ELSEVIER

The Spine Journal 000 (2021) 1–6

THE SPINE JOURNAL

Clinical Study

An independent inter- and intraobserver agreement assessment of the AOSpine sacral fracture classification system

Julio Urrutia, MD[a,*], Arturo Meissner-Haecker, MD[a], Nelson Astur, MD[b], Manuel Valencia, MD[c,d], Ratko Yurac, MD[c], Gaston Camino-Willhuber, MD[e], Marcelo Valacco, MD[f]

[a] *Department of Orthopaedic Surgery, School of Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile*
[b] *Santa Casa de Misericordia de São Paulo, Sao Paulo, Brazil and Hospital Israelita Albert Einstein, Morumbi, Sao Paulo, Brazil*
[c] *Department of Orthopaedic Surgery, Clinica Alemana de Santiago, Chile*
[d] *Hospital Mutual de Seguridad, Santiago, Chile*
[e] *Institute of Orthopedics "Carlos E. Ottolenghi," Hospital Italiano de Buenos Aires, Buenos Aires, Argentina*
[f] *Hospital Churruca Visca, Buenos Aires, Argentina*

Received 17 September 2020; revised 24 December 2020; accepted 6 February 2021

Abstract

BACKGROUND CONTEXT: The AOSpine sacral classification scheme was recently described. It demonstrated substantial interobserver and excellent intraobserver agreement in the study describing it; however, an independent assessment has not been performed.

PURPOSE: To perform an independent inter- and intraobserver agreement evaluation of the AOSpine sacral fracture classification system.

STUDY DESIGN: Agreement study.

METHODS: Complete computerized tomography (CT) scans, including axial images, with coronal and sagittal reconstructions of 80 patients with sacral fractures were selected and classified using the morphologic grading of the AOSpine sacral classification system by six evaluators (from three different countries). Neurological modifiers and case-specific modifiers were not assessed. After a four-week interval, the 80 cases were presented to the same raters in a random sequence for repeat assessment. We used the Kappa coefficient (κ) to establish the inter- and intraobserver agreement.

RESULTS: The interobserver agreement was substantial when considering the fracture severity types (A, B, or C), with κ=0.68 (0.63−0.72), but moderate when considering the subtypes: κ=0.52 (0.49−0.54). The intraobserver agreement was substantial considering the fracture types, with κ=0.69 (0.63−0.75), and considering subtypes, κ=0.61 (0.56−0.67).

CONCLUSION: The sacral classification system allows adequate interobserver agreement at the type level, but only moderate at the subtypes level. Future prospective studies should evaluate whether this classification system allows surgeons to decide the best treatment and to establish prognosis in patients with sacral fractures.

Keywords: Sacral fracture; Sacrum; Agreement study; Spinal injury classification system, Fracture classification

FDA device/drug status: Not applicable.

Author Disclosures: ***JU:*** Nothing to disclose. ***AMH:*** Nothing to disclose. ***NA:*** Nothing to disclose. ***MV:*** Nothing to disclose. ***RY:*** Nothing to disclose. ***GCW:*** Nothing to disclose. ***MV:*** Nothing to disclose.

Funding Disclosures: No funding was used to perform this study.

Institutional Review Board approval was obtained to conduct this study.

The manuscript submitted does not contain information about medical device(s)/drug(s).

*Corresponding author. Department of Orthopaedic Surgery, School of Medicine, Pontificia Universidad Catolica de Chile, Diagonal Paraguay 362, Santiago, Chile. Tel.: (56-2)23543467; fax: (56-2)23546847.

E-mail address: jurrutia@med.puc.cl (J. Urrutia).

https://doi.org/10.1016/j.spinee.2021.02.005
1529-9430/

Introduction

Fracture classifications should be comprehensive, easy to apply, and should demonstrate adequate inter- and intraobserver agreement. Thus, they can allow communication among physicians to identify and group comparable cases, guide the best treatment choices, and standardize research terminology. Recently, the Arbeitsgemeinschaft für Osteosynthesefragen (AO) Spine Knowledge Forum described a new sacral injury classification system [1]; this scheme was developed to classify the different patterns of sacral fractures and is intended to be used by spine and trauma surgeons. The AOSpine sacral classification is a morphology-based system developed considering the specific patterns of injuries that are observed in the sacrum [2], similar to the new AO Spine thoraco-lumbar spine injury classification system [3] and the subaxial cervical spine injury classification [4]. Furthermore, the AOSpine sacral fracture classification system also evaluates the degree of neurologic deficit, and the presence of case-specific modifiers (the presence of soft tissue damage, metabolic bone disease, anterior pelvic ring injury, and sacroiliac joint injury).

