Spine

SPINE Volume 46, Number 24, pp 1705–1713
© 2021 The Author(s). Published by Wolters Kluwer Health, Inc.

DIAGNOSTICS

OPEN

The Influence of Surgeon Experience and Subspeciality on the Reliability of the AO Spine Sacral Classification System

Brian A. Karamian, MD,[a] Gregory D. Schroeder, MD,[a] Hanna A. Levy, BS,[a] Jose A. Canseco, MD, PhD,[a] AO Spine Sacral Classification Group Members, Lorin M. Benneker, MD,[b] Frank Kandziora, MD, PhD,[c] Shanmuganathan Rajasekaran, PhD,[d] F. Cumhur Öner, MD, PhD,[e] Klaus J. Schnake, MD,[f,g] Christopher K. Kepler, MD, MBA,[a] and Alexander R. Vaccaro, MD, PhD, MBA[a]

Study Design. Cross-sectional survey.

Objective. To determine the influence of surgeons' level of experience and subspeciality training on the reliability, reproducibility, and accuracy of sacral fracture classification using the Arbeitsgemeinschaft für Osteosynthesefragen Spine Sacral Classification System.

Summary of Background Data. A surgeons' level of experience or subspecialty may have a significant effect on the reliability and accuracy of sacral classification given various levels of comfort with imaging assessment required for accurate diagnosis and classification.

Methods. High-resolution computerized tomography (CT) images from 26 cases were assessed on two separate occasions by 172 investigators representing a diverse array of surgical subspecialities (general orthopedics, neurosurgery, orthopedic spine, orthopedic trauma) and experience (<5, 5–10, 11–20, >20 yrs). Reliability and reproducibility were calculated with Cohen kappa coefficient (k) and gold standard classification agreement was determined for each fracture morphology and subtype and stratified by experience and subspeciality.

Results. Respondents achieved an overall $k = 0.87$ for morphology and $k = 0.77$ for subtype classification, representing excellent and substantial intraobserver reproducibility, respectively. Respondents from all four practice experience groups demonstrated excellent interobserver reliability when classifying overall morphology ($k = 0.842/0.850$, Assessment 1/Assessment 2) and substantial interobserver reliability in overall subtype ($k = 0.719/0.751$) in both assessments. General orthopedists, neurosurgeons, and orthopedic spine surgeons exhibited excellent interobserver reliability in overall morphology classification and substantial interobserver reliability in overall subtype classification. Surgeons in each experience category and subspecialty correctly classified fracture morphology in over 90% of cases and fracture subtype in over 80% of cases according to the gold standard. Correct overall classification of fracture morphology (Assessment 1: $P = 0.024$, Assessment 2: $P = 0.006$) and subtype ($P_2 < 0.001$) differed significantly by years of experience but not by subspecialty.

Conclusion. Overall, the Arbeitsgemeinschaft für Osteosynthesefragen spine sacral classification system appears to be universally applicable among surgeons of various subspecialties and levels of experience with acceptable reliability, reproducibility, and accuracy.

Keywords: AO spine, classification, pelvic fracture, sacral fracture, spine trauma, subspecialty, surgeon experience, validation

Level of Evidence: 4

Spine 2021;46:1705–1713

From the [a]Rothman Orthopaedic Institute at Thomas Jefferson University, Philadelphia, PA; [b]Spine Service, Orthopaedic Department, Sonnenhofspital, Bern, Switzerland; [c]Center for Spine Surgery and Neurotraumatology, BG Unfallklinik Frankfurt am Main, Frankfurt, Germany; [d]Department of Orthopaedics, Trauma and Spine Surgery, Ganga Hospital, Coimbatore, Tamil Nadu, India; [e]University Medical Center, Utrecht, the Netherlands; [f]Center for Spinal Surgery, Malteser Waldkrankenhaus St. Marien, Erlangen, Germany; and [g]Department of Orthopedics and Traumatology, Paracelsus Private Medical University Nuremberg, Nuremberg, Germany.

Acknowledgment date: March 2, 2021. Acceptance date: March 30, 2021.

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

This study was organized and funded by AO Spine through the AO Spine Knowledge Forum Trauma, a focused group of international Trauma experts.

Relevant financial activities outside the submitted work: consultancy, grants, royalties, stocks, payment for lecture.

Address correspondence and reprint requests to Brian A. Karamian, MD, Department of Orthopaedic Surgery, Rothman Institute at Thomas Jefferson University Hospital, 925 Chestnut St, 5th Floor, Philadelphia, PA 19107; E-mail: brian.karamian@rothmanortho.com

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/BRS.0000000000004199

Despite the increasing prevalence of sacral fractures, controversy ensues regarding the appropriate management of these injuries.[1–3] The absence of an appropriate conceptual framework for the classification of sacral fractures has constrained the communication, education, and research necessary to evaluate clinical outcomes

after injury. Though numerous sacral fracture classifications have been created, none have been universally adopted due to various limitations and drawbacks.[4–10] According to Maurice Müller, "A classification is useful only if it considers the severity of the bone lesion and serves as a basis for treatment and for evaluation of the results."[11] Previous classification systems are bereft of prognosis, focusing on injury location, morphology, or mechanism. Their inability to inform management has resulted in limited high-quality evidence standardizing the treatment of sacral fractures.

