

A retrospective evaluation of reliability and reproducibility of Arbeitsgemeinschaft für Osteosynthesefragen classification and Fernandez classification for distal radius fracture

Yan Yinjie, DO, MD^a, Wen Gen, DO, MD^b, Wan Hongbo, DO, MD^a, Xu Chongqing, MD^b, Zhaoxiang Fan, DO, MD^a, Feng Yanqi, DO, MD^a, Wu Xuequn, DO, MD^{a,*}, Mo Wen, DO, MD^{a,*}

Abstract

This is a retrospective evaluated. The objective of this study was to test the interobserver reliability and intraobserver reproducibility of fracture classification with Arbeitsgemeinschaft für Osteosynthesefragen (AO) system and Fernandez system used by 5 senior orthopedic surgeons.

Anteroposterior and lateral radiographs of 160 patients hospitalized for displaced distal radius fracture were examined. Independently, 5 orthopedic surgeons evaluated the radiographs according to 2 different distal radius classification systems (3 types of results). Three statistical tools were used to measure interobserver reliability and intraobserver reproducibility. The intraclass correlation coefficient and kappa coefficient (k) were used to assess both interobserver and intraobserver agreement of AO and Fernandez. Kappa value indicated poor agreement (<0), slight (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), good (0.61–0.80), and perfect (0.81–1.00).

The intraobserver reproducibility of AO system (9 types) and Fernandez system were moderate with a value of 0.577 and 0.438. The intraobserver reproducibility of AO system (27 subtypes) was 0.286. The interobserver reliability of AO system (9 types) was moderate with a value of 0.469 and that of Fernandez was moderate with a value of 0.435. The interobserver reliability of AO system (27 subtypes) was 0.299.

Neither of the 2 systems can give us a satisfactory agreement between interobserver reliability and intraobserver reproducibility. In AO system, the interobserver reliability and intraobserver reproducibility of the 9 types decreased with the increase of subgroups.

Abbreviation: AO = Arbeitsgemeinschaft für Osteosynthesefragen.

Keywords: classification, distal radius fracture, reliability, reproducibility

1. Introduction

Distal radial fracture is a common in orthopedics, making up 15% of all adult fractures.^[1] Currently, there are at least 20

Editor: Dennis Enix.

YY, WG and XC are co-first authors.

The authors have no conflicts of interest to disclose.

^aDepartment of Orthopaedics, LongHua Hospital, Shanghai University of Traditional Chinese Medicine, ^bDepartment of Orthopaedics, Shanghai University of Medicine & Health Sciences Affiliated Sixth People's Hospital East Campus, Shanghai, China.

*Correspondence: Wu Xuequn, Department of Orthopaedics, LongHua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China (e-mail: wxqlzj@163.com); Mo Wen, Department of Orthopaedics, LongHua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China (e-mail: mw2218@126.com).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Yinjie Y, Gen W, Hongbo W, Chongqing X, Fan Z, Yanqi F, Xuequn W, Wen M. A retrospective evaluation of reliability and reproducibility of Arbeitsgemeinschaft für Osteosynthesefragen classification and Fernandez classification for distal radius fracture. *Medicine* 2020;99:2(e18508).

Received: 4 April 2019 / Received in final form: 18 November 2019 / Accepted: 26 November 2019

<http://dx.doi.org/10.1097/MD.00000000000018508>

classification systems on distal radius fractures, but none can well describe the details, treatment, and prognosis of the fracture. The objective of this study was to test the interobserver reliability and intraobserver reproducibility of fracture classification with Arbeitsgemeinschaft für Osteosynthesefragen (AO) system and Fernandez system used by 5 senior orthopedic surgeons.