Several sacral fracture classifications have been previously described. Denis et al. [5] classified sacral fractures based on the location relative to the foramina, with a hierarchy founded on the risk of neurological injury; in that classification, zone-I fractures are those lateral to the foramina, zone-II fractures go through the foramina (but lateral to the spinal canal), and zone-III fractures extend medial of the foramina into the spinal canal. Roy-Camille et al. proposed a morphological classification for transverse sacral fractures, hypothesizing that the fracture pattern was determined by the presence of flexion or extension of the lumbar spine at the time of injury [6]. Isler et al. proposed a classification of pelvic ring injuries based on the integrity of the posterior sacroiliac ring segment; it described that a vertical fracture of the sacrum exiting medial to the articular process of S1 causes a subluxation or dislocation of the ipsilateral L5−Sl joint, thus causing a significant instability [7]. Also, Vaccaro et al. proposed a morphological sacral fracture classification, including combinations of horizontal and vertical patterns [8]. However, these classifications are not comprehensive enough to include the entire range of sacral fractures, and they have not achieved a universal use. Additionally, they lack independent validations, and they are not prognosis-based schemes allowing treatment guidelines.

The AOSpine sacral fracture classification was described by a group of world leaders in spine and orthopedic trauma; they obtained moderate interobserver agreement applying it, with substantial agreement assessing levels of severity (ie, A, B, or C), and substantial intraobserver agreement (with excellent agreement evaluating levels of severity) [1]. However, any classification requires independent validations before it is universally accepted and reaches wide clinical use. The objective of this study was to perform an independent inter- and intraobserver agreement evaluation of the AOSpine sacral fracture classification system.

Materials and methods

Institutional review board approval was obtained to perform the study. One author, who later did not participate in the classification phase of this study, retrospectively selected 80 cases with sacral fractures from a large patient database treated in two tertiary care centers from 2011 to 2020. The author who selected the cases was independent from the raters who assessed the cases.

The inclusion criterion was the presence of any sacral fracture evaluated with a CT scan; exclusion criteria were the presence of osteosynthesis in the sacrum without preoperative images (seven cases), pathological fracture (four cases), and old fractures without acute images (two cases). To be included in the study, patients were required to have complete computerized tomography (CT) scans, including axial images, with coronal and sagittal reconstructions. Coronal and sagittal images were reformatted from axial images using bone windows; a section thickness of 1 mm was used to reformat the images. The CT scans were reviewed using the Impax Web3000 program (Agfa-Gevaert, Mortsel, Belgium) by six fellowship-trained orthopedic spine surgeons, from five different centers in three countries; all raters are consultants in their centers. The raters were trained in this new classification scheme through an online session to discuss it and to clarify doubts before performing the evaluations to standardize the rating process [9]. Additionally, they were provided with the original article by Vaccaro et al. [1] to resolve any doubt at the time of assessment. The evaluators were unaware of the patients' identification or the treatments they received.

The six raters classified the fractures according to the morphologic grading of the AOSpine sacral classification system into types A, B, or C injuries (Fig. 1). They also performed an analysis for subtypes at the same time they determined the main type. The neurological modifiers and case-specific modifiers were not included in the agreement study (alike the study by Vaccaro et al. [1]) because no clinical information was provided.

Interobserver agreement was determined by comparing the initial responses of the six evaluators. Intraobserver agreement was established by comparing the same rater's responses between two evaluations of the same cases. The assessments were separated by a 4-week interval and cases were presented randomly to avoid recall bias.

R software (The R Project for Statistical Computing, Vienna, Austria) was used for sample size estimation. Considering the data from Vaccaro et al. [1], a confidence interval approach to sampling size estimation for interobserver agreement studies with multiple raters was used as reported by Rotondi et al. [10]. For 6 evaluators, with a 95% confidence interval, a lower limit of 0.6 and an upper limit of 0.8 (an expected substantial reliability),

