


Global Validation of the AO Spine Upper Cervical Injury Classification: Geographic Region Affects Reliability and Reproducibility

Global Spine Journal
2022, Vol. 0(0) 1–9
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/21925682221124100
journals.sagepub.com/home/gsj


Mark J. Lambrechts, MD¹ , Gregory D. Schroeder, MD¹, Brian A. Karamian, MD¹ , Jose A. Canseco, MD, PhD¹ , Richard Bransford, MD² , Cumhur Oner, MD³, Lorin M. Benneker, MD⁴, Frank Kandziora, MD PhD⁵, Rajasekaran Shanmuganathan, MD, PhD⁶ , Rishi Kanna, MD⁶, Andrei F. Joaquim, MD PhD⁷ , Jens R Chapman, MD⁸, Emiliano Vialle, MD⁹ , Mohammad El-Sharkawi, MD¹⁰ , Marcel Dvorak, MD¹¹, Klaus Schnake, MD^{12,13}, Christopher K. Kepler, MD, MBA¹, and Alexander R. Vaccaro, MD, MBA, PhD¹, for the AO Spine Upper Cervical Injury Classification International Members

Abstract

Study Design: Global Survey

Objective: To determine the accuracy, interobserver reliability, and intraobserver reproducibility of the AO Spine Upper Cervical Injury Classification System based on surgeons' AO Spine region of practice (Africa, Asia, Central/South America, Europe, Middle East, and North America).

Methods: A total of 275 AO Spine members assessed 25 upper cervical spine injuries and classified them according to the AO Spine Upper Cervical Injury Classification System. Reliability, reproducibility, and accuracy scores were obtained over two assessments administered at three-week intervals. Kappa coefficients (k) determined the interobserver reliability and intraobserver reproducibility.

Results: On both assessments, participants from Europe and North America had the highest classification accuracy, while participants from Africa and Central/South America had the lowest accuracy ($P < .0001$). Participants from Africa (assessment 1 (AS1):k = .487; AS2:0.491), Central/South America (AS1:k = .513; AS2:0.511), and the Middle East (AS1:0.591; AS2: .599)

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

² Department of Orthopaedic and Sports Medicine, Harborview Medical Center, University of Washington, Seattle, WA, USA

³ Department of Orthopedic Surgery, University Medical Center, University of Utrecht, Utrecht, the Netherlands

⁴ Spine Unit, Sonnenhof Spital Bern, University of Bern, Bern, Switzerland

⁵ Unfallklinik Frankfurt Am Main, Frankfurt, Germany

⁶ Department of Orthopedics and Spine Surgery, Ganga Hospital, Coimbatore, India

⁷ Department of Neurology, Neurosurgery Division, State University of Campinas, Campinas, Sao Paulo, Brazil

⁸ Swedish Medical Center, Swedish Neuroscience Institute, Seattle, WA, USA

⁹ Cajuru University Hospital, Catholic University of Paraná, Curitiba, Brazil

¹⁰ Department of Orthopaedic and Trauma Surgery, Assiut University, Assiut, Egypt

¹¹ University of British Columbia, Vancouver, BC, Canada

¹² Center for Spinal and Scoliosis Surgery, Malteser Waldkrankenhaus St. Marien, Erlangen, Germany

¹³ Department of Orthopedics and Traumatology, Paracelsus Private Medical University Nuremberg, Nuremberg, Germany

Corresponding Author:

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

Email: mark.lambrechts@rothmanortho.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>).

achieved moderate reliability, while participants from North America (AS1:k = .673; AS2:0.648) and Europe (AS1:k = .682; AS2: 0.681) achieved substantial reliability. Asian participants obtained substantial reliability on AS1 (k = .632), but moderate reliability on AS2 (k = .566). Although there was a large effect size, the low number of participants in certain regions did not provide adequate certainty that AO regions affected the likelihood of participants having excellent reproducibility ($P = .342$).

Conclusions: The AO Spine Upper Cervical Injury Classification System can be applied with high accuracy, interobserver reliability, and intraobserver reproducibility. However, lower classification accuracy and reliability were found in regions of Africa and Central/South America, especially for severe atlas injuries (IIB and IIC) and atypical hangman's type fractures (IIIB injuries).