AO classification and Fernandez classification are commonly used. As one of many classifications systems for long bone fractures, AO was first published in 1987.^[2] This classification system is based on the severity of distal radius fractures. AO classifies distal radius fractures into the following types:

- (1) Type A (extra-articular): A1, ulna fractured and radius intact; A2, simple or impacted metaphyseal radial fracture; A3, comminuted metaphyseal radial fracture.
- (2) Type B (partial articular): B1, sagittal in radius; B2, frontal and dorsal radius; B3, frontal and volar radius.
- (3) Type C (complete joint): C1, simple joint and simple metaphysis; C2, simple joint and comminuted metaphysis; C3, multi-fragmented joint.

Each type is divided into 3 subtypes, then 27 subtypes in total (Fig. 1).

Fernandez system^[3] classifies distal radius fracture into 5 types according to the radiographic findings, the stress direction of fracture, and the mechanism of fracture. Type I, bending fracture of metaphysis; Type II, shearing fracture of the joint surface; Type



Figure 1. AO classification. AO = Arbeitsgemeinschaft für Osteosynthesefragen.

III, compression fracture of joint surface; Type IV, avulsion fracture or radiocarpal fracture dislocation; Type V, combined fracture associated with high-velocity injuries. The purpose of this study was to analyze the 2 methods' interobserver reliability and intraobserver reproducibility, their differences, and their use in clinical diagnosis (Fig. 2).

2. Subjects and methods

2.1. Patients

The research followed the principles of the Declaration of Helsinki and was approved by our hospital ethics committee. Database records of patients treated in our hospital for distal radius fractures between 2010 and 2017 were retrospectively collected and analyzed. Among these patients, 64 are males, 86

are females. At surgery, mean age was 49.5 years (range, 18–87 years). Anteroposterior and lateral radiographs of 160 hospitalized patients were examined (nondisplaced fractures were excluded). Radiographs were taken from a distance of 105 cm by a professional radiologist using the same radiographic apparatus. Orthopedic surgeons classified all 160 fractures with AO (types and subtypes) and Fernandez to determine intraobserver reproducibility (in different intervals by the same orthopedic surgeon) and interobserver reliability. For inclusion in this study, all the patients were required to have complete imaging studies and available clinical data. Complete and available clinical data obtained included demographic characteristics, chief complaint, complications, and surgical treatment. Because of the retrospective nature of the study, patient consent for inclusion was waived.

