

Poor reliability and reproducibility of 3 different radiographical classification systems for distal ulna fractures

Maria MOLONEY¹, Jan KÅREDAL², Tomas PERSSON², Simon FARNEBO^{1,4}
and Lars ADOLFSSON^{3,4}



Acta Orthopaedica

¹ Department of Plastic Surgery, Hand Surgery, and Burns, Linköping University; ² Department of Radiology, Hospital of Motala; ³ Department of Orthopaedics, Linköping University; ⁴ Department of Clinical and Experimental Medicine, Linköping University, Sweden

Correspondence: maria.moloney@regionostergotland.se

Submitted 2021-10-26. Accepted 2022-03-09.

Background and purpose — Classification of fractures can be valuable for research purposes but also in clinical work. Especially with rare fractures, such as distal ulna fractures, a treatment algorithm based on a classification can be helpful. We compared 3 different classification systems of distal ulna fractures and investigated their reliability and reproducibility.

Patients and methods — patients with 97 fractures of the distal ulna, excluding the ulnar styloid, were included. All fractures were independently classified by 3 observers according to the classification by Biyani, AO/OTA 2007, and AO/OTA 2018. The classification process was repeated after a minimum of 3 weeks. We used Kappa value analysis to determine inter- and intra-rater agreement.

Results — The inter-rater agreement of the AO/OTA 2007 classification was judged as fair, κ 0.40, whereas the agreement of AO/OTA 2018 and Biyani was moderate at κ 0.42 and 0.43 respectively. The intra-rater agreement was judged as moderate for all classifications.

Interpretation — The differences between the classifications were small and the overall impression was that neither of them was good enough to be of substantial clinical value. The Biyani classification, being developed specifically for distal ulna fractures, was the easiest and most fitting for the fracture patterns seen in our material, but lacking options for fractures of the distal diaphysis. Standard radiographs were considered insufficient for an accurate classification. A better radiographic method combined with a revised classification might improve accuracy, reliability, and reproducibility.

Falling on an outstretched arm with an extended wrist is the typical trauma causing a distal radius fracture. More rarely it also results in a fracture of the distal ulna. The incidence of distal ulna fractures, with or without a concomitant radius fracture, is 74/100,000 person-years (1). Metaphyseal fractures of the distal ulna can be found in 6–8 % of distal radius fractures (2,3). While a lot of research has focused on the treatment of distal radius fractures, there has been much less interest in distal ulna fractures. The recently published Swedish national guidelines for treatment of distal radius fractures concluded that there is not enough evidence to make recommendations on treatment of associated distal ulna fractures (4). Patients with concomitant distal radius and ulna fractures have been found to score a substantially higher DASH (score 20 for surgically treated and 25 for non-surgically treated) compared with patients with only a distal radius fracture (score 9) (5). In a previous study we found that most extra-articular distal ulna fractures seem to be best treated without internal fixation (6).

Classification of fractures can be valuable for research purposes, but for a classification to be clinically relevant and routinely used it should preferably help to guide the clinician in choosing the accurate treatment for a specific fracture. An accurate classification should therefore be both reliable and reproducible, meaning that different users achieve the same result and that the result does not change over time. It should also be all-inclusive and not leave any fractures unclassified, as this will reduce its possible clinical value. The most well-known and extensive classification of fractures of the long bones was developed by Müller and collaborators in 1990 (7). This formed the base for the widely used classification produced by the Arbeitsgemeinschaft für Osteosynthesefragen (AO) Foundation/Orthopaedic Trauma Association (OTA),

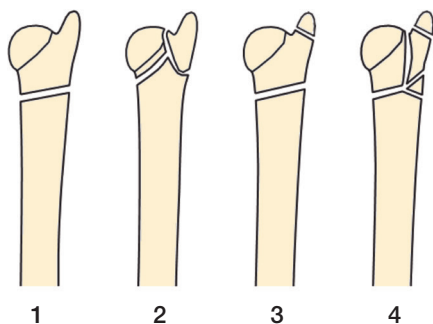


Figure 1. Classification by Biyani et al. 1995 (2).

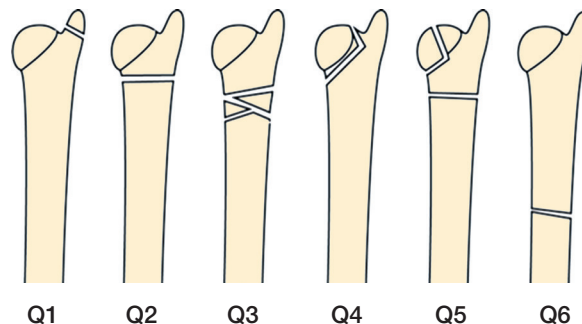


Figure 2. AO/OTA classification 2007 (7).

published as a compendium to the Journal of Orthopaedic Trauma (JOT) in 1996 (7). This classification has since then been revised and improved, most recently in 2007 and 2018. Not many other classification systems have been developed specifically for distal ulna fractures, with the exception of the classification published by Biyani and colleagues in 1995 (2).

We compared 3 different classification systems of distal ulna fractures and investigated their reliability and reproducibility.

Patients and methods

Data from all patients in Östergötland county, with a population of approximately 465,000 inhabitants, treated for a fracture of the distal ulna, isolated or in combination with a fracture of the distal radius, 2010–2014, was collected. All patients who had visited 1 of the 3 orthopedic departments in the area (Linköping University Hospital, Vrinnevi Hospital of Norrköping, and the Hospital of Motala) received an ICD-10 diagnostic code in the digital journal system. In 2015 we searched the central database for all codes of a distal forearm fracture (S52.50, S52.51, S52.60, S52.61, S52.20, S52.21, S52.80, S52.81) during 2010–2012 and in 2019 we extended the search to also include fractures sustained during 2013 and 2014. The radiographs of all patients who had received one of these codes were screened by one of the authors (MM), who was not going to be an observer, in the digital radiology system PACS/IDS7 to identify all who had suffered a fracture of the distal third of the ulna during the defined time period. Patients under the age of 18 and those who had sustained a fracture of the ulnar styloid tip only (with or without a concomitant distal radius fracture) were excluded from the study, as well as patients who were deceased or had emigrated from Sweden. Of 191 possible distal ulna fractures, 96 patients with 97 fractures were included.

All fractures were diagnosed by initial plain radiographs taken with an anteroposterior (AP) and a lateral projection. These images were used for classification by 3 observers with experience of examining wrist radiographs: 2 senior specialists in radiology (JK and TP) and 1 senior orthopedic surgeon (LA). Observer 1 had been a specialist in radiology for 36 years, but

for the last 5 years had been partly retired from clinical work. Observer 2 had been a specialist in radiology for 25 years and with a special interest in skeleton radiology. Observer 3, the orthopedic surgeon, was specialized in upper extremity surgery and had been a specialist for 30 years. Classification was performed according to the 2018 AO/OTA classification (8), the 2007 AO/OTA classification (7), and the classification by Biyani et al. (2). Classification was performed independently by the 3 observers, and all 97 fractures were classified according to the 3 different classification systems. The observers had never used these classifications on a regular basis. They were all provided with the same text and pictures describing the different classification systems and for the 2018 AO/OTA classification they could also use the AO/OTA classification application for smartphones (9). The classification was repeated after a minimum of 3 weeks by all 3 observers.

Classifications

The Biyani classification was developed in 1995 for fractures of the distal ulna associated with distal radius fractures, with or without associated fractures of the ulnar styloid. This was based on a radiographic review of 19 fractures of the distal radius and ulna. It divides the distal ulna fractures into 4 different fracture patterns, named types 1–4 (2) (Figure 1).

In the original AO classification, distal ulna fractures associated with distal radius fractures were classified with a Q modifier. There were 6 different Q classes, of which Q1 referred to styloid fractures (7). The AO and OTA classification committee revised the classification in 2007 and also in 2018. In the latest review it had become evident that it was more accurate to separately code radius and ulna fractures (8). The distal segment of the ulna was now classified as 2U3 followed by A for extra-articular, B for partial articular, and C for complete articular fractures. 2U3A was further subdivided into 1 for styloid process (subdivided into (1) tip and (2) base), 2 for simple (subdivided into (1) spiral, (2) oblique, and (3) transverse) and 3 for multifragmentary. This means that there are 8 possible fracture classes of distal ulna fractures. There are also universal modifiers that can be added to the end of the fracture code, for example impaction or dislocation (8). The AO/OTA classifications can be seen in Figures 2 and 3.

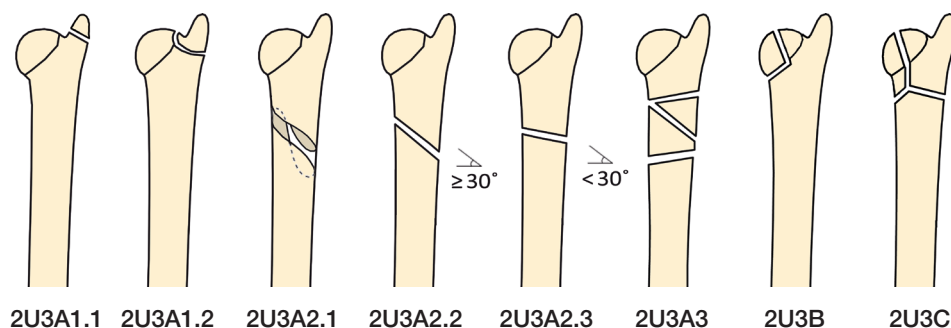


Figure 3. AO/OTA classification 2018 (8).



Figure 4. Example of distal ulna fracture difficult to classify.

Table 1. Inter-rater agreement of AO/OTA 2007

AO/OTA 2007	Observers	Kappa	Agreement
First examination	1 vs. 2	0.55	Moderate
	1 vs. 3	0.32	Fair
	2 vs. 3	0.39	Fair
	Mean	0.42	Moderate
Second examination	1 vs. 2	0.43	Moderate
	1 vs. 3	0.26	Fair
	2 vs. 3	0.45	Moderate
	Mean	0.38	Fair
Total	Mean	0.40	Fair

Table 2. Inter-rater agreement of AO/OTA 2018

AO/OTA 2018	Observers	Kappa	Agreement
First examination	1 vs. 2	0.54	Moderate
	1 vs. 3	0.36	Fair
	2 vs. 3	0.35	Fair
	Mean	0.42	Moderate
Second examination	1 vs. 2	0.46	Moderate
	1 vs. 3	0.44	Moderate
	2 vs. 3	0.40	Fair
	Mean	0.43	Moderate
Total	Mean	0.42	Moderate

Both the Biyani classification and AO/OTA 2007 were constructed for concomitant distal radius and ulna fractures. Despite this fact we chose to include isolated distal ulna fractures because the AO/OTA 2018 includes all distal ulna fractures, stating that it is more accurate to classify ulnar fractures separately. The Biyani classification was based on metaphyseal fractures of the distal ulna. Since we found that distal diaphyseal fractures occasionally also engage the metaphysis and that these fractures are included in the other systems we chose to include the Biyani classification because the investigation comprised fractures of the distal third of ulna.

Statistics

Kappa value analysis was used to determine inter- and intra-rater agreement: 0 (less agreement than is expected by chance alone) to 1 (perfect agreement) (10). To interpret the κ values we used the criteria of Landis and Koch: ≤ 0.2 , slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and ≥ 0.81 , almost perfect agreement (11). Intra-rater agreement was calculated for each observer. From these κ values a mean was calculated for each classification system. Statistical analysis was performed using IBM SPSS Statistics 27 (IBM Corp, Armonk, NY, USA).

Ethics, funding, and potential conflicts of interest

The study protocol was reviewed and approved by the Regional Ethical Review Board (Dnr 2014/200-31) and informed con-

sent from all patients participating in this study was obtained. The authors received no financial support, and declare no conflicts of interest.

Results

The mean age at the time of injury was 63 years (SD 15) and 79 patients out of 96 were women. 1 patient had bilateral distal ulna fractures, and both fractures were included in the study. The total mean κ value of the AO/OTA 2007 classification was fair, κ 0.40, whereas the agreement of AO/OTA 2018 and Biyani was moderate at κ 0.42 and 0.43 respectively. The intra-rater agreement was moderate for all classifications. The intra-rater agreement was fair for observer 1, moderate for observer 2, and substantial for observer 3 (Tables 1–4).

One type of fracture was considered impossible to fit into either classification system. This was a fracture of a 1.5 cm long fragment of the ulnar border including the styloid (Figure 4).

In the AO/OTA 2007 classification a quite common disagreement between the observers was between Q2 and Q6, where no clear anatomical landmark can separate the 2 classes. In the Biyani classification the observers noted that a problem was that there was no option for a fracture extending into the distal diaphysis. The combination of a subcapitular fracture of the ulna in combination with a fracture of the ulnar styloid was common. This combination could be found in Biyani but not in the 2 AO classifications. In the AO/OTA 2018 the

Table 3. Inter-rater agreement of Biyani

Biyani	Observers	Kappa	Agreement
First examination	1 vs. 2	0.51	Moderate
	1 vs. 3	0.43	Moderate
	2 vs. 3	0.51	Moderate
	Mean	0.49	Moderate
Second examination	1 vs. 2	0.30	Fair
	1 vs. 3	0.25	Fair
	2 vs. 3	0.53	Moderate
	Mean	0.36	Fair
Total	Mean	0.43	Moderate

class 2U3A2.1, spiral fractures, and isolated partial articular fracture, 2U3B, were particularly difficult to identify with the radiographs to hand.

Discussion

To our knowledge there have been no previous investigations of reliability and reproducibility of different classifications of distal ulna fractures. When assessing the reliability and reproducibility of the AO/OTA 2018 classification compared with the previous AO/OTA 2007 and Neers classification for proximal humeral fractures a substantial mean inter- and intra-rater agreement was seen for the AO/OTA 2018 system. The inter-rater agreement was substantially higher for the AO/OTA 2018 compared with the AO/OTA 2007 classification (12).

None of the classifications for distal ulna fractures showed an impressive result of inter- or intra-rater agreement. The differences between the 3 systems were small and the overall impression was that none of them was good enough to be of substantial clinical value. Both the Biyani and the 2007 AO/OTA classifications are constructed for distal ulna fractures associated with distal radius fractures. Despite this we chose to include isolated ulna fractures in our material since these are not separated in the 2018 AO classification and the other two systems have no classification for isolated fractures. The AO/OTA 2007 and the AO/OTA 2018 were developed to have the same structure for all the long bones of the body, and they do not consider how the ulna is commonly fractured. This might be the reason why there was no optimal option for several fractures found in our material, such as for example fractures of the metaphysis combined with styloid fractures. The Biyani classification on the other hand was developed by examining the fracture patterns of the ulna in a number of distal forearm fractures. This system, however, was based on only a small number of fractures and all classes are very distal, in the epi-/metaphysis. With the Biyani classification the main problem is that fractures involving the distal diaphysis are lacking, perhaps because they are not as common in conjunction with distal radius fractures. The AO/

Table 4. Intra-rater agreement of the classification systems

	Observers	Kappa	Agreement
AO/OTA 2007	1	0.28	Fair
	2	0.65	Substantial
	3	0.65	Substantial
	Mean	0.53	Moderate
AO/OTA 2018	1	0.38	Fair
	2	0.52	Moderate
	3	0.66	Substantial
	Mean	0.52	Moderate
Biyani	1	0.42	Moderate
	2	0.64	Substantial
	3	0.68	Substantial
	Mean	0.58	Moderate

OTA 2018 classification is extensive and complex compared with the other 2 classifications. This makes it difficult to use, since implementation takes time to learn. It also includes universal modifiers, such as for example impaction or dislocation. This makes the different options for classes extensive and harder to handle. For this reason we chose not to include the modifiers in this comparison. The importance of classifying the ulna fracture as located in the metaphysis or diaphysis is not specified in either of the classifications and is up to the examiner. We chose to use the classifications slightly outside their original intention. AO/OTA 2007 was not designed for isolated distal ulna fractures and the Biyani classification not for fractures of the distal diaphysis. The reason we chose to adapt these classification systems to our material was that the AO/OTA 2018 is the newest of the classification systems and covers both these aspects and that our material is a large cohort of fractures representative of distal ulna fractures as a whole (6). This decision could, however, have an impact on the result where fractures did not fit into the specified classes. We consider this to be valuable information as we believe a classification of a rare fracture, such as distal ulna fracture, should comprise all the commonly existing fracture patterns.

