Validation of the AO Spine Sacral Classification System: Reliability Among Surgeons Worldwide

Gregory D. Schroeder, MD, a Brian A. Karamian, MD, a Jose A. Canseco, MD, PhD, a Luiz R. Vialle, MD, PhD, b Frank Kandziora, MD, PhD, c Lorin M. Benneker, MD, d Shanmuganathan Rajasekaran, PhD, e Jörg H. Holstein, MD, PhD, f Klaus J. Schnake, MD, g,h Mark F. Kurd, MD, a Marcel F. Dvorak, MD, i Emilioano N. Vialle, MD, b Andrei F. Joaquim, MD, PhD, j Rishi M. Kanna, MS, MRCS, FNB, e Michael Fehlings, MD, PhD, k Jefferson R. Wilson, MD, PhD, k Jens R. Chapman, MD, l James C. Krieg, MD, a Conor P. Klewenko, MD, m Reza Firoozabadi, MD, m F. Cumhur Öner, MD, PhD, n Christopher K. Kepler, MD, MBA, a and Alexander R. Vaccaro, MD, PhD, MBA a

Objectives: To (1) demonstrate that the AO Spine Sacral Classification System can be reliably applied by general orthopaedic surgeons and subspecialists universally around the world and (2) delineate those injury subtypes that are most difficult to classify reliably to refine the classification before evaluating clinical outcomes.

Design: Agreement study.

Setting: All-level trauma centers, worldwide.

Participants: One hundred seventy-two members of the AO Trauma and AO Spine community.

Intervention: The AO Sacral Classification System was applied by each surgeon to 26 cases in 2 independent assessments performed 3 weeks apart.

Main Outcome Measurements: Interobserver reliability and intraobserver reproducibility.

Results: A total of 8097 case assessments were performed. The kappa coefficient for interobserver agreement for all cases was 0.72/0.75 (assessment 1/assessment 2), representing substantial reliability. When comparing classification grading (A/B/C) regardless of subtype, the kappa coefficient was 0.84/0.85, corresponding to excellent reliability. The kappa coefficients for interobserver reliability were 0.95/0.93 for type A fractures, 0.78/0.79 for type B fractures, and 0.80/0.83 for type C fractures. The overall kappa statistic for intraobserver reliability was 0.82 (range 0.18–1.00), representing excellent reproducibility. When only evaluating morphology type (A/B/C), the average kappa value was 0.87 (range 0.18–1.00), representing excellent reproducibility.

Conclusion: The AO Spine Sacral Classification System is universally reliable among general orthopaedic surgeons and subspecialists worldwide, with substantial interobserver and excellent intraobserver reliability.

Key Words: AO Spine Classification, international, interobserver, reliability, spine trauma, sacral fracture, pelvis fracture

INTRODUCTION

The management of high-energy sacral fractures remains challenging for both spine surgeons and...
traumatologists. Difficulties with clinical decision making are often attributable to the lack of a comprehensive sacral fracture classification system. Current classification schemes are mainly descriptive, lack clinical correlations regarding prognosis and treatment, and are nonvalidated.\(^1\)–\(^7\) The lack of an appropriate conceptual framework in classifying sacral fractures has resulted in barriers to communication, research, education, standardized care, and evaluation of outcomes.

