ELSEVIER

The SPINE JOURNAL

The Spine Journal 000 (2022) 1–6

Clinical Study

An independent inter- and intra-observer agreement assessment of the AOSpine upper cervical injury classification system

Julio Urrutia, MD[a], Byron Delgado, MD[a], Gaston Camino-Willhuber, MD[b,c], Alfredo Guiroy, MD[d,e], Nelson Astur, MD[f], Marcelo Valacco, MD[g], Juan José Zamorano, MD[h,i], Catalina Vidal, Bkin[a], Ratko Yurac, MD[i,j,*]

[a] *Department of Orthopaedic Surgery, School of Medicine, Pontificia Universidad Catolica de Chile*
[b] *Hospital Italiano de Buenos Aires, Institute of Orthopedics "Carlos E. Ottolenghi, Buenos Aires, Argentina*
[c] *UCI Medical Center, University of California, Irvine, CA, USA*
[d] *Orthopedic Department, Spanish Hospital, Mendoza, Argentina*
[e] *Elite Spine Health and Wellness Center, Fort Lauderdale, FL, USA*
[f] *Hospital Israelita Albert Einstein, Morumbi, Sao Paulo, Brazil*
[g] *Hospital Churruca Visca, Buenos Aires, Argentina*
[h] *Spine Unit, Hospital del Trabajador-ACHS, Santiago, Chile*
[i] *Department of of Orthopaedic Surgery, Clinica Alemana de Santiago, Santiago, Chile*
[j] *Department of Orthopedic Surgery, School of Medicine, University del Desarrollo, Santiago, Chile*

Received 25 July 2022; revised 25 October 2022; accepted 8 November 2022

Abstract

BACKGROUND CONTEXT: The complex anatomy of the upper cervical spine resulted in numerous separate classification systems of upper cervical spine trauma. The AOSpine upper cervical classification system (UCCS) was recently described; however, an independent agreement assessment has not been performed.
PURPOSE: To perform an independent evaluation of the AOSpine UCCS.
STUDY DESIGN: Agreement study.
PATIENT SAMPLE: Eighty four patients with upper cervical spine injuries.
OUTCOME MEASURES: Inter-observer agreement; intra-observer agreement.
METHODS: Complete imaging studies of 84 patients with upper cervical spine injuries, including all morphological types of injuries defined by the AOSpine UCCS were selected and classified by six evaluators (from three different countries). The 84 cases were presented to the same raters randomly after a 4-week interval for repeat evaluation. The Kappa coefficient (κ) was used to determine inter- and intra-observer agreement.
RESULTS: The interobserver agreement was almost perfect when considering the fracture site (I, II or III), with κ=0.82 (0.78−0.83), but the agreement according to the site and type level was moderate, κ=0.57 (0.55−0.65). The intra-observer agreement was almost perfect considering the injury, with κ=0.83 (0.78−0.86), while according to site and type was substantial, κ=0.69 (0.67−0.71).
CONCLUSIONS: We observed only a moderate inter-observer agreement using this classification. We believe our results can be explained because this classification attempted to organize many different injury types into a single scheme.

Keywords: Agreement study; Classification system; Fracture classification; Spine; Upper cervical spine fracture

FDA device/drug status: Not applicable.

JU: The Spine Journal (NASS): level B. ***BD:*** Nothing to disclose. ***GC-W:*** Nothing to disclose. ***AG:*** Nothing to disclose. ***NA:*** Nothing to disclose. ***MV:*** Nothing to disclose. ***JJZ:*** Nothing to disclose. ***CV:*** Nothing to disclose. ***RY:*** Nothing to disclose.

*Corresponding author. Department of Orthopaedic Surgery, Universidad del Desarrollo, Vitacura 5951, Vitacura, Santiago, Chile, Tel.: (56-2) 23543467; fax: (56-2)23546847.

E-mail address: ryurac@gmail.com (R. Yurac).

https://doi.org/10.1016/j.spinee.2022.11.005
1529-9430/

Introduction

The objectives of any classification scheme should be to enable communication among physicians through the identification and grouping of comparable cases, guide the best treatment options, and standardize research terminology. Ideally, a classification system should also allow the comparison of different treatment outcomes for similar injuries. Moreover, a classification should be comprehensive, easy to apply and demonstrate adequate inter- and intra-observer agreement to attain these objectives.

The complex anatomy of the upper cervical spine, with a broad spectrum of injuries affecting the occipital condyles, C1, the odontoid process, or the posterior arch of C2, has resulted in numerous separate classification systems of upper cervical spine trauma (UCT) [1–4]. These historical classification systems covered similar injury patterns and were primarily published before the advent of computed tomography (CT) scan reconstructions, generally using plain radiographs images [5–7].

