


Development of Online Technique for International Validation of the AO Spine Subaxial Injury Classification System

Global Spine Journal
2022, Vol. 0(0) 1–9
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/21925682221098967
journals.sagepub.com/home/gsj


Mark J. Lambrechts, MD¹ , Gregory D. Schroeder, MD¹, Brian A. Karamian, MD¹ , Jose A. Canseco, MD, PhD¹ , Cumhur Oner, MD², Emiliano Vialle, MD³ , Shanmuganathan Rajasekaran, PhD⁴ , Olesja Hazenbiller, MS⁵, Marcel R. Dvorak, MD⁶, Lorin M. Benneker, MD⁷, Frank Kandziora, MD⁸, Klaus Schnake, MD⁹, Christopher K. Kepler, MD¹, and Alexander R. Vaccaro, MD, MBA, PhD¹, for the AO Spine Subaxial Classification Group Members

Abstract

Study Design: Global cross-sectional survey

Objective: To develop and refine the techniques for web-based international validation of fracture classification systems.

Methods: A live webinar was organized in 2018 for validation of the AO Spine Subaxial Injury Classification System, consisting of 35 unique computed tomography (CT) scans and key images with subaxial spine injuries. Interobserver reliability and intraobserver reproducibility was calculated for injury morphology, subtype, and facet injury according to the classification system. Based on the experiences from this webinar and incorporating rater feedback, adjustments were made in the organization and techniques used and in 2020 a repeat validation webinar was performed, evaluating images of 41 unique subaxial spine injuries.

Results: In the 2018 session, the AO Spine Subaxial Injury Classification System demonstrated fair interobserver reliability for fracture subtype ($\kappa = 0.35$) and moderate reliability for fracture morphology and facet injury ($\kappa=0.45, 0.43$, respectively). However, in 2020, the interobserver reliability for fracture morphology ($\kappa = 0.87$) and fracture subtype ($\kappa = 0.80$) was excellent, while facet injury was substantial ($\kappa = 0.74$). Intraobserver reproducibility for injury morphology ($\kappa = 0.49$) and injury subtype/facet injury were moderate ($\kappa = 0.42$) in 2018. In 2020, fracture morphology and subtype reproducibility were excellent ($\kappa = 0.85, 0.88$, respectively) while reproducibility for facet injuries was substantial ($\kappa = 0.76$).

¹Rothman Institute at Thomas Jefferson University, Philadelphia, PA, USA

²Department of Orthopedic Surgery, University Medical Center, University of Utrecht, Utrecht, Netherlands

³Spine Surgery Group, Department of Orthopaedics, Cajuru University Hospital, Catholic University of Parana, Curitiba, Brazil

⁴Department of Orthopedics and Spine Surgery, Ganga Hospital, Coimbatore, India

⁵Research Department, AO Spine, AO Foundation, Davos, Switzerland

⁶Division of Spine, University of British Columbia, Vancouver, BC, Canada

⁷Department of Orthopaedic Surgery, Inselspital, University of Bern, Bern, Switzerland

⁸Unfallklinik Frankfurt am Main, Frankfurt, Germany

⁹Center for Spinal and Scoliosis Surgery, Malteser Waldkrankenhaus St. Marien, Erlangen, Germany

¹⁰Department of Orthopedics and Traumatology, Paracelsus Private Medical University Nuremberg, Nuremberg, Germany

Corresponding Author:

Mark J. Lambrechts, Rothman Institute at Thomas Jefferson University, 925 Chestnut St. 5th floor Philadelphia, PA, USA 19107 Philadelphia, PA, USA.

Email: mark.lambrechts2016@gmail.com



Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Conclusion: With optimized webinar-based validation techniques, the AO Spine Subaxial Injury Classification System demonstrated vast improvements in intraobserver reproducibility and interobserver reliability. Stringent fracture classification methodology is integral in obtaining accurate classification results.

Keywords

cervical spine, injury classification, CT, vertebral body fracture, methodology

Introduction

Classification systems are communication instruments designed to characterize injuries and simplify the transmission of information between providers. These systems are also used as tools to help guide clinical management. Ideal classification systems are both concise and comprehensive, and must be reliably applied by all care providers. A classification system that lacks reliability or has poor validity will have minimal clinical value as it may serve as a biased predictor of patient outcomes.¹ At best, this may confound scientific research and at worst, it may negatively influence patient management.²

Historically, validation of classification systems have occurred during national or international conferences; however, the coronavirus disease-2019 (COVID-19) pandemic precipitated utilization of online platforms to communicate ideas.³ Although potential drawbacks with videoconference platforms including quickened mental fatigue and information overload, they can also provide unparalleled opportunity for global communication and inclusion of additional participants.⁴ Previous attempts at validating the AO Spine Subaxial Injury Classification have been based on the reliability and reproducibility of only a few physicians, bringing into question the generalizability of these injury classification systems.⁵⁻⁷ Therefore, a protocol which facilitates identifying the reliability and reproducibility of an injury classification system internationally with the inclusion of hundreds of participants may elucidate pitfalls of the current AO Spine Subaxial Injury Classification.

The value of evaluating a classification system in this manner lies in its ability to include a variety of surgeon characteristics that can determine if the classification system can be globally implemented, throughout levels of training, and across multiple specialties. Accordingly, the aim of this study was to perform an advanced validation of the AO Spine Subaxial Injury Classification System using surgeons from around the world through various years of practice. We have therefore developed an online validation technique, which can be used for these kinds of studies.

Methods

The AO foundation has relationships with multiple institutions internationally. In order to receive anonymous injury films, IRB approval is required by each of the centers providing the

images. Additionally, AO has research imaging associates who confirm image anonymity to ensure there is no breach of patient confidentiality. Once anonymity of the films is confirmed, injury CT and radiographs are deposited in a secured database for future validation studies.

An online live webinar format was utilized to maximize the inclusion of AO Spine Subaxial Injury Classification System validation participants with an emphasis on inclusion of participants from different parts of the world with varying degrees of experience. The first of these webinars was conducted in 2018. All AO Spine members were invited to participate in the webinar. A total of 264 AO Spine members joined the session (assessment 1) which was repeated after 3 weeks (assessment 2). The webinar session was conducted in English. Prior to assessment of the injury classification, all participants watched a live tutorial video by one of the developers of the classification system during a training session. Radiographic images for 35 distinct injuries were then presented for review (all injury types had 3 distinct injury CT videos reviewed except for F4 injuries, which only had two distinct injuries, and B1 injuries, which were underrepresented in the database and were thus chosen not to be included in the validation). The case order was randomized during both case assessments. Computed tomography (CT) scans of the entire cervical spine including high-resolution axial, sagittal, and coronal videos were viewed by each participant at a rate of 10 frames/second. Additionally, key images (reported by participants to be difficult to visualize due to their small size) were also available for viewing simultaneously. Respondents were asked to classify fractures based on the AO Spine Subaxial Injury Classification System, which included injury location (C3-C7), morphology (A, B, C), subtype (A1, A2, B2, etc.), and presence of a facet injury (Figure 1). The validation members were also asked if the facet injury was unilateral or bilateral and a primary or secondary injury. All answers were recorded in an online survey using Research Electronic Data Capture (REDCap) that required a classification be recorded for each injury without the ability for the participant to skip cases. Therefore, if a participant had internet connection issues and was unable to view the CT scan of an injury, they were required to put random answers to resume the presentation (e.g. if five CT scans were missed due to connection issues, five random classifications were required to be put into the survey). It should also be noted that if the online survey was closed prior to completion of the validation, the

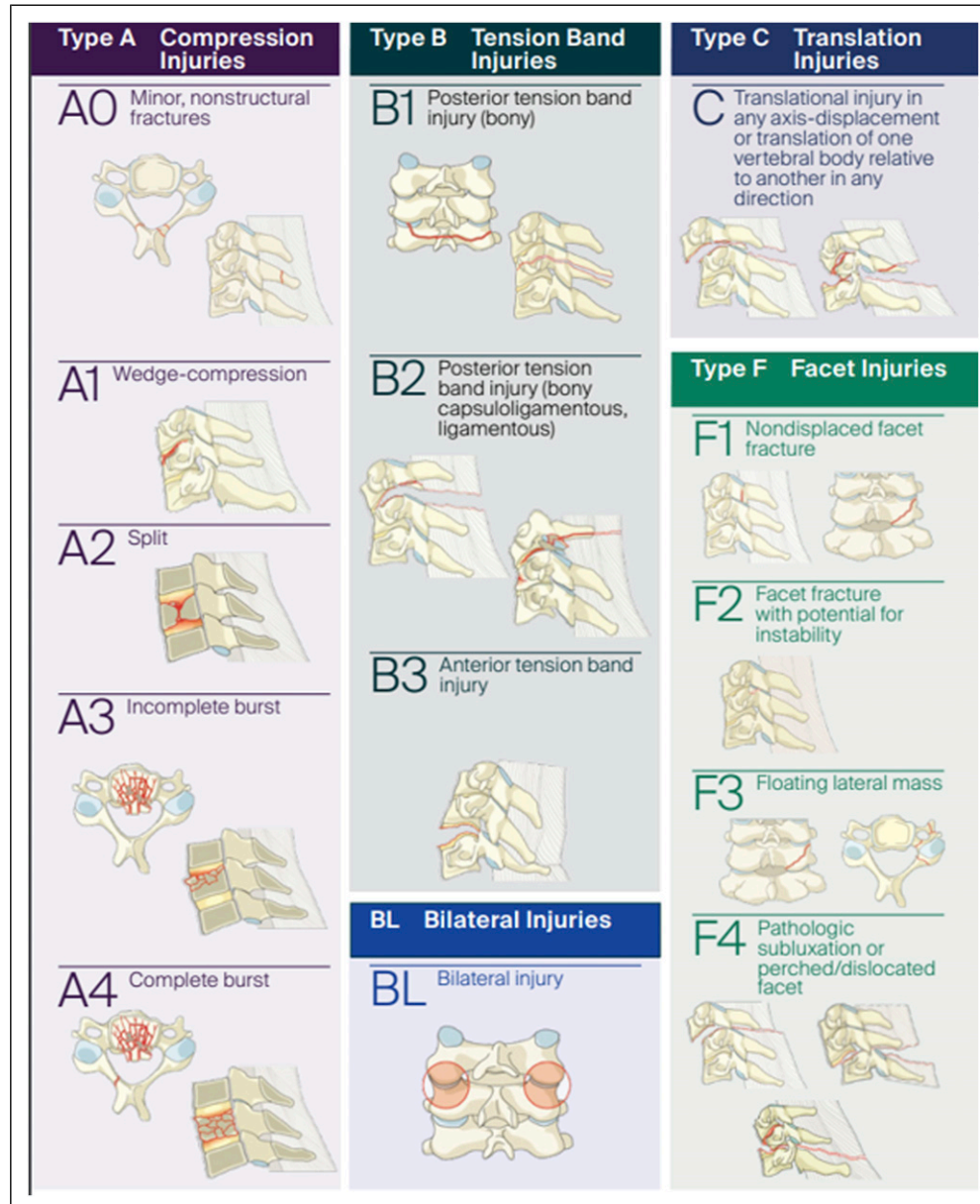


Figure 1. Pictorial demonstration of the AO Spine Subaxial Injury Classification. Permission to use this image granted by the AO Foundation®, AO Spine, Switzerland.

answers were not saved and random answers were required as input prior to resuming the validation.

Because this was a novel way of conducting a fracture classification validation, we asked the participants to provide feedback and make suggestions for improvement. All the comments from the validation members were reviewed. Based on the comments, changes in the construction of the webinar were implemented including reducing the frame rate speed of all CT videos and limiting the CT video to the segment of injury. Additionally, we decided to focus the classification on injury type (A,B,C) and fracture subtype with separate facet

injury scoring. We also modified the survey to tolerate missed questions due to internet connection issues (Figure 2).

With this modified approach, we repeated the study in 2020. In this live webinar, 203 AO Spine members from six different geographic regions participated. Similar to the previous iteration, all participants attended the live tutorial video and training session. However, CT images presented during this assessment were trimmed to the region of interest and were played once at a rate of 2 frames/second. Additionally, key images were presented both before and after CT images, and were standardized in size and quality. In this validation,

NOTE: Please remember, a facet injury is always considered the secondary injury, unless it occurs in isolation.

This question requires an answer. ←

* 10. List the primary injury

Primary Injury: Injury Level [dropdown] Main Injury [dropdown]

11. List secondary injury:

Additional Injury: Injury Level [dropdown] Secondary Injury [dropdown]

12. List the facet injury if it is the primary injury or if it is the major cause of the primary injury.

Facet Injury: Level Specification [C3-C4] Facet Injury [F2] Bilateral [BL]

Prev Next

A

Assessment 1

Please classify the fracture type from A-C according to the AO Spine Subaxial Injury Classification System?

Note: If you do not see any injury, or you think the injury is an isolated F1 or F2 or F3 injury, please select A0.

* must provide value

A0
 A1
 A2
 A3
 A4
 B1
 B2
 B3
 C

Does the injury involves the facets? Yes No

* must provide value

Please specify the facet injury type?

* must provide value ←

F1
 F2
 F3
 F4

Submit

B

Completed Welcome - Intro

Completed Identification - Intro

Begin survey Training Session - TR - Case 1

Begin survey Training Session - TR - Case 2

Begin survey Training Session - TR - Case 3

Begin survey Training Session - TR - Case 4

Begin survey Training Session - TR - Case 5

Begin survey Training Session - TR - Case 6

Begin survey Training Session - TR - Case 7

Begin survey Training Session - TR - Case 8

Begin survey Training Session - TR - Case 9

Begin survey Training Session - TR - Case 10

Begin survey Assessment 1 - Case 1

C Begin survey Assessment 1 - Case 2

Figure 2. (A). Example of the 2018 survey, which requires a primary injury selection in order to proceed to the next case. The black arrow demonstrates facet injury selection alone will not let the rater proceed to the next question. (B). The 2020 survey does not require entry of a primary or secondary injury and allows for easier selection of facet injuries. The black arrow indicates this question only populates with selection of “yes” to a facet injury. (C). The 2020 survey did not require completing surveys on all previous cases before proceeding to a specific case assessment.

cases with facet injuries were presented separately from morphologic A, B, C classification and were not required to be indicated as primary vs secondary injuries or distinguished as unilateral vs bilateral injuries. Additionally, validation members were no longer required to identify the level of injury. Forty-one cases were included in these 2020 sessions.

Demographic data for all validation members was recorded in both 2018 and 2020 based on geographic region (North America, Central and South America, Europe, Africa, Middle East, and Asia and the Pacific) and surgical subspecialty (orthopaedic spine, neurosurgery, or other), but number of years in practice (<5, 5-10, 11-20, and >20) was only recorded in 2020.

AO Spine Knowledge Forum Trauma

Members of the AO Spine Knowledge Forum Trauma committee picked images chosen for inclusion in the

validation. The decision for selecting of each image is based on if the image is representative of the classification description. For example, does an A4 injury accurately represent a complete burst fracture or is it difficult to tell if there it is an incomplete burst fracture with questionable extension into the inferior endplate. This might be difficult to detect given the time constraints for the validation, since each injury evaluated is seen live without the ability to review the image at a later time period.

Gold Standard Committee

The gold standard committee assigned the AO Spine Subaxial Injury Classification System grade to each distinct injury. The committee was comprised of five surgeons who are members of the AO Spine Knowledge Forum Trauma committee, which also consisted of the original creators of the classification schema. Each gold standard committee member rated each

injury individually. Any image with disagreement then required a formal discussion to resolve the dispute. Each injury film selected for validation inclusion ultimately achieved unanimous agreement by the committee and this was determined the final gold-standard classification by which the validation participants would be graded.

Statistical Analysis

Chi-square test was used to evaluate the significance of change in demographic data between the 2018 and 2020 webinar cohorts. Agreement percentages were used to compare the validation members classification grade to the “gold standard”, defined by a panel of expert spine surgeons and traumatologists who came to unanimous agreement on the classification of the injury.

Per consultation with our statistician, a minimum of three cases was required to achieve adequate power to compare intraobserver reproducibility and interobserver reliability ratings based on each injury. Cohen’s Kappa (κ) statistic was used to assess the reproducibility and reliability of the injury morphology (A, B, or C), injury subtype (A1, A2, A3, etc.), and facet injury (F1, F2, F3, or F4) classification between independent observers (interobserver reliability) and the reproducibility of the injury classification over two assessments (intraobserver reproducibility). The κ coefficients were interpreted using the Landis and Koch grading system.⁸ A κ coefficient of less than 0.2 was defined as slight, between 0.2 and 0.4 as fair, between 0.4 and 0.6 as moderate, between 0.6 and 0.8 as substantial, and greater than 0.8 as excellent reliability or reproducibility.

Results

A total of 264 validation members participated in the webinar in 2018, while 203 validation members participated in 2020. The

validation members in 2018 and 2020 were primarily from Europe (2018: 34.1%, 2020: 40%), Central and South America (2018: 24.2%, 2020: 16.7%), and Asia (17.4%, 2020: 24.6%) with the remaining from North America (2018: 11%, 2020: 8.9%), the Middle East (2018: 7.8%, 2020: 7.4%), and Africa (2018: 5.7%, 2020: 2.5%). There was no significant difference in the validation members’ geographic region between the 2018 and 2020 validations ($P = 0.067$). Both validation cohorts consisted predominantly of orthopaedic spine surgeons (2018: 42.4%, 2020: 60.6%) and neurosurgeons (2018: 28.8%, 2020: 36.9%), but “other” physicians did comprise a much larger percentage of the 2018 group compared to 2020 (2018: 28.8%, 2020: 2.5%), which was significant ($P < 0.001$). The “other” group consisted predominantly of general orthopaedic traumatologists, radiologists, and residents. Demographic data for the AO Spine Subaxial Injury Classification System validation members are presented in Table 1.

Percent Agreement with Gold Standard

Percent agreement for fracture morphology on assessment 1 (AS1) and assessment 2 (AS2) in 2018 (AS1: 54.5%, AS2: 54%) was over 40% lower than the percent agreements noted in 2020 (AS1: 95.4%, AS2: 94.7%). Fracture subtype percent agreement in 2018 (AS1: 40.9%, AS2: 40.5%) was 56% less than the 2020 validation group (AS1: 91.7%, AS2: 90.6%), and percent agreement on facet injury in 2018 (AS1: 50.6%, AS2: 62.8%) was 30% lower than the 2020 validation group (AS1: 88.6%, AS2: 91.3%). Additionally, there was high variability in percent agreement between fracture morphology (“C” types: 32.3% vs “A” types: 82.9%), fracture subtype (“B2” type: 7.73% vs “A2” type: 89.2%), and facet injury (“F2” type: 44.1% vs “F4” type: 79.3%) in 2018. However, the 2020 validation demonstrated minimal variability and a high overall accuracy in fracture morphology [range, 87.2-97.3%],

Table 1. Summary of Surgeon Respondent Demographics.

Category	Characteristic	Respondent (%) 2018 N = 264	Respondent (%) 2020 N = 203	P-value
Geographic Region	North America	29 (11.0)	18 (8.9)	0.067
	Central and South America	64 (24.2)	34 (16.7)	
	Europe	90 (34.1)	81 (40.0)	
	Africa	15 (5.7)	5 (2.5)	
	Asia	46 (17.4)	50 (24.6)	
	Middle East	20 (7.8)	15 (7.4)	
Number of Years in Practice	<5 years	X	33 (16.3)	N/A
	5-10 years	X	64 (31.5)	
	11-20 years	X	69 (34.0)	
	>20 years	X	37 (18.2)	
Surgical Subspecialty	Neurosurgeon	76 (28.8)	123 (36.9)	<0.001
	Orthopaedic Spine	112 (42.4)	75 (60.6)	
	Other*	76 (28.8)	5 (2.5)	

*Excluded from further analyses

Table 2. Percent Agreement With Gold Standard Fracture Classification.

Subaxial Cervical Classification percent correct		2018 Validation		2020 Validation	
		Assessment 1 (%)	Assessment 2 (%)	Assessment 1 (%)	Assessment 2 (%)
Fracture Morphology	A	76.1	82.9	97.3	97.1
	B	41.9	42.0	90.1	87.2
	C	32.3	32.3	94.0	93.3
	Combined	54.5	54.0	95.4	94.7
Fracture Subtype	A0	40.7	49.1	94.1	93.6
	A1	65.9	76.8	88.3	84.2
	A2	80.7	89.2	95.6	94.5
	A3	74.7	87.5	90.9	92.0
	A4	31.8	28.0	83.8	82.7
	B2	7.55	7.73	80.3	78.4
	B3	72.7	79.3	94.5	90.9
	C	32.3	21.5	94.0	93.3
	Combined	40.9	40.5	91.7	90.6
Facet Injury	F1	41.9	58.6	88.2	92.1
	F2	27.4	44.1	86.3	89.4
	F3	44.8	44.9	87.2	89.8
	F4	69.5	79.3	91.4	93.0
	Combined	50.6	62.8	88.6	91.3

Table 3. Interobserver Reliability in Fracture Classification.

Subaxial Cervical Classification		2018 Validation		2020 Validation	
		Assessment 1 (k)	Assessment 2 (k)	Assessment 1 (k)	Assessment 2 (k)
Fracture Morphology	A	0.51	0.61	0.87	0.84
	B	0.32	0.40	0.80	0.74
	C	0.18	0.13	0.89	0.87
	Combined	0.40	0.45	0.86	0.87
Fracture Subtype	A0	0.26	0.38	0.89	0.88
	A1	0.38	0.51	0.77	0.67
	A2	0.37	0.48	0.84	0.84
	A3	0.39	0.49	0.77	0.72
	A4	0.41	0.35	0.77	0.73
	B2	0.10	0.03	0.73	0.68
	B3	0.52	0.66	0.91	0.84
	C	0.18	0.13	0.89	0.87
	Combined	0.32	0.35	0.84	0.80
Facet Injury	F1	0.26	0.32	0.65	0.75
	F2	0.20	0.26	0.59	0.69
	F3	0.22	0.19	0.77	0.75
	F4	0.46	0.55	0.68	0.76
	Combined	0.36	0.43	0.67	0.74

subtype [range, 78.4-95.6%], and facet injury [range, 86.3-93.0%] (Table 2).

Interobserver Reliability

The overall interobserver reliability demonstrated disparate results when comparing the 2018 and 2020 validations. The

overall interobserver reliability for fracture morphology (AS1: $\kappa = 0.40$, AS2: $\kappa = 0.45$), fracture subtype (AS1: $\kappa = 0.32$, AS2: $\kappa = 0.35$), and facet injury (AS1: $\kappa = 0.36$, AS2: $\kappa = 0.43$) ranged from fair to moderate in 2018. However, in 2020 the interobserver reliability of fracture morphology (AS1: $\kappa = 0.86$, AS2: $\kappa = 0.87$) and fracture subtype (AS1: $\kappa = 0.84$, AS2: $\kappa = 0.80$) were excellent on both assessments, while facet

Table 4. Intraobserver Reproducibility in Fracture Classification.

Subaxial Cervical Classification	2018 Validation	2020 Validation
	Kappa (mean)	Kappa (mean)
Fracture Morphology	0.49	0.85
Fracture Subtype	0.42	0.88
Facet Injury		0.76

injury classification was substantial (AS1: $\kappa = 0.67$, AS2: $\kappa = 0.74$) (Table 3).

Intraobserver Reproducibility

Intraobserver reproducibility for fracture morphology (Type A, B, and C) was moderate ($\kappa = 0.49$) and excellent ($\kappa = 0.85$) in 2018 and 2020, respectively. Intraobserver reproducibility for fracture subtype (A1-C) and facet injury (F1-F4) were combined in 2018 due to survey methodology and we were unable to analyze separately; however, the combined fracture subtype and facet injury intraobserver validation was moderate ($\kappa = 0.42$). In 2020, separate calculations were performed for fracture subtype intraobserver reproducibility ($\kappa = 0.88$), and facet injury reproducibility ($\kappa = 0.76$), which were excellent and substantial, respectively (Table 4).

Discussion

In an attempt to involve surgeons from around the world with varying degrees of experience, the AO Spine knowledge forum trauma attempted a new technique to validate the AO Spine Subaxial Injury Classification System. In our first trial in 2018 we came across some technical and organizational problems. The results were disappointing and the feedback from the participants was overwhelmingly negative. While the initial results demonstrated lower than expected reliability, user comments indicated that this was likely a technical issue with the validation and not an issue with the classification itself. The results from the 2018 validation demonstrated “fair” interobserver reliability and “moderate” intraobserver reproducibility. This was much lower than the AO Spine preliminary subaxial validation by the knowledge forum members.⁹ Based on the participant’s responses, this lower reliability was largely attributed to the difficulty in fracture visualization during the live conference format and survey question structure. After internal discussions we modified our approach. The frames viewed per second were reduced five-fold in the 2020 validation and CT scans were trimmed to the area of injury rather than displaying the entire cervical spine to minimize rater fatigue during the assessment.¹⁰ Additionally, raters were not required to determine the vertebral injury level or to decipher between primary vs secondary injuries for fractures with both A, B, C and facet components. As a result, the 2020 validation demonstrated

excellent interobserver reliability and intraobserver reproducibility for fracture morphology and subtype, with substantial interobserver reliability and intraobserver reproducibility for facet injuries.

In an attempt to limit validation member recall of each injury, two separate live webinar events were held three weeks apart in 2018. This ensured all validation members would have an equal length of time (three weeks) between each subaxial cervical spine injury evaluation. However, the frame rate speed (10 frames/second) during the 2018 validation likely contributed to study bias due to different validation members’ injury identification and classification processing speeds, which may be affected by age, experience level, or English language proficiency.¹¹⁻¹³ Therefore, when performing radiographic validation studies with a large number of participants in a live webinar format, our study identified slow frame rate speed is imperative to allow adequate time for imaging review, as the live format does not allow for raters to visualize the scan with multiple passes. This may be exponentially important when relying on international validation team members who have variable broadband internet speeds and variable English proficiency.¹⁴ Additionally, validation participants should be given adequate time between cases to record the injury classification so as not to interrupt viewing of subsequent injury films.

While the validation of any classification system is essential to establish its clinical utility, the literature is currently bereft of recommendations for large scale advanced validation testing. The results of this study highlight the importance of proper methodology for international classification validation. Although both validation studies were performed by AO members using the exact same classification system, disparate validation results were obtained. This ultimately highlights how suboptimal methodology can serve as a biased predictor of outcomes in research. Part of the importance of this study is building the foundation of the methodology required to obtain accurate fracture classification validations. We recommend displaying standardized high-resolution key images before and after viewing CT imaging. CT images may be played through once only to limit time spent to one minute on each injury evaluation, but they should be trimmed to the region of injury and played at a rate of no faster than 2 frames/second. Ideally, surveys should also have at minimum 15 seconds to allow for answer selection before displaying the subsequent injury for rating, and survey questions should not have mandatory answers before proceeding to the next injury due to potential internet-connectivity issues. Live technical support familiar with the webinar software should be present throughout the validation to assist when necessary. Finally, a trial tutorial of selected cases should be instituted prior to the official validation to help guide the rater in the workflow necessary for validation completion.

In order to establish whether methodological issues alone were causative of the lower classification accuracy of the 2018 validation, we examined the effect of regional participation

between the 2018 and 2020 webinars. Although there was a slight increase in participants from Asia and Europe in 2020, this did not reach statistical significance. Further, previous literature evaluating AO Spine fracture classification systems has not demonstrated geographic region or surgeon experience accounts for differences in radiographic classification of thoracolumbar fractures. Therefore, we believe it unlikely surgeon demographics alone accounted for the large variability in the subaxial cervical spine classification accuracy.¹⁵

In order to further identify the role of methodology in variations between the 2018 and 2020 iterations of the fracture classification, we compared both groups to an established baseline as determined by the AO Spine knowledge forum trauma. The knowledge forum trauma previously demonstrated an interobserver reliability for injury subtype of 0.64 and an intraobserver reproducibility of 0.75 indicating substantial reliability and reproducibility of the AO Spine Subaxial Injury Classification System.⁹ It should be noted, only 10 participants were included in the study, so poor reliability or reproducibility from a single member would have large effects on the results of their validation.⁹ Although the 2018 validation demonstrated fair to moderate injury subtype and morphology reliability and reproducibility, the 2020 iteration demonstrated excellent intraobserver reproducibility and interobserver reliability, thus comparing favorably to the results obtained by the knowledge forum trauma.⁹ Although it is unclear why the 2020 classification results actually had improved reliability and reproducibility compared to the knowledge forum trauma, this may be a result of having over 200 participants, which allowed for some variability in incorrectly classifying injuries. Additionally, in 2020 the classification system had been present for 5 years prior to the validation, giving participants the opportunity to practice classifying subaxial cervical spine fractures prior to the study.

There were multiple limitations present in the design of our study. Although a side-to-side comparison of the 2018 and 2020 interobserver reliability and intraobserver reproducibility appears to indicate study design alteration after 2018 led to improved interobserver reliability and intraobserver reproducibility, this may be partly due to confounding on multiple levels. First, although regional demographics were not significantly different between the two validation groups, we were unable to collect surgeon experience during the 2018 validation. Additionally, since there was a significantly larger group of non-spine surgeons in 2018, this may have led to worse interobserver reliability and intraobserver reproducibility in the 2018 participation as opposed to true differences due to validation methodology alone. Second, if validation members had internet connection problems, inadequate time to classify injuries between cases, or they did not visualize the fracture during the first viewing of the CT scan, they were unable to review the CT scan for either the 2018 or 2020 iteration of the validation study. In clinical practice, there is no limitation to the time you are able to view a CT scan and subsequently classify the injury. This may have artificially

suppressed the percent correct compared to the gold standard and the interobserver reliability and intraobserver reproducibility for both validation studies. Third, the study did not use identical injury films during the 2018 and 2020 validation periods since the initial primary goal of the study was to obtain an accurate depiction of the interobserver reliability and intraobserver reproducibility of the AO Spine Subaxial Injury Classification System, not to validate if the 2020 methodology was superior to the 2018 version. Finally, injury films evaluated by validation participants were limited to what was available in the AO Spine repository. Due to the limited number of F4 injuries available ($N = 2$), the interobserver reliability and intraobserver reproducibility may have been underpowered. Further, there were no B1 injuries evaluated. Given the lower reliability and reproducibility of B type injuries, this may have artificially raised the reliability and reproducibility of our results in both 2018 and 2020. It is worth noting that B1 injuries were not evaluated in either version of the validation, so this would not affect the comparison of the reliability and reproducibility results between the 2018 and 2020 validations.

Conclusion

The AO Spine Subaxial Injury Classification System demonstrated substantial reliability for facet fracture injury, and excellent reliability for fracture morphology and subtype among surgeons worldwide based on the most recent 2020 validation assessment. The results highlight the importance of survey methodology in classification validation. Imaging speed, video length, survey construction, and web conference interface had significant impact in the dissimilar results found between the 2018 and 2020 validation assessments.

Acknowledgments

The authors of the manuscript would like to thank Christian Knoll and Janik Hilse from the AO Innovation Translation Center for their support with the statistical analysis, developing the methodology, and database structure.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the The AO Spine Injury Classification Systems were developed and funded by AO Spine through the AO Spine Knowledge Forum Trauma, a focused group of international spine trauma experts. AO Spine is a clinical division of the AO Foundation, which is an independent medically-guided not-for-profit


organization. Study support was provided directly through the AO Spine Research Department.

ORCID iDs

Mark J. Lambrechts  <https://orcid.org/0000-0002-9106-2228>

Brian A. Karamian  <https://orcid.org/0000-0003-0512-6019>

Jose A. Canseco  <https://orcid.org/0000-0002-2152-5725>

Emiliano Vialle  <https://orcid.org/0000-0003-1157-4889>

Shanmuganathan Rajasekaran  <https://orcid.org/0000-0001-6043-006X>

References

- Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma*. 2005; 19(6):401-406. doi:10.1097/01.bot.0000155310.04886.37
- Roselaar N, Cuadra MI, Lyman S, Bias and Confounding. In: *Basic Methods Handbook for Clinical Orthopaedic Research*. Berlin, Heidelberg: Springer; 2019. DOI: 10.1007/978-3-662-58254-1_3
- Karl KA, Peluchette JV, Aghakhani N. Virtual work meetings during the COVID-19 Pandemic: The good, bad, and ugly. *Small Group Research*. 2021. Advanced online publication. 10.1177/10464964211015286
- Bailenson J. Nonverbal Overload: A theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior*. 2021;2(1):■■■. DOI: 10.1037/tmb0000030. DOI: 10.1037/tmb0000030
- Kanagaraju V, Yelamarthy PKK, Chhabra HS, et al. Reliability of Allen Ferguson classification versus subaxial injury classification and severity scale for subaxial cervical spine injuries: A psychometrics study. *Spinal Cord*. 2019;57(1):26-32. doi:10.1038/s41393-018-0182-z
- Lee WJ, Yoon SH, Kim YJ, Kim JY, Park HC, Park CO. Interobserver and intraobserver reliability of sub-axial injury classification and severity scale between radiologist, resident and spine surgeon. *J Korean Neurosurg Soc*. 2012;52(3): 200-203. doi:10.3340/jkns.2012.52.3.200
- Silva OT, Sabba MF, Lira HI, et al. Evaluation of the reliability and validity of the newer AOSpine subaxial cervical injury classification (C-3 to C-7). *J Neurosurg Spine*. 2016;25(3): 303-308. doi:10.3171/2016.2.SPINE151039
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174
- Vaccaro AR, Koerner JD, Radcliff KE, et al. AOSpine subaxial cervical spine injury classification system. *Eur Spine J*. 2016; 25(7):2173-2184. doi:10.1007/s00586-015-3831-3
- Faber LG, Maurits NM, Lorist MM. Mental fatigue affects visual selective attention. *PLoS One*. 2012;7(10):e48073. doi: 10.1371/journal.pone.0048073
- Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychol Rev*. 1996;103(3):403-428. doi: 10.1037/0033-295x.103.3.403
- Eckert MA, Keren NI, Roberts DR, Calhoun VD, Harris KC. Age-related changes in processing speed: unique contributions of cerebellar and prefrontal cortex. *Front Hum Neurosci*. 2010; 4:10. doi:10.3389/neuro.09.010.2010
- Barrón-Estrada ML, Zatarain-Cabada R, Romero-Polo JA, Monroy JN. Patrony: A mobile application for pattern recognition learning. *Educ Inf Technol (Dordr)*. 2021;27: 1237-1260. Epub ahead of print. doi:10.1007/s10639-021-10636-7
- Speedtest Global Index. *Ranking mobile and fixed broadband speeds from around the world on a monthly basis*. Seattle, WA: Speedtest Global Index. 2021. <https://www.speedtest.net/global-index>. <https://www.speedtest.net/global-index>. (Accessed September 27, 2021)
- Schroeder GD, Kepler CK, Koerner JD, et al. Is there a regional difference in morphology interpretation of A3 and A4 fractures among different cultures? *J Neurosurg Spine*. 2016;24(2): 332-339. doi:10.3171/2015.4.SPINE1584