Classification systems are instruments used by surgeons to improve communication, guide treatment, and understand prognosis and are ideal when simple, reproducible, and comprehensive information can be transmitted to guide clinical management.\(^8\) Current sacral fracture classification schemes are based on various independent parameters such as injury location, fracture morphology, mechanism of injury, and associated injuries. The existence of numerous systems, however, is an indication that none have been universally adopted. The first developed classifications of sacral fractures focused on the location and pattern of injury.\(^4\)–\(^6\)\(^9\) The most notable and widely used, the Denis 3-zone classification, is based on the location of the fracture with reference to the sacral neuroforamina.\(^2\) The Roy-Camille classification and its most recent modification were developed to further characterize transverse fractures of the sacrum, the most common type of Denis zone 3 injuries.\(^1\)–\(^4\) Finally, the Isler\(^3\) classification was devised to address lumbosacral instability pertaining to fractures involving the superior S1 facet. All these classifications, however, are suboptimal regarding true purpose of a classification system and lack associated management recommendations. The Denis classification remains overly simplistic and is likely still used primarily because of its ease of application. Although the zone of sacral fracture was noted to have implications regarding neurologic injury, this relationship has been called into question as of late.\(^10\) The modified Roy-Camille and Isler classifications are aimed to help guide management but are too specific regarding fracture type and are not sufficiently generalizable. Most recently, the Lumbosacral Injury Classification System incorporated fracture morphology, posterior ligamentous complex integrity, and neurologic injury status regarding lumbosacral injuries.\(^5\) Although this system does incorporate a severity score to help guide surgical decision making, it focuses specifically on those fractures resulting in lumbosacral instability and has not achieved widespread adoption.

In an attempt to overcome the current limitations of sacral fracture classifications, all members of AO Spine and AO Trauma were surveyed by researchers from the AO Spine Knowledge Forum Trauma regarding management of the most controversial aspects of sacral fractures.\(^11\) The result was the AO Spine Sacral Classification System, a concise and comprehensive scheme to be used universally among surgeons (see Supplemental Digital Content 1, http://links.lww.com/JOT/B416). This sacral classification is separated into 3 subtypes based on morphological characteristics: type A (lower sacroccygeal injuries), type B (posterior pelvic injuries), and type C (spino pelvic injuries). It is a hierarchical system in which each morphologic type is subdivided into increasing numerical subtypes based on the energy of injury, with higher numbers ascribed to increased injury severity. Case-specific and neurologic modifiers, not evaluated in this study, use clinical information to help further plan management based on relevant clinical information.\(^12\) Regional differences in the care and treatment of patients, including varying indications for operative management, must be taken into consideration during the development of a universal classification scheme. The proposed classification scheme, therefore, is based on morphology and was created independent of treatment algorithms to account for such international variability of surgical indications.

Validation of classification systems remains crucial because a poorly validated system will be a biased predictor of patient outcomes.\(^13\) Incorrect classification of fracture severity may result in inappropriate surgical management and can potentially lead to undue harm. Furthermore, lack of appropriate validation can result in bias in scientific research that relies on categorization. The classification system presented in this study has completed pilot agreement validation as previously published, demonstrating excellent interobserver reliability for fracture severity and substantial reliability for all fracture subtypes among expert spine and trauma surgeons from around the world.\(^11\)–\(^12\) Although promising, these results must be verified in the broader clinical context, incorporating surgeons of all types from around the world to evaluate generalizability. This study, therefore, is essential to determine whether a sacral fracture classification system can be reliably adopted on a worldwide scale.

Accordingly, the aims of this study were to (1) establish that the AO Spine Sacral Classification System can be reliably applied by surgeons of all types universally across the world and (2) delineate those injury types that are most difficult to classify reliably to refine the classification before evaluating clinical outcomes.

**METHODS**

The background methodology for the development and description of the AO Spine Sacral Classification System has been previously described.\(^11\)–\(^12\) High-resolution computed tomography images from 26 cases with sacral fractures representing all subtypes of the classification were assessed by 172 investigators. An email request was sent out to all members of the AO Spine and AO Trauma community for surgeons who routinely treat patients with sacral fractures and who would like to be involved in the validation of the sacral fracture classification. The investigators represented 6 world regions (Europe, North America, Latin and South America, the Middle East, Africa, and Asia). The investigators included orthopaedic spine surgeons, general orthopaedic surgeons, neurosurgeons, and orthopaedic traumatologists. Before the assessment, all participants participated in a video tutorial of the classification and a 10-case training session. The assessments were then performed through web conference, where both key high-resolution images of the fracture and the entire axial computed tomography scan, and, if necessary, the sagittal and coronal sequences, were analyzed. Two assessments were performed by each investigator independently 3 weeks apart from one another. The case order was scrambled so that a consecutive series was not presented, given the hierarchical nature of the classification system.
For reliability analysis, surgeons were provided with a video instruction of the classification system and a reference guide to injury types, which included both written and iconic descriptions with drawings of clinical images (see Supplemental Digital Content 1, http://links.lww.com/JOT/B416). To minimize linguistic and semantic confusion across different regions, verbal descriptions of the injury patterns were combined with standardized iconic images of each injury type to establish a rigorous visual and linguistic descriptive understanding of each injury pattern.