Recently, the AOSpine Knowledge Forum released the Upper Cervical Classification System (UCCS) [8–10]; this scheme was developed to classify the different patterns of UCT into one system. This scheme is based on the other AOSpine classification systems; however, the UCCS divides injuries according to the affected region(s) (injury sites). These sites are I, injuries involving the occipital condyles and the craniocervical junction; II, injuries involving the atlas and the atlantoaxial joint; and III, injuries involving the axis and the C2-3 joint. Similar to the AOSpine thoraco-lumbar, subaxial cervical, and sacral fracture classification schemes, the AOSpine UCCS subdivides these three injury sites into three types: type A (stable), with bony injuries only, without significant ligamentous, tension band, or disc injury; type B (stable or unstable), having tension band or ligamentous injury, with or without bony injury, but without complete separation of anatomic integrity; and type C injuries (unstable), which have a significant translation in any directional plane and separation of anatomic integrity. [8,9,11].

The AOSpine UCCS was developed by a group of world leaders in spine and orthopaedic trauma; studies involving those authors obtained substantial inter-observer agreement classifying the different types of injuries and almost perfect agreement assessing the injury site [9,11]. However, any classification requires further independent validations before being universally accepted for clinical and educational use. Therefore, the objective of this study was to perform an independent inter- and intra-observer agreement evaluation of the AOSpine UCCS.

Material and methods

Institutional review board approval was obtained to perform this study. One author, who later did not participate in the classification phase of this study, retrospectively selected 84 cases with upper cervical injuries from a large patient database treated in two tertiary care centers. That author included all morphological types of injuries defined by the AOSpine UCCS to perform an adequate agreement study.

For inclusion in this study, the patients had to have complete cervical 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. Patients with non-traumatic upper cervical fractures (eg fractures associated with spinal tumors and infections) were excluded.

The CT scans were reviewed using the Impax Web4000 program (Agfa-Gevaert, Mortsel, Belgium) by six spine surgeons from six different centers in three countries. The raters were trained in this new classification scheme before performing their evaluations. They were provided with the original article to resolve any doubt at the assessment time. Additionally, they watched the full video of the AOSpine UCCS explanation given by Dr. Vaccaro before starting the evaluation [10]. The evaluators were unaware of the patients' identification or treatments they received.

The six raters classified the fractures according to the AOSpine UCCS. Injuries were classified according to location (I, occipital and craniocervical region; II, atlas and atlantoaxial joints; and III, axis and C2-3 joints) and injury type (A, B, and C), as shown in Figure. We did not assess the case-specific modifiers or the neurological status modifiers.

Inter-observer agreement was determined by comparing the initial responses of the six evaluators. Intra-observer agreement was established by comparing the same rater's responses between two evaluations of the same cases. A 4-week interval separated the assessments, and cases were presented randomly to avoid recall bias.

R (The R Project for Statistical Computing, Vienna, Austria) was used for sample size estimation. Considering the data from Maeda et al. [9], a confidence interval approach to sampling size estimation for inter-observer agreement studies with multiple raters was used as reported by Rotondi et al. [12]. For six evaluators, with a 95% confidence interval, 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 patients to 84 to have enough cases for each subtype.

Statistical analysis was conducted with SPSS version 17 (SPSS, Chicago, IL, USA). We used the Kappa coefficient (κ) to determine the inter-observer and intra-observer agreement; κ values were expressed with a 95% confidence interval (CI). The agreement was first measured at the injury site and then at the injury type. Levels of agreement for κ were determined as proposed by Landis et al. [13], with κ values 0.00 to 0.20 considered slight agreement; 0.21 to 0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement.