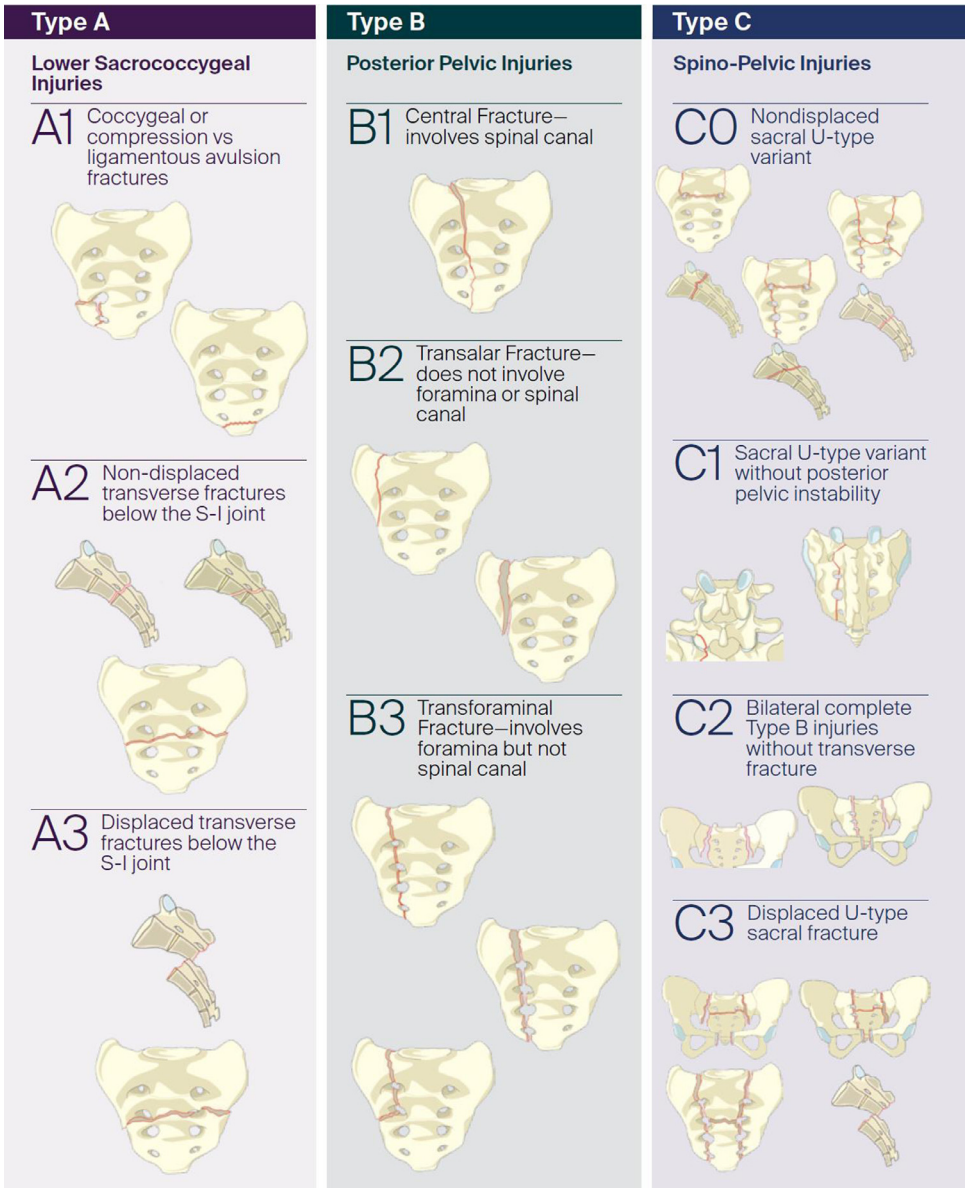

Figure 1. Main characteristics of injury types and subtypes. Acknowledgement of Copyright – AO Foundation.

we determined 65 cases as the required sample. However, considering that some subtypes are infrequent, we increased the number of cases to 80 patients to assure enough cases of each subtype.

Statistical analysis was conducted with SPSS version 17 (SPSS, Chicago, IL, USA). We used the Kappa coefficient (κ) to determine the interobserver and intraobserver agreement; κ values were expressed with a 95% confidence interval (CI). The agreement was first measured at the type level (A, B, or C type), and then at the subtype level. Levels of agreement for κ were determined as proposed by Landis et al. [11] with κ values 0.00 to 0.20 considered slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1.00, almost perfect agreement.

Results

At least one evaluator identified all three of the different major fracture severity patterns described by the AOSpine sacral classification system (A, B, and C), as well as all the different subtypes of injuries. Of the 480 classifications, 56 (11.7%) were described as type A fractures, 301 (62.7%), as type B injuries, and 123 (25.6%) as type C lesions (Table 1).

