Reliability Evaluation of the New AO Spine-DGOU Classification for Osteoporotic Thoracolumbar Fractures

Guisela Quinteros[1,2], Juan P. Cabrera[3], Julio Urrutia[4], Charles A. Carazzo[5], Alfredo Guiroy[6], Bartolomé Marré[7,8], Andrei Joaquim[9], Ratko Yurac[7,8]

■ OBJECTIVES: **To perform an interobserver and intraobserver agreement evaluation of the new AO Spine-DGOU classification system for osteoporotic thoracolumbar fractures (OFc).**

■ METHODS: **Complete imaging studies of 97 patients (radiographs, computed tomography scans, and magnetic resonance imaging) with osteoporotic thoracolumbar fractures were selected and classified using the OFc by 6 spine surgeons (3 senior surgeons with more than 15 years of experience and 3 surgeons with less than 15 years). After a 4-week interval, the same cases were presented to the same evaluators in a random sequence for a new classification assessment. The weighted kappa coefficient (wK) was used to determine the interobserver and intraobserver agreement.**

■ RESULTS: **The interobserver agreement was moderate, wK = 0.59 (95% confidence interval 0.54–0.64). The intraobserver agreement was fair, wK = 0.35 (95% confidence interval 0.29–0.40). Interobserver agreement slightly improved for junior staff between first and second evaluation, suggesting a learning effect. Better agreement was obtained by senior staff at the interobserver and intraobserver agreement.**

■ CONCLUSIONS: **This independent assessment demonstrated that new OFc allows moderate interobserver agreement and fair intraobserver agreement. Further studies are necessary prior to its widespread adoption.**

INTRODUCTION

The elderly population is increasing worldwide, along with an increased life expectancy. In parallel, the incidence of osteoporotic vertebral fracture (OVF) is increasing, causing significant morbidity and mortality[1]; however, no classification of OVF has obtained international recognition and acceptance. Recently, the Spinal Section of the German Orthopedic and Trauma Society (DGOU) developed the osteoporotic fracture (OF) Classification, also adopted by AO Spine (AO Spine-DGOU Osteoporotic Fracture Classification System), a morphologic classification of the different types of OVF.[2] The OF classification graded thoracolumbar OVFs according to their morphologic and deformity components into 5 types, progressively more severe. These 5 degrees of severity range from no deformation with vertebral body edema (OF 1), deformation of 1 endplate without or with minimal minor posterior wall involvement (OF 2), deformation of 1 endplate with distinct posterior wall involvement (OF 3), deformation of both endplates with/without posterior wall involvement (OF 4), and injuries with anterior or posterior tension band injuries (OF 5).

This classification was validated by the proponent authors with substantial interobserver agreement, with an overall kappa

Key words
- Agreement study
- Osteoporotic vertebral fractures
- Osteoporotic fracture classification

Abbreviations and Acronyms
CI: Confidence interval
CT: Computed tomography
DGOU: Spinal Section of the German Orthopedic and Trauma Society
MRI: Magnetic resonance imaging
OF: Osteoporotic fracture
OVF: Osteoporotic vertebral fracture
wK: weighted kappa

From the [1]Spine Unit, Orthopedics Department, Hospital Regional de Talca, Talca, Chile; [2]Facultad de Medicina, Universidad Católica del Maule, Talca, Chile; [3]Department of Neurosurgery, Hospital Clínico Regional de Concepción, Concepción, Chile; [4]Department of Orthopaedic Surgery, School of Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile; [5]Department of Neurosurgery, University of Passo Fundo, São Vicente de Paulo Hospital, Passo Fundo, Rio Grande do Sul, Brazil; [6]Spine Unit, Orthopedic Department, Hospital Español de Mendoza, Mendoza, Argentina; [7]Department of Orthopedic and Traumatology, University del Desarrollo, Santiago, Chile; [8]Spine Unit, Department of Traumatology, Clínica Alemana, Santiago, Chile; and [9]Department of Neurosurgery, University of Campinas (UNICAMP), Campinas-SP, Brazil

To whom correspondence should be addressed: Guisela Quinteros, M.D. [E-mail: Guisela.quinterosr@gmail.com]

Citation: World Neurosurg. (2022) 161:e436-e440. https://doi.org/10.1016/j.wneu.2022.02.029