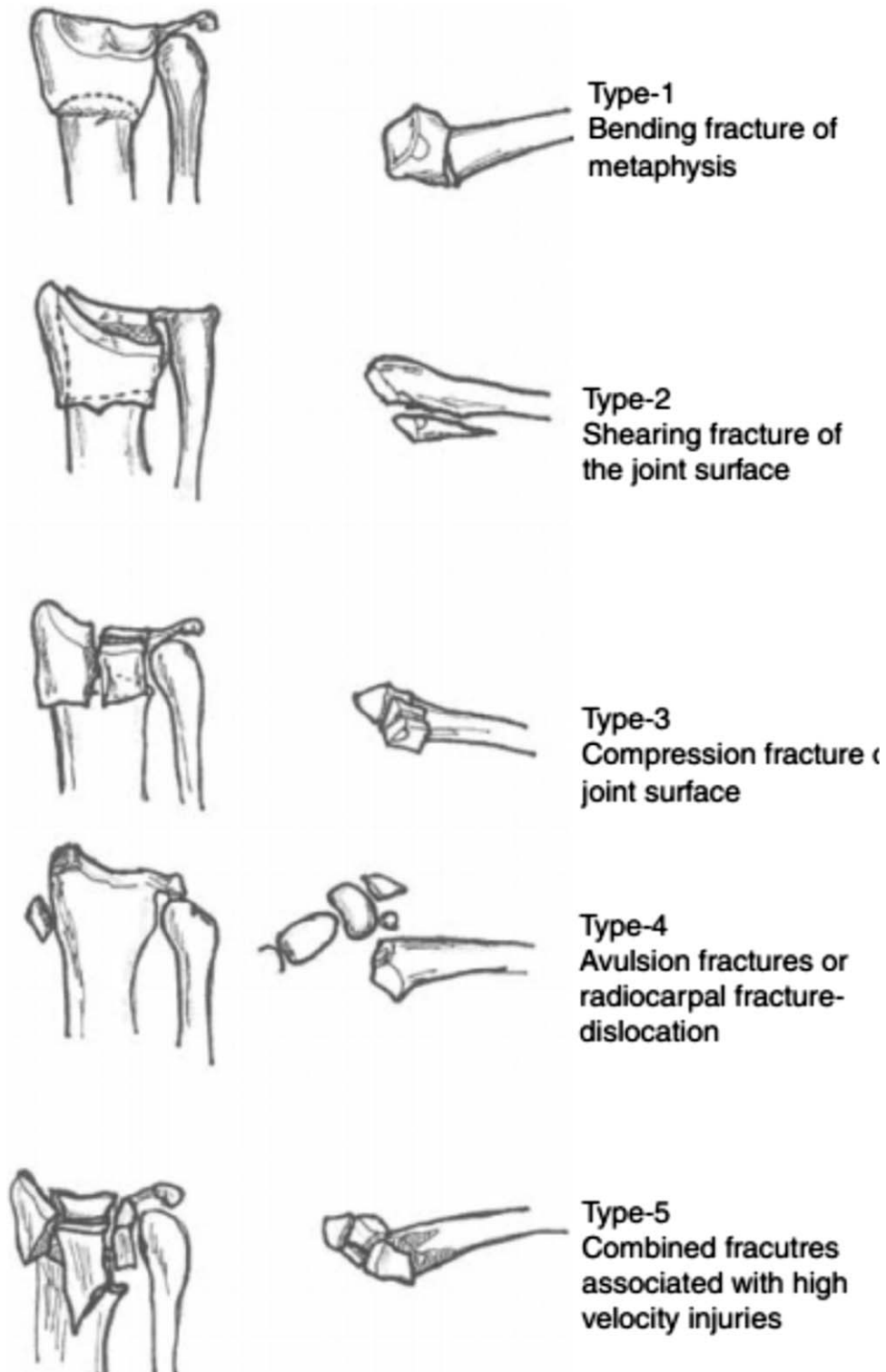


Figure 2. Fernandez classification.

Five orthopedic surgeons participated in this study. The radiographs were evaluated a first assessment according to 2 systems (3 types of results) by each specialist independently. Three months apart, the radiographical data were classified and scored again. The 2 sets of classification responses were recorded by a

physician who was not involved in the study, and then the reliability and reproducibility of both classification systems were analyzed.

Interobserver reliability was evaluated by comparing the initial classification of the 5 evaluators. Intraobserver reproducibility was evaluated by comparing 1 evaluator's classification to the

same case, with a 3-month interval and in a random order to minimize the recall bias.

2.2. Evaluators

All the 5 participants were senior orthopedic surgeons with more than 5 years' clinical experience and solid orthopedic knowledge. Scientific curiosity, and camaraderie were the only incentives for participation.

2.3. Inclusion criteria

- (1) Adult patients (>18 years old) admitted to our hospital between June 2010 and June 2017 for distal radius fractures; having complete imaging examination records and clinical data.
- (2) Closed fracture.
- (3) Having suffered fracture for less than 3 weeks.

2.4. Exclusion criteria

- (1) Incomplete clinical data.
- (2) Pathological fracture.
- (3) Having disorders of calcium and phosphorus metabolism.
- (4) Long-term use of hormones.

