



# Evaluation of Intra- and Interobserver Reproducibility of the New AO/OTA Classification for Distal Radius Fractures Compared with the Fernandez Classification

## *Avaliação da reprodutibilidade intra- e interobservador da nova classificação AO/OTA para fraturas distais do rádio comparada com a classificação de Fernandez*

Beatriz Canhoto Carula<sup>1,2</sup> Giana Silveira Giostri<sup>1</sup> Fabiana Beber<sup>1</sup>  
Marianna Fergutz dos S. Batista<sup>2</sup> Flaviana Busignani da Silva<sup>1</sup> Eduardo Murilo Novak<sup>1</sup>

<sup>1</sup>Hospital Universitário Cajuru, Curitiba, PR, Brasil

<sup>2</sup>Hospital Universitário Evangélico Mackenzie, Curitiba, PR, Brasil

Address for correspondence Beatriz Canhoto Carula, MD, Rua Desembargador Mota, 1648, Curitiba, PR, Brazil (e-mail: carula.beatriz@gmail.com).

Rev Bras Ortop 2022;57(6):917–923.

### Abstract

**Objective** To evaluate the inter- and intraobserver reliability and reproducibility of the new AO/OTA 2018 classification for distal radius fractures and to compare it with the Fernandez classification system.

**Method** A questionnaire was applied in the Qualtrics software on 10 specialists in hand surgery who classified 50 radiographs of distal radius fractures according to the Fernandez and AO/OTA 2018 classifications and, subsequently, indicated their treatment. The questionnaire was applied in time *T0* and repeated after 4 weeks (*t1*). The mean agreement between the answers, and the reliability and inter- and intraobserver reproducibility were analyzed using kappa indexes.

**Results** The mean interobserver agreement in the Fernandez classification was 76.4, and it was 59.2% in the AO/OTA 2018 classification. The intraobserver agreements were 77.3 and 56.6%, respectively. The inter- and intraobserver kappa indexes for the Fernandez classification were 0.57 and 0.55, respectively, and, in the AO/OTA 2018 classification, they were 0.34 and 0.31, respectively.

**Conclusion** The AO/OTA 2018 classification showed a low intra- and interobserver reproducibility when compared with the Fernandez classification. However, both

### Keywords

- ▶ radius fractures/ classification
- ▶ wrist injuries
- ▶ reproducibility of results
- ▶ surveys and questionnaires

Work developed at the Hospital Universitário Cajuru, Curitiba, PR, Brazil

received  
March 23, 2021  
accepted after revision  
August 13, 2021  
published online  
September 26, 2022

DOI <https://doi.org/10.1055/s-0041-1740471>.  
ISSN 0102-3616.

© 2022. Sociedade Brasileira de Ortopedia e Traumatologia. All rights reserved.

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Thieme Revinter Publicações Ltda., Rua do Matoso 170, Rio de Janeiro, RJ, CEP 20270-135, Brazil

## Resumo

### Palavras-chave

- ▶ fraturas do rádio/classificação
- ▶ traumatismos do punho
- ▶ reprodutibilidade dos testes
- ▶ inquéritos e questionários

classifications have low intra- and interobserver indexes. Although the Fernandez classification did not obtain excellent results, it remains with better agreement for routine use.

**Objetivo** Avaliar a confiabilidade e a reprodutibilidade inter- e intraobservadores da nova classificação AO/OTA 2018 para fraturas distais do rádio e compará-la com o sistema classificatório de Fernandez.

**Métodos** Foi aplicado um questionário no software Qualtrics em 10 especialistas em cirurgia da mão que classificaram 50 radiografias de fraturas distais de rádio de acordo com as classificações de Fernandez e AO/OTA 2018 e, posteriormente, indicaram seu tratamento. Esse questionário foi aplicado em tempo *T0* e repetido após 4 semanas (*t1*). Analisou-se a média de concordância entre as respostas e confiabilidade e reprodutibilidade inter- e intraobservadores utilizando os índices kappa.

**Resultados** A concordância média interobservador para a classificação de Fernandez foi de 76,4, e de 59,2% para a AO/OTA 2018. A concordância intraobservador foi de 77,3 e 56,6%, respectivamente. O índice de kappa inter- e intraobservador para a classificação de Fernandez foram de 0,57 e de 0,55, respectivamente, e a classificação AO/OTA 2018 obteve 0,34 e 0,31, respectivamente.