*Journal homepage: **www.journals.elsevier.com/world-neurosurgery***

*Available online: **www.sciencedirect.com***

1878-8750/$ - see front matter

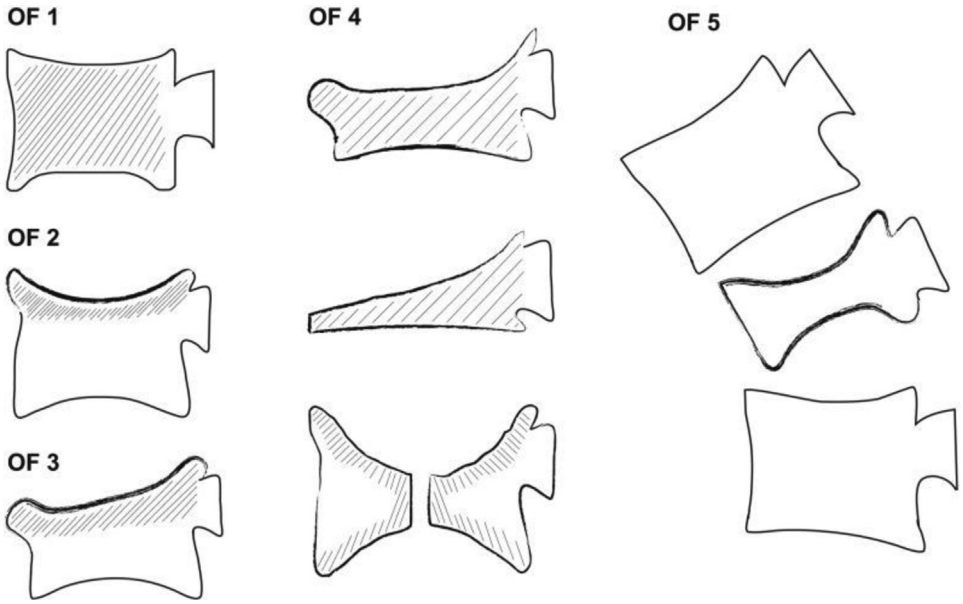

Figure 1. Schematic representation of the 5 OF subtypes (OF 1—OF 5). Reprinted with permission waived from Schnake et al.[2]

coefficient of 0.63[3]; however, this classification has not been validated by an independent external group. Reliability of a classification system is of paramount importance prior to its use and may be influenced by clinical scenarios, level of surgeon's experience, and geographical differences of health care systems, among other factors. If a classification is not reliable, it is difficult to accept its use, as it may not reflect a real-world scenario.[4]

The aim of the present study was to evaluate the agreement using the OF Classification system.

METHODS

Study Design and Data Collection

Institutional review board approval was obtained prior to perform this study. One fellowship-trained spine surgeon retrospectively selected 97 cases of patients who had osteoporosis and thoracolumbar fractures treated in 1 center from 2015 to 2020, excluding traumatic injuries in young patients. An online survey (SurveyMonkey) with the complete imaging studies (plain radiographs, computed tomography (CT) scan and magnetic resonance imaging (MRI) of the affected segment/level) of patients with osteoporotic fractures from T1 to L5 was e-mailed to 6 evaluators in August 2020.

Evaluators classified the fractures according to the OF Classification, which is described in Figure 1.

After a 4-week interval, the 97 cases were emailed again in a random sequence for a new evaluation to assess intraobserver agreement. The evaluators were 6 spine surgeons from 5 Latin American centers, who were divided according to their years of experience in 2 groups: 3 surgeons with less than 15 years of experience and 3 surgeons with more than 15 years of experience.

Before classifying the injuries, all surgeons had discussed in an online meeting the original article by Schnake et al.[2] The survey data were compiled in a Microsoft Excel spreadsheet (Microsoft Corporation 2020; Redmond, WA).

Table 1. The Distribution of Fractures According to the OF Type

OF Type	Fracture Type Distribution (%)
OF 1	1 (1%)
OF 2	26 (27%)
OF 3	27 (27%)
OF 4	30 (31%)
OF 5	13 (14%)

Inclusion and Exclusion Criteria

We included patients with acute OVF (<4 weeks) having complete available imaging studies, including anteroposterior and lateral plain radiographs, CT scan, and MRI. The exclusion criteria were patients with nonosteoporotic traumatic thoracolumbar fractures, metastatic vertebral fractures, and patients without complete imaging studies.

Statistical Analysis

Considering the data from Schnake et al.,[3] a confidence interval approach to sampling size estimation for interobserver

Table 2. Interobserver Agreement (wκ)

OF Type	wκ Value	95% CI	Agreement
Overall OF	0.59	0.54–0.64	Moderate

agreement studies with multiple raters was used, as reported by Rotondi et al.[5] For 6 evaluators, with a 95% confidence interval (CI), a lower limit of 0.6 and an upper limit of 0.8 (an expected substantial reliability), we determined 65 cases as the required sample. However, considering that some subtypes are infrequent, we increased the number of cases to 97 patients to ensure enough cases of each subtype.