2.5. Statistical analysis

Three statistical tools were used to measure interobserver reliability and intraobserver reproducibility. The intraclass correlation coefficient (ICC) and kappa coefficient (*k*) were used to assess the inter- and intra- observer agreement of AO and Fernandez classification system (2-way mixed effect model in which people's effects are random and measures' effects are fixed).^[4] Reliability was tested by first evaluating whether the 5 assessors agreed with each other's assessments in both cycles of the study independently. This was cross-checked by another kappa test for ICC. Intraobserver reproducibility was evaluated further by comparing each individual assessor's performance between the first and second rounds using the Cohen kappa test.^[5]

We calculated the Cohen kappa coefficient with 95% confidence interval (CI) for all 3 classification systems (3 types of results). Kappa is a measurement used for determining the level of agreement in categorical variables corrected for chance. Kappa coefficients were evaluated according to Landis and Koch classification. According to this classification, Kappa value indicated poor agreement (<0), slight (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), good (0.61–0.80), and perfect (0.81–1.00).^[6] SPSS 20.0 statistical software was used for the

Table 1

Level of agreement.

K value	Level of agreement
0.00	Poor agreement
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Good
0.81–1.00	Near perfect

Table 2

Reproducibility of AO and Fernandez.

Evaluator	K value of AO system (9 groups)	K value of AO system (27 subgroups)	K value of Fernandez system
A	0.631	0.318	0.434
B	0.612	0.395	0.511
C	0.53	0.197	0.449
D	0.593	0.241	0.407
E	0.522	0.283	0.391
Mean value	0.577	0.286	0.438

AO = Arbeitsgemeinschaft für Osteosynthesefragen.

statistical analysis, all the values were expressed with a 95% CI and *P*-values of <.05 were considered to be statistically significant (Table 1).

3. Results

3.1. Intraobserver reproducibility

The intraobserver reproducibility of AO system (9 types) classification was moderate with a value of 0.577. The intraobserver reproducibility of the Fernandez system was moderate with a value of 0.438. These values meant a moderate agreement. But the intraobserver reproducibility of AO system (27 subtypes) was 0.286, suggesting a fair agreement (Table 2).

3.2. Interobserver reliability

The interobserver reliability of AO s (9 types) classification was moderate with a value of 0.469. And that of Fernandez was moderate with a value of 0.435. These values meant a moderate agreement. But the interobserver reliability of AO system (27 subtypes) was 0.299, meaning a fair agreement (Table 3).

4. Discussion

4.1. Main findings of the present study

AO/The Association for the Study of Internal Fixation (ASIF) is the most detailed and comprehensive classification method that digitizes and systemizes the condition of fracture. As many as 27 subtypes of distal radius fractures have been classified. As the

Table 3

Reliability of AO classification systems and Fernandez classification systems.

Evaluator	K value of AO (9 types)	K value of AO (27 subtypes)	K value of Fernandez system
A-B	0.512	0.328	0.54
A-C	0.446	0.337	0.613
A-D	0.463	0.275	0.421
A-E	0.398	0.267	0.338
B-C	0.371	0.314	0.417
B-D	0.507	0.368	0.451
D-E	0.418	0.276	0.39
C-D	0.572	0.31	0.431
C-E	0.611	0.204	0.337
D-E	0.396	0.311	0.408
Average value	0.469	0.299	0.435

AO = Arbeitsgemeinschaft für Osteosynthesefragen.

number of groups and subgroups increase – that is the fracture morphology becomes more complex further investigations are needed to classify the fracture, making it less clinically applicable. In our research, the reproducibility and reliability of AO/ASIF were 0.577 and 0.469 among 3 types, and 0.286 and 0.299 among all subtypes. A declining trend was obvious. A related research also shows that the good reproducibility of the AO/ASIF classification of the distal radius fractures is only manifested in type A, B, and C, not in the subtypes.^[7]

Fernandez system is also a common radial distal fracture typing that illustrates the stress direction of fracture and the underlying mechanism. According to the injury severity of the bone and surrounding soft tissues, a clear classification from type I to type V conveniences the diagnosis and treatment.^[8] However, this system does not cover the severity of comminuted articular fractures, not to mention its diagnosis and treatment. In this study, the reproducibility and reliability of the Fernandez classification were 0.438 and 0.435. This result is not as satisfactory as the AO classification.