Keywords

AO spine, cervical, validation, injury classification

Introduction

Classification schema are increasingly shared on global platforms; thus, international reliability and reproducibility analysis should be undertaken to see if limitations or modifications of the classification are necessary prior to widespread implementation. The strategy for validation attempts is well outlined by Audigé.¹ In short, a panel of experts evaluate the classification to determine its potential suitability for widespread use. Subsequently, international participants evaluate the classification to determine its generalizability. If both rounds of analysis demonstrate high levels of reliability, injury severity scores and injury management algorithms can be proposed as has been done for the AO Spine Thoracolumbar Injury Classification System.^{2,3}

Similar to the AO Spine Sacral Injury Classification System, the AO Spine Upper Cervical Injury Classification System is reliant on computed tomography (CT) scans to classify upper cervical spine injuries.⁴ This was a primary goal of AO Spine given the well-documented global inequalities in advanced imaging sites, especially the lack of access to magnetic resonance imaging (MRI) in regions like Western Africa and Central/South America.^{5,6} Therefore, while MRIs may have additional prognostic value when compared to CT scans for cervical spine trauma, injury classifications based on CT scans may allow for greater global application of the injury classification.⁷

Unlike previous upper cervical spine injury classifications, which have focused on isolated areas of the upper cervical spine, the AO Spine Upper Cervical Injury Classification System unifies all upper cervical spine segments from the occipital condyles to the C2-3 joint.⁸⁻¹⁷ Although each upper cervical motion segment has unique properties based on its anatomical function, they share certain inherent biomechanics properties under physiologic and posttraumatic conditions.¹⁸ Thus, injury classification principles may be sharable across upper cervical spine segments. Given the lack of an international validation for this newly proposed classification, the purpose of our study was to determine the intraobserver reproducibility and interobserver reliability of the AO Spine Upper Cervical Injury Classification System based on an international audience.

Methods

AO Spine Upper Cervical Injury Classification

This classification was designed to provide a simplistic easy-to-use schema. The injury is first classified based on the anatomical location of the injury, as designated by roman numerals (i.e. I. Craniocervical junction and occipital condyles; II. Atlas and C1-C2 joint; III. C2 ring, dens and C2-C3 joint). Injury types are then classified based on their level of stability and are assigned alphabetical grades. Type A — Purely bony injuries without any ligamentous injury or vertebral body translation; Type B — Tension band failures or ligamentous disruption/avulsion injuries without vertebral translation in any plane; Type C — Vertebral body translation in any plane consistent with subluxation or dislocation. Consistent with previous AO Spine classifications, neurologic injuries and injury modifiers are assigned, but they were not evaluated during this validation (Figure 1).

Reproducibility and Reliability Evaluation

A total of 275 AO Spine members responded favorably to participating in an international validation of the AO Spine Upper Cervical Injury Classification System. Prior to participation, each study participant watched a video in English (https://www.youtube.com/watch?v=KyUYfa_JMb4), which described the classification and informed participants how to correctly classify a variety of upper cervical spine injuries. Each study participant was given the opportunity to ask questions to the instructor (one of the original creators of the classification schema) before participating in a sample validation of three upper cervical spine injuries. Another opportunity to ask the instructor questions was allotted to the study participants after the sample validation, but prior to participating in the reliability and reproducibility analysis.

The official reliability and reproducibility analysis of the AO Spine Upper Cervical Injury Classification System was conducted via a live, online webinar format as previously described.¹⁹ A total of 25 CT videos of upper cervical spine injuries were evaluated with the webinar conducted in English. Key images of the injury, which were concurrently

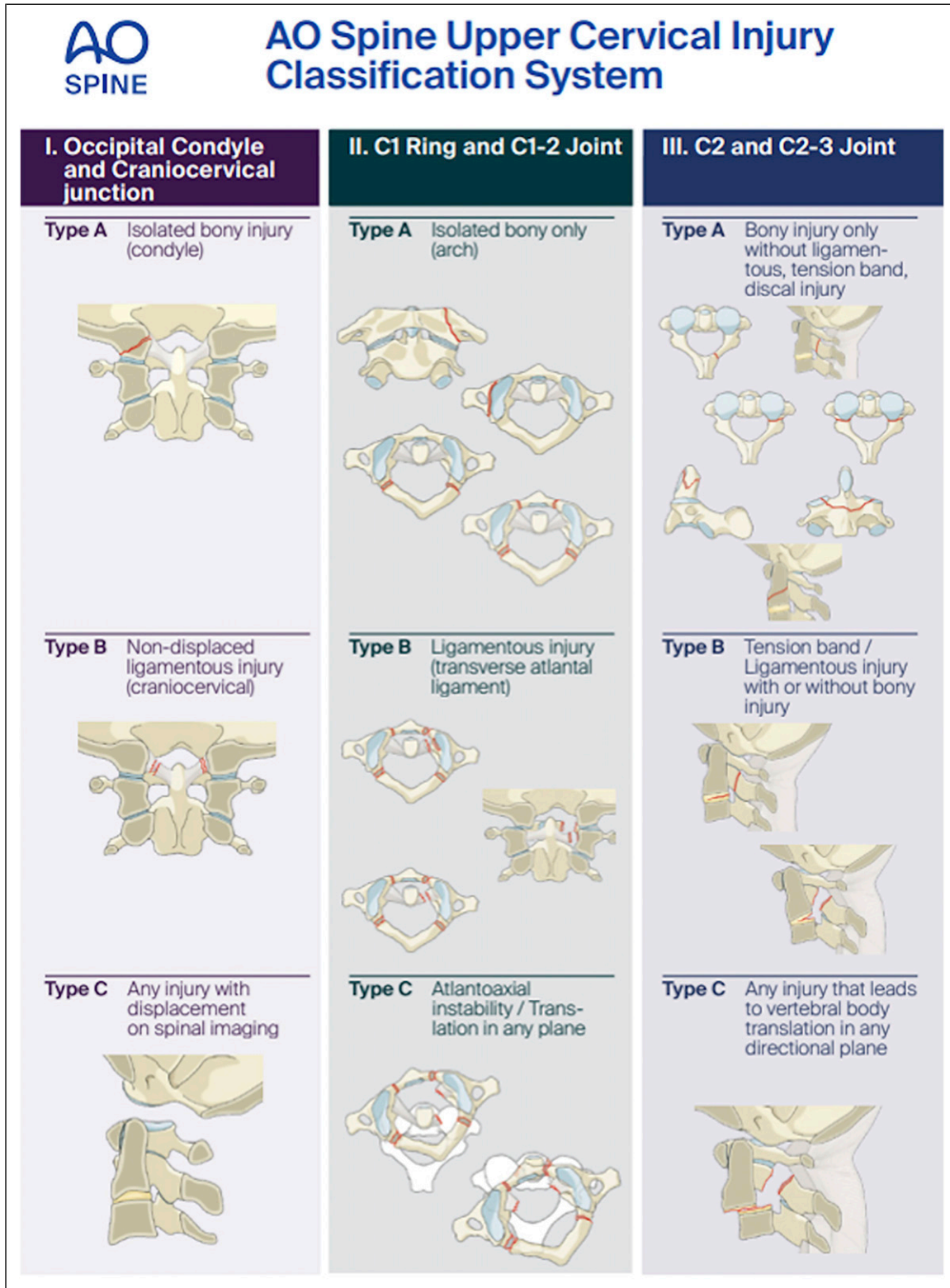


Figure 1. Pictorial demonstration of the AO Spine Upper Cervical Injury Classification. The classification is based on injury location (I, II, or III) and injury type (A, B, C). Permission to use this figure was granted by the AO Foundation®, AO Spine, Switzerland.

provided during the CT viewing, consisted of axial, sagittal, and coronal CT scans. Each video was played once at a rate of 2 frames/second. An online REDCap survey was then used to capture the study participants' classification answers. The second assessment, where cases were re-randomized and re-assessed, took place three weeks after the first assessment.

Gold Standard Committee

The original creators of the classification system (all members of the AO Spine Knowledge Forum Trauma) provided the official classification of each injury film. Each injury film reached unanimous agreement within the gold standard committee before the study participants evaluated the injury.

Statistics

Percent agreement between study participants and the gold-standard committee was tabulated for anatomic location (I, II, or III), injury type (A, B, or C), and the combination of anatomic location and injury type. Intergroup comparisons based on AO region (Africa, Asia, Central/South America, Europe, Middle East, and North America) were performed.

Kappa coefficients (κ) were calculated based on the agreement between different study participants (interobserver reliability) and the same participant after a three-week interval (intraobserver reproducibility). Interobserver reliability and intraobserver reproducibility were calculated for anatomical injury location, injury type, and overall classification. All of the reported kappa values utilized Fleiss' Kappa coefficient, which allows for missed ratings and comparisons between more than two validation members.²⁰ Interpretation of the reliability and reproducibility were based on the Landis and Koch convention, which categorized Kappa values as "slight" (<.2), "fair" (.2 - .4), "moderate" (.41 - .60), "substantial" (.61 - .8), and "excellent" (.81-1.0).²¹ Pearson's chi square test or Fisher's exact test was utilized to screen for potentially relevant associations.

Results

A total of 25 upper cervical spine injuries were evaluated by 275 participating international AO Spine members. The most common regions of participants were Europe (N=105, 38.2%) and Asia (N=73, 26.5%), while the regions with the lowest participation were the Middle East (N=20, 7.3%) and Africa (N=14, 5.1%) (Table 1). All injury types were evaluated at least twice with the exception of IB injuries, which were not evaluated due to their lack of inclusion in the AO Spine imaging database (Supplemental Appendix A).

Classification Accuracy

There were significant differences in classification accuracy when evaluating participants by AO global region. On the first assessment, participants from Europe (82.6%) and North

Table 1. Breakdown of AO Spine Members Who Participated in the AO Spine Upper Cervical Injury Classification including their AO Region, Subspecialty, and Surgical Experience.

| Survey Demographics | | N (%) |
|---------------------|---------------------------|------------|
| AO Region | # of participants | 275 (100) |
| | Africa | 14 (5.1) |
| | Asia | 73 (26.5) |
| | Central/South America | 36 (13.1) |
| | Europe | 105 (38.2) |
| | Middle East | 20 (7.3) |
| | North America | 27 (9.8) |
| Subspecialty | # of participants | 275 (100) |
| | Neurosurgery | 100 (36.4) |
| | Orthopaedic Spine Surgery | 168 (61.1) |
| | Other | 7 (2.5) |
| Surgical Experience | # of participants | 275 (100) |
| | <5 years | 71 (25.8) |
| | 5-10 years | 77 (28) |
| | 11-20 years | 82 (29.8) |
| | >20 years | 45 (16.4) |

America (82.5%) had the highest classification accuracy, while participants from Africa (70.1%) and Central and South America (72.0%) had the lowest classification accuracy ($P < .0001$). Similar findings were identified on assessment two with Europe (82.8%) and North America (81.7%) having the highest classification accuracy and Africa (70.9%) and Central and South America (72.2%) obtaining the lowest classification accuracy ($P < .0001$). When evaluating injury subtypes, there were significant differences for all injury subtypes evaluated on assessment one, but regional variation in classification accuracy were only identified for Type IC ($P = .027$), IIC ($P = .0378$), IIIA ($P = .0143$), and IIIB ($P < .0001$) on the second assessment (Table 2).

When stratifying by injury location alone and by injury type alone, significant regional variability was noted for both location ($P < .0001$) and injury type ($P < .0001$) on assessment one and two. In general, accurate classifications were more likely for injury location than injury type on both assessments with Type B injuries being the least likely to be correctly classified (Supplemental Appendix B).

Interobserver Reliability

On the first assessment, study participants from Africa ($\kappa = .487$), Central/South America ($\kappa = .513$), and the Middle East ($\kappa = .591$) achieved moderate reliability. North American ($\kappa = .673$), European ($\kappa = .682$), and Asian ($\kappa = .632$) participants achieved substantial reliability. On the second assessment only study participants from North America ($\kappa = .648$) and Europe ($\kappa = .681$) received substantial reliability grades. The remaining AO regions achieved moderate reliability. Large differences in reliability were especially notable for IIB, IIC, and IIIB injuries on assessments one and two (Table 3).

Identification of the injury location was excellent on assessment one for Asia ($\kappa = .859$), Europe ($\kappa = .902$), the

Table 2. Proportion and Percent of Correctly Classified Injuries Based On AO Spine region.

| First Assessment | | | | | | | |
|-----------------------|------------------------|------------------|-----------------------------|------------------|-------------------|---------------------|---------------------|
| Injury Classification | AO Region of the World | | | | | | P-value |
| | Africa P (%) | Asia P (%) | Central/South America P (%) | Europe P (%) | Middle East P (%) | North America P (%) | |
| Overall | 218/311 (70.1) | 1279/1603 (79.8) | 551/765 (72.0) | 2043/2472 (82.6) | 341/435 (78.4) | 504/611 (82.5) | <.0001* |
| IA | 16/25 (64.0) | 112/130 (86.2) | 47/61 (77.0) | 169/198 (85.4) | 28/34 (82.4) | 43/48 (89.6) | .0457* |
| IC | 19/26 (73.1) | 117/128 (91.4) | 57/62 (91.9) | 188/198 (94.9) | 32/35 (91.4) | 44/49 (89.8) | .0260* |
| IIA | 36/48 (75.0) | 193/257 (75.1) | 87/121 (71.9) | 329/394 (83.5) | 57/71 (80.3) | 81/96 (84.4) | .0221* |
| IIB | 28/51 (54.9) | 174/257 (67.7) | 85/124 (68.5) | 292/396 (73.7) | 47/70 (67.1) | 62/99 (62.6) | .0494* |
| IIC | 13/26 (50.0) | 89/128 (69.5) | 27/62 (43.5) | 127/198 (64.1) | 23/34 (67.6) | 36/50 (72.0) | .0056* |
| IIIA | 43/50 (86.0) | 210/256 (82.0) | 93/123 (75.6) | 346/396 (87.4) | 56/68 (82.4) | 89/99 (89.9) | .0188* |
| IIIB | 18/36 (50.0) | 139/192 (72.4) | 49/91 (53.8) | 220/297 (74.1) | 34/53 (64.2) | 54/73 (74) | .0005* |
| IIIC | 45/49 (91.8) | 245/255 (96.1) | 106/121 (87.6) | 372/395 (94.2) | 64/70 (91.4) | 95/97 (97.9) | .0129* |
| Second Assessment | | | | | | | |
| Injury Classification | AO Region of the World | | | | | P-value | |
| | Africa P (%) | Asia P (%) | Central/South America P (%) | Europe P (%) | Middle East P (%) | | North America P (%) |
| Overall | 222/313 (70.9) | 996/1305 (76.3) | 496/687 (72.2) | 1801/2176 (82.8) | 298/377 (79.0) | 446/546 (81.7) | <.0001* |
| IA | 18/26 (69.2) | 87/104 (83.7) | 41/55 (74.5) | 144/176 (81.8) | 24/30 (80.0) | 36/44 (81.8) | .5146 |
| IC | 20/25 (80.0) | 87/104 (83.7) | 48/55 (87.3) | 163/176 (92.6) | 28/30 (93.3) | 42/43 (97.7) | .0270* |
| IIA | 39/50 (78.0) | 164/209 (78.5) | 84/110 (76.4) | 295/344 (85.8) | 50/60 (83.3) | 76/88 (86.4) | .1002 |
| IIB | 30/50 (60.0) | 139/208 (66.8) | 66/107 (61.7) | 244/350 (69.7) | 43/62 (69.4) | 61/87 (70.1) | .5316 |
| IIC | 11/25 (44.0) | 63/105 (60.0) | 29/55 (52.7) | 117/174 (67.2) | 20/30 (66.7) | 33/43 (76.7) | .0378* |
| IIIA | 35/48 (72.9) | 164/209 (78.5) | 92/112 (82.1) | 306/346 (88.4) | 51/60 (85.0) | 73/87 (83.9) | .0143* |
| IIIB | 26/39 (66.7) | 112/156 (71.8) | 40/83 (48.2) | 213/264 (80.7) | 33/45 (73.3) | 51/66 (77.3) | <.0001* |
| IIIC | 43/50 (86.0) | 180/210 (85.7) | 96/110 (87.3) | 319/346 (92.2) | 49/60 (81.7) | 74/88 (84.1) | .0585 |

Abbreviations: Proportion (P) – total number of correct responses / total number of injury films evaluated.

*Indicates statistical significance with $P < .05$.

Middle East ($k = .805$), and North America ($k = .880$), whereas it was substantial for Africa ($k = .793$) and Central/South America ($k = .724$). The results were similar for assessment two, but Asia fell from excellent reliability ($k = .859$) to substantial ($k = .753$). When evaluating injury type, Africa ($k = .413$), Asia ($k = .580$), Central/South America ($k = .470$), and the Middle East ($k = .546$) had moderate reliability, while North America ($k = .651$) and Europe ($k = .642$) had substantial reliability. No group changed their overall Landis and Koch rating between the first and second assessment. Type B injuries had the lowest reliability on both assessment one and two with reliability ranging from slight to moderate depending on the AO region ([Supplemental Appendix C](#)).

Intragroup Reproducibility

When evaluating intraobserver reproducibility for the overall classification, members from Africa ($k = .64 \pm .19$) achieved similar scores compared to Asia ($k = .68 \pm .23$), Central/South

America ($k = .61 \pm .25$), Europe ($k = .74 \pm .14$), the Middle East ($k = .72 \pm .21$), and North America ($k = .75 \pm .15$). When evaluating by anatomical location, Africa also obtained similar reproducibility ($k = .83 \pm .16$) as Asia ($k = .86 \pm .25$), Central/South America ($k = .80 \pm .27$), Europe ($k = .92 \pm .11$), the Middle East ($k = .82 \pm .24$), and North America ($k = .92 \pm .11$). When evaluating reproducibility by injury type, similar trends were observed with Africa obtaining a reproducibility of ($k = .59 \pm .23$), which was comparable to Asia ($k = .63 \pm .28$), Central/South America ($k = .58 \pm .23$), Europe ($k = .71 \pm .16$), the Middle East ($k = .75 \pm .20$), and North America ($k = .70 \pm .18$).

These values resulted in Africa having 25% of study participants achieve excellent intraobserver reproducibility compared to 34.7% of participants from Asia, 24% from Central/South America, 44.7% from Europe, 42.9% from the Middle East, and 47.6% from North America. Although the differences in effect size between regions was large, due to the low absolute number of participants in some regions,

Table 3. Interobserver Reliability of the AO Spine Upper Cervical Injury Classification System Based on AO Spine Region.

| First Assessment | | | | | | |
|-----------------------|------------------------|----------|---------------------------|------------|-----------------|-------------------|
| Injury Classification | AO Region of the World | | | | | |
| | Africa (k) | Asia (k) | Central/South America (k) | Europe (k) | Middle East (k) | North America (k) |
| Overall | .487 | .632 | .513 | .682 | .591 | .673 |
| IA | .467 | .787 | .646 | .806 | .693 | .754 |
| IC | .615 | .850 | .814 | .904 | .842 | .885 |
| IIA | .438 | .551 | .531 | .660 | .550 | .583 |
| IIB | .322 | .445 | .400 | .546 | .434 | .463 |
| IIC | .327 | .466 | .268 | .492 | .446 | .468 |
| IIIA | .642 | .686 | .538 | .734 | .619 | .777 |
| IIIB | .296 | .559 | .301 | .604 | .426 | .649 |
| IIIC | .745 | .826 | .657 | .817 | .755 | .873 |

| Second Assessment | | | | | | |
|-----------------------|------------------------|----------|---------------------------|------------|-----------------|-------------------|
| Injury Classification | AO Region of the World | | | | | |
| | Africa (k) | Asia (k) | Central/South America (k) | Europe (k) | Middle East (k) | North America (k) |
| Overall | .491 | .566 | .511 | .681 | .599 | .648 |
| IA | .504 | .698 | .574 | .763 | .743 | .763 |
| IC | .726 | .732 | .776 | .903 | .893 | .948 |
| IIA | .473 | .569 | .505 | .653 | .571 | .620 |
| IIB | .331 | .420 | .344 | .495 | .417 | .496 |
| IIC | .315 | .417 | .314 | .540 | .461 | .551 |
| IIIA | .549 | .605 | .593 | .750 | .638 | .709 |
| IIIB | .390 | .485 | .322 | .636 | .525 | .604 |
| IIIC | .701 | .721 | .670 | .852 | .770 | .702 |

Table 4. Intraobserver reproducibility of the AO Spine Upper Cervical Injury Classification System based on the AO Region of the World.

| Level of Agreement | AO Region of the World | | | | | | P-value |
|------------------------|------------------------|---------------|---------------------------------------|-----------------|----------------------|------------------------|---------|
| | Africa N (%) | Asia N (%) | Central and South America N (%) | Europe N (%) | Middle East N (%) | North America N (%) | |
| Slight (<.2) | 0 | 3 (6.12) | 2 (8.0) | 0 | 0 | 0 | .0825 |
| Fair (.2-.40) | 2 (16.67) | 1 (2.04) | 2 (8.0) | 3 (3.53) | 2 (14.29) | 0 | .0845 |
| Moderate (>.40-.60) | 3 (25.0) | 7 (14.29) | 6 (24.0) | 11 (12.94) | 1 (7.14) | 4 (19.05) | .5797 |
| Substantial (>.60-.80) | 4 (33.33) | 21 (42.86) | 9 (36.0) | 33 (38.82) | 5 (35.71) | 7 (33.33) | .9717 |
| Excellent (>.80-1.0) | 3 (25.0) | 17 (34.69) | 6 (24.0) | 38 (44.71) | 6 (42.86) | 10 (47.62) | .3420 |

the significance of this finding is uncertain ($P = .342$) (Table 4).

Discussion

In order for an injury classification to receive widespread adoption, the classification should be comprehensible by an international group of surgeons with varying baseline demographics. To formally study whether the AO Spine Upper Cervical Injury Classification System met this threshold, we evaluated the impact of the surgeons' AO region and its effect

on the accuracy, interobserver reliability, and intraobserver reproducibility of the classification. In general, the classification demonstrated high accuracy, interobserver reliability, and intraobserver reproducibility, although there were some notable differences in the accuracy and reliability scores based on AO region. In particular, surgeons from Africa and Central/South America had lower accuracy and reliability scores, especially when evaluating moderate to severe injuries of the atlas (IIB and IIC) and atypical hangman's type fractures (IIIB injuries).

One important distinction that differentiates this classification from previous AO Spine classifications, especially the

AO Spine Thoracolumbar Injury Classification, is its reliance on CT scans instead of magnetic resonance imaging (MRI).²² Previous literature has demonstrated inequalities in the access to MRI, which are especially prevalent in low-income countries. While limited access to CT scans also exist, they are available at a greater rate than MRIs.²³ The designers of the classification understand that access to CT scans can be problematic in lower income countries; however, CT imaging for high-energy trauma, especially spinal trauma has become standard of care.²⁴ Technological advances do appear to be on the horizon (digitalization of imaging), which may help equalize the current inequality gaps surrounding access to imaging, but this is unlikely to occur in the next decade.²² The lower rates of advanced imaging (both CT and MRI) in Central/Latin America and Africa may result in classification accuracy rates lower than the remaining AO regions of the world.^{5,6} Although speculative, we hypothesize these AO regions may be more reliant on plain radiographs to identify spine trauma. Thus, even though CT scans are more ubiquitous than MRIs, they may still be underutilized in some regions of the world due to accessibility issues. In hindsight, a survey question asking each participant whether radiographs, CT, MRI, or a combination of imaging modalities was commonly used in their practice when identifying upper cervical spine injuries would have been highly instructive and valuable to identify potential gaps in imaging accessibility.

Hopefully, global access to imaging improves and CT and MRI scans become more prevalent in lower income countries, which may allow physicians to become familiar with classifying upper cervical spine injuries based on CT and MRI scans. In this event, incorporation of MRI images with CT scans for suspected upper cervical spine injuries is advantageous since MRI identifies injuries in 12% of obtunded patients with normal CT scans, albeit only 6% of patients have treatment plan alterations when an MRI is available in addition to a CT scan.⁷ It is worth noting that although the validation was performed without MRI images, the ability to supplement CT scans with MRI images is already built into the classification system. In fact, the addition of MRI would likely improve the classification reliability of type B injuries, which can be identified by ligamentous disruption (MRI only) or tension band failure (MRI and/or CT scan).

One notable area for potentially improving the accuracy and reliability of this classification is through global education. In this regard, this study has provided a platform for bringing awareness to the challenges in identifying potential mechanisms and modes of failure of moderate to severe atlas injuries (type IIB and IIC) and atypical hangman equivalent injuries (IIIB). However, accurate classification and identification of these injuries is not a new problem given the relative rarity of these injury types. The Gehweiler classification was recently evaluated and had a similar moderate overall interobserver reliability ($k = .50$).²⁴ While the AO Spine Upper Cervical Injury Classification System had similar interobserver reliability, it may be simpler to use, while also unifying

the entire upper cervical spine into a singular classification system. The complexity of atypical hangman fracture variants is also well described with the mode of vertebral body fracture dependent on the injury mechanism.⁸ Given the number of fracture variants present in IIIB injuries, an exhaustive validation of the different fracture patterns is not possible, nor is it likely necessary, since these injury types are predominantly tension band failures with questionable stability. The crux of the AO Spine Upper Cervical Injury Classification System is built on identifying whether an injury is stable or not.²⁵ Purely bony injuries (type A) can commonly be treated with non-surgical management (unless the injury is a modifier – for example, a dens fracture at the watershed line), injuries with questionable stability (tension band failures or ligament disruptions/avulsions) may be treated either nonsurgical or operative, while subluxation or dislocation of a vertebrae is treated with reduction and operative stabilization or Halo-immobilization.²⁶ Widespread acceptance of this hierarchical injury progression may aid future validation attempts across AO global regions.

As previously discussed, classification validations should be performed in a stepwise fashion.¹ The relatively high accuracy and interobserver reliability of this classification, which was performed on an international level, allows the classification to progress to injury severity score establishment. This process is dependent on the hierarchical progression of injury types (A, B, and C).^{2,26} Once injury severity scores are established, a treatment algorithm may be proposed and validated through a modified Delphi approach. This process will allow surgeons to grade the severity of an upper cervical spine injury and then determine appropriate management of the injury (operative or conservative management).³

There were multiple limitations present throughout this study, which require further discussion. First, multiple different surgeon demographics may have affected the accuracy and reliability of a surgeons' classification including subspecialty, practice location, work setting, trauma center involvement, and familiarity with reading CT scans. These demographic factors may overlap and provide some study confounding, which could not be completely accounted for. Second, study participants were comprised of AO Spine members who may understand the principles of AO classifications better than non-AO Spine members. Third, only injury scans uploaded to the de-identified AO Spine database were evaluated. Thus, no type IB injuries were examined and the reliability and accuracy of classifying these injuries is unknown.

Conclusion

The AO Spine Upper Cervical Injury Classification System found that, in general, the classification can be applied with high accuracy, interobserver reliability, and intraobserver reproducibility. However, differences in classification accuracy and reliability are present based on the surgeons AO

region. Surgeons from Africa and Central/South America have lower accuracy and reliability classification scores on the overall classification, but the differences in the reliability and accuracy were most notable when classifying moderate to severe atlas injuries (IIB and IIC) and atypical hangman's type fractures (IIIB injuries). Additional education on properly classifying these rare and severe injury types is warranted to improve classification accuracy and reliability.

Acknowledgments

The authors of the manuscript would like to thank Olesja Hazenbiller for her assistance in developing the methodology and providing support during the validation. We would also like to thank Hans Bauer, senior biostatistician at Staburo GmbH for his assistance with the statistical analysis.









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

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>
 Richard Bransford  <https://orcid.org/0000-0003-2293-5774>
 Rajasekaran Shanmuganathan  <https://orcid.org/0000-0001-6043-006X>
 Andrei F. Joaquim  <https://orcid.org/0000-0003-2645-0483>
 Emiliano Vialle  <https://orcid.org/0000-0003-1157-4889>
 Mohammad El-Sharkawi  <https://orcid.org/0000-0001-6177-7145>

Supplemental Material

Supplemental material for this article is available online.

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.
- Kepler CK, Vaccaro AR, Schroeder GD, et al. The Thoracolumbar AOSpine Injury Score. *Global Spine J*. 2016;6(4): 329-334. doi:10.1055/s-0035-1563610.
- Vaccaro AR, Schroeder GD, Kepler CK, et al. The surgical algorithm for the AOSpine thoracolumbar spine injury classification system. *Eur Spine J*. 2016;25(4):1087-1094. doi:10.1007/s00586-015-3982-2.
- Vaccaro AR, Schroeder GD, Divi SN, et al. Description and Reliability of the AOSpine Sacral Classification System. *J Bone Joint Surg Am*. 2020;102(16):1454-1463. doi:10.2106/JBJS.19.01153.
- Ogbole GI, Adeyomoye AO, Badu-Peprah A, Mensah Y, Nzeh DA. Survey of magnetic resonance imaging availability in West Africa. *Pan Afr Med J*. 2018;30:240. doi:10.11604/pamj.2018.30.240.14000.. Published.
- Volpi G. Radiography of diagnostic imaging in Latin America. *Nucl Med Biomed Imaging*. 2016;1:10-12.
- Schoenfeld AJ, Bono CM, McGuire KJ, et al. Computed tomography alone versus computed tomography and magnetic resonance imaging in the identification of occult injuries to the cervical spine: a meta-analysis. *J Trauma*. 2010;68(1):109-113. PMID: 20065765. doi:10.1097/TA.0b013e3181c0b67a.
- Benzel EC, Hart BL, Ball PA, Baldwin NG, Orrison WW, Espinosa M. Fractures of the C-2 vertebral body. *J Neurosurg*. 1994;81(2):206-212. doi:10.3171/jns.1994.81.2.0206.
- Jefferson G. Fracture of the atlas vertebra: report of four cases, and a review of those previously recorded. *Br J Surg Lond*. 1920;7:407-422.
- Anderson PA, Montesano PX. Morphology and treatment of occipital condyle fractures. *Spine (Phila Pa 1976)*. 1988;13: 731-736.
- Tuli S, Tator CH, Fehlings MG, Mackay M. Occipital condyle fractures. *Neurosurgery*. 1997;41:368-376.
- Traynelis VC, Marano GD, Dunker RO, Kaufman HH. Traumatic atlanto-occipital dislocation. Case report. *J Neurosurg*. 1986;65:863-870. doi:10.3171/jns.1986.65.6.0863.
- Bellabarba C, Mirza SK, West GA, et al. Diagnosis and treatment of craniocervical dislocation in a series of 17 consecutive survivors during an 8-year period. *J Neurosurg Spine*. 2006;4: 429-440. doi:10.3171/spi.2006.4.6.429.
- Dickman CA, Greene KA, Sonntag VK. Injuries involving the transverse atlantal ligament: classification and treatment guidelines based upon experience with 39 injuries. *Neurosurgery*. 1996;38(1):44-50. doi:10.1097/00006123-199601000-00012.
- Anderson LD, D'Alonzo RT. Fractures of the odontoid process of the axis. *J Bone Joint Surg Am*. 1974;56:1663-1674.
- Effendi B, Roy D, Cornish B, Dussault RG, Laurin CA. Fractures of the ring of the axis. A classification based on the analysis of 131 cases. *J Bone Joint Surg Br*. 1981;63-B:319-327.
- Levine AM, Edwards CC. The management of traumatic spondylolisthesis of the axis. *J Bone Joint Surg Am*. 1985;67: 217-226.
- White AA 3rd, Johnson RM, Panjabi MM, Southwick WO. Biomechanical analysis of clinical stability in the cervical spine. *Clin Orthop Relat Res*. 1975;109:85-96. doi:10.1097/00003086-197506000-00011.
- Lambrechts MJ, Schroeder GD, Karamian BA, et al. Development of Online Technique for International Validation of the AO Spine Subaxial Injury Classification System. *Global Spine J*.

- 2022:21925682221098967. Online ahead of print. doi:[10.1177/21925682221098967](https://doi.org/10.1177/21925682221098967).
20. Gwet KL. Intrarater reliability. In: RB D'Agostino, L Sullivan, J Massaro, eds. *Wiley encyclopedia of clinical trials*. Hoboken, NJ: John Wiley & Sons, Inc.; 2008:473-485.
 21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
 22. Hricak H, Abdel-Wahab M, Atun R, et al. Medical imaging and nuclear medicine: a Lancet Oncology Commission. *Lancet Oncol*. 2021;22(4):e136-e172. doi:[10.1016/S1470-2045\(20\)30751-8](https://doi.org/10.1016/S1470-2045(20)30751-8).
 23. Holmes JF, Akkinapalli R. Computed tomography versus plain radiography to screen for cervical spine injury: a meta-analysis. *J Trauma*. 2005;58(5):902-905. doi:[10.1097/01.ta.0000162138.36519.2a](https://doi.org/10.1097/01.ta.0000162138.36519.2a).
 24. Laubach M, Pishnamaz M, Scholz M, et al. Interobserver reliability of the Gehweiler classification and treatment strategies of isolated atlas fractures: an internet-based multicenter survey among spine surgeons. *Eur J Trauma Emerg Surg*. 2020;48(1):601-611. doi:[10.1007/s00068-020-01494-y](https://doi.org/10.1007/s00068-020-01494-y).
 25. Vaccaro AR, Karamian BA, Levy HA, et al. Update on Upper Cervical Injury Classifications: The New AO Upper Cervical Spine Classification System. *Clin Spine Surg*. 2021. Epub ahead of print. doi:[10.1097/BSD.0000000000001215](https://doi.org/10.1097/BSD.0000000000001215).
 26. Schroeder GD, Canseco JA, Patel PD, et al. Establishing the Injury Severity of Subaxial Cervical Spine Trauma: Validating the Hierarchical Nature of the AO Spine Subaxial Cervical Spine Injury Classification System. *Spine (Phila Pa 1976)*. 2021;46(10):649-657. doi:[10.1097/BRS.0000000000003873](https://doi.org/10.1097/BRS.0000000000003873).