Statistical analysis used agreement percentages to evaluate the correct classification of each rating in comparison with the gold standard, as defined by a panel of expert spine surgeons and traumatologists. Cohen kappa (k) statistic was used to assess the reliability of classification between independent observers (interobserver agreement) and reproducibility between classifications of the same observer on separate assessments (intraobserver reproducibility). The k coefficients were interpreted using the Landis and Koch grading system. A k coefficient of less than 0.2 was defined as slight agreement or reproducibility, between 0.2 and 0.4 as fair agreement or reproducibility, between 0.4 and 0.6 as moderate agreement or reproducibility, between 0.6 and 0.8 as substantial agreement or reproducibility, and greater than 0.8 as excellent agreement or reproducibility. The k coefficients were calculated for fracture morphology type (A, B, or C) and classification of subtypes (eg, A1, A2, or A3).

### RESULTS

A total of 172 surgeons were included in the study, resulting in 8097 case assessments. Practice setting for surgeons included 41% academic, 48% hospital employed, and 11% private practice. Baseline characteristics including region, years of practice, and surgical specialty are summarized in Table 1. The 26 cases for review consisted of 7 (26.9%) type A, 8 (30.8%) type B, and 11 (42.3%) type C fractures. Overall, 94.0% of assessments were correctly classified by morphology type. When stratifying by morphologic group, 98.3% were correctly classified as type A, 92.5% were correctly classified as type B, and 92.3% were correctly classified as type C fractures. Subtype A1 had the highest agreement percentage (95.4%), whereas fracture subtypes C1 and C3 had the lowest agreement percentages (67.6% and 77.5%, respectively). Correct classification stratified by years of experience, specialty, and region are provided in Table 2.

#### Interobserver Reliability

Because observers evaluated 26 cases and performed 2 separate assessments, k values are presented as assessment 1 k/assessment 2 k. The overall k coefficient for all cases was 0.72/0.75, representing substantial reliability. When comparing classification grading (A/B/C) regardless of subtype, k was 0.84/0.85, corresponding to excellent reliability. The k coefficients for interobserver reliability were 0.95/0.93 for type A fractures, 0.78/0.79 for type B fractures, and 0.80/0.83 for type C fractures (Table 3). These values represent substantial reliability for type B fractures and excellent reliability for type A and C fractures. When evaluating by subtype, A1 had the highest level of agreement with k of 0.88/0.89, and fractures C1 and C0 had the lowest level of agreement with k of 0.58/0.66 and 0.60/0.68, respectively (Table 4).

#### Intraobserver Reliability

The overall k was 0.82 (range 0.18–1.00), representing excellent reproducibility. When only evaluating morphology type (A/B/C), the average k was 0.87 (range 0.18–1.00), representing excellent reproducibility.

### TABLE 2. Correct Classification of Fracture Stratified by Years of Experience, Specialty, and Region

| Years of experience | % Correct |
|---------------------|-----------|
| <5                  | 84.6      |
| 5–10               | 82.8      |
| 11–20              | 87.4      |
| >20                | 82.0      |
| All                | 84.7      |
| Specialty          |           |
| General orthopaedics | 86.0 |
| Neurosurgery       | 84.3      |
| Orthopaedic spine  | 83.1      |
| Orthopaedic trauma | 84.5      |
| All                | 84.7      |
| Region             |           |
| Africa             | 80.6      |
| Asia               | 82.1      |
| Europe             | 89.3      |
| Latin and South America | 81.2 |
| Middle East        | 87.5      |
| North America      | 92.0      |
| All                | 84.7      |