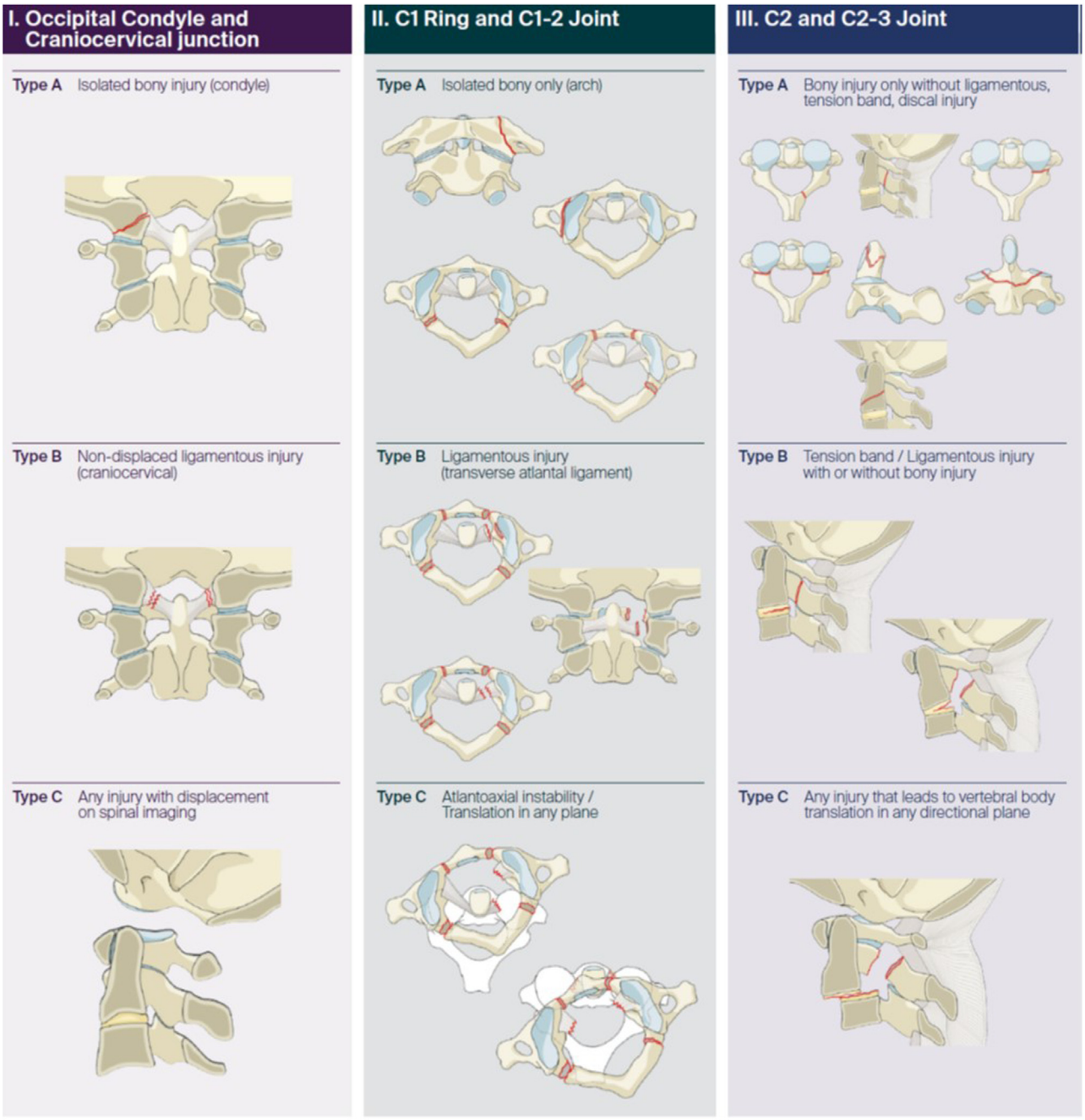

Fig. Main characteristics of injury sites and types. Acknowledgement of Copyright – AO Foundation

Results

At least one evaluator identified all three of the different injury sites described by the AOSpine UCCS (I, II, and III) and all the types (A, B, and C).

Of the 504 responses from six evaluations of the 84 cases, 12.30% were identified as involving the condyles or atlanto-occipital joints (site I), 18.85% as occurring at the atlas and atlanto-axial joints (site II) and 68.85% as affecting the axis and C2-3 disc and joints (site III lesions), as shown in Table 1.

The proportion of injuries classified in the different combinations of sites and types is shown in Table 2.

Inter-observer agreement

We observed an almost perfect inter-observer agreement (κ=0.82 [95% CI 0.78–0.83]), when assessing the fracture site (I, II, III). The κ value for each fracture site was 0.80 for type I fractures, 0.74 for type II, and 0.68 for type III injuries. (Table 3).

Table 1
Distribution of responses of injury sites

Type	N°	%
I	62	12.30
II	95	18.85
III	347	68.85
Total	504	100

Table 2
Distribution of responses of injury sites and types

Subtype	N°	%
IA	57	11.30
IB	3	0.60
IC	2	0.40
IIA	63	12.50
IIB	15	2.98
IIC	17	3.37
IIIA	289	57.34
IIIB	25	4.96
IIIC	33	6.55
Total	504	100

Table 3
Inter-observer agreement for each injury site

Type	κ	95% CI
Overall	0.82	0.78–0.83
I	0.80	0.74–0.86
II	0.75	0.69–0.81
III	0.88	0.83–0.94

According to the site and type level, the inter-observer agreement was moderate (κ=0.57 [95% CI 0.55–0.65]). The detailed interobserver agreement by site and type is shown in Table 4.