**Conclusão** A classificação AO/OTA 2018 mostrou uma reprodutibilidade intra- e interobservadores baixa quando comparada à classificação de Fernandez. Porém, ambas as classificações apresentam índices intra- e interobservadores baixos. Embora a classificação de Fernandez não tenha obtido resultados excelentes, ela permanece com melhor concordância para o uso rotineiro.

## Introduction

Distal radius fracture (DRF) is one of the most common fractures, representing 12% of all fractures in the Brazilian population, associated or not with ulna fractures.<sup>1-5</sup> It is considered a public health problem because it affects young men due to high-energy trauma, and the elderly due to bone fragility.<sup>5-8</sup> In most cases, for the correct diagnosis, radiographs of the wrist in posteroanterior (PA), lateral (L), and oblique<sup>8</sup> incidences are sufficient to establish the appropriate treatment without excessively burdening the health system.<sup>5,8</sup>

To be clinically useful for radiography evaluation, a classification system should be comprehensive and simple, besides having intraobserver reliability and interobserver reproducibility.<sup>2,9-12</sup> In 1967, the iconic classification by Frykman was published, based on simple features of radiographic anatomy. Subsequently, a series of classification systems for DRFs followed, including the classifications proposed by Melone,<sup>13</sup> Fernández,<sup>14</sup> Universal (Cooney, 1993)<sup>15</sup>, and the AO group (2007),<sup>16</sup> which basically ordered the radiographic characteristics of these lesions.<sup>10</sup>

Currently, due to the particularities of DRFs, there is no consensus on what the best classification would be.<sup>6,17-19</sup> The AO/OTA classification (2007) is widespread among specialists and, perhaps, the most cited in articles in the literature. Easy to use, it orders the most common possibilities of DRFs without relating to the mechanism of trauma or offering a prognostic idea for the lesions. In this sense, the classification proposed by Fernandez brings elements that establish this connection with the initial trauma and the

prognosis, tending to identify DRFs more widely; however, it seems to present low reproducibility in the communication between specialists.<sup>4</sup>

In 2018, the AO/OTA group updated their classification with the addition of qualifiers and modifiers in each subtype to offer more possibilities in the identification of DRFs.<sup>9,20</sup> This new AO classification was more complete, but increased the complexity in its application.<sup>21</sup>

The purpose of the present study is to evaluate the reproducibility of the AO/OTA 2018 classification among experienced surgeons and to compare it to the Fernandez classification, which is the one already used by the group of authors. Present study also has the objective of evaluating the influence of the use of these classifications on decision-making for the treatment of DRFs.

## Materials and Method

A questionnaire was applied, through the online application *Qualtrics*, containing questions about 50 radiographs of distal radius fractures that should be analyzed using 2 different methods, the Fernandez and the AO/OTA 2018 systems, in order to classify the radiographs and choose the treatment for each one.

Radiographic images were taken retrospectively from the medical records of a trauma reference hospital. They were digital with good resolution and were standardized in the anteroposterior, profile, pronated oblique, and supinated oblique incidences of acute fractures of the distal third of the radius (with or without associated ulna fracture) in

patients > 18 years of age with mature skeleton between October 2018 and March 2020. Radiographs of patients < 18 years old, with previous fractures, and poor-quality radiographs were excluded.

The images were randomly selected by hand surgeon orthopedists who were not evaluators. The collected images had their identification and date hidden throughout the questionnaire, being identified only with a number, and were randomly sampled to reduce the bias of the evaluators regarding the intraobserver reproducibility test.

Ten observers specialized in hand surgery, from different regions of Brazil and with > 10 years of training in the specialty, were invited to voluntarily evaluate the images. Initially, 11 specialists started the study, but only 10 completed all stages. A copy of both classifications was made available for consultation (► **Appendix 1**, supplementary material)). Each evaluator individually answered a block of 5 questions for each of the 50 radiographs, without access to the answers of the others. ► **Figures 1** and **2** contain an example of the guides provided.

The questionnaire started with the mandatory filling of an Informed Consent Form (TCLE) and identification of the evaluator. Each of the 50 radiographs belonged to a block with 5 questions (► **Appendix 2**, supplementary material) asking the type of fracture according to the Fernandez classification, containing 5 alternatives and a single answer (types 1, 2, 3, 4, and 5). Next, the choice of preference for treatment was asked, also with a single response, containing the following alternatives: (1) Reduction and conservative treatment (plaster); (2) Reduction and percutaneous pinning; (3) Surgical reduction and fixation with palmar "T" plate associated or not with Kirschner wires; (4) Reduction and fixation with locked volar plate; (5) Surgical reduction and fixation with blocked dorsal plate; (6) Others.

