

# The Comprehensive AO CMF Classification System for Mandibular Fractures: A Multicenter Validation Study

Paul A. Mittermiller, MD<sup>1</sup> Serena S. Bidwell, MPH<sup>1</sup> Florian M. Thieringer, MD, DDS<sup>2</sup>  
 Carl-Peter Cornelius, MD, DDS, PhD<sup>3</sup> Amber W. Trickey, PhD, MS, CPH<sup>1</sup> Risto Kontio, MD, DDS, PhD<sup>4</sup>  
 Sabine Girod, MD, DDS, PhD<sup>1</sup> and the AO Trauma Classification Study Group\*

<sup>1</sup>Stanford-Surgery Policy Improvement Research and Education (S-SPIRE) Center, Department of Surgery, Stanford University, Stanford, California

<sup>2</sup>Department of Cranio-Maxillofacial Surgery, University Hospital of Basel, Basel, Basel-Stadt, Switzerland

<sup>3</sup>Department of Oral and Maxillofacial Surgery and Facial Plastic Surgery, Ludwig-Maximilian-University Munich, Munich, Germany

<sup>4</sup>Department of Oral and Maxillofacial Surgery, Helsinki University Hospital, Helsinki, Finland

Address for correspondence Sabine Girod, MD, DDS, PhD, Department of Surgery, 770 Welch Road, #400, Palo Alto, CA 94304 (e-mail: sgirod@stanford.edu).

Craniomaxillofac Trauma Reconstruction 2019;12:254–265

## Abstract

The AO CMF has recently launched the first comprehensive classification system for craniomaxillofacial (CMF) fractures. The AO CMF classification system uses a hierarchical framework with three levels of growing complexity (levels 1, 2, and 3). Level 1 of the system identifies the presence of fractures in four anatomic areas (mandible, midface, skull base, and cranial vault). Level 2 variables describe the location of the fractures within those defined areas. Level 3 variables describe details of fracture morphology such as fragmentation, displacement, and dislocation. This multiplanar radiographic image-based AO CMF trauma classification system is constantly evolving and beginning to enter worldwide application. A validation of the system is mandatory prior to a reliable communication and data processing in clinical and research environments. This interobserver reliability and accuracy study is aiming to validate the three current modules of the AO CMF classification system for mandible trauma in adults. To assess the performance of the system at the different precision levels, it focuses on the fracture location within the mandibular regions and condylar process subregions as core components giving only secondary attention to morphologic variables. A total of 15 subjects individually assigned the location and features of mandibular fractures in 200 CT scans using the AO CMF classification system. The results of these ratings were then statistically evaluated for interobserver reliability by Fleiss' kappa and accuracy by percentage agreement with an experienced reference assessor. The scores were used to determine if the variables of levels 2 and 3 were appropriate tools for valid

## Keywords

- ▶ mandibular fractures
- ▶ classification
- ▶ midface
- ▶ fractures
- ▶ computed tomography
- ▶ validation
- ▶ interobserver reliability

\* AO Trauma Classification Study Group: Dana Johns, MD, Florian Probst, MD, DMD, JiaQiao, MD, Johanna Snall, MD, DDS, PhD, Karri Mesimaki, MD, DDS, Moritz Bader, MD, DDS, Philipp Goetz, MD, DMD, Stacey Moon, DDS, Tommy Wilkmann, MD, DDS, Wenko Smolka, MD, DDS, PHD.

received  
 September 14, 2018  
 accepted  
 November 4, 2018  
 published online  
 January 31, 2019

Copyright © 2019 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.  
 Tel: +1(212) 584-4662.

DOI <https://doi.org/10.1055/s-0038-1677459>.  
 ISSN 1943-3875.

classification. Interobserver reliability and accuracy were compared by hierarchy of variables (level 2 vs. level 3), by anatomical region and subregion, and by assessor experience level using Kruskal–Wallis and Wilcoxon’s rank-sum tests. The AO CMF classification system was determined to be reliable and accurate for classifying mandibular fractures for most levels 2 and 3 variables. Level 2 variables had significantly higher interobserver reliability than level 3 variables (median kappa: 0.69 vs. 0.59,  $p < 0.001$ ) as well as higher accuracy (median agreement: 94 vs. 91%,  $p < 0.001$ ). Accuracy was adequate for most variables, but lower reliability was observed for condylar head fractures, fragmentation of condylar neck fractures, displacement types and direction of the condylar process overall, as well as the condylar neck and base fractures. Assessors with more clinical experience demonstrated higher reliability (median kappa high experience 0.66 vs. medium 0.59 vs. low 0.48,  $p < 0.001$ ). Assessors with experience using the classification software also had higher reliability than their less experienced counterparts (median kappa: 0.76 vs. 0.57,  $p < 0.001$ ). At present, the AO CMF classification system for mandibular fractures is suited for both clinical and research settings for level 2 variables. Accuracy and reliability decrease for level 3 variables specifically concerning fractures and displacement of condylar process fractures. This will require further investigation into why these fractures were characterized unreliably, which would guide modifications of the system and future instructions for its usage.

Trauma classification systems are essential for providing reliable and reproducible documentation of fracture patterns and their extent. Appropriate classification systems may allow for more effective clinical communication and support for decision making when formulating treatment plans. Trauma often requires healthcare providers from different specialties to work together during a patient’s treatment course. This stresses the need for a common language to facilitate professional exchange.

Currently, there is a multitude of existing classification systems for mandibular trauma. These systems can vary with regard to how they define topographic mandibular regions and often lack clear definitions and details. Some of these inconsistencies may arise as the result of historical limitations in imaging, which have improved with the development of modern cross-sectional imaging techniques.<sup>1</sup>

Due to these drawbacks, the Arbeitsgemeinschaft für Osteosynthesefragen (AO) developed a new and comprehensive craniomaxillofacial (CMF) trauma classification system for adult craniofacial trauma.<sup>2</sup> This AO CMF classification system is multispecialty in scope (plastic and reconstructive surgery, otorhinolaryngology, oral and maxillofacial surgery, and neurosurgery) and surveys the cranial vault, skull base, midface, and mandible in a total of four anatomic modules.<sup>2</sup>

To establish a mainstream trauma classification system with standardized rules and conventions that is universally employed by the global medical community, it is fundamental to collect and stratify data according to comparable categories for subsequent evaluation on pertinent criteria.<sup>3</sup> As with all modern fracture classification systems, a distinct methodologic approach is crucial to come up with a scientifically sound validation.<sup>4</sup> In iterative cycles, this classification design was refined until it reached robust performance in terms of accuracy, reliability, and reproducibility.<sup>2</sup>

The current AO CMF classification system for mandible fractures has been developed through several revisions.<sup>1,5,6</sup> The developmental process involved international expert groups of variable size (4–18 individuals) and background (CMF surgery, radiology, basic science, applied biostatistics) and started with a series of pilot agreement studies that resulted in a preliminary proposition of fracture classification.<sup>7</sup> The kappa statistic ( $k$ ) measuring the chance-corrected proportional interobserver agreement of this first-generation scheme as well as of a successor model persistently indicated shortfalls in the acceptable strength of agreement within internal follow-up studies. A detailed analysis of the raw data identified an overzealous complexity of the proposed model, which made an attempt to comply with the tripartition fracture severity concept advocated in the AO long bone fracture classification system.<sup>8</sup>

Instead of adhering to an overly complex system, the current AO CMF trauma classification system aims to create a workable solution in the form of three hierarchical precision levels (elementary, basic, and focused) which represents a scale of increasing complexity. Notably the mandible fracture classification was reconfigured under almost ideal circumstances. Expert groups were primarily involved in the definition and redesign of the schemes. They were therefore highly cognizant of the options available when creating the classification system and of the limitations of the system.

To ease application of the classification system, the internal developmental phase included focusing on an updated software package, the AO Comprehensive Injury Automatic Classifier (AOCOIAC) version 4.0 (AO Foundation, Dübendorf, Switzerland; [www.aofoundation.org/aocoiac](http://www.aofoundation.org/aocoiac)). This software allows for straightforward documentation and easy fracture coding.

The comprehensive AO CMF trauma classification system for adults, approved and propagated by the AO, is presently on its way into more widespread use. However, there is the foremost need to conduct second phase validation studies in terms of interobserver reliability and accuracy that replicate realistic clinical encounters. In other words, surgeons in different stages of training and experience who may use the classification schemes must try the classification software. The goal of any injury classification is to create a common language to serve as the basis for communicating between healthcare providers and for evaluating treatments and their outcomes to assist with future clinical decisions. To that end, the AO CMF classification needs to be tested to demonstrate its validity. The purpose of this study was to evaluate the interobserver reliability and accuracy of the AO CMF trauma classification system and to investigate relationships between scoring reliability and rater experience level.

## Methods

### Imaging Case Series Database

To test the AO CMF trauma classification system, a database of 200 consecutive de-identified computed tomographic (CT) scans of mandibular fractures was created using the Stanford Translational Research Environment (STRIDE).<sup>9</sup> The de-identification process was presented to the Privacy Office and received approval. The Stanford University Institutional Review Board deemed the study exempt from review because all identifiable patient health information was removed. Inclusion criteria for the database were as follows: (1) patient older than 18 years, (2) patient sustained a mandible fracture, and (3) available pretreatment CT (helical or cone beam).

Using the Cohort Discovery Tool within STRIDE, an initial search revealed 450 cases since 2010 that met the inclusion criteria. CT scans from patients were retrospectively added to the database and screened by one surgeon (S.G.) to confirm that they met inclusion criteria. This was performed until a cohort of 200 consecutive cases was assembled. This cohort was representative of all fracture types and locations. For each case, an image folder was created that contained de-identified three-dimensional reconstructions of the radiographic imaging data and any additional CT images relevant to the fracture. The reconstructions were created from the DICOM (Digital Imaging and Communications in Medicine) data at the Department of Radiology, Stanford University (Stanford, CA). The folders were then shared with surgeons at four CMF surgery centers (Stanford University; Universitäts-spital Basel, Switzerland; Ludwig Maximilian University, Munich, Germany; and Helsinki University Hospital, Finland). Assessors at each site classified the 200 CT scans using the AO AOCIOAC software, version 4.0.<sup>10</sup> The assessors were given a manual (Craniomaxillofacial Fracture Classification Module User Manual version 4.0.0) to support their understanding of the classification software and were allowed to complete the classifications in as much time as needed. Additional images were provided upon request. The 200 fracture cases were evaluated by 15 assessors, resulting in a total of 3,000 assessments of mandibular fracture patterns.

### Overview of Variables

The AO CMF classification modules for mandible fractures have been described previously.<sup>1,5,6</sup> They are based on a system with three levels of increasing detail. Level 1 identifies the fracture within one of four regions (mandible, mid-face, skull base, and cranial vault). For mandible fractures, level 2 variables describe the location of the fracture within the mandibular regions. Level 3 variables then describe details about the fracture morphology, including fragmentation, displacement, and dislocation. A brief synopsis of the categories and fracture variables is shown in ▶Table 1.

It is most important to first distinguish the location of a fracture in one or multiple anatomic regions or subregions prior to determining the morphologic properties of each fracture. Level 2 classification involves defining a fracture within nine previously defined topographical regions (▶Fig. 1a).<sup>1</sup> A total of four “transitional zones” are interposed between the mandibular regions and form corridors approximately the width of the canine or the third molar. The transitional zones allow for the clear-cut allocation of fracture lines entering into them or passing through them into adjacent mandibular regions. A few specific rules have been defined that allow a fracture to be categorized as either “confined” to a single region or “not confined” to a single region, meaning the fracture extends over at least two adjacent regions (e.g., the symphysis, the right or left body, and the right or left angle and ramus; ▶Fig. 1a).

Level 3 classification of mandible fractures in the non-condylar regions of the mandible involves evaluating tooth injuries, periodontal trauma, involvement of the alveolar process, fracture fragmentation severity (none, minor, major), and determining whether there is bone loss.<sup>5</sup>

A particular level 3 classification applies to fractures within the condylar process (CP).<sup>6</sup> Condylar fractures are allocated to one of three subregions: the condylar head (CH), the condylar neck (CN), or the condylar base (CB). The borders are defined by three horizontally arranged reference lines (▶Fig. 1b).<sup>6</sup> A CH fracture involves the area superior to the CH reference line. A CN fracture is affirmed if more than one-third of the fracture is higher than the sigmoid notch line. Finally, a CB fracture corresponds to a fracture line where more than two-thirds of its courses extend below the sigmoid notch line and the fracture exits posteriorly above the masseteric notch line (▶Fig. 1b). Fractures within the CH are further described in relation to the lateral condylar pole zone (▶Fig. 1c). Fractures are medial to the pole zone (m-) only if all fracture lines pass medial to the pole zone. These differ from pole zone fractures (p-), which include fractures that run within or lateral to the pole zone. If a p-fracture is present, it is the preponderant feature for classifying the fracture and therefore a concomitant m-fracture is simply considered a fragmentation variable (▶Fig. 1c).

Level 3 variables to define CP fracture morphology include variables for fragmentation and displacement. With regard to fracture fragmentation, the fracture is defined as having none, minor, or major fragmentation of the CH, CN, or CB.

Level 3 variables that focus on fracture displacement describe features of the CP overall in addition to features within its subregions (CH, CN, CB).

**Table 1** AO classification system for mandible fractures—overview of its variables

Noncondylar regions	Condylar process		Condylar process subregions	
Angle/Ramus Body Symphysis Coronoid			Condylar head	Condylar neck and condylar base
<b>Fracture location</b>				
0 = not fractured 1 = fractured 2 = undetermined	0 = not fractured 1 = fractured 2 = undetermined		0 = not fractured 1 = fractured 2 = undetermined	0 = not fractured 1 = fractured 2 = undetermined
			m = medial to pole zone p = within or lateral to pole zone	
<b>Fracture morphology</b>				
<b>Fragmentation</b>				
0 = none 1 = minor 2 = major			0 = none 1 = minor 2 = major	0 = none 1 = minor 2 = major
<b>Displacement</b>				
	Condylar head/cranial fragment related to articular fossa		Vertical apposition/fragment contact	Sideward displacement
	0 = no displacement 1 = displacement 2 = dislocation	Direction: l = lateral m = medial a = anterior p = posterior	0 = complete 1 = partial 2 = lost	0 = none 1 = partial 2 = full
	Ramus end/caudal fragment related to fossa			Direction: l = lateral m = medial a = anterior p = posterior
	0 = no displacement 1 = displacement	Direction: l = lateral m = medial a = anterior p = posterior		Angulation
	Distortion of condylar head			0 = none (< 5) 1 = ≥ 5–45 2 = > 45
	0 = orthotopic 1 = dystopic			Direction: l = lateral m = medial a = anterior p = posterior
	Overall loss of vertical ramus height			
	0 = no change of height 1 = loss of height 2 = increase of height			