Intra-observer agreement

We observed an almost perfect intra-observer agreement considering the injury site (I, II, or III) with a κ=0.83 (95% CI 0.78–0.86). (Table 5). The intra-observer agreement according to site and type was substantial (κ=0.69 [95% CI 0.67–0.71]); the detailed intra-observer agreement is shown in Table 6.

Table 4
Inter-observer agreement for each injury site plus type

Subtype	κ	95% CI
Overall	0.57	0.55–0.65
IA	0.81	0.74–0.88
IB	−0.006	−0.06–0.05
IC	0.19	0.14–0.24
IIA	0.68	0.62–0.74
IIB	0.41	0.35–0.47
IIC	0.40	0.35–0.45
IIIA	0.66	0.60–0.72
IIIB	0.15	0.09–0.21
IIIC	0.20	0.15–0.25

Table 5
Intra-observer agreement for each injury site

Type	κ	95% CI
Overall	0.83	0.78–0.86
I	0.80	0.77–0.83
II	0.72	0.69–0.75
III	0.82	0.79–0.85

Table 6
Intra-observer agreement for each injury site plus type

Subtype	κ	95% CI
Overall	0.69	0.67–0.71
IA	0.79	0.76–0.82
IB	−0.006	−0.03–0.02
IC	0.04	0.01–0.07
IIA	0.65	0.62–0.68
IIB	0.34	0.31–0.37
IIC	0.39	0.36–0.42
IIIA	0.66	0.63–0.69
IIIB	0.15	0.12–0.18
IIIC	0.29	0.26–0.32

Discussion

The upper cervical spine has a complex and unique anatomy. Considering the complexity of this segment and different injury patterns, no previous universal classification system of UCT has been adopted by the spine community. Using the AOSpine UCCS, our panel exhibited almost perfect interobserver agreement detecting the fracture site, with κ=0.82. However, assessing the site and fracture type, we observed only a moderate agreement (k=0.57).

The purposes of an injury classification system should be to facilitate communication between peers, provide information on their severity, and improve decision-making regarding treatment, outcome, and prognosis [14]. Ideally, a classification scheme should also allow the comparison of different treatment outcomes for similar injuries. According to our findings, a sub-optimal agreement using a classification was observed, meaning that a significant fraction of injuries will be classified differently by various assessors; hence, many patients presenting the same injury could undergo different treatments. Such insufficient agreement represents a limitation of the AOSpine UCCS as, to obtain widespread use, a new classification must demonstrate adequate agreement among the authors who developed it and among independent assessors to confirm the classification's external validity.

We believe our results can be explained because the AOSpine UCCS attempted to organize many different injury types into a single scheme. As opposed to the other AOSpine injury classification systems, which described different injuries occurring in vertebrae or vertebral segments that are similar, the unique bone anatomy of the occipital condyles, atlas, and axis plus the capsuloligamentous complex of this region makes organizing the wide spectrum of injuries affecting this region into a single scheme a difficult task. As an example, an un-displaced fracture of the posterior arch of C2, with a very good prognosis treated with external immobilization, should be classified as a site III type A fracture; paradoxically, a displaced fracture of the base of the odontoid process, which is a completely different injury (and which has a different prognosis and requires a different treatment) is also classified as a site III type A fracture.

Furthermore, part of the insufficient inter-observer agreement using this classification could be explained because of some injuries' low prevalence, since the outcome's prevalence strongly influences κ value. Despite our strategy to increase the number of cases for assessment even beyond the sample size estimation to have injuries of all sites and types, several sites plus type injuries were identified in less than 10% of cases, as shown in Table 2. As expected, those injuries exhibited lower agreement among the different raters, as displayed in Table 4.

Of note, the intra-observer agreement according to site and type was substantial (k=0.69). However, the intra-observer agreement depends on each assessor's individual interpretation of the classification, which reflects either a steadily correct or a steadily incorrect appraisal by an assessor, independent of agreement with other evaluators.

Any new classification system should undergo a three-step validation process: 1) A description of categories by experts, 2) A multicenter agreement assessment done by a representative panel of potential users of the scheme, and 3) A prospective clinical study to evaluate its clinical value [15]. Our study contributes to the second step of this process through a multicenter agreement evaluation performed by spine surgeons from six centers in three countries.