Interobserver agreement

We observed full interobserver agreement in 51 of the 80 cases (64%) when comparing morphologic grading by fracture severity type (A, B, or C), with a κ value of 0.68 (95% CI: 0.63–0.72), considered as substantial agreement.

Table 1
Distribution of responses

	Our study			Vaccaro et al. [1]	
Type/Subtype	N	% (subtype)	% (type)	% (subtype)	% (type)
A1	16	3.3	11.7	2.1	15.8
A2	32	6.7		8.6	
A3	8	1.7		5.1	
B1	4	0.8	62.7	3.9	47.7
B2	205	42.7		28.4	
B3	92	19.2		15.4	
C0	34	7.1	25.6	6.3	35.3
C1	26	5.4		7.2	
C2	23	4.8		4.9	
C3	40	8.3		16.9	
Missing	0	0	0	1.2	1.2
Total	480	100	100	100	100

The κ values for each fracture type were 0.79 (95% CI: 0.73−0.85) for type A fractures, 0.65 for type B injuries (95% CI: 0.59−0.71), and 0.64 (95% CI: 0.59−0.70) for type C injuries Table 2). These values are considered substantial agreement for type A, B, and C injuries.

When we assessed the level of agreement according to the subtype level, a full agreement was observed in 26 of the cases (32.5%); the κ value was 0.52 (95% CI: 0.49−0.54), which is considered moderate agreement. The detailed inter-rater agreement by subtype level is shown in Table 3.

Intraobserver agreement

In the repeated assessment four weeks after the first evaluation, we observed 83.5% full intraobserver agreement according to the fracture severity level (A, B, or C); the κ value was 0.69 (95% CI: 0.63−0.75), which is considered substantial agreement. The detailed intraobserver agreement by type level is shown in Table 4.

Full agreement at the subtype level was obtained in 70.6% of the evaluations, with a κ value of 0.61 (95% CI: 0.56−0.67), which is also a substantial agreement. The detailed intraobserver agreement by subtype level is shown in Table 5.

Discussion

Our study showed substantial interobserver agreement at the fracture type level (κ=0.68), and moderate agreement at the subtype level (κ=0.52) using the morphologic grading of the AOSpine sacral fracture classification system. This finding is similar to the agreement recently published by the authors who described this classification [1]; however, to obtain a widespread use, a new classification scheme must demonstrate adequate inter- and intraobserver agreement not only among the authors who developed it, but also among independent evaluators to corroborate the classification's external validity.

A sacral fracture classification system must be comprehensive to include all the different patterns of trauma, and it should also demonstrate adequate inter- and intraobserver agreement. While some previous classifications have analyzed sacral fractures as part of pelvic injury classification systems [7,12,13], other classifications have focused on developing more specific classification patterns of sacral fractures [5,6,14,15]. Nevertheless, those classifications did not present a comprehensive scheme including all types of sacral injuries, and they do not allow an adequate evidence-based treatment algorithm. Moreover, to the best of our knowledge, none of them has been evaluated with an

Table 2
Interobserver agreement (κ) by type

Type	K	95% CI
A	0.79	0.73−0.85
B	0.65	0.59−0.71
C	0.64	0.59−0.70

Table 3
Inter-observer agreement (κ) by subtype

Subtype	κ	95% CI
A1	0.33	0.27 - 0.38
A2	0.56	0.5 - 0.62
A3	0.19	0.13 - 0.24
B1	-0.01	- 0.07 - 0.0
B2	0.70	0.64 - 0.75
B3	0.49	0.43 - 0.55
C0	0.42	0.36 - 0.47
C1	0.24	0.18 - 0.29
C2	0.33	0.28 - 0.39
C3	0.56	0.5 - 0.61

Table 4
Intra-observer agreement (κ) by type

Type	κ	95% CI
A	0.71	0.62 - 0.80
B	0.67	0.58 - 0.76
C	0.70	0.61 - 0.79

Table 5
Intra-observer agreement (κ) by subtype

Subtype	κ	95% CI
A1	0.43	0.34 - 0.52
A2	0.54	0.45 - 0.62
A3	0.53	0.44 - 0.62
B1	0.28	0.19 - 0.37
B2	0.70	0.61 - 0.79
B3	0.60	0.51 - 0.69
C0	0.62	0.53 - 0.71
C1	0.46	0.38 - 0.55
C2	0.48	0.39 - 0.57
C3	0.74	0.65 - 0.83

inter- and intraobserver agreement study. Therefore, no universally accepted classification scheme has been adopted for sacral fractures. The AOSpine sacral classification proposes a hierarchical system based on an increasing grade of posterior pelvic and spino-pelvic instability. Thus, type-A fractures do not present posterior pelvic or spino-pelvic instability; type-B fractures are vertical injuries with posterior pelvic instability but not spino-pelvic instability, and type C injuries are those exhibiting spinopelvic instability.