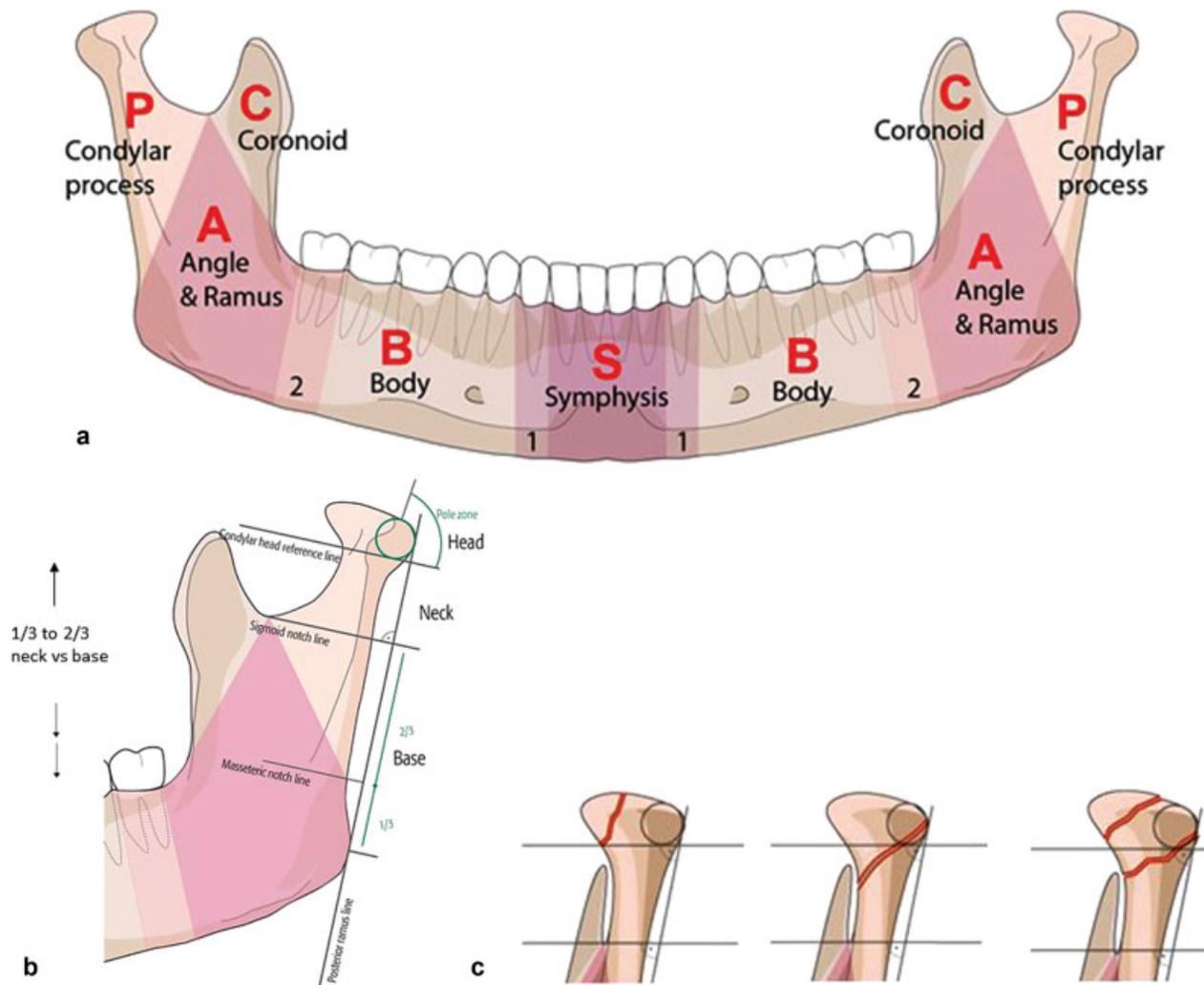
Aspects concerning the overall CP fragment refer to the displacement/dislocation of the CH in relation to the fossa, as well as the displacement of the ramus or caudal fragment end in relation to the fossa. Moreover, the distortion of the condyle bearing fragment, and the change of the vertical ramus height, is described. In CH fractures, the displacement is characterized by the vertical apposition of the medial fragment.

CN and CB fractures are detailed in terms of sideward displacement (degree and direction) and angulation (degree and direction; ► **Table 1**).

**Statistical Analysis**

The classification software automatically saved input from each assessor. Results were aggregated in a REDCap database and then imported to the statistical software RStudio version

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.



**Fig. 1** (a) Panoramic view of the mandible with a full set of permanent teeth. The mandible is divided into nine topographical regions with the symphysis (S) in the anterior central position. All other regions, including the body (B), angle/ramus (A), coronoid (C), and condylar process (P) are symmetric. Two pairs of transitional zones are assembled between the S and B (1 = anterior transition zone) and between the B and A (2 = posterior transition zone). (b) Three reference lines are oriented perpendicular to the posterior ramus to define condylar process subregions. The condylar head reference line is a tangent line caudal to a sphere around the lateral pole zone and separates the condylar head (CH) from the condylar neck (CN). The sigmoid notch line runs through the deepest point of the sigmoid notch and separates the condylar neck (CN) from the condylar base (CB). The masseteric notch line is located one-third of the distance from the most prominent point of the posterior border of masseteric tuberosity to the sigmoid notch line and this line defines the inferior extent of the condylar base. (c) Fractures of the condylar head are defined based on whether the fracture line courses within the pole (p-fracture) or medial to the pole (m-fracture). A combination of fractures that includes a pole fracture and fracture medial to the pole are described as a p-fracture with fragmentation. Adapted from CMTR 2014;7(S1).<sup>1,6</sup>

1.0.153 (RStudio Team 2016. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA; URL <http://www.rstudio.com/>) for further analysis. Of the 172 variables collected during the classification process, 86 were used for analysis. The first 86 variables asked assessors to classify fractures located within any of the nine topographical regions (CP right and left, coronoid right and left, angle/ramus right and left, body right and left, and symphysis; ▶ **Fig. 1**). Assessments specific to dentition, edentulousness, bone atrophy, and alveolar process fractures had low frequencies of occurrence and limited data; therefore, these measures were excluded from the final analyses.

Fleiss' kappa coefficients were used to evaluate interobserver reliability among the fifteen assessors for each variable of the classification software.<sup>11</sup> Kappa coefficients

compute the degree of agreement between all assessors that exceeds agreement due to chance alone. One kappa coefficient was calculated for each of the 86 evaluated variables. The authors evaluated reliability as follows: <0 as indicating no agreement, 0–0.20 as slight agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1.0 as almost perfect agreement.<sup>12</sup>

Accuracy was measured by comparing each of the fourteen assessors to the one assessor who had the most experience using the classification software (6 years) and the highest level of clinical experience treating mandibular fractures (>100 fractures). The percentage agreement between each assessor and the reference assessor was calculated for every variable across all 200 cases. Accuracy was

measured for each assessor. Therefore, 14 values of agreement with the reference assessor were calculated per variable and then averaged for each variable.

Interobserver reliability and accuracy were compared by hierarchy of variables (level 2 vs. level 3), by anatomical region and subregion (within the CP—CH, CN, and CB), and by assessor experience level (previous experience with the classification system and clinical experience). Level 2 variables represent more basic variables within the CMF classification, such as assessment of fracture location in each of the nine anatomical regions, while level 3 variables are more complex. Kruskal–Wallis and Wilcoxon’s rank-sum tests were used to evaluate differences by group within each of these comparisons. The Kruskal–Wallis test is a nonparametric method for analyzing rank-order differences between three or more groups of an independent variable. The Wilcoxon rank-sum test is similar to the Kruskal–Wallis test but is intended for comparing population mean-rank differences between only two groups. This same method of evaluating interobserver reliability and accuracy was used for all other variables in the study (e.g., location of fracture line in CH, fragmentation in condylar subregions, CP displacement).

To determine if there were any differences by anatomic laterality, the kappa coefficients from the left side of the mandible were compared with the kappa coefficients from the right side of the mandible using Wilcoxon’s rank-sum tests.

The 15 assessors had various levels of experience both in treating mandibular fractures and in using the CMF fracture classification system. We compared interobserver reliability and accuracy of the assessors by levels of clinical experience and AO CMF fracture classification experience. The assessors were divided into three groups based on the number of treated mandibular fractures (low <50, mid 50–100, high >100). A Kruskal–Wallis test was used to examine differences in reliability and accuracy according to treatment experience. Only three of the fifteen assessors had prior experience with the classification system software, while the remaining twelve had not used the software previously. Comparisons of reliability and accuracy of those with and without prior classification experience were evaluated by Wilcoxon’s rank-sum test.

## Results

All 200 cases were evaluated by each of 15 assessors. Fracture location and morphology, specifically fragmentation and displacement, were key features of the analysis. There were 14 basic order variables (level 2) and 72 variables that asked for more detailed and difficult to define fracture information (level 3). The level 2 variables had significantly higher interobserver reliability than level 3 variables (median kappa: 0.69 vs. 0.59,  $p < 0.001$ ). Accuracy was also significantly higher among the level 2 variables compared with level 3 (median agreement: 94 vs. 91%,  $p < 0.001$ ).

### Fracture Location within all Mandible Regions Is Reliably Defined

The level 2 variables indicated fractures at each of the anatomical regions. Interobserver reliability was substantial at each of the level 2 fracture locations, with the highest reliability for identification of a fracture in the CP (→Fig. 2). Identification of fractures in the CP had significantly higher interobserver reliability than fractures in the noncondylar regions (0.83 vs. 0.69,  $p = 0.04$ ). Accuracy for every anatomical region was greater than 50%, with the highest accuracy for identification of a fracture in the coronoid (86%).

Fracture locations were also identified within the three subregions of the CP—head (CH), neck (CN), and base (CB) (→Fig. 3a). The interobserver reliabilities of these fracture locations were also substantial (all  $k \geq 0.73$ ). The reliability of fracture identification in the head was the highest (left and right  $k = 0.82$ ), followed by the base (left  $k = 0.78$ , right  $k = 0.79$ ). The neck had the lowest reliability, though still acceptably high (left  $k = 0.73$ , right  $k = 0.75$ ). Accuracy for the CP subregion locations was moderate, ranging from 57 to 71%. There was no significant difference in the reliability or accuracy of fracture location variables by CP subregion (reliability  $p = 0.10$ , accuracy  $p = 0.71$ ).

Within the CH, the location of the fracture line was further delineated with the option to specify if the course was medial to the pole zone or within or lateral to the pole zone on both the right and left sides (→Fig. 3b). Interobserver reliability measures were lower than the previous location variables in CH, CN, and CB, ranging from 0.38 to 0.59 or fair to moderate, but accuracy measures were quite high, 84 to 92%.

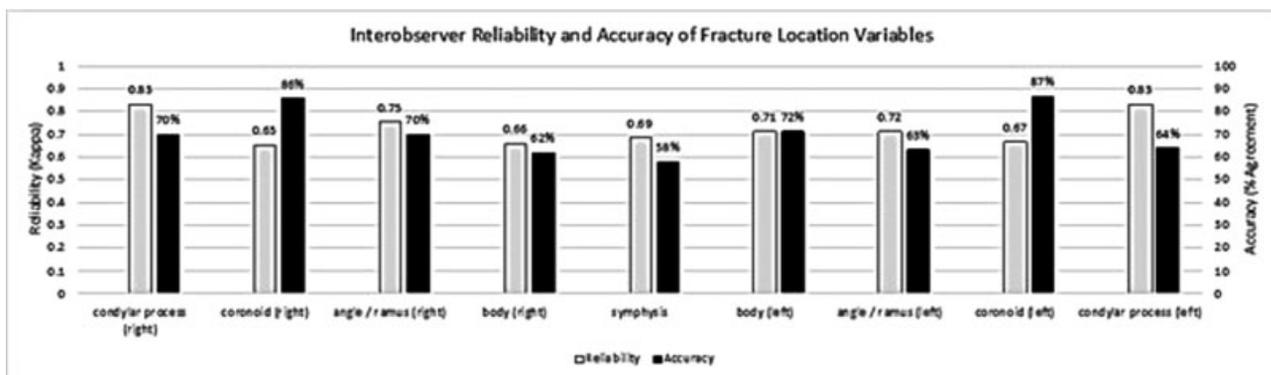
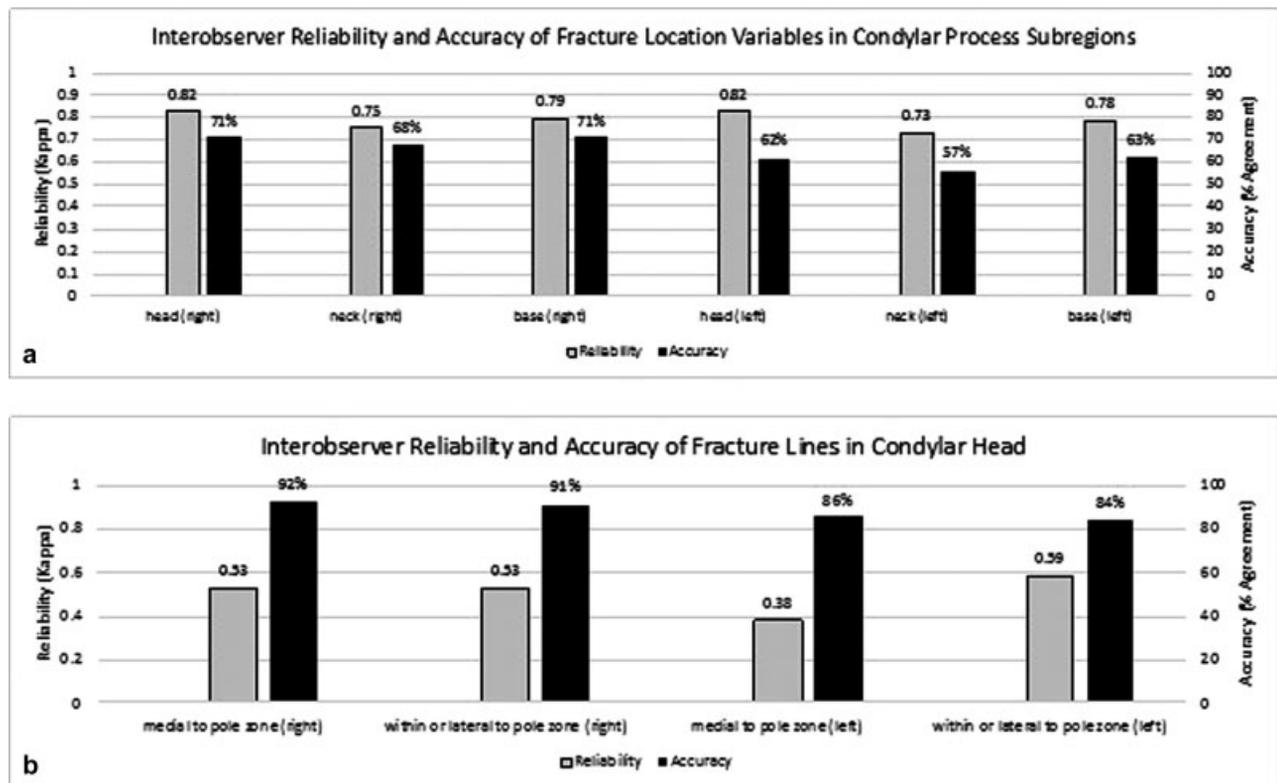


Fig. 2 Level 2—Interobserver reliability and accuracy of fracture location in the mandibular regions.



**Fig. 3** (a) Level 3—Interobserver reliability and accuracy of fracture location in condylar subregions. (b) Level 3—Interobserver reliability and accuracy of fracture location within the condylar head (m-type fractures vs. p-type fractures).

### Fracture Fragmentation Had Moderate to High Reliability and/or Accuracy

Fragmentation, one of the two key features of fracture morphology, was evaluated by asking the assessors to mark the degree of fragmentation (none, minor, or major) for each fracture pattern. There was moderate reliability for fragmentation in each of the noncondylar regions (level 2), with the highest reliability in the angle/ramus ( $k = 0.64$ ), and lowest reliability in the body ( $k = 0.59$ ). Accuracy was lowest in the symphysis (►Fig. 4a).

Within the CP subregions, classifications of fragmentation in the base had the highest reliability, followed by the head and then the neck. Accuracy was highest for fragmentation in the CH and lowest in the CB (►Fig. 4b).

### Type of Displacement Was Identified with Varied Reliability and Moderate to High Accuracy

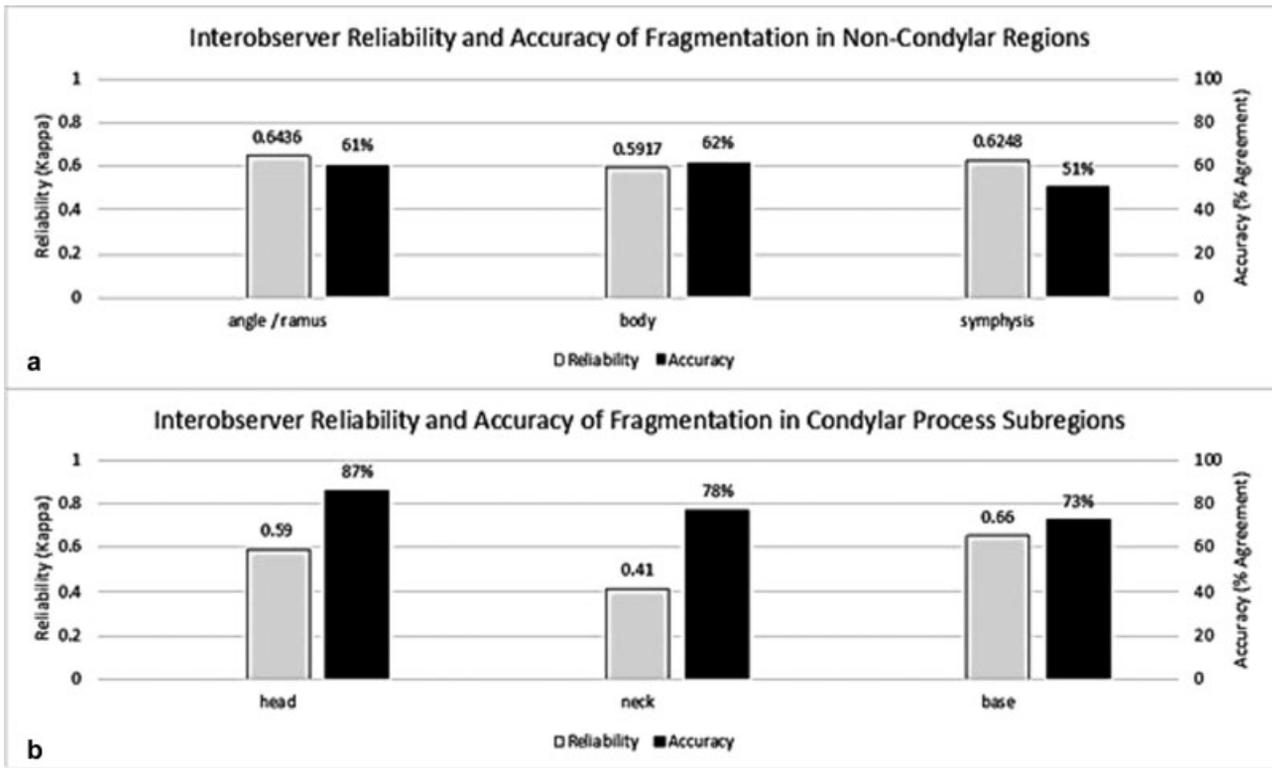
Besides fragmentation, displacement is the other key feature describing fracture morphology. For the CP, displacement is an umbrella term encompassing displacement/dislocation of the CH constituent of the CP fragment in relation to the fossa, displacement of the caudal mandibular end or ramus fracture end in relation to the fossa, as well as distortion of the CH constituent, and change of vertical ramus height. These displacement type variables for the overall CP (►Fig. 5a) reached moderate reliability ( $k$  range: 0.62–0.74). Measures of accuracy were also in a moderate range from 61 to 66% agreement (►Fig. 5a).

Assessors also evaluated a series of specific displacement types within each of the CP subregions. For example, in CH fractures, the degree of vertical apposition was assessed; in CN and CB fractures, the sideward displacement direction, the angulation of the condyle bearing fragment, and the override/shortening were assessed (►Fig. 5b). Reliability across the CH, CN, and CB ranged from 0.37 to 0.65. CN displacement had significantly lower reliability compared with CH and CB displacement ( $p = 0.002$ ). Accuracy was moderate to high (73–87%), and there was a significant difference in accuracy by each CP subregion ( $p = 0.038$ ).

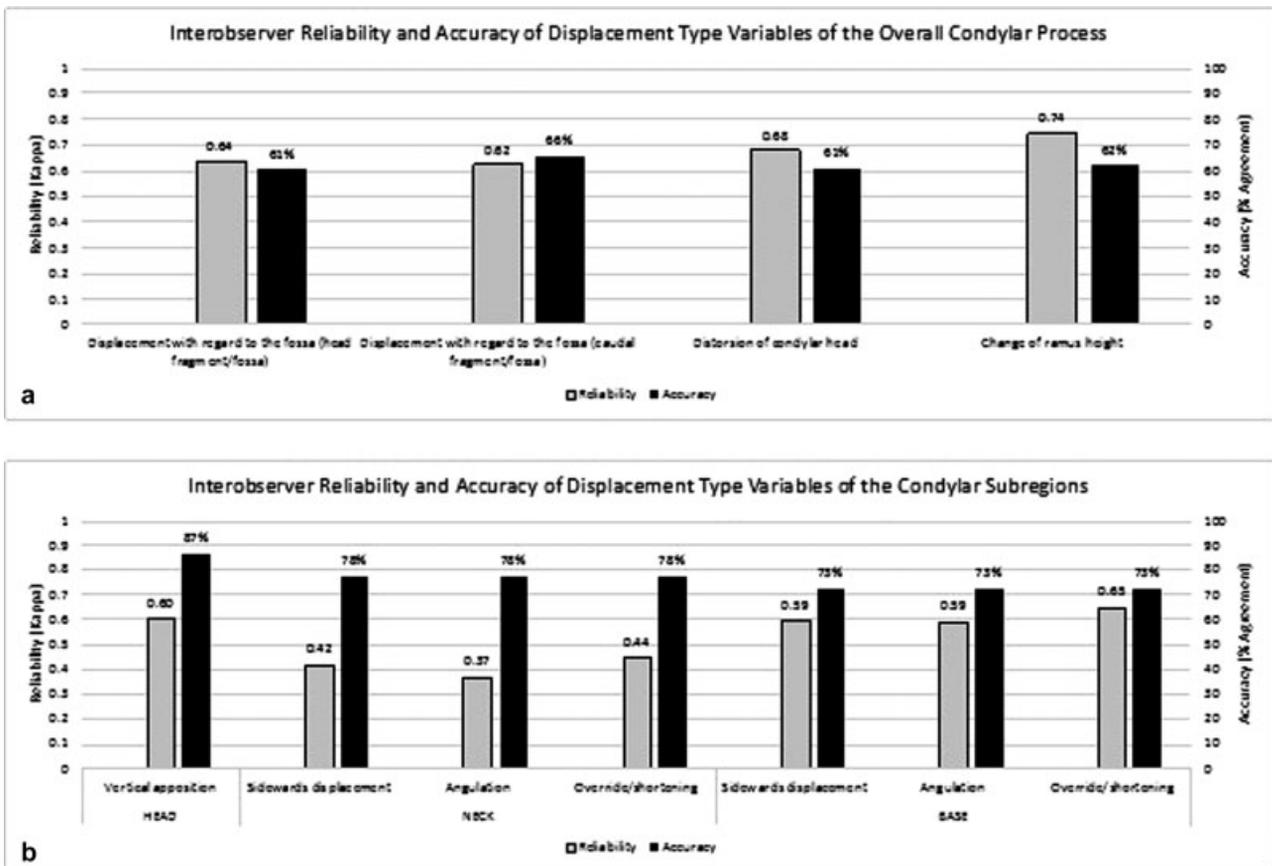
### Direction of Displacement in the Condylar Process Was Identified with Fair to Moderate Reliability and Moderate to High Accuracy

In the CP, reliability ranged from 0.29 when assessing the direction of displacement of the caudal fragment to 0.63 for direction of displacement of the CH fragment relative to the fossa (►Fig. 6a). Accuracy was moderately high (67–73% agreement).

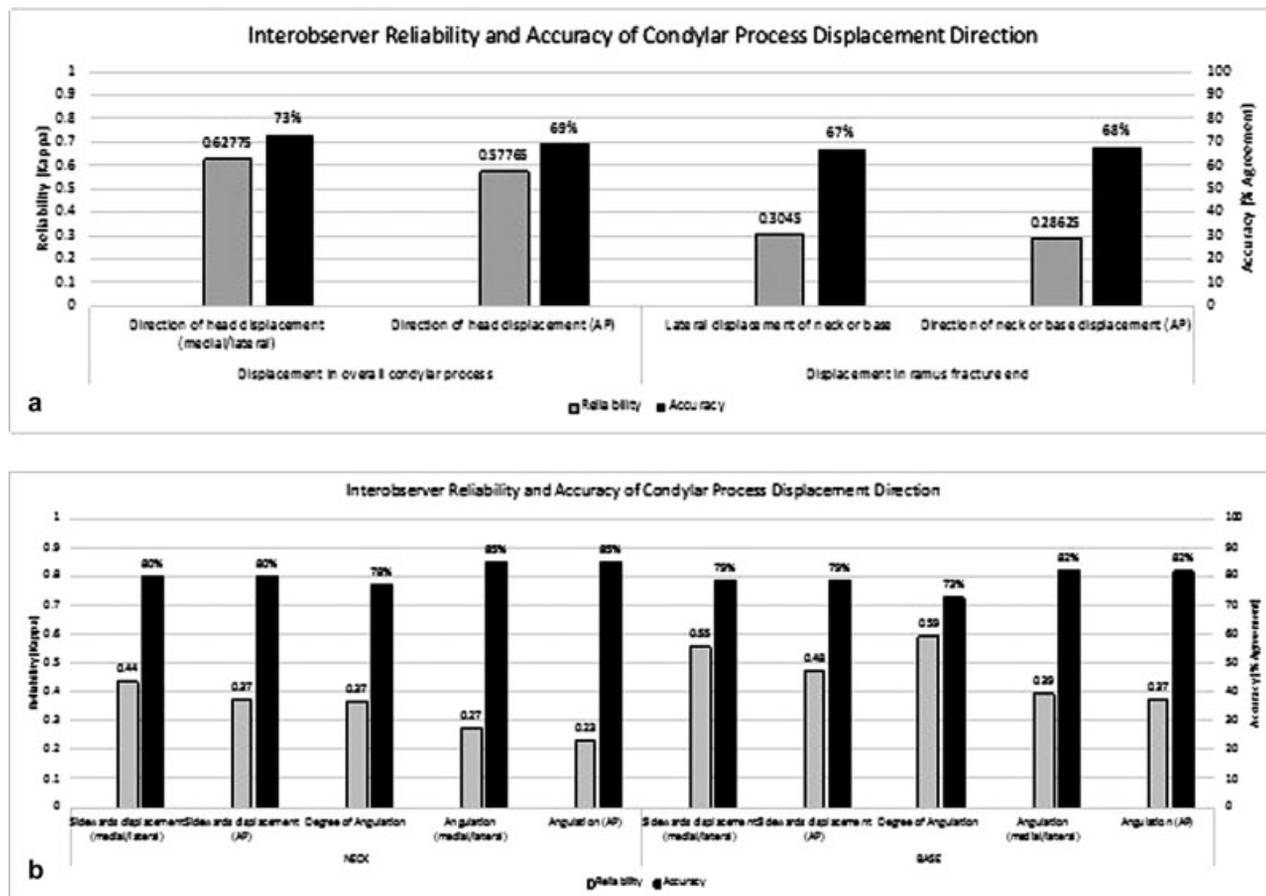
Within the CP subregions of the neck or base, reliability values for displacement direction variables were fair to moderate ( $k = 0.23$ –0.59; ►Fig. 6b). Reliability was the lowest when evaluating angulation of the CN and highest when assessing degree of angulation and direction of sideward displacement in the CB. Accuracy was relatively high across all variables in CN and CB (agreement = 73–85%).



**Fig. 4** (a) Level 3—Interobserver reliability and accuracy of fragmentation in noncondylar mandible regions. (b) Level 3—Interobserver reliability and accuracy of fragmentation in condylar subregions. Left and right measures have been averaged.



**Fig. 5** (a) Level 3—Interobserver reliability and accuracy of displacement-type variables of the overall condylar process. (b) Level 3—Interobserver reliability and accuracy of displacement-type variables within the condylar subregions. Left and right measures have been averaged.



**Fig. 6** (a) Level 3—Interobserver reliability and accuracy of displacement direction of the condylar head fragment (left) and the caudal fragment (right). Left and right measures have been averaged. (b) Level 3—Interobserver reliability and accuracy of displacement direction variables of condylar process subregions. Left and right measures have been averaged.

### Assessors with More Experience Show Higher Reliability and Accuracy

Assessors who have treated more mandibular fractures (clinical experience) and have previously used the AO CMF classification software (classification experience) have consistently higher reliability and accuracy than assessors with less experience. Across all 86 variables used in analysis, there was an increasing trend of interobserver reliability among the assessors who had more clinical experience. Assessors with the lowest experience treating mandibular fractures had a median kappa value of 0.48, while the assessors with medium clinical experience had a median kappa value of 0.59, and those with the highest clinical experience had a median kappa value of 0.66 ( $p < 0.001$ , **Fig. 7**). Assessors who had prior experience using the classification software also had a significantly higher median kappa value (0.76) compared with the assessors who did not have experience with the software (0.57,  $p < 0.001$ ). The same trends were noted with accuracy by experience. Those with more clinical experience had significantly higher measures of accuracy ( $p < 0.001$ ). Similarly, assessors with prior AO CMF classification experience had higher percentage agreement with the reference; those with no prior classification experience had a median of 90.5% agreement, while those with prior experience had a median of 93.5% agreement ( $p < 0.001$ , **Fig. 7**).

Among the nine level 2 fracture location variables, more clinical exposure was generally associated with higher interrater reliability (**Fig. 8**). Reliability ranged from moderate to very high kappa values. Similarly, assessors who had prior experience using the AO CMF fracture classification system also had higher measures of interrater reliability, except for when evaluating the fracture location in the right coronoid (**Fig. 9**). Reliability was moderate to very high, ranging from 0.57 to 0.93.

### Data Quality

There were 200 fracture cases evaluated by 15 assessors, making a total of 3,000 evaluations. Only 3.1% of these 3,000 evaluations had known classification errors in which the assessor marked a fracture as being unconfined to a particular region without indicating any of the adjacent regions as also having an unconfined fracture. In other words, these cases were erroneously identified as both overlapping with more than one region and also being confined to a single region. This error rate is fairly small and includes errors made by both more experienced and less experienced assessors.

### Discussion

Accurate and consistent assessment of CMF fractures is essential for communication both in the clinic and research

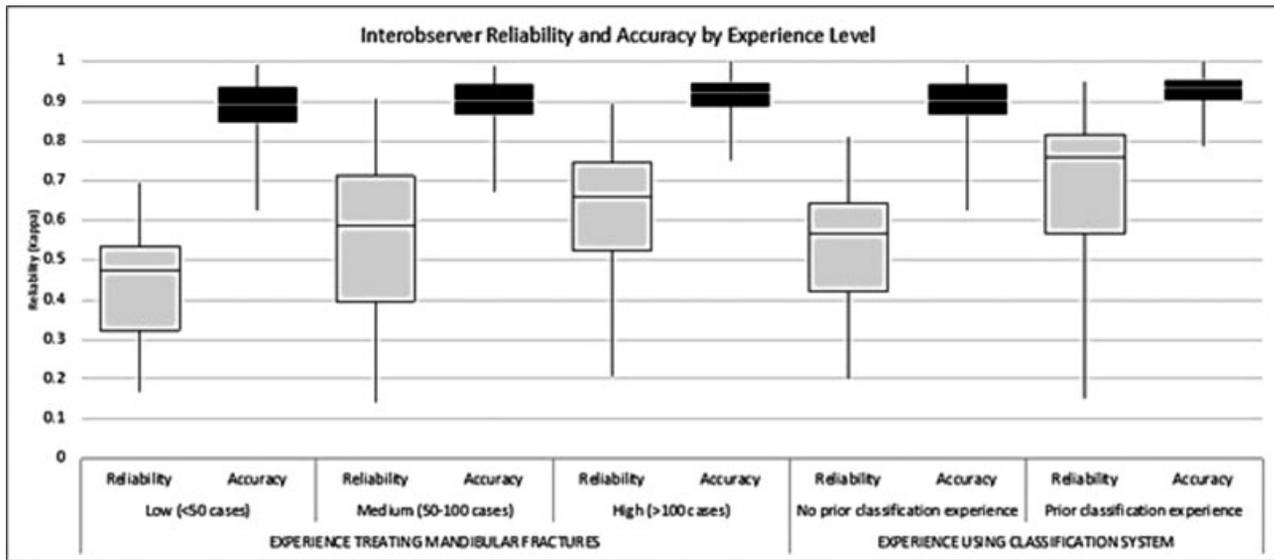


Fig. 7 Median reliability and accuracy across all 86 variables of level 2 and level 3 variables by assessor experience.

settings. The AO has developed a comprehensive craniofacial fracture classification system for adults that aims to meet the high demands of modern visual coding as well as verbal and nonverbal communication. So far the system has not been validated in a phase 2 setting, that is, in multicenter agreement studies.<sup>4</sup> The purpose of this study was to validate the three modules of the AO CMF classification system for mandible fractures by a group of assessors from four centers in high-resource countries.<sup>1,5,6</sup>

Historically, several schemes were published to classify mandible fractures. Most of them have primarily focused on the topographical location of the fractures, with varying definitions of the mandibular regions and subregions.<sup>1</sup> Additionally, some have included characteristics that require patient examination with regard to occlusion and soft-tissue involvement. In contrast, the AO CMF classification system is

based entirely on tomographic radiographic imaging (i.e., CT and cone beam CT). The purpose of this classification scheme was to create a method for classifying all mandible fractures of varying complexity using only radiographic images.

An ideal classification scheme is comprehensive, relevant to the clinical situation, and structured in a logical fashion. The AO classification system has been designed to allow for rapid classification of fractures using level 2 variables. It also allows for classifying more complex fracture patterns using level 3 details. These variables aim to classify fractures based on location and fracture morphology. It has also been integrated into a computer program (AOCOIA 4.0) to facilitate application and use.

This study found that the overall reliability and accuracy of the AO mandible fracture classification system were adequate for both fracture location and morphology with

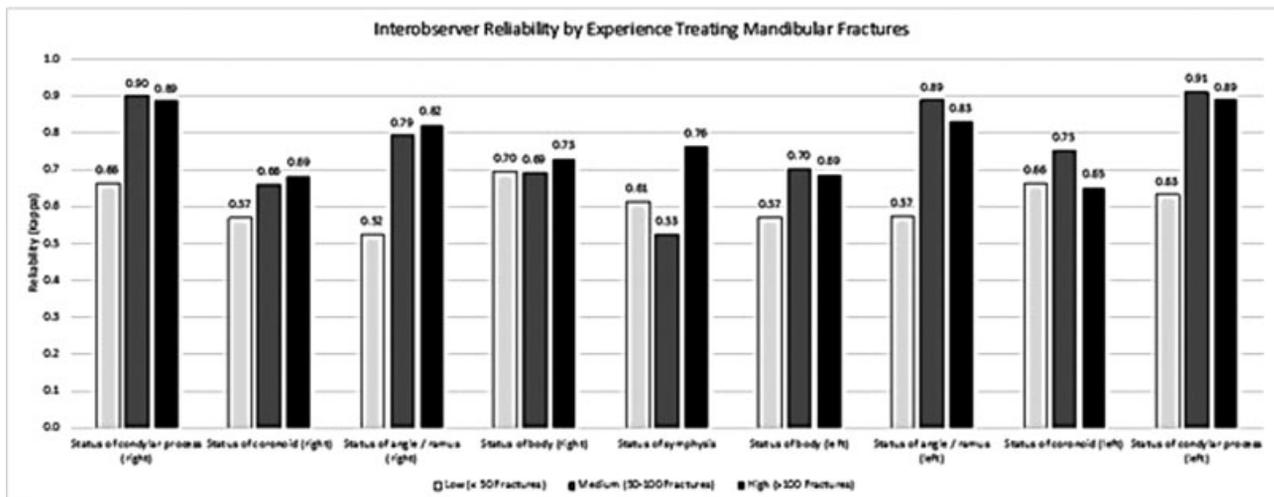
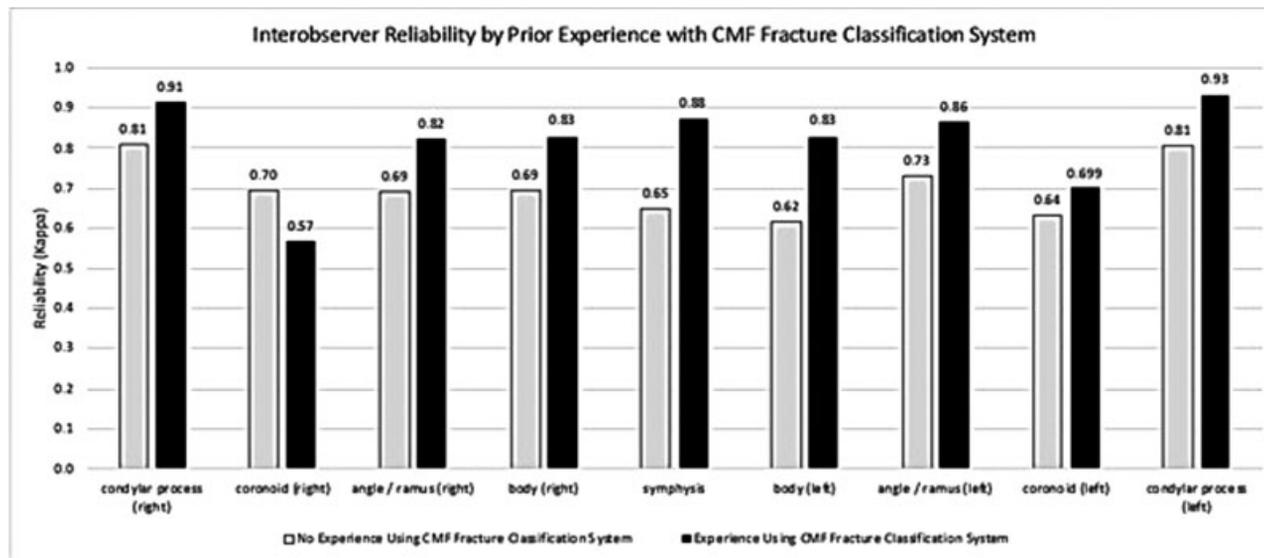


Fig. 8 Interobserver reliability of fracture location in the mandibular regions (level 2) by clinical experience treating mandibular fractures.



**Fig. 9** Interobserver reliability of fracture location in the mandibular regions (level 2) by prior experience with classification system.

regard to most level 2 and level 3 variables. With regard to level 2 variables, reliability was highest for characterizing fractures of the CP and lowest for the coronoid (►Fig. 2). Accuracy was highest when identifying fractures of the coronoid and lowest for the symphysis. Overall, both reliability and accuracy decrease when moving from level 2 to level 3 variables. When focusing on level 3 fracture location variables, reliability remains high with regard to fracture location within the CP (0.73–0.82) and drops when looking at more specific variables such as m-type versus p-type location (►Fig. 3) within the CH (0.38–0.59). Accuracy remains high both when looking at fracture location in the CP (57–71%) and within the CH (84–92%).

Level 3 fragmentation variables have similar reliability and accuracy measures for both condylar and noncondylar regions (►Fig. 4). Reliability ranges from 0.59 to 0.64 in noncondylar regions and 0.41 to 0.66 in condylar regions. Accuracy ranges from 51 to 62% in noncondylar regions and 73 to 87% in condylar regions. Reliability is worst (kappa of 0.41) when evaluating fragmentation of CN fractures. Reliability of the classification system for CN fractures is also poor when looking at displacement variables, signifying either general difficulty of classifying CN fractures or challenges with applying the classification system.

Level 3 displacement variables describing the overall CP (►Fig. 5a) remain at acceptable levels in the study. Reliability ranges from 0.62 to 0.74, while accuracy ranges from 61 to 66%. The reliability decreases as one focuses on displacement within the condylar subregions (►Fig. 5b). These values are lower particularly for CN fractures. Reliability drops to 0.37 when evaluating angulation of the neck. It is also low when looking at sideward displacement (0.42) and override/shortening (0.44) of CN fractures. As mentioned earlier, this may be due to the difficulty of classifying neck fractures as even a small degree of displacement in this region can result in what appears to be notable changes at the neck. In other words, there is a component of

subjectivity when the assessor is determining whether a miniscule degree of CN displacement, angulation, or shortening is considered to be occurring. The remainder of level 3 displacement type variables (i.e., variables for the head and base) maintain adequate reliability values, ranging from 0.59 to 0.65. Despite the low reliability, accuracy is acceptable for all level 3 displacement type variables within the condylar subregions, ranging from 73 to 87%.

Level 3 displacement variables focusing on displacement direction for the CP as an overall fragment (►Fig. 6a, left) are adequately reliable and accurate when looking at the displacement or dislocation of the CH in relation to the fossa (0.58–0.63 and 60–73%, respectively). Accuracy is high when looking at the direction of displacement of the caudal fragment (►Fig. 6a, right; 67–68%), but reliability drops significantly (0.29–0.30). This trend remains when looking at displacement direction within the condylar subregions. Accuracy is adequate for neck or base fractures (73–85%), but reliability is low for neck fractures (0.23–0.44). However, it is slightly higher when looking at displacement direction of base fractures with angulation having the worst values (0.37–0.39), which improves slightly when looking at displacement direction of base fractures and their degree of angulation (0.48–0.59).

When evaluating interobserver reliability and accuracy, there is a general trend of increased reliability and accuracy for assessors who have treated more fractures. Additionally, the assessors have higher interobserver reliability and accuracy if they have had prior experience with the CMF fracture classification system. This fits the intuitive assumption that individuals are better at classifying fractures if they have treated fractures and have had more exposure to the classification system.

There are a few limitations of the study that are inherent to the design. A limited number of assessors (15) were used to validate the system.<sup>13–16</sup> Although this number of assessors was adequate for drawing several conclusions, a more

granular evaluation of reliability and accuracy could be determined with more assessors at each level of training and while in practice.

As there is no easily applicable method for obtaining a gold standard definition of the fracture patterns in each case, the author with the most experience with this classification system was used as the benchmark against which to measure assessor accuracy. One could potentially resolve this issue in the future by discussing the true nature of each individual fracture in a group setting and agree upon the fracture as a whole. However, with 200 cases, many of which had multiple fractures, this would be impractical.

There are some classification variables with wide differences between reliability and accuracy. Most commonly in this study, there is a relatively high accuracy with occasional low reliability. An important consideration of the reliability measure, the kappa coefficient, is the subtracted level of agreement due to chance alone. When the proportion of agreement due to chance alone is high, the result will be a low kappa coefficient, irrespective of the level of agreement with the selected reference observer.<sup>17</sup>

One method that could be considered for improving the reliability and accuracy of the classification scheme would be to add objective measurements to the system. Zhou et al, for example, measured the degree of angulation of condylar fractures and the amount of ramus height reduction observed in condylar fractures.<sup>18</sup> The AO CMF classification scheme had difficulty maintaining high reliability with some variables such as CN fractures, possibly due to subjective evaluations of the variables, including the degree of angulation, shortening, and displacement. An objective measurement could help improve classification of those variables and others.

This classification system and software will hopefully be used both in clinical and research settings to improve communication and presentation of information on fractures. It has been shown in this article to be both accurate and reliable to individuals at varied levels in their training or before their surgical training. It will ideally improve documentation, communication between teams, and clinical decision making. More studies will be needed to evaluate how this impacts the quality of patient care.<sup>4</sup>

## Conclusion

The AO has developed a tool that is comprehensive, clinically relevant, and easy to use. This study demonstrates that the mandibular fracture classification system is also both accurate and reliable for level 2 variables. These values decrease when evaluating level 3 variables, in particular reliably identifying the location of fractures within the CH and when describing the displacement morphology of fractures within the CN. This may be improved upon through efforts to increase training and improve classification instructions. Additionally, the data recording and entry could be improved by implementing an input process that automatically checks for plausibility.

## Acknowledgments

The authors thank the Arbeitsgemeinschaft Osteosynthese Cranio-Maxillo-Facial (AOCMF) for sponsoring the trial and the Arbeitsgemeinschaft Osteosynthese Clinical Investigation and Documentation (AOCID) for serving as contract research organization.

## References

- Cornelius CP, Audigé L, Kunz C, et al. The comprehensive AOCMF classification system: mandible fractures- Level 2 tutorial. *Cranio-maxillofac Trauma Reconstruction* 2014;7(Suppl 1):S15-S30
- Audigé L, Cornelius CP, Di Ieva A, Prein J, Group CC; CMF Classification Group 6. The first AO classification system for fractures of the craniomaxillofacial skeleton: rationale, methodological background, developmental process, and objectives. *Craniomaxillofac Trauma Reconstr* 2014;7(01, Suppl 1):S006-S014
- Manson PN, Hollier L, Schubert W. CMF classification. *Craniomaxillofac Trauma Reconstr* 2014;7(01, Suppl 1):S001-S003
- Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma* 2005;19(06):401-406
- Cornelius CP, Audigé L, Kunz C, et al. The comprehensive AOCMF classification system: mandible fractures-Level 3 tutorial. *Craniomaxillofac Trauma Reconstr* 2014;7(01, Suppl 1):S031-S043
- Neff A, Cornelius CP, Rasse M, et al. The comprehensive AOCMF classification system: condylar process fractures - level 3 tutorial. *Craniomaxillofac Trauma Reconstruction* 2014; 7(Suppl 1):S44-S58
- Buitrago-Téllez CH, Audigé L, Strong B, et al. A comprehensive classification of mandibular fractures: a preliminary agreement validation study. *Int J Oral Maxillofac Surg* 2008;37(12):1080-1088
- Müller ME, Nazarian S, Koch P, Schatzker J. *The Comprehensive Classification of Fractures of Long Bones*. Berlin, Heidelberg: Springer; 1990
- Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE—An integrated standards-based translational research informatics platform. *AMIA Annu Symp Proc* 2009;2009:391-395
- Audigé L, Cornelius CP, Kunz C, Buitrago-Téllez CH, Prein J. The comprehensive AOCMF classification system: classification and documentation within AOCOIAC software. *Craniomaxillofac Trauma Reconstr* 2014;7(01, Suppl 1):S114-S122
- Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76(05):378-382
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(01):159-174
- Slongo T, Audigé L, Schlickewei W, Clavert JM, Hunter J; International Association for Pediatric Traumatology. Development and validation of the AO pediatric comprehensive classification of long bone fractures by the Pediatric Expert Group of the AO Foundation in collaboration with AO Clinical Investigation and Documentation and the International Association for Pediatric Traumatology. *J Pediatr Orthop* 2006;26(01):43-49
- Kreder HJ, Hanel DP, McKee M, Jupiter J, McGillivray G, Swiontkowski MF. Consistency of AO fracture classification for the distal radius. *J Bone Joint Surg Br* 1996;78(05):726-731
- Schipper IB, Steyerberg EW, Castelein RM, van Vugt AB. Reliability of the AO/ASIF classification for pertrochanteric femoral fractures. *Acta Orthop Scand* 2001;72(01):36-41
- Martin JS, Marsh JL, Bonar SK, DeCoster TA, Found EM, Brandser EA. Assessment of the AO/ASIF fracture classification for the distal tibia. *J Orthop Trauma* 1997;11(07):477-483
- Rigby AS. Statistical methods in epidemiology. v. Towards an understanding of the kappa coefficient. *Disabil Rehabil* 2000;22(08):339-344
- Zhou Z, Li Z, Ren J, et al. Digital diagnosis and treatment of mandibular condylar fractures based on Extensible Neuro imaging Archive Toolkit (XNAT). *PLoS One* 2018;13(02):e0192831-e0192814