The Arbeitsgemeinschaft für Osteosynthesefragen (AO) Spine Knowledge Forum Trauma recently developed the AO spine sacral classification system in an attempt to provide a concise yet comprehensive classification scheme for the standardization of treatment and prognostication of outcomes after sacral fractures.[12] This system is separated into three main morphologic fracture types: type A (lower sacrococcygeal injuries), type B (posterior pelvic injuries), and type C (spino-pelvic injuries). Similar to previous AO Spine classification systems, it is designed in a hierarchical manner in which each morphologic type is subdivided into increasing numerical subtypes based on the severity of injury.[13,14] Case specific modifiers and neurologic injury at the time of examination incorporate patient specific data to individualize management within a universally applicable scheme (Figure 1).

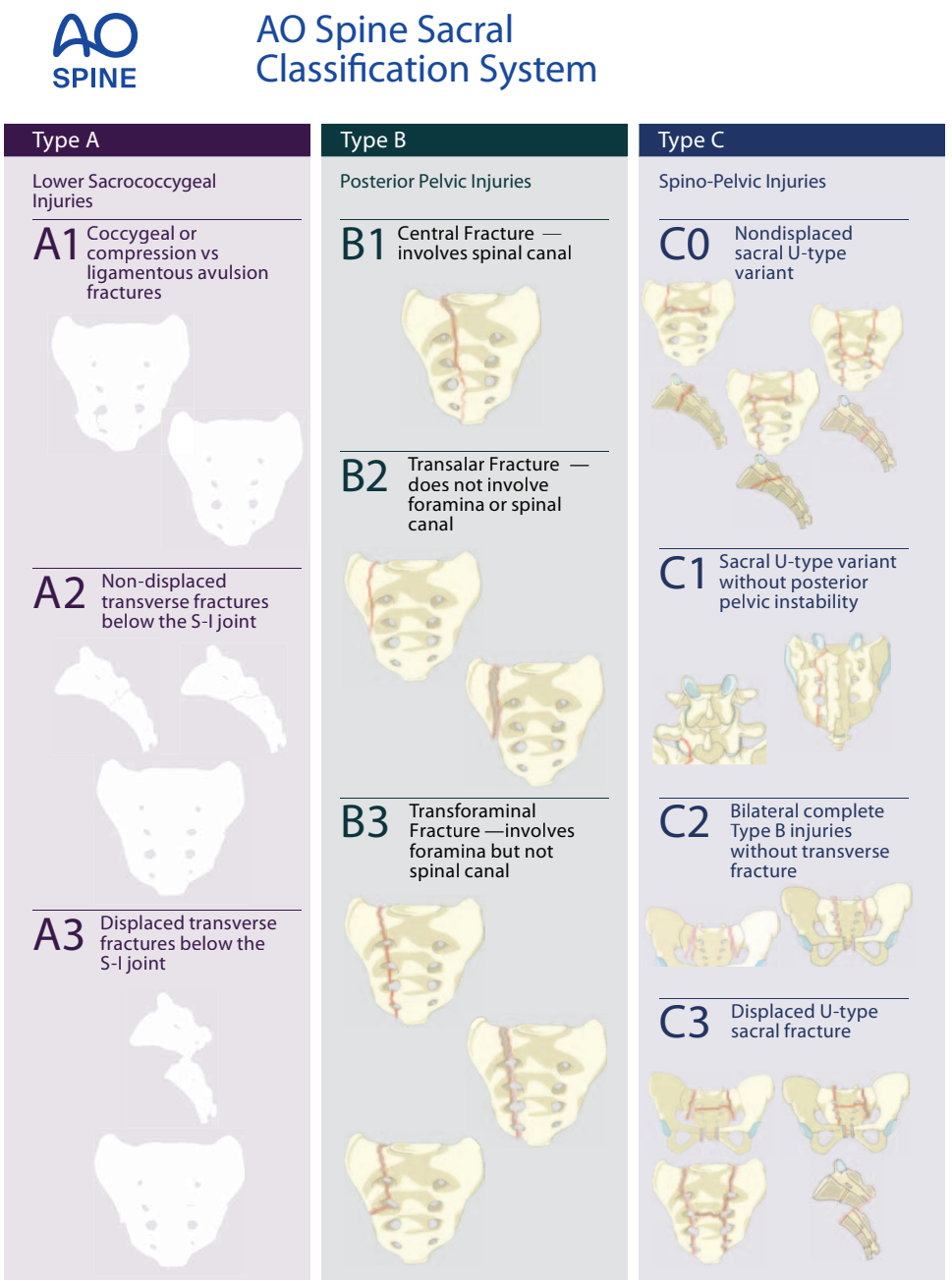


Figure 1. The AO spine sacral classification system.

An ideal classification system should be easily comprehensible and reliable amongst the diverse group of surgeons involved in the diagnosis and management of sacral fractures. A surgeons' level of experience may have a significant effect on the reliability and accuracy of a classification system. Moreover, surgeons of different subspecialities may have various levels of comfort with imaging assessment of sacral injuries required for accurate diagnosis and classification. Accordingly, the current study aimed to investigate the influence of the surgeons' level of experience and subspeciality training on (1) the reliability and reproducibility of the novel AO Spine Sacral Classification System, and (2) the appropriate classification of sacral fractures according to this system.

METHODS

The methodology for the development and description of the AO Spine Sacral Classification System has been previously described.[12,15] A request was sent out to all members of the AO Spine and AO Trauma community to recruit surgeons who routinely treat patients with sacral fractures for participation in the validation of the sacral fracture classification. Previously obtained imaging was reviewed and classified by members of the AO Knowledge Forum Trauma. Cases with complete agreement were deemed acceptable for use in the validation as the gold standard. High-resolution computerized tomography (CT) images from 26 cases were assessed by 172 investigators representing a diverse array of surgical subspecialities and experience. At minimum, two cases of each fracture subtype were included. Prior to the validation, a training session including a video introduction, visual and verbal definitions of the classification system, and a 10-case practice assessment. Validation assessments were performed via web conference where both key high-resolution images, as well as axial/sagittal/coronal CT scan sequences of the fracture, were presented. Two assessments were performed by each investigator independently 3 weeks apart from one another. The case order was randomized in both assessments such that a consecutive series was not presented given the hierarchical nature of the classification system.

Statistical Analysis

Respondents were divided into groups based on their years of experience in practice (<5, 5–10, 11–20, >20 yrs) and surgical subspecialty (general orthopedics, neurosurgery, orthopedic spine, orthopedic trauma). 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 during separate evaluations (intraobserver reproducibility). Reliability and reproducibility were determined for each fracture morphology (A, B, C) and fracture subtype (A1, A2, A3, B1, B2, B3, C0, C1, C2, C3) and stratified by experience and subspeciality. The k coefficients were interpreted using the Landis and Koch grading system with kappa less than 0.20 defined as slight reliability/reproducibility, 0.20 to 0.40 as fair reliability/reproducibility, 0.40 to 0.60 as moderate reliability/reproducibility, 0.60 to 0.80 as substantial reliability/reproducibility, and more than 0.80 as excellent reliability/reproducibility.[16]

Accuracy of classification was calculated through percentage agreement with the predetermined gold standard fracture type (morphology and subtype) for each assessment. Cases with incomplete or poor-quality imaging were excluded from use. Gold standard agreement was stratified by surgical experience and subspecialty and compared via Fisher exact test. General orthopedic and orthopedic trauma specialties were combined for the analysis of gold standard agreement due to the low number of participants in these groups. Statistical significance was defined at P value $<.05$ and no adjustment for multiplicity was performed.

RESULTS

Overall, 172 surgeons were invited to participate. Respondent demographic characteristics, including the number of years in practice and subspecialty, are shown in Table 1. The majority of respondents were fellowship trained orthopedic spine surgeons (66.3%), followed by a large minority of neurosurgeons (23.3%). Orthopedic traumatologists and general orthopedists represented 5.2% and 4.7% of respondents, respectively (Table 1). Of the 172 surgeons, 158 completed the first assessment and 162 completed the

TABLE 1. Summary of Surgeon Respondent Demographics

Category	Characteristic	Respondent%
Number of years in practice n = 171	<5 years	22.8%
	5–10 years	24%
	11–20 years	33.3%
	>20 years	19.9%
Surgical subspecialty n = 172	General orthopedics	4.7%
	Neurosurgeon	23.3%
	Orthopedic spine	66.3%
	Orthopedic trauma	5.2%
	*Other	0.6%

*Excluded from further analyses.

TABLE 2. Intraobserver Reproducibility Mean Kappa Values by Surgeon Experience and Subspecialty

	Years in Practice				Surgical Subspecialty				
	<5 years	5–10 years	11–20 years	>20 years	General Orthopedics	Neurosurgery	Orthopedic Spine	Orthopaedic Trauma	All Participants
Participant Grouping	**n = 39**	**n = 41**	**n = 57**	**n = 34**	**n = 8**	**n = 40**	**n = 114**	**n = 9**	**n = 172**
Fracture morphology	0.87	0.83	0.92	0.85	0.82	0.89	0.87	0.84	0.87
Fracture subtype	0.76	0.76	0.80	0.73	0.75	0.78	0.77	0.78	0.77

second assessment for a total of 8320 case assessments. The 26 cases for review consisted of 7 (26.9%) type A, 8 (30.8%) type B, and 11 (42.3%) type C fractures with two cases each representing A1, A2, B1, C0 subtypes, and three cases each representing A3, B2, B3, C1, C2, C3 subtypes.

Intraobserver Reproducibility

Surgeons from the various subspecialities and years in practice achieved an overall $k = 0.87$ for morphology and $k = 0.77$ for subtype classification, representing excellent and substantial reproducibility, respectively (Table 2). Classification reproducibility was comparable across all surgical subspecialties and years of practice experience.

Interobserver Reliability

Respondents from all four practice experience groups (<5, 5–10, 11–20, >20 yrs) demonstrated excellent interobserver reliability when classifying overall morphology ($k = 0.842/0.850$, Assessment 1/Assessment 2) and substantial reliability in overall subtype ($k = 0.719/0.751$) in both assessments (Table 3). Across all experience groups, type A classification was associated with excellent reliability whereas type B and C classifications demonstrated both substantial and excellent reliability, respectively. Surgeons with 11 to 20 years of practice experience were the only group to display excellent interobserver reliability for type A ($k = 0.944/0.969$), B ($k = 0.823/0.848$), and C ($k = 0.858/0.875$) morphologies. Surgeons with 11 to 20 years of experience were also the only group with at least substantial interobserver reliability for all fracture subtypes in both assessments (Table 3). Fracture subtype A1 was classified with excellent reliability for all experience groups, whereas B1, C0, C1, C2, and C3 demonstrated moderate reliability in one or more experience groups (B1: >20; C0: >20; C1: <5, 5–10, >20; C3: <5, 5–10, >20) (Table 3).

General orthopedists, neurosurgeons, and orthopedic spine surgeons exhibited excellent interobserver reliability in overall morphology classification and substantial reliability in overall subtype classification (Table 3). Orthopedic trauma surgeons classified overall morphology ($k = 0.748$) and subtype ($k = 0.618$) less reliably than the other subspecialists in the first assessment. All surgical subspecialties

TABLE 3. Interobserver Reliability by Surgeon Experience and Surgical Subspecialty

	Surgeon Experience								Surgical Subspecialty									
Characteristic	<5 years		5–10 years		11–20 years		>20 years		General Orthopedics		Neurosurgeon		Orthopedic Spine		Orthopedic Trauma		All Participants	
Assessment Number	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
Fracture morphology																		
A	0.967	0.871	0.928	0.889	0.944	0.969	0.960	0.963	0.972	0.809	0.962	0.919	0.952	0.932	0.804	1.000	0.948	0.927
B	0.780	0.763	0.762	0.768	0.823	0.848	0.704	0.761	0.778	0.813	0.809	0.807	0.765	0.781	0.725	0.871	0.775	0.794
C	0.781	0.823	0.786	0.777	0.858	0.875	0.753	0.802	0.771	0.802	0.825	0.829	0.802	0.822	0.716	0.887	0.802	0.828
Overall/ Combined	0.842	0.819	0.825	0.812	0.875	0.897	0.806	0.842	0.840	0.808	0.865	0.852	0.840	0.845	0.748	0.919	0.842	0.850
Fracture subtype																		
A1	0.984	0.843	0.870	0.887	0.826	0.916	0.869	0.921	1.000	0.844	0.908	0.878	0.868	0.894	0.812	1.000	0.881	0.894
A2	0.783	0.676	0.638	0.711	0.771	0.770	0.779	0.696	0.845	0.862	0.736	0.721	0.758	0.716	0.496	0.644	0.743	0.719
A3	0.814	0.657	0.733	0.677	0.793	0.773	0.742	0.714	0.944	0.730	0.766	0.731	0.773	0.705	0.595	0.675	0.773	0.713
B1	0.684	0.667	0.691	0.627	0.709	0.691	0.586	0.507	0.716	0.803	0.773	0.667	0.644	0.599	0.596	0.749	0.678	0.634
B2	0.751	0.754	0.740	0.786	0.782	0.886	0.868	0.782	0.687	0.782	0.782	0.817	0.791	0.800	0.631	1.000	0.779	0.815
B3	0.690	0.847	0.803	0.787	0.861	0.857	0.668	0.781	0.676	0.862	0.736	0.844	0.798	0.810	0.707	0.906	0.771	0.826
C0	0.602	0.672	0.579	0.649	0.649	0.740	0.548	0.647	0.718	0.647	0.625	0.621	0.594	0.694	0.416	0.756	0.598	0.681
C1	0.587	0.685	0.517	0.614	0.659	0.702	0.529	0.606	0.520	0.683	0.575	0.666	0.579	0.653	0.733	0.731	0.584	0.663
C2	0.782	0.847	0.787	0.777	0.808	0.902	0.768	0.772	0.944	0.843	0.784	0.774	0.780	0.851	0.770	0.906	0.791	0.836
C3	0.575	0.751	0.593	0.674	0.610	0.779	0.590	0.681	0.562	0.782	0.546	0.684	0.626	0.739	0.426	0.680	0.593	0.729
Overall/ Combined	0.725	0.740	0.695	0.719	0.747	0.802	0.695	0.711	0.761	0.784	0.723	0.740	0.721	0.746	0.618	0.805	0.719	0.751

achieved excellent interobserver reliability for type A morphology and substantial interobserver reliability for type B and C morphologies. Subtype A1 had excellent classification reliability for all subspecialties. However, subtypes B1, B2, C0, C1, and C3 were all associated with a moderate reliability for all subspecialties (Table 3). Orthopedic traumatologists had the highest frequency of moderate interobserver subtype reliability (six occasions: A2, A3, B1, B2, C0, C3).

Gold Standard Agreement

Surgeons in each experience category and subspecialty correctly classified fracture morphology in over 90% of cases and fracture subtype in over 80% of cases according to the gold standard (Table 4). Correct overall classification of fracture morphology ($P_1 = 0.024$, $P_2 = 0.006$; $P_1 =$ Assessment 1, $P_2 =$ Assessment 2) and subtype ($P_2 < 0.001$) differed significantly by years of experience but not by subspecialty. Classification of type A morphology demonstrated a significant difference ($P_2 \leq 0.001$) when comparing surgical experience of the raters. Fracture subtype classification also varied significantly by surgeon experience for subtypes A2 ($P_1 = 0.015$), B1 ($P_1 = 0.046$), and B3 ($P_1 = 0.012$) injury patterns. Classification of type A ($P_1 = 0.003$) and B ($P_1 = 0.007$) morphologies demonstrated a significant difference when comparing the surgical subspecialty of the raters. Within type B morphology, classification of subtype B1 ($P_1 = 0.005$) differed significantly by surgical subspecialty.

DISCUSSION

The ultimate goal of any classification system is to set the framework for evidence-based algorithms in the treatment of the pathology being classified. In order to do so, the classification system must facilitate communication requiring all users to be able to accurately and reliably apply the scheme. In this study of 8320 case assessments, we investigated the effects of surgeon experience and subspecialty training on the reliability, reproducibility, and accuracy of sacral fracture classification using the novel AO spine sacral classification system. Our results suggest that the AO spine sacral classification system can be universally applied by surgeons of various subspecialties and differing levels of experience with satisfactory results.

In this study, surgeon years of experience was partitioned into four groups (<5, 5–10, 11–20, >20 yrs) with a relatively even distribution of participants. When evaluating interobserver reliability, those with >20 years of experience were found to classify fracture morphology and subtype with less reliability. When evaluating the accuracy of classification overall, significant differences were found between groups in both assessments for fracture morphology and in the second assessment for fracture subtype. Surgeons with 5 to 10 years of experience had increased difficulty correctly classifying fractures of A and C morphologies and surgeons with more than 20 years of experience had increased difficulty correctly classifying all fracture subtypes overall in comparison to the other groups. Similar results were seen in the intraobserver reproducibility for fracture subtype overall where surgeons with more than 20 years of experience were found to perform less reliable than those with less years of experience. These findings mirror the validation results of the AO spine thoracolumbar injury classification system where fractures were most frequently misclassified by the most experienced surgeons.[17,18] A plausible explanation for these observations is that more experienced surgeons are less inclined to learn and follow a new classification system due to their familiarity with prior systems. However, surgeons with slightly less experience (11–20 yrs) were the only group to display excellent interobserver reliability for all morphologies and at minimum substantial interobserver reliability for all fracture subtypes, indicating the importance of a certain degree of experience in correctly classifying fractures. Additionally, experienced (albeit slightly less experienced) surgeons may still be adept at adopting and applying new classification systems. Despite these significant differences, it is important to note that the overall reliability in this study remains substantial at minimum with high accuracy regardless of surgeon level of experience.

Overall, surgeon subspecialty did not appear to have a significant effect on the classification of sacral fractures. The interobserver reliability demonstrated that orthopedic traumatologists performed less reliable in classifying fracture morphology and fracture subtype on the first assessment. However, meaningful improvement was seen on the second assessment in both categories, ultimately obtaining excellent reliability. General orthopedists had the lowest reproducibility scores for fracture morphology and fracture subtype, an observation that is unlikely to be clinically significant as all subspecialities demonstrated excellent and substantial reproducibility for fracture morphology and subtype, respectively. While there was no significant difference in the combined accuracy of all assessments between subspecialties, few significant differences were found in the analysis for certain morphology and fracture subtypes. Specifically, orthopedic spine surgeons had the greatest difficulty correctly classifying type B morphology fractures, and specifically B1 fractures, despite excellent and substantial reliability, respectively. This may be due to the relative rarity of B1 fracture type cases and possible inexperience in diagnosing isolated longitudinal fractures medial to the foramen on CT.[19] Additionally, general orthopedists and orthopedic trauma surgeons had a lower (94.5%) accuracy in diagnosing fractures of A morphology. However, the excellent overall accuracy achieved calls into question the clinical relevance of these significant differences.

Not surprisingly, the simplest fracture patterns (type A) and those injuries that were most stable (lower sacrococcygeal injuries) demonstrated the highest reliability, reproducibility, and accuracy among all subspecialities and levels of experience. Spinopelvic injuries (type C) were the most challenging fracture morphology to accurately diagnose. Within that, the most challenging subtypes include C0, C1, and C3. Whereas displaced lumbopelvic dissociations may be visible on plain radiographs, non-displaced spinopelvic injuries can be obscured by the relatively cephalad

TABLE 4. Percent Agreement With Gold Standard Fracture Classification by Surgeon Experience and Surgical Subspecialty

	Surgeon Experience										Surgical Subspecialty									
Characteristic	<5 years		5–10 years		11–20 years		>20 years		*P*-Value		General Orthopedics and Orthopedic Trauma		Neurosurgeon		Orthopedic Spine		*P*-Value		All Participants	
Assessment Number	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Fracture morphology																				
A	98.8%	96.2%	97.3%	96.8%	98.6%	100%	99%	99.1%	0.5216	**0.0001**	94.5%	96.2%	98%	97.7%	99.1%	98.7%	**0.0027**	0.1291	98.4%	98.2%
B	93.8%	90.4%	94.9%	91.6%	93.8%	94%	89.2%	89.6%	0.0550	0.1721	92.6%	94.1%	97.2%	92.1%	91.8%	91.3%	**0.0074**	0.5723	93.3%	98.2%
C	90.5%	93.9%	88.8%	91.1%	94%	94.8%	89.6%	93.3%	**0.0220**	0.1402	90.6%	95.8%	90.2%	94%	91.4%	92.9%	0.7625	0.3415	91.1%	93.5%
Overall/ Combined	94.3%	93.5%	93.7%	93.2%	95.5%	96.3%	92.6%	94%	**0.0241**	**0.0055**	92.5%	95.3%	95.2%	94.6%	94.1%	94.3%	0.3012	0.5322	94.2%	94.5%
Fracture subtype																				
A1	98.6%	96%	92%	97.2%	91.3%	97.3%	91.4%	100%	0.1784	0.5459	93.8%	96.7%	91.8%	96.1%	93.5%	98.1%	0.8835	0.4109	93.2%	97.5%
A2	92.8%	91.9%	82.4%	90.1%	91.2%	95.4%	98.3%	85.7%	**0.0146**	0.1620	81.2%	90%	88.9%	90.7%	93%	92%	0.0837	0.8557	90.8%	91.5%
A3	88.3%	71.9%	85%	74.1%	89.3%	83.5%	80.5%	76.3%	0.2387	0.1029	86.7%	77.8%	87.2%	81.4%	86%	75.7%	0.9498	0.4582	86.4%	77.4%
B1	82.4%	83.8%	88%	88.9%	87%	85.5%	71.7%	74.6%	**0.0460**	0.1392	80.6%	93.3%	94.6%	85.5%	79.3%	81.7%	**0.0049**	0.2559	83.3%	83.8%
B2	92.3%	85.5%	91.9%	86%	87.6%	93.7%	96.5%	88.2%	0.1195	0.1101	82.6%	95.6%	91.6%	89.4%	92.7%	87.8%	0.0753	0.3400	91.5%	89%
B3	87.5%	93.6%	94.5%	87.7%	96.7%	92.5%	89.5%	86%	**0.0216**	0.1684	93.2%	90.7%	93.5%	91.2%	92.3%	90.1%	0.9592	0.9686	92.7%	90.5%
C0	75%	93.2%	73%	90.3%	84%	97.3%	73.7%	93.7%	0.2645	0.2261	74.2%	96.7%	76.7%	92.1%	77.9%	94.4%	0.8909	0.6973	77.4%	94.1%
C1	73.1%	84.8%	67.3%	76.9%	78.3%	83.4%	73.6%	78.5%	0.2553	0.3491	76.6%	84.4%	67.9%	81.6%	74.7%	80.8%	0.3373	0.8353	73.4%	81.4%
C2	88.2%	88.5%	85.6%	87.9%	92.1%	92.6%	83.5%	81.9%	0.1991	0.0826	93.3%	95.6%	89.8%	86.8%	86.5%	88%	0.3963	0.2721	88.1%	88.5%
C3	59.2%	79.5%	61.4%	70.1%	58%	79.6%	56.3%	71%	0.8995	0.1585	59.6%	77.8%	50%	72.8%	61.5%	76.5%	0.1122	0.6920	58.9%	75.9%
Overall/ Combined	83.7%	86.9%	82.1%	84.9%	85.5%	90.1%	81.5%	83.6%	0.0645	**0.0000**	82.2%	89.8%	83.2%	86.8%	83.7%	86.5%	0.674	0.1825	83.6%	87%

Bold = p<0.05

location of the fracture and superimposed ilia making it more challenging to diagnose without a high index of suspicion.[20] While CT imaging of fractures were presented during the validation, MRI has been demonstrated to have an evolving role in the diagnosis of sacral insufficiency and stress fractures which could have improved the reliability and accuracy results for spinopelvic injuries (type C).[2]

The respondents demonstrated improvements in interobserver reliability and accuracy between first and second assessments across most sacral injury patterns with few exceptions. The largest increases in both reliability and accuracy assessments were noted for C0, C1, and C3 subtypes, which were previously been noted to be most difficult to classify across all raters. This improvement across assessments underscores a potential "learning effect." As surgeons become more familiar with the classification system and incorporate its use into their daily practice, the reliability and accuracy may continue to improve. However, taking into account the first assessment alone, the results demonstrate acceptable accuracy, which underscores the applicability of the classification scheme to even the naïve surgeon.

This study is not without limitations. This investigation was performed in a retrospective manner based on previously obtained images. The true reliability and accuracy of a classification system is measured through its prospective application in real time. However, the logistics of performing over 8000 case assessments worldwide pose significant hurdles. Accordingly, we have performed an assessment we believe to be as close as possible to the "real-life scenario" taking into account these practical obstacles. For example, using a live web conference with a single pass through CT sections, raters had limited time to diagnose and classify fractures. Additionally, raters were not able to return to previous answers once submitted, a relative strength compared with surveys where raters are able to compare fractures and classifications against one another by returning to previous cases/answers. Moreover, while the levels of experience were relatively evenly distributed, a higher proportion of orthopedic spine respondents compared with other subspecialties participated. As a result, general orthopedic and orthopedic trauma subspecialities were required to be combined for statistical analysis. This is consistent with AO Spine membership demographics. However, high agreement regarding management of patients has been previously shown between orthopedic spine and other subspecialty surgeons.[21,22] Lastly, study participants were all members of AO Spine and/or AO Trauma which may impart a participation bias towards academic and hospital employed surgeons who may be more familiar with AO classification systems.

Overall, the AO spine sacral classification system appears to be universally applicable among surgeons of various subspecialties and levels of experience with acceptable reliability, reproducibility, and accuracy. Future prospective clinical studies are needed to evaluate the clinical relevance and the usefulness of classification categories before the system can be used as a management tool.

➢ Key Points

- ❑ Respondents from all levels of practice experience demonstrated excellent interobserver reliability when classifying overall morphology and substantial interobserver reliability in overall subtype.
- ❑ General orthopedists, orthopedic traumatologists, neurosurgeons, and orthopedic spine surgeons exhibited excellent interobserver reliability in overall morphology classification and substantial interobserver reliability in overall subtype classification.
- ❑ The AO spine sacral classification system appears to be universally applicable among surgeons of various subspecialties and levels of experience with acceptable reliability, reproducibility, and accuracy.

Acknowledgments

This study was organized and funded by AO Spine International through the AO Spine Knowledge Forum Trauma, a focused group of international spinal trauma experts acting on behalf of AO Spine. AO Spine is a clinical division of the AO Foundation which is an independent medically-guided non-profit organization. Study support was provided directly through the AO Spine Research Department. The authors would like to thank Olesja Hazenbiller (AO Spine) for her editorial and administrative assistance.

AO Spine Sacral Classification Group Members:
Aaron Hilis
Alfredo Guiroy
Aaron Hockley
Ahmed Shawky Abdelgawaad
Akbar Jaleel Zubairi
Alarcon Olger
Alberto Campione
Alecio Barcelos
Alex del Arco
Alexander Weening
Ali Haghnegahdar
Ali Öner
Alqatub Ahmed
Alsammak Wael
Alvin Pun
Amit Bhandutia
Andjel German
Antonio Sanchez Rodriguez
Ashraf El Naga
Barbeiro Gonçalves
Bas Teresa
Bernhard Ullrich
Bernucci Claudio
Bruno Saciloto

Carlos Pernal
Charalampidis Anastasios
Christina Cheng
Ciccioli Nicolas
Cumhur Kilicer
Daniel Cruz
Daniela Linhares
Dare Ashaolu
Das Gurudip
Dave Anthony Dizon
De Falco Giovanni
Dewan Shamsul Asif
Dihan Aponso
Dimitros Patronis
Duchén Rodríguez Luis Miguel
Elias Enmanuel Javier Martinez
Enrico Pola
Erol Gercek
Fabii Pozzi
Frank Lyons
Geoffrey Tipper
Glukhov Dmitrii
Gupta Anuj
Hai Nie
Hamouda Waeel
Hany Soliman
Heiller Torres Valencia
Huiren Tao
Ichiro Okano
Ignacio Fernández Bances
Igor Paredes
Janardhana Aithala Parampalli
Jason Pittman
Javier Lecaros
Javier Matta
Jeronimo Milano
Joana Freitas
Joana Ganau
Joep Kitzen
Jorge Alves
Jose Joefrey Arbatin
Juan Carlos Ramos Torres
Juan Esteban Muñoz Montoya
Juan Lourido
Jun Zhang
Khatri Manoj
Khattab Mohamed
Kimmatkar Nitin
Konstantinos Margetis
Kumar Vineet
Lady Lozano C
Leandro Badalassi
Lykourgos Kollintzas
Maarten de Boer
Manabu Ito
Martin Estefan
Martin Holas
Martinelli Federico
Matias Pereira Duarte
Matti Scholz
Maximo-Alberto Diez-Ulloa
Medina Matías
Michel Triffaux
Mohammad El-Sharkawi
Morakis Andreas
Mostafa Abdelhafez
Muthu Sathish
Nasser Khan
Nicassio Nicola
Nishanth Ampar
Nuno Neves
Olga Morillo
Osmar Moraes
Osundina Ahmad
Paul van Urk
Pedro Bazán
Pershin Andrey
Peter Ferlic
Peter Loughenbury
Philippe Bancel
Popescu Eugen Cezar
Pragnesh Bhatt
Prakash Sitoula
Przemyslaw Kolakowski
Purnajyoti Banerjee
Rabindra Pradhan
Rajesh Bahadur Lakhey
Rakesh Kumar
Ramieri Alessamdro
Reddy Karthik
Rian Vieira
Ricardo Frada
Ricardo Rodrigues-Pinto
Ricky Rasschaert
Russo Fabrizio
Salah Alakkad
Samer Samy
Sbaffi Pier Filippo
Schnake, Klaus
Sebastián Anibal Kornfeld
Sergio Zylbersztejn
Sícoli Alfredo
Siegmund Lang
Stacey Darwish
Stipe Corluka
Subbiah Jayakumar
Subramaniam Macherla haribabu
Sudhir Ganesan
Taha Karimjee
Takuya Nagai
Tarek Elhewala
Tejeda Martin
Thami Benzakour
Toluse Adetunji

Tsuyoshi Okudaira
Tulio Rangel
Umit Guler
Vijay Kumar Loya
Vishwajeet Singh
Yetkin Söyüncü
Zacharia Silk
Zagorac Slavisa
Zdenek Klez

References

1. Bydon M, De la Garza-Ramos R, Macki M, et al. Incidence of sacral fractures and in-hospital postoperative complications in the United States. *Spine (Phila Pa 1976)* 2014;39:E1103–9.
2. Wagner D, Ossendorf C, Gruszka D, Hofmann A, Rommens PM. Fragility fractures of the sacrum: how to identify and when to treat surgically?. *Eur J Trauma Emerg Surg* 2015;41:349–62.
3. Rodrigues-Pinto R, Kurd M, Schroeder G, et al. Sacral fractures and associated injuries. *Global Spine J* 2017;7:609–16.
4. Strange-Vognsen HH, Lebech A. An unusual type of fracture in the upper sacrum. *J Orthop Trauma* 1991;5:200–3.
5. Denis F, Davis S, Comfort T. Sacral fractures: an important problem. Retrospective analysis of 236 cases. *Clin Orthop Relat R* 1988;227:67–81.
6. Isler B. Lumbosacral lesions associated with pelvic ring injuries. *J Orthop Trauma* 1990;4:1–6.
7. Roy-Camille R, Saillant G, Gagna G, et al. Transverse fracture of the upper sacrum. *Spine (Phila Pa 1976)* 1985;10:838–45.
8. Lehman RA, Kang DG, Bellabarba C. A new classification for complex lumbosacral injuries. *Spine J* 2012;12:612–28.
9. Bonnin JG. Sacral fractures and cauda equina lesions. *Med World* 1947;67:140–4.
10. Tile M. Pelvic ring fractures: should they be fixed?. *J Bone Joint Surg Br* 1988;70:1–12.
11. Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma* 2005;19: 404–409.
12. Vaccaro AR, Schroeder G, Divi S, et al. Description and reliability of the AOSpine sacral classification system. *J Bone Joint Surg* 2020;16:1454–63.
13. Vaccaro AR, Oner FC, Kepler C, et al. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. *Spine (Phila Pa 1976)* 2013;38:2028–37.
14. Vaccaro AR, Koerner J, Radcliff K, et al. AOSpine subaxial cervical spine injury classification system. *Eur Spine J* 2016;25: 2173–2184.
15. Schroeder GD, Kurd M, Kepler C, et al. The development of a universally accepted sacral fracture classification: a survey of AOSpine and AOTrauma members. *Global Spine J* 2016;6: 686–694.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
17. Sadiqi S, Oner FC, Dvorak M, et al. The influence of spine surgeons' experience on the classification and intraobserver reliability of the novel AOSpine thoracolumbar spine injury classification system-An International Study. *Spine (Phila Pa 1976)* 2015;40:E1250–6.
18. Rajasekaran S, Kanna R, Schroeder G, et al. Does the spine surgeon's experience affect fracture classification, assessment of stability, and treatment plan in thoracolumbar injuries?. *Global Spine J* 2017;7:309–16.
19. Ebraheim NA, Biyani A, Salpietro B. Zone III fractures of the sacrum. *Spine (Phila Pa 1976)* 1996;21:2390–6.
20. Porrino JA, Kohl CA, Holden D, et al. The Importance of Sagittal 2D reconstruction in pelvic and sacral trauma: avoiding oversight of U-shaped fractures of the sacrum. *Am J Roentgenol* 2010;194:1065–71.
21. Grauer JN, Vaccaro AR, Beiner JM, et al. Similarities and differences in the treatment of spine trauma between surgical specialties and location of practice. *Spine (Phila Pa 1976)* 2004;29:685–96.
22. Canseco JA, Schroeder G, Patel P, 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; 30:517–23.