The following question contained the same x-rays, now to be classified according to the AO/OTA 2018 classification, with a single answer (2R3A1, 2R3A2, 2R3A3, 2R3B1, 2R3B2, 2R3B3, 2R3C1, 2R3C2, and 2R3C3), as well as an answer for ulna fracture, in case it was present, with the options: 2U3A1, 2U3A2, 2U3A3, 2U3B, and 2U3C or "Does not apply".

Next, the 2018 AO/OTA classification modifiers were asked, with the following options: (0) Does not apply; (1) Not diverted; (2) Diverted, (3a) Joint impaction; (3b) Metaphyseal impaction; (4) Not impacted; (5a) Previous diverted (Volar); (5b) Posterior diverted (Dorsal); (5c) Diverted ulnar; (5d) Radial diverted; (5e) Multidirectional diverted; (6a) Subluxation – Volar Ligament Instability; (6b) Subluxation

– Dorsal Ligament Instability; (6c) Subluxation – Ulnar Ligament Instability; (6d) Subluxation – Radial Ligament Instability; (6e) Subluxation – Multidirectional Ligament Instability; (7) Diaphysary Extension; (8) Low Bone Quality. In this question, the evaluator could choose more than one alternative. The number of options marked for further analysis was considered: (1) Only one modifier option; (2) Two modifier options; (3) Three modifier options; (4) Four or more modifier options.

The last question of each block of radiographs questioned whether, after classifying the same fracture using the AO/OTA 2018 method, the specialist would change or maintain the initial treatment they had chosen after classifying the same radiography according to the Fernandez classification.

The first application of the questionnaire occurred concomitantly with the 10 evaluators, considered as time  $t_0$ . After 4 weeks, the evaluators answered the questionnaire again, with the same 50 radiographs in a different order from the previous one, in what was termed as time  $t_1$ .

The mean agreement between the observers in the answers to each question alone was considered, and whether the change in classifications implied the change of conduct and treatment in the interobserver analysis, in times  $t_0$  and  $t_1$ . The index was defined as excellent if > 75% of the participants agreed with the same answer, as satisfactory if the agreement ranged from 50 to 75%, and as unsatisfactory if < 50% of the participants agreed.<sup>21</sup>

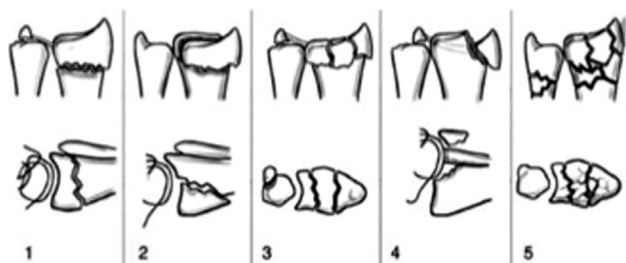
To evaluate the reliability and the reproducibility of the classifications and to compare their applicability, we tested the interobserver reproducibility, which analyzed the agreement between the 10 evaluators regarding the same fracture in relation to the chosen classifications, comparing the answers of all questions, and of all observers, in both cycles ( $t_0$  and  $t_1$ ).

The intraobserver reproducibility was tested by comparing the level of agreement of the same observer when answering the same questions at two different times (times  $t_0$  and  $t_1$ ). Considering that one evaluator was absent from the questionnaire in time  $t_1$ , we had the comparative analysis of only nine examiners in this second stage.

The evaluation of the consistency of the inter- and intra-observer responses was used using two parameters: the proportion of agreement and the kappa index. The first is the average percentage of cases on which the evaluators agreed. The second is used to evaluate the agreement between the observers and involves adjusting the observed proportion of agreement by correcting the proportion of agreement that arises in each case.<sup>1,7,17,23</sup>

The calculation of kappa indexes was performed in Online Kappa Calculator (Justus Randolph) with data analysis using *free-marginal kappa*, since the evaluators remained free to choose the answers.<sup>20</sup> Traditionally, the kappa coefficient values, interpreted by Landis et al., range from 0 to 1, with 1 being equal to perfect agreement and 0 corresponding to no agreement, as specified in ► **Table 1**.<sup>1,7,23,24</sup>

The present study was approved by the Ethics Committee of the institution under the number CAAE 22570419.0.0000.0020.