We did not assess the case-specific or the neurological status modifiers, which could be a limitation of our study. However, we wanted to replicate the methodology used in the studies involving the authors who developed this classification [9,11]; thus, we could make the results of the two studies comparable. On the other side, agreement of other AOSpine injury classifications did not use modifiers either [16–18]. We also acknowledge that using magnetic resonance imaging (MRI) scans could increase agreement using this classification, as MRI has the potential to assess ligamentous injuries better. Nevertheless, the standard for managing spine trauma still uses CT scans in most trauma centers worldwide. Finally, the low prevalence of some injuries, as already discussed, affects κ values. Although this prevalence corresponds to the real-world scenario of the UCT, it could be considered a methodological limitation of our study.

While independent agreement studies usually show a lower agreement than what is observed by the developers of a new scheme [16–18], the level of agreement we observed using the AOSpine UCCS makes us recommend being cautious in using this classification in practice and research. Considering that the availability of MRI scanners has increased with time, future studies should evaluate if MRI contributes to a better agreement using this classification by independent observers, as MRI can favor the diagnosis of the ligamentous component of the injury. Moreover, as a recent study about new training methods showed better interrater agreement [19], upcoming agreement studies using approaches such as that one should be warranted. Finally, future prospective studies are required to establish whether this classification allows surgeons to build an evidence-based treatment algorithm.

Funding

None.

Declaration of competing interest

The authors have no conflict of interest to disclose.

References

[1] Anderson PA, Montesano PX. Morphology and treatment of occipital condyle fractures. Spine (Phila Pa 1976) 1988;13(7):731–6.
[2] Levine AM, Edwards CC. Fractures of the atlas. J Bone Joint Surg Am 1991;73(5):680–91.
[3] Anderson LD, D'Alonzo RT. Fractures of the odontoid process of the axis. J Bone Joint Surg Am 1974;56(8):1663–74.
[4] Levine AM, Edwards CC. The management of traumatic spondylolisthesis of the axis. J Bone Joint Surg Am 1985;67(2):217–26.
[5] Shafafy R, Valsamis EM, Luck J, Dimock R, Rampersad S, Kieffer W, et al. Predictors of mortality in the elderly patient with a fracture of the odontoid process. Bone Joint J 2019;101-B(3):253–9.
[6] Effendi B, Roy D, Cornish B, Dussault RG, Laurin CA. Fractures of the ring of the axis. A classification based on the analysis of 131 cases. J Bone Joint Surg Br 1981;63-B(3):319–27.
[7] Traynelis VC, Marano GD, Dunker RO, Kaufman HH. Traumatic atlanto-occipital dislocation. Case report. J Neurosurg 1986;65(6):863–70.
[8] Vaccaro AR, Karamian BA, Levy HA, Canseco Jose, Rajasekaran Shanmuganathan, Benneker Lorin, et al. Update on upper cervical injury classifications: the new AO upper cervical spine classification system. Clin Spine Surg 2022;35(6):249–55.
[9] Maeda FL, Formentin C, de Andrade EJ, Sousa Pedro, Goyal Dhruv, Shroeder Gregory, et al. Reliability of the New AOSpine classification system for upper cervical traumatic injuries. Neurosurgery 2020;86(3):E263–E70.
[10] AOSpine. AOSpine upper cervical classification system. Available at: www.youtubecom/watch?v=KyUYfa_JMb4. 2018. Accessed July 20, 2022.
[11] Vaccaro AR, Lambrechts MJ, Karamian BA, Canseco Jose, Oner Cumhur, Vialle Emiliano, et al. AO spine upper cervical injury classification system: a description and reliability study. Spine J 2022. In press.

[12] 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.

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

[14] Urrutia J, Zamora T, Campos M, Yurac Ratko, Palma Joaquin, Mobarec Sebastian, 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 (7):2185–92.

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

[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 2015;40(1):E54–8. Phila Pa 1976.

[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 2017;42(5):298–303. Phila Pa 1976.

[18] Urrutia J, Meissner-Haecker A, Astur N, Valencia M, Yurac R, Camino-Willhuber G, et al. An independent inter- and intraobserver agreement assessment of the AOSpine sacral fracture classification system. Spine J 2021;21(7):1143–8.

[19] Karamian BA, Schroeder GD, Levy HA, Canseco Jose, Benneker Lorin, Kandziora Frank, et al. The Influence of Surgeon Experience and Subspeciality on the Reliability of the AO Spine Sacral Classification System. Spine (Phila Pa 1976) 2021;46(24):1705–13.