Our study showed substantial inter-observer agreement at the fracture severity level, comparable to the agreement observed by the authors who described the classification, who reported a κ=0.75 [1]. Likewise, we obtained a moderate interobserver agreement at the fracture subtype level, which is equivalent to the agreement obtained in the study performed by Vaccaro et al. (κ=0.58) [1]. Moreover, our results are also comparable with previous independent validations for the AOSpine thoracolumbar spine (κ=0.62 for the main types and κ=0.55 for subtype) and subaxial cervical spine (κ=0.61 for the main types and κ=0.57 for subtypes) classifications [16–18], which also used six raters to assess each classification. However, like for the cervical and thoraco-lumbar injury classifications, such agreements reveal that an important proportion of injuries will be classified differently by various evaluators; therefore, many patients presenting the same injury could receive different treatments. This represents a limitation of classification schemes, and every effort to have classifications allowing adequate communication among treating physicians should be encouraged.

Similar to Vaccaro et al. [1], the highest interobserver agreement at the fracture severity level was obtained with type A fractures; however, we found lower agreement rating fracture subtype B1 (while Vaccaro et al. had their lowest agreement rating A1 injuries). We believe the low agreement evaluating B1 fractures could result from the low number of cases identified as B1 fractures, since the prevalence of the outcome strongly influences κ value. Nonetheless, the interobserver agreement for type B and type C fractures was substantial (κ=0.65 and 0.64, respectively).

Noteworthily, our assessors evaluated the entire imaging files on each patient, which is a strength of our study, as it reproduces a more real clinical practice. However, it may explain why we obtained a lower intraobserver agreement than the study by Vaccaro et al. [1]. Although it is usual that independent studies show lower agreement than the studies from the original groups that developed the classifications [19,20], the study from Vaccaro et al. used key images showing the fracture morphology for each case. It is possible that by having to decide the most important images from the entire imaging files, our raters might have changed their answers from the first to the second assessment. Furthermore, another strength is that we included evaluators from five different centers; therefore, we minimized an institutional bias assessing these injuries. While an agreement study could have a larger number of raters, the benefit of performing a sample size estimation is not only to have enough cases (or raters) to avoid an underpowered study. It also avoids an unnecessary number of cases (or evaluators) to reveal the effect under investigation. In an agreement study, a larger number of evaluators is used to decrease the number of cases to be assessed; however, as some sacral injury subtypes are infrequent, it does not seem possible to decrease the number of cases to be assessed. Moreover, independent validations of the AOSpine thoracolumbar and subaxial cervical spine classifications have been performed with the same number of raters [16–18].

We detected certain limitations of the sacral fracture classification system when we applied it to our cases. First, even though this is a hierarchical scheme, several traumatic B fractures (especially B3 fractures) are more severe than C0 insufficiency fracture; maybe, this scheme should be restricted only to traumatic injuries, as other AOSpine classification systems. Second, the classification does not specify how many millimeters or degrees of angulation are required to consider a fracture as displaced; this may explain a difficulty in differentiating A2 from A3, or C0 from C3 injuries. Additionally, we found difficulties identifying C1 injuries; while type C fractures designate injuries resulting in spinopelvic instability, C1 fractures can be easily confused with type B3 injuries. A disagreement classifying a fracture as type B3 or C1 would be important, as it would be a disagreement at the main type of injury. Finally, it is unclear if a unilateral alar fracture plus a transverse fracture corresponds to a type B or type C injury; although it is an unusual fracture pattern, these injuries may be difficult to classify.

It has been recommended that a new classification should undergo a three-step validation process [21]: (1) A definition of categories by experts, (2) A multicenter agreement evaluation performed by a representative panel of future users of the classification, and (3) A prospective clinical study to assess its clinical usefulness. Our study contributes to this classification's validation process with a multicenter agreement assessment performed by spine surgeons from five centers in three countries. Nevertheless, future studies should also include pelvic trauma surgeons as raters, as they also are potential users of this scheme; moreover, sacral fractures are frequently part of a pelvic ring injury.