---

*Schroeder et al. J Orthop Trauma • Volume 35, Number 12, December 2021

---

Copyright © 2021 The Author(s). Published by Wolters Kluver Health, Inc.
DISCUSSION

This article demonstrates the applicability of the AO Spine Sacral Classification System among surgeons from a variety of subspecialties in orthopaedics and neurosurgery from various regions of the world. The overall interobserver reliability of the classification scheme was found to be substantial at minimum, and the overall intraobserver reproducibility was found to be excellent. Those fractures types that had the highest interobserver reliability were A1, B2/3, and C2 fractures, demonstrating excellent reliability. Most challenging in classifying were C0/1/3 fractures. C1/3 fractures had the lowest agreement percentages for correct classification; however, the reliability of C3 fractures improved from moderate to substantial from the first to the second assessments. C0/1 fractures demonstrated substantial reliability on both assessments. Given the lack of displacement and the difficulty in identifying nondisplaced fracture lines of C0/1, these fractures may be more challenging to identify. This subtype of fracture may require particular emphasis in teaching to facilitate knowledge translation. However, no fracture subtypes were found to have poor reliability of classification.

There were no major differences when analyzing fracture classification by years of experience or by surgical subspecialty, underscoring the utility of the classification scheme to the less experienced surgeon. Classification of fracture pattern based on geography showed regional variations. Observers from Africa, Latin and South America, and Asia had a greater proportion of incorrect classifications compared with those from Europe, the Middle East, and North America. This may be due to worldwide variations in access to advanced imaging, with decreased availability in rural areas of middle-income and low-income countries.\(^{15,16}\) When comparing the first and second assessments, there was a slight “learning effect” in which interobserver reliability improved during the second assessment in most of the fracture subtypes. This is most notable in type C fractures. For example, 77.4% of C0 fractures were correctly identified on the first assessment, whereas the second assessment demonstrated correct identification of 94.1% of C0 fractures. Despite such learning effect, there remains high overall consistency between assessments. Even when taking into account the first assessment alone and negating the learning effect, there remains substantial interobserver reliability.

This is the first study for the advanced validation of the AO Spine Sacral Classification System. In addition, this is also the first study to validate a comprehensive sacral fracture classification system. There are no previous studies to compare reliability or demonstrate noninferiority. However, previous validation of this classification among expert spine surgeons and traumatologists demonstrated interobserver k values of 0.58 for overall classification and 0.75 for classification of morphology only.\(^{12}\) Interobserver reproducibility among experts had a k value of 0.71 for overall classification and 0.83 for morphology only. Interestingly, this study notes superior reliability when the classification system was applied among a larger group of nonexpert surgeons from across the globe. This underscores the generalizability of the classification to less experienced surgeons despite demographic, cultural, or training differences. When comparing the current reliability results with other widely adopted classification systems for fractures outside the sacrum, the AO Spine Sacral Classification System has either similar or superior reliability grading. For instance, the Thoracolumbar Injury Classification and Severity Score widely used for fracture care of the thoracolumbar spine was found to have an overall k of 0.63 (substantial) among an expert group of surgeons.\(^{17,18}\) Thoracolumbar Spine Injury Classification and the AO Spine Subaxial Cervical Spine Injury Classification systems have overall interobserver reliabilities with k of 0.56 (moderate) and 0.64 (substantial), respectively.\(^{19–21}\) Therefore, in this worldwide sample of surgeons of all types, we demonstrate at minimum substantial reliability that is comparable with current widely adopted classification schemes used in the treatment of spine patients.

Communication, research, education, and appropriate evidence-based treatment of patients are all dependent on the ability to comprehensively classify their injuries or conditions, as is the accuracy of information that surgeons are able to impart to patients about prognosis. Reliability is essential to communicate, classify, and appropriately guide treatment of relevant pathology. Without establishing such reliability, the education of surgeons becomes impractical. Furthermore, the validity of any research regarding the treatment and outcomes of sacral fractures can be called into question. The development process for the AO Spine classification systems is based on the evaluation of numerous cases in multiple sessions by a group of investigators and follows an iterative process.
Adding to the uncertainty, most of the fracture subtypes in the classification system exhibit substantial interobserver and excellent intraobserver reliabilities. The results of this study are promising, demonstrating acceptable accuracy in the classification of fractures among surveyed orthopaedic, spine, and trauma subspecialists with various levels of experience. Further studies are required to assess the reliability of the classification system among a more generalizable population. Future clinical studies, including application of the classification system prospectively, will need to be conducted to further evaluate the clinical relevance and the usefulness of classification categories in a true pragmatic approach.

Acknowledgments

AO Spine Sacral Classification Group Members: Aaron Hils, Alfredo Guiroy, Aaron Hockley, Ahmed Shawkyl Abdelgawaad, Akbar Jaleel Zubair, Marcon Olger, Alberto Campione, Alexio Barcelos, Alex del Arco, Alexander Weening, Ali Haghnejehgdar, Ali Oner, Alqatab Ahmed, Alsammak Wael, Alvin Pun, Amit Bhandutia, Andelle German, Antonio Sanchez Rodriguez, Ashraf El Naga, Barbeiro Gonçalves, Bas Teresa, Bernhard Ulrich, Bernucci Claudio, Bruno Saciloto, Carlo Bellabarra, Carlos Pernal, Charalampdis Anastasios, Christina Cheng, Ciccioli Nicolas, Cumhur Kilicler, Daniel Cruz, Daniela Linhares, Dave Ashauola, Das Gurdip, Dave Anthony Dizon, De Falco Giovanni, Dewan Shamsul Asif, Dihan Aponso, Dimitros Patronis, Duchen Rodriguez Luís Miguel, Elias Enmanuel Javier Martinez, Enrico Pera, Erol Gercek, Fabii Pozzi, Frank Lyons, Geoffrey Tipper, Glukhov Dmitrii, Gupta Anuj, Hai Nie, Hamouda Wael, Hany Soliman, Heiller Torres Valencia, Huiren Tao, Ichiro Okano, Ignacio Fernández Bances, Igor Paredes, Janardhana Aithala Paramalli, Jason Pittman, Javier Lecaros, Javier Matta, Jeronimo Milano, Joana Freitas, Joana Ganau, Joep Kizzen, Jorge Alves, Jose Joffrey Arbatin, Juan Carlos Ramos Torres, Juan Esteban Muñoz Montoya, Juan Lourido, Jun Zhang, Khatri Manoaj, Khattab Mohamed, Kimmakkar Ninit, Konstantinos Margetis, Kumar Vinnet, Lady Lazano C, Leandro Badalassi, Lykourgos Kollintzas, Maarten de Boer, Manabu Ito, Martin Estefan, Martin Holas, Martinelli Federico, Mathias Pereira Duarte, Matti Scholz, Maximio-Alberto Diez-Ulloa, Medina Matias, Michel Triffaux, Mohammad El-Sharkawi, Morakis Andreas, Mostafa Abdelhafez, Muthu Sathish, Nasser Khan,Nicassio Nicola, Nishanth Anpanr, Nuno Neves, Olga Morillo, Osmar Moraes, Osmudina Ahmad, Paul van Urk, Pedro Bazén, Pershin Andrey, Peter Fersack, Peter Loughenbury, Philippe Bancel, Popescu Eugen Cezar, Pragnesh Bhatt, Prakash Sitoula, Przemyslaw Kolakowski, Purnajoyi Banerjee, Rabindra Pradhan, Rajesh Bahadur Lakhey, Rakesh Kumar, Ramieri Alessandro, Reddy Karthik, Rian Vieira, Ricardo Frada, Ricardo Rodrigues-Pinto, Ricky Rasschaert, Russo Fabrizio, Salah Alakkad, Samer Samy, Shaffi Pier Filippo, Schnake, Klaus, Sebastian Anibal Kornfeld, Sergio Zybersztejn, Sicoli Alfredo, Siegmund Lang, Stacey Darvish, Stipe Corluka, Subbhallah Jayakumar, Subramaniam Macherla haribabu, Sudhir Ganesan, Taha Karimjke, Takaya Nagai, Tarek Elhewala, Tejeda Martin, Thami Benzakour.
REFERENCES

1. Strange-Vognsen HH, Lebech A. An unusual type of fracture in the upper sacrum. J Orthop Trauma. 1991;5:200–203.
2. Denis F, Davis S, Comfort T. Sacral fractures: an important problem. Retrospective analysis of 236 cases. Clin Orthop Relat Res. 1988;227:67–81.
3. Isler B. Lumbosacral lesions associated with pelvic ring injuries. J Orthop Trauma. 1990;4:1–6.
4. Roy-Camille R, Saillant G, Gagna G, et al. Transverse fracture of the upper sacrum. Spine. 1985;10:838–845.
5. Lehman RA, Kang DG, Bellabarba C. A new classification for complex lumbosacral injuries. J Orthop Trauma. 1990;4:1–6.
6. Schnake KJ, Schroeder GD, Vaccaro AR, et al. AOSpine classification systems (subaxial, thoracolumbar). J Orthop Trauma. 2017;31(Suppl 4):S14–S23.
7. Sabiston CP, Wing PC. Sacral fractures: classification and neurologic implications. J Trauma. 1986;26:1113–1115.
8. Khan JM, Marquez-Lara A, Miller AN. Relationship of sacral fractures to nerve injury: is the Denis classification still accurate?. J Orthop Trauma. 2016;31:181–184.
9. Schroeder GD, Kurd MF, Kepler CK, et al. The development of a universally accepted sacral fracture classification: a survey of AOSpine and AO Trauma members. Glob Spine J. 2016;6:686–694.
10. Vaccaro AR, et al. Description and reliability of the AOSpine sacral classification system. J Bone Joint Surg. 2020;102:1454–1463.
11. Audigé L, Bhandari M, Hanson B, et al. A concept for the validation of fracture classifications. J Orthop Trauma. 2005;19:404–409.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–174.
13. Ngoya PS, Muhogora WE, Pitcher RD. Defining the diagnostic divide: an analysis of registered radiological equipment resources in a low-income African country. Pan Afr Med J. 2016;25:99.
14. Volpi G. Radiography of diagnostic imaging in Latin America. Nucl Med Biomed Imaging. 2016;1:10–12.
15. Vaccaro AR, Lehman RA, Hurlbert RJ, et al. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. Spine (Phila Pa 1976). 2005;30:2325–2333.
16. Whang PG, Vaccaro AR, Poelstra KA, et al. The influence of fracture mechanism and morphology on the reliability and validity of two novel thoracolumbar injury classification systems. Spine (Phila Pa 1976). 2007;32:791–794.
17. Vaccaro AR, Oner C, Kepler CK, et al. AOSpine thoracolumbar spine injury classification system. Spine. 2013;38:2028–2037.
18. McCarthy PJ, Vaccaro AR, Koerner JD, Radcliff KE, et al. AOSpine subaxial cervical spine injury classification system. Eur Spine J. 2016;25:1082–1086.
19. Bydon M, De la Garza-Ramos R, Macki M, et al. Incidence of sacral fractures and in-hospital postoperative complications in the United States: an analysis of 2002-2011 data. Spine (Phila Pa 1976). 2014;39:1103–1109.
20. Grauer JN, Vaccaro AR, Reiner JM, et al. Similarities and differences in the treatment of spine trauma between surgical specialties and location of practice. Spine (Phila Pa 1976). 2004;29:685–696.
21. 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. 2020;30:517–523. doi:10.1007/s00586-020-06555-2.