The intra-rater agreement showed considerable differences between the observers. We chose observers who had worked for a long time in their respective field because we believed that classification might be a difficult process and requires that the observers have a great deal of experience in examining radiographs. The orthopedic surgeon's (observer 3) overall classification showed a substantial reproducibility. This might be due to his special interest in fractures of the upper limb but also to him being more used to using different classification systems for other orthopedic injuries and being used to the AO/OTA system. The radiologists had a moderate and fair intra-rater agreement respectively. The radiologist who was more specialized in skeleton radiology (observer 2) had the better result, possibly because of examining more radiographs of the wrist in his day-to-day work. None of the observers

had used these classifications before this study. It could be considered a weakness of this study that the intra-rater agreement varied for the 3 observers, although for a classification to be of value it should have high reproducibility for all observers intended to use it, for example radiologists and orthopedic surgeons. Considering this, maybe more observers should be used in future studies and observers of different experience within these specialties. We could not see that there was any improvement in the inter-rater agreement between the first and the second examination but results might improve with training. However, if a classification were to be used in their daily work by radiologists examining unselected radiographs referred from the ER, they could not be expected to have special training in a certain fracture classification.

The radiographs used for the classifications were performed with the aim of finding fractures after trauma to the wrist. Only in a very small number of cases was there a suspicion of a more proximal injury, thus pictures of the whole forearm were few. The standard examination of the wrist includes an anteroposterior and a lateral projection where the radius should constitute the dorsal contour. In the 2018 AO/OTA classification oblique radiographs are also recommended, but this was not routinely done in the radiology departments involved in this study. When examining the original radiologists' reports, associated radius fractures almost always dominate the text, and the ulna fractures are only sparsely described. The ulna is in most cases projected behind the radius in the lateral projection and because it is common with more or less dorsal angulation of the fractures the verdict on whether or not the distal ulna fracture is intra- or extra-articular was difficult for the observers. Even though the classifications are made for plain radiographs, this is perhaps not sufficient for distal ulna fractures, or the AP and lateral views should be complemented with oblique views. Alternatively, computed tomography (CT), or cone beam CT (CBCT) if available, would show more details of the fracture's exact location and extensions. This would possibly make the classification more accurate. CBCT, originally used in orthodontics, has a higher spatial resolution than conventional CT and utilizes a lower radiation dose, comparable to that of 2–3 plain radiographs (13). Better imaging, as through a CT, could possibly improve the inter- and intra-rater agreement and, even more importantly, might improve accuracy of the classification. An additional CT has been shown by some authors to increase the intra-rater agreement (but not the inter-rater agreement) of several classification systems for distal radius fractures (14), while others have not found any improvements in reliability or reproducibility of the same classification systems (15). These studies, however, included only cases where CT was performed for planning of the treatment or in cases of a questionable indication for surgery. This results in a selection bias where more complicated fractures are included, in which the classification might be more difficult than in simpler fractures. Kleinlugtenbelt et al. also used only 2D-CT images whereas Harness et al.

showed that 3D-CT increased both reliability and reproducibility of radiographic characterization of intra-articular distal radius fractures compared with 2D-CT images (14,16). A CT of the ulna could more accurately identify whether the fractures are intra- or extra-articular, an important aspect of all of the investigated classification systems. Whether an oblique projection should be added or CT recommended remains to be further investigated.

More detailed imaging could also aid in improving a classification system. In order to make a classification more clinically relevant it would be a starting point to define where the distal ulna starts, and also to consider the surrounding stabilizing soft tissues that can affect the outcome of the fracture treatment and rehabilitation. When considering this, the Biyani classification seems to be the most suitable of the 3 we examined. It would, however, be of interest to include distal diaphyseal fractures, as they are not uncommon in conjunction with distal radius fractures as a result of the same trauma. The Biyani classification was considered by our 3 observers the easiest to use and the most fitting for distal ulna fractures. A modification of the Biyani classification to include fractures of both the metaphysis and the distal diaphysis seems like an alternative for a more reliable description of distal ulna fractures.

Our conclusion is that a better radiographical method and an improved classification is likely to result in higher accuracy, reliability, and reproducibility. A better classification with high accuracy, inter- and intra-rater agreement could help in evaluations of treatment algorithms and methods of internal fixation.

All authors contributed to the study and read and approved the finished manuscript before submission. MM, SF, and LA designed the study. LA, JK, and TP were the three observers classifying the fractures. MM collected and analyzed the data, and wrote the manuscript. SF, LA, JK, and TP contributed to analyzing the data and revising the manuscript.

The authors would like to thank Lars Valter for the help with the statistical analysis.

Acta thanks Marcus Landgren, Mats Wadsten, and other anonymous reviewers for help with peer review of this study.

1. **Moloney M, Farnebo S, Adolfsson L.** Incidence of distal ulna fractures in a Swedish county: 74/100,000 person-years, most of them treated non-operatively. *Acta Orthop* 2020; 91(1): 104-8.
2. **Biyani A, Simison A J, Klenerman L.** Fractures of the distal radius and ulna. *J Hand Surg Br* 1995; 20(3): 357-64.
3. **Herzberg G, Castel T.** [Incidence of distal ulna fractures associated with distal radius fractures: Treatment options]. *Hand Surg Rehabil* 2016; 35S: S69-S74.
4. **Nationellt vårdprogram för behandling av distala radiusfrakturer.** Available from: <https://dd2flujgsl7escs.cloudfront.net/external/Nationellt+vårdprogram+för+behandling+av+distala+radiusfrakturer.pdf>
5. **Landgren M, Abramo A, Geijer M, Kopylov P, Tägil M.** Similar 1-year subjective outcome after a distal radius fracture during the 10-year-period 2003–2012. *Acta Orthop* 2017; 88(4): 451-6.

6. **Moloney M, Farnebo S, Adolfsson L.** Distal ulna fractures in adults: subcapitular, transverse fractures did not benefit from surgical treatment. *Arch Orthop Trauma Surg* 2022 Jan 21. doi: 10.1007/s00402-022-04336-1. Online ahead of print.
7. **Müller M E, K P, Nazarian S, Schatzker J.** Radius/ulna = 2. In: *The comprehensive classification of fractures of long bones*. 1990, Berlin, Heidelberg: Springer.
8. **Radius and ulna.** *J Orthop Trauma* 2018; 32(Suppl. 1): S21-S32.
9. **AO/OTA Fracture Classification** [1.3.1]; 2018. Available from: https://www2.aofoundation.org/AOFileServerSurgery/MyPortalFiles?FilePath=/Surgery/en/_docs/AOTA%20Classification%20Compendium%202018.pdf
10. **Audige L, Bhandari M, Kellam J.** How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop Scand* 2004; 75(2): 184-94.
11. **Landis, J R, Koch G G.** The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1): 159-74.
12. **Marongiu G, Leinardi L, Congia S, Frigau L, Mola F, Capone A.** Reliability and reproducibility of the new AO/OTA 2018 classification system for proximal humeral fractures: a comparison of three different classification systems. *J Orthop Traumatol* 2020; 21(1): 4.
13. **Pallaver A, Honigmann P.** The role of cone-beam computed tomography (CBCT) scan for detection and follow-up of traumatic wrist pathologies. *J Hand Surg Am* 2019; 44(12): 1081-7.
14. **Kleinlugtenbelt Y V, Groen SR, Ham SJ, Kloen P, Haverlag R, Simons MP, et al.** Classification systems for distal radius fractures. *Acta Orthop* 2017; 88(6): 681-7.
15. **Arealis G, Galanopoulos I, Nikolaou VS, Lacon A, Ashwood N, Kitsis C.** Does the CT improve inter- and intra-observer agreement for the AO, Fernandez and Universal classification systems for distal radius fractures? *Injury* 2014; 45(10): 1579-84.
16. **Harness N G, Ring D, Zurakowski D, Harris G J, Jupiter J B.** The influence of three-dimensional computed tomography reconstructions on the characterization and treatment of distal radial fractures. *J Bone Joint Surg Am* 2006; 88(6): 1315-23.