Our results have clinical connotations; the interobserver agreement obtained at the main type level indicates that this classification allows adequate communication, and it may be used in clinical practice; however, at the subtype level, the inter-observer agreement is only moderate. From a research perspective, the substantial agreement obtained using this classification should allow clinical studies using this system to be considered reliable (as far as it includes the three main type levels), particularly as regards multicenter studies or meta-analyses. However, only future prospective studies will establish whether this classification allows surgeons to decide the best treatment for patients with sacral injuries.

Conflict of interest

The authors have no conflict of interest to disclose.

References

[1] Vaccaro AR, Schroeder GD, Divi SN, Kepler CK, Kleweno CP, Krieg JC, et al. Description and Reliability of the AOSpine Sacral Classification System. J Bone Joint Surg Am 2020;102:1454–63.

[2] Schroeder GD, Kurd MF, Kepler CK, Krieg JC, Wilson JR, Kleweno CP, et al. The Development of a Universally Accepted Sacral Fracture Classification: A Survey of AOSpine and AOTrauma Members. Global Spine J 2016;6:686–94.

[3] 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 (Phila Pa 1976) 2013;38:2028–37.

[4] Vaccaro AR, Koerner JD, Radcliff KE, Oner FC, Reinhold M, Schnake KJ, et al. AOSpine subaxial cervical spine injury classification system. Eur Spine J 2016;25:2173–84.

[5] Denis F, Davis S, Comfort T. Sacral fractures: an important problem. Retrospective analysis of 236 cases. Clin Orthop Relat Res 1988;227:67–81.

[6] Roy-Camille R, Saillant G, Gagna G, Mazel C. Transverse fracture of the upper sacrum. Suicidal jumper's fracture. Spine (Phila Pa 1976) 1985;10:838–45.

[7] Isler B, Ganz R. Classification of pelvic ring injuries. Injury 1996;27 (Suppl 1). S-A3-12.

[8] Vaccaro AR, Kim DH, Brodke DS, Harris M, Chapman JR, Schildhauer T, et al. Diagnosis and management of sacral spine fractures. Instr Course Lect 2004;53:375–85.

[9] Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 2011;64(1):96–106.

[10] 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(7):778–84.

[11] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.

[12] Tile M. Pelvic ring fractures: should they be fixed? J Bone Joint Surg Br 1988;70(1):1–12.

[13] Letournel E. [Surgical fixation of displaced pelvic fractures and dislocations of the symphysis pubis (excluding acetabular fractures) (author's transl). Rev Chir Orthop Reparatrice Appar Mot 1981;67 (8):771–82.

[14] Strange-Vognsen HH, Lebech A. An unusual type of fracture in the upper sacrum. J Orthop Trauma 1991;5(2):200–3.

[15] Sabiston CP, Wing PC. Sacral fractures: classification and neurologic implications. J Trauma 1986;26(12):1113–5.

[16] Urrutia J, Zamora T, Yurac R, Campos M, Palma J, Mobarec S, et al. An independent interobserver reliability and intraobserver reproducibility evaluation of the new AOSpine Thoracolumbar Spine Injury Classification System. Spine (Phila Pa 1976) 2015;40:E54–8.

[17] Urrutia J, Zamora T, Yurac R, Campos M, Palma J, Mobarec S, et al. An Independent Inter- and Intraobserver Agreement Evaluation of the AOSpine Subaxial Cervical Spine Injury Classification System. Spine (Phila Pa 1976) 2017;42:298–303.

[18] Urrutia J, Zamora T, Campos M, Yurac R, Palma J, Mobarec S, et al. A comparative agreement evaluation of two subaxial cervical spine injury classification systems: the AOSpine and the Allen and Ferguson schemes. Eur Spine J 2016;25:2185–92.

[19] Oner FC, Ramos LM, Simmermacher RK, Kingma PT, Diekerhof CH, Dhert WJ, 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.

[20] Blauth M, Bastian L, Knop C, Lange U, Tusch G. [Inter-observer reliability in the classification of thoraco-lumbar spinal injuries]. Der Orthopade 1999;28(8):662–81.

[21] Audige L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. J Orthop Trauma 2005;19(6):401–6.