4.2. Comparison with other studies

Before this research, we reviewed the literature on the reliability and reproducibility of distal radius fracture. Few results were found. The smallest research only involves 5 cases, and the biggest involved 96 cases. Their researchers include doctors, radiologists, general surgeons, and other professionals. But some of them lack of a complete grasp of the orthopedic knowledge system. So in this study, we maximized the sample size as possible to reduce statistical errors. Five orthopedic surgeons were invited into the research because they had more than 5 years' clinical experience and an in-depth understanding of the mechanism and types of fractures. To better reflect the interobserver reliability and intraobserver reproducibility of participants. We deliberately analyzed the data of AO-classified fractures (9 types and 27 subtypes) and compared them with the results of Fernandez system. Then the trend of reliability and repeatability of the same classification with more subtypes was studied. The limitations of this paper are as follows:

- (1) No more orthopedic surgeons are invited to participate in the assessment.
- (2) There are too few types of classification for the study, and it is not possible to comprehensively evaluate the common classification of distal radius fractures.

4.3. Implication and explanation of findings

A scientific fracture classification system should be simple and informative enough to define the mechanism, guide the treatment, predict the prognosis, and show a perfect agreement between interobserver reliability and intraobserver reproducibility.^[9]

Through our research, we found that the above 2 systems cannot bring a satisfactory agreement between interobserver reliability and intraobserver reproducibility. And in the AO/ASIF classification, with the increase of subgroups, its reliability and reproducibility decreased. In this study, interobserver reliability of AO (9 types) and Fernandez was fair, which is consistent with the results of former literature.^[10]

4.4. Conclusion, recommendation, and future directions

We believe that the classification of the distal radius fractures simply based on X-ray films is not applicable enough and limitation. We speculate that computed tomography (CT) scanning should be introduced. Through the combination of digital multi-slice spiral CT, a more scientific classification system should be developed to reveal the multi-planar and 3-dimensional characteristics of the fractures, and eventually improve the typing, diagnosis, and treatment of clinical physicians. Because In fact at times a final classification cannot be made till after the fracture is visualized at surgery.

Author contributions

Conceptualization: Yan Yinjie, Zhaoxiang Fan, Mo Wen.

Data curation: Wen Gen, Zhaoxiang Fan, Feng Yanqi.

Formal analysis: Zhaoxiang Fan, Feng Yanqi.

Funding acquisition: Yan Yinjie.

Investigation: Wen Gen, Wan Hongbo.

Methodology: Wu Xuequn.

Software: Wan Hongbo.

Supervision: Wu Xuequn.

References

- [1] Davis DI, Baratz M. Soft tissue complications of distal radius fractures. *Hand Clin* 2010;2:229–35.
- [2] Müller ME, Nazarian S, Koch P. Classification AO Des Fractures: Tome I: Les Os Longs. 1st ed. Berlin, Germany: Springer-Verlag; 1987. Part 2.
- [3] Fernandez DL, Jupiter J. Fractures of the Distal Radius: A Practical Approach to Management. New York, NY: Springer Verlag; 1996.
- [4] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
- [5] Naqvi SGA, Reynolds T, Kitsis C. Interobserver reliability and intraobserver reproducibility of the fernandez classification for distal radius fractures. *J Hand Surg* 2009;34:483–5.
- [6] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [7] vanLeerdam RH, Souer JS, Lindenhovius ALC, et al. Agreement between initial classification and subsequent reclassification of fractures of the distal radius in a prospective cohort study. *Hand* 2010;5:68–71.
- [8] Fernandez DL. Distal radius fracture: the rationale of a classification. *Chirurgie de la Main* 2001;20:411–25.
- [9] Kural C, Sungur I. Evaluation of the reliability of classification systems used for distal radius fractures. *Orthopedics* 2010;11:801.
- [10] Belloti JC, Tamaoki MJ, Franciozi CE, et al. Are distal radius fracture classifications reproducible Intra and interobserver agreement. *Sao Paulo Med J* 2008;126:180–5.