**Fig. 1** Fernandez classification.

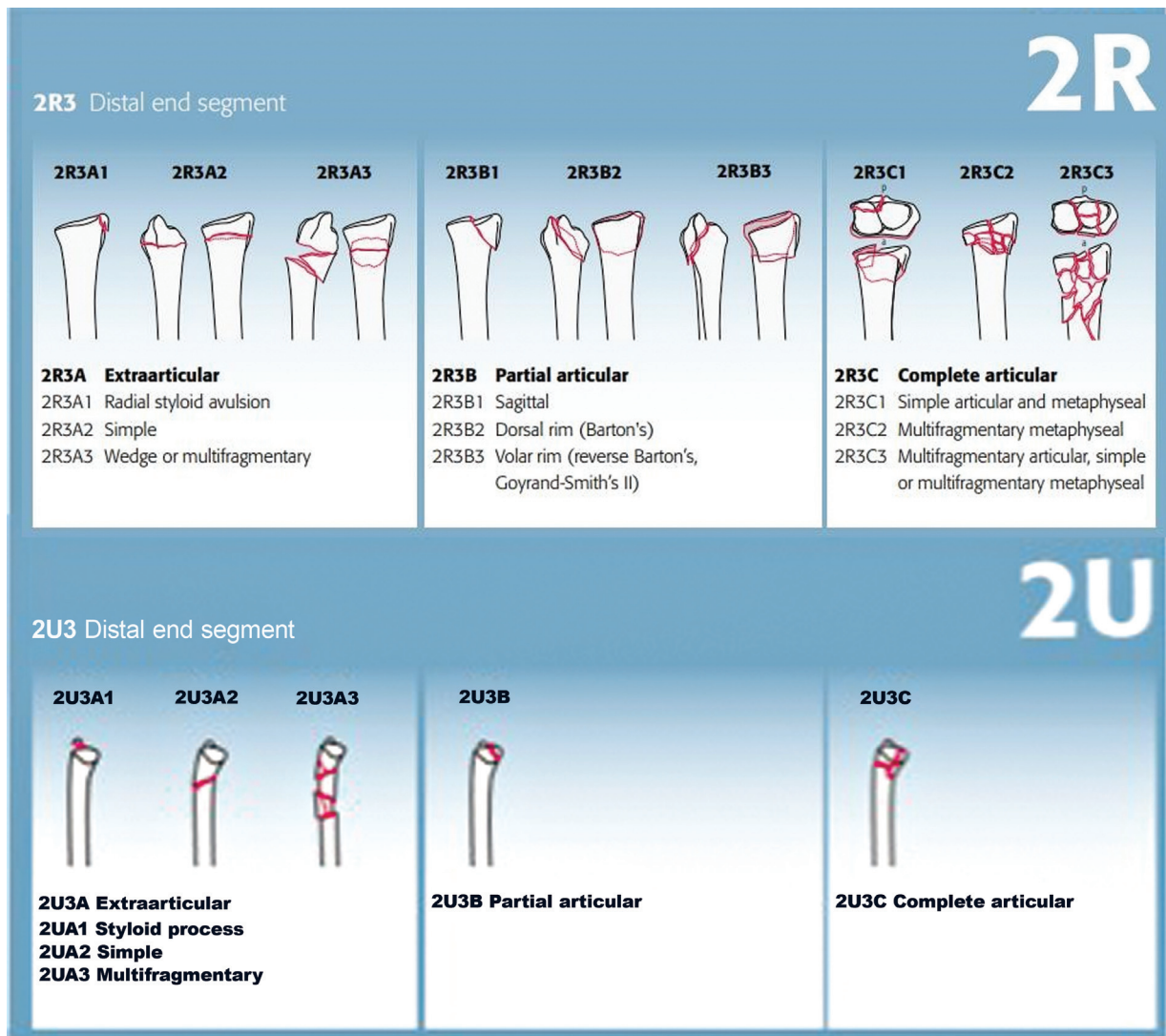


Fig. 2 AO/ OTA 2018 classification.

**Results**

The analysis of the responses of the initial block referring to the Fernandez classification showed an inter- and intraobserver mean agreement, respectively, of 76.40 and 77.33%. In terms of maintaining the same treatment in both questionnaires (*t0* and *t1*), agreement was 62% in the interobserver assessment and 64.7% intraobserver. In the AO/OTA 2018 classification, the inter- and intraobserver agreement for the radius fracture segment were 59.2 and 56.66%, respectively. In the ulna fractures segment, 81.2 and 80.44% intra- and interobserver fractures were obtained, respectively. And for the segment concerning the number of modifiers used, the agreement was 52.6 and 49.55%. After classification by the AO/OTA system in the radio, ulna and modifier segments, 95.4% of the evaluators maintained the treatment indication based on the Fernandez classification. Considering the intraobserver agreement for this item, 94.17% