 to 0.89) in the intra- and inter-observer assessment using the AOSpine classification.

In the present study, participants were asked for their personal opinion regarding the applicability of Magerl AO and AOSpine classification systems. Most of the participants (68.2%) believed that AOSpine had better applicability.

In fact, AOSpine was shown to be a reproducible and valid classification system, easily understood and readily applicable in clinical practice. A worldwide accepted classification system is important for surgeons and researchers to reach a standardized diagnosis and treatment for thoracolumbar fractures.

The authors' efforts to exclude low-contrast or poorly positioned radiographs do not reflect the way these exams are used in clinical practice. Moreover, the lack of further imaging tests in the present study, such as computed tomography and magnetic resonance imaging, may cause failure to diagnose posterior complex injuries and limit the results.

Although this is a promising system for classifying fractures, the authors suggest that future studies should assess the other variables of the AOSpine system, i.e., the neurological state and the key modifiers, as well as evaluating their relationship with the morphological classification types.

Conclusion

It was observed that the AOSpine classification system for thoracolumbar fractures presented better reliability and reproducibility when compared with the Magerl AO classification system. The AOSpine was shown to be a good system to classify thoracolumbar fractures regarding their morphology, justifying its standardization for use with these fractures.

Conflicts of interest

The authors declare no conflicts of interest.

REFERENCES

1. Kepler CK, Vaccaro AR. Thoracolumbar spine fractures and dislocations. In: Court-Brown CM, Heckman JD, McQueen MM, Ricci WM, Tornetta P, editors. *Rockwood and Green's fractures in adults*. 8th ed. Philadelphia: Wolters Kluwer; 2015. p. 1757-94.
2. Meena S, Sharma P, Chowdhury B. Management of thoracolumbar fractures. *Indian J Neurosurg*. 2015;(4):56-62.
3. 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.
4. Heinzelmann M, Wanner GA. Thoracolumbar spinal injuries. In: BOOS N, Aebi M, editors. *Spinal disorders, fundamentals of diagnosis and treatment*. New York: Springer; 2008. p. 883-924.
5. Reinhold M, Audige L, Schnake KJ, Bellabarba C, Dai L, Oner FC. AO spine injury classification system: a revision proposal for the thoracic and lumbar spine. *Eur Spine J*. 2013;(22):2184-201.
6. Audige L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? *Acta Orthop Scand*. 2004;75(2):184-94.
7. Audige L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma*. 2005;(19):404-9.
8. Vaccaro AR, Oner C, Kepler CK, Dvorak M, Schnake K, Bellabarba C, et al. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. *Spine*. 2013;38(23):2028-37.
9. Azam Q, Sadat-Ali M. The concept of evolution of thoracolumbar fracture classifications helps in surgical decisions. *Asian Spine J*. 2015;9(6):984-94.
10. Kepler CK, Vaccaro AR, Koerner JD, Dvorak MF, Kandziora F, Rajasekaran S, et al. Reliability analysis of the AOSpine thoracolumbar spine injury classification system by a worldwide group of naive spinal surgeons. *Eur Spine J*. 2016;25(4):1082-6.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;(33):159-74.
12. Oner FC, Ramos LMP, Simmermacher RKJ, Kingma PTD, Diekerhof CH, Dhert WJA, 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(3):235-45.
13. Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB, Mehbood AA. Assessment of two thoracolumbar fracture

- classification systems as used by multiple surgeons. *J Bone Joint Surg Am.* 2005;87(7):1423-9.
14. Maçaneiro CH, Miyamoto RK, Lauffer RF, Larsen RV. Avaliação da reprodutibilidade entre duas classificações de fraturas da coluna tóracolombar e suas correlações com o tratamento. *Coluna/Comumna.* 2008;7(2):153-9.
 15. Martin JS, Marsh JL, Bonar SK, Decoster TA, Found EM, Brandser EA. Assessment of the AO/ASIF fracture classification for the distal tibia. *J Orthop Trauma.* 1997;11(7):477-83.
 16. Sanders RW. The problem with apples and oranges. *J Orthop Trauma.* 1997;(11):465-6.
 17. Marques CAC, Graells XS, Kulcheski AL, Meurer G, Benato M, Santoro PG. Reliability of the AO classification of thoracolumbar fractures compared to TLICS and Magerl. *Coluna/Columna.* 2017;16(1):56-9.
 18. Azimi P, Mohammadi HR, Azhari S, Alizadeh P, Montazeri A. The AOSpine thoracolumbar spine injury classification system: a reliability and agreement study. *Asian J Neurosurg.* 2015;10(4):282-5.