The analyzes were performed with RStudio software version 1.1.383 (https://www.rstudio.com/products/rstudio/download/) using the *irr*, *raters*, and *psych* packages. We applied the weighted kappa statistics (wκ) for 2-way agreements. wκ allows measuring agreement with multiple response levels when not all disagreements are equally important; weight was set linearly. Levels of agreement for κ were determined as proposed by Landis et al.[6] with κ values 0.00–0.20 considered slight agreement; 0.21–0.40 fair agreement; 0.41–0.60 moderate agreement; 0.61–0.80 substantial agreement; and 0.81–1.00 almost perfect agreement.

RESULTS

We evaluated 97 patients, with a mean age of 75 ±14 years. Eighty-six fractures were in the thoracolumbar area (T11-L2) (89%), and 11 in the lumbar spine (L3-L5) (11%).

There were a total of 1164 responses, 582 in each evaluation. The distribution of the fractures classified was 1 (1%) OF 1, 26 (27%) OF 2, 27 (27%) OF 3, 31 (30%) OF 4, and 13 (14%) OF 5 (Table 1).

Most of the cases were OF 2, OF 3, and OF 4.

Interobserver Agreement

There was a moderate interobserver agreement, wκ = 0.59 (95% CI 0.54–0.64; Table 2).

When we compared the wκ from the first evaluation with the second evaluation, there were some improvements for less-experienced surgeons, changing from an acceptable agreement (first evaluation: wκ = 0.31; 95% CI 0.20–0.40) to a considerable agreement (second evaluation: wκ = 0.61; 95% CI 0.51–0.70); the agreement did not change between senior surgeons from the first to the second assessment (Tables 3 and 4)

Table 3. Interobserver Agreement (wκ) Obtained at First Evaluation for Senior and Junior Assessors

OF Type	First Evaluation wκ Values	95% CI	Agreement
All OF types	0.57	0.50–0.64	Moderate
Senior staff	0.62	0.52–0.70	Substantial
Junior staff	0.31	0.20–0.40	Fair

Table 4. Interobserver Agreement (κ) Obtained at Second Evaluation for Senior and Junior Assessors

OF Type	Second Evaluation wκ Values	95% CI	Agreement
All OF types	0.62	0.50–0.64	Substantial
Senior Staff	0.61	0.52–0.70	Substantial
Junior Staff	0.61	0.51–0.70	Substantial

Specifically assessing some fracture types, a higher level of agreement was obtained when classifying OF 2, OF 4, and OF 5 types (Table 5).

Intraobserver Agreement

Intraobserver agreement was fair, wκ = 0.35 (95% CI 0.29–0.40).

When comparing intraobserver agreement according to the evaluators' experience, it was observed that senior surgeons had a nonsignificant higher agreement (Table 6).

DISCUSSION

Considering the importance of this new classification system, independent agreement evaluation is an essential step to corroborate its external reproducibility. Our study reported moderate interobserver agreement (wκ = 0.59) and fair intraobserver agreement (wκ = 0.35) using this classification of OVF. The interobserver agreement was inferior to those obtained by the proponents of the classification, who reported a substantial agreement, with a kappa value of 0.63.[3] The intraobserver agreement obtained by our raters requires caution prior to adopting this new system.

Some surgeons extrapolate the traditional systems used for nonosteoporotic trauma to evaluate OVF in clinical practice.[7] The Genant semiquantitative assessment[8] and the Sugita osteoporotic classification[9] are classification schemes based on lateral radiographs. Kanchiku et al.[10] proposed a classification of OVF using MRI and radiographs. Recently, a new classification system for chronic symptomatic osteoporotic thoracolumbar fractures was described; it is based on dynamic radiography, CT scans, and MRI.[11] The OF classification, on the other hand, is based on plain radiographs, CT scans, and MRI.[2] This new system proposed a crescent degree of instability, with a clear

Table 5. Interobserver Agreement (κ) by OVF Type

OF Type	wκ	95% CI	Agreement
OF 1	0.23	0.05–0.41	Fair
OF 2	0.53	0.45–0.60	Moderate
OF 3	0.33	0.26–0.39	Fair
OF 4	0.50	0.42–0.57	Moderate
OF 5	0.54	0.42–0.65	Moderate

Table 6. Intraobserver Agreement (κ) According to Surgeon Experience

OF Type	WK	95% CI	Agreement
Overall OF	0.35	0.29—0.40	Fair
Senior staff	0.43	0.35—0.50	Moderate
Junior staff	0.31	0.24—0.37	Fair

progression of bone destruction from type 1 to 5, as well illustrative and descriptive scenarios of each morphology.

We observed that a better interobserver agreement was obtained assessing OF 2, OF 4, and OF 5. We believe that the low agreement evaluating the OF 1 subtype may be explained by the low number of cases identified as OF 1 fractures, since a low prevalence affects kappa values. Higher agreement between more severe injury morphologies is also expected. The intermediary forms may be more affected by minor morphologic changes that may be difficult to assess, leading to classification disagreement. We also must emphasize the improvement in the agreement rate of the junior staff in the second round, which strongly suggested that they learned how to use the system—an effect of practice.

An adequate classification of OVF should not only help to perform a proper diagnosis; more importantly, it should guide treatment. Managing patients with OVF can be complex considering the altered physiology of old age, frailty, comorbidities, cognitive dysfunction, and multi-pharmacy, among other factors.[12] OF classification has gained international visibility because it suggests a treatment algorithm for each type of fracture.[13] Recently, the authors of this classification have described that a higher preoperative OF subgroup shows a higher postoperative benefit regarding radiologic parameters (vertebral body height and local kyphotic angle) after kyphoplasty.[14] Nevertheless, based on our results we cannot state that this classification meets, to date, a sufficient level of agreement to be established as a universal method of communication among spine surgeons. Future prospective studies are required to determine the actual clinical utility of this classification scheme. Likewise, no matter what type of classification the spine surgeon adopts, the individual decision depends upon fracture type, degree of instability, bone quality, and patient general health status.[15-21]

Some of the strengths of this study are that the classification's reproducibility process was performed by 6 spine surgeons from 5 centers in 3 different countries. Additionally, our assessors had different years of experiences, which is the most common clinical setting. Therefore, our methodology meets the criteria suggested by Audigé et al. for reliability studies of fracture classifications.[22,23] Another strength of this study is the sample size estimation, which allows us to have enough cases (and raters) to avoid an underpowered study; moreover, like the original study describing this classification, it also used 6 raters.[3] However, our study has limitations. Despite our efforts for a rigorous methodology, a first limitation is the computational platform, which does not allow for the performance of exact measurements of vertebral body height, which can lead to confusion when choosing between OF 2 and OF 3 types, potentially affecting reliability. Another limitation is that, like the original authors' study describing this classification, 85% of fractures were OF 2, OF 3, or OF 4 types—with only one case of OF 1—limiting the interpretation of our findings. Finally, validity (therapy recommendations) is not permissible once we only evaluated OF Classification—not validating treatment recommendations.

This study provided an independent international evaluation of the agreement using this classification. However, we observed only a moderate interobserver agreement and a fair intraobserver agreement. Future prospective studies should establish whether this classification meets the standards to be used in the clinical management of OVF.

CONCLUSIONS

The new OF Classification system had moderate interobserver agreement and fair intraobserver agreement in this international external validation study. Further studies evaluating the reliability of this new system are necessary prior to its use in clinical studies.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Guisela Quinteros: Writing — review & editing, Methodology, Formal analysis. **Juan P. Cabrera:** Writing — review & editing. **Julio Urrutia:** Writing — review & editing, Methodology. **Charles A. Carazzo:** Writing — review & editing, Supervision. **Alfredo Guiroy:** Writing — review & editing, Formal analysis. **Bartolomé Marre:** Writing — review & editing, Methodology, Supervision. **Andrei Joaquim:** Writing — review & editing, Formal analysis, Language revision. **Ratko Yurac:** Writing — review & editing, Methodology, Supervision.

REFERENCES

1. Svedbom A, Hernlund E, Ivergård M, et al. Osteoporosis in the European Union: a compendium of country-specific reports. *Arch Osteoporos*. 2013;8:137.

2. Schnake KJ, Blattert TR, Hahn P, et al. Classification of osteoporotic thoracolumbar spine fractures: recommendations of the Spine Section of the German Society for Orthopaedics and Trauma (DGOU). *Global Spine J*. 2018;8(2 suppl):46S-49S.

3. Schnake KJ, Hahn P, Franck A, et al. Development of a classification system (OF-classification) and of a score for therapeutic decision making (OF-score) for osteoporotic thoracolumbar fractures [abstract]. *Eur Spine J*. 2013;22:2590.

4. Canseco JA, Schroeder GD, Patel PD, et al. Regional and experiential differences in surgeon preference for the treatment of cervical facet injuries: a case study survey with the AO Spine Cervical Classification Validation Group. *Eur Spine J*. 2021;30:517-523.

5. Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol*. 2012;65:778-784.

6. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

7. Walter J, Haciyakupoglu E, Waschke A, Kalff R, Ewald C. Cement leakage as a possible complication of balloon kyphoplasty—is there a difference between osteoporotic compression fractures (AO type A1) and incomplete burst fractures (AO type A3.1)? *Acta Neurochir (Wien)*. 2012;154:313-319.

8. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*. 1993;8:1137-1148.

9. Sugita M, Watanabe N, Mikami Y, Hase H, Kubo T. Classification of vertebral compression fractures in the osteoporotic spine. *J Spinal Disord Tech*. 2005;18:376-381.

10. Kanchiku T, Taguchi T, Kawai S. Magnetic resonance imaging diagnosis and new classification of the osteoporotic vertebral fracture. *J Orthop Sci*. 2003;8:463-466.

11. Hao DJ, Yang JS, Tuo Y, et al. Reliability and application of the new morphological classification system for chronic symptomatic osteoporotic thoracolumbar fracture. *J Orthop Surg Res*. 2020;15:348.

12. Rajasekaran S, Kanna RM, Schnake KJ, et al. Osteoporotic thoracolumbar fractures-how are they different? Classification and treatment algorithm. *J Orthop Trauma*. 2017;31(suppl 4):S49-S56.

13. Blattert TR, Schnake KJ, Gonschorek O, et al. [Nonsurgical and surgical management of osteoporotic vertebral body fractures: Recommendations of the Spine Section of the German Society for Orthopaedics and Trauma (DGOU)] [published correction appears in Orthopade. 2019 Jan 21;:]. *Orthopade*. 2019;48:84-91.

14. Palmowski Y, Balmer S, Hu Z, et al. Relationship between the OF classification and radiological outcome of osteoporotic vertebral fractures after kyphoplasty [published online ahead of print, 2020 Nov 2]. *Global Spine J*. 2020, 2192568220964051.

15. Gonschorek O, Hauck S, Weiß T, Bühren V. Percutaneous vertebral augmentation in fragility fractures—indications and limitations. *Eur J Trauma Emerg Surg*. 2017;43:9-17.

16. Marongiu G, Congia S, Verona M, Lombardo M, Podda D, Capone A. The impact of magnetic resonance imaging in the diagnostic and classification process of osteoporotic vertebral fractures. *Injury*. 2018;49(suppl 3):S26-S31.

17. Dreimann M, Stangenberg M, Eicker SO, Frosch KH, Viezens L. [Minimally invasive posterior and anterior stabilization of the thoracolumbar spine after traumatic injuries]. *Unfallchirurg*. 2020;123:752-763.

18. Goldstein S, Smorgick Y, Mirovsky Y, Anekstein Y, Blecher R, Tal S. Clinical and radiological factors affecting progressive collapse of acute osteoporotic compression spinal fractures. *J Clin Neurosci*. 2016;31:122-126.

19. Liu L, Wang Q, Cheng S, et al. Coronal imaging changes associated with recollapse of injured vertebrae after percutaneous vertebroplasty or percutaneous kyphoplasty treatment for osteoporotic thoracolumbar fracture. *World Neurosurg*. 2018;120:e737-e744.

20. Pizones J, Izquierdo E, Alvarez P, et al. Impact of magnetic resonance imaging on decision making for thoracolumbar traumatic fracture diagnosis and treatment. *Eur Spine J*. 2011;20(suppl 3):390-396.

21. Smorgick Y, Geftler A, Goldstein S, Mirovsky Y, Blecher R, Anekstein Y. Determination of any correlation between sagittal spinopelvic configuration and progressive collapse of acute osteoporotic compression spine fractures: a retrospective radiological analysis. *Asian Spine J*. 2020;14:872-877.

22. Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma*. 2005;19:401-406.

23. Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop Scand*. 2004;75:184-194.

Conflict of interest statement: This study was organized by the AO Spine Latin America Trauma Study Group. AO Spine is a clinical division of the AO Foundation, which is an independent medically guided not-for-profit organization. Study support was provided directly through AO Spine Latin America regarding data collection, data analysis, and proofreading.

Received 21 January 2022; accepted 7 February 2022

Citation: World Neurosurg. (2022) 161:e436-e440. https://doi.org/10.1016/j.wneu.2022.02.029

Journal homepage: ***www.journals.elsevier.com/world-neurosurgery***

Available online: ***www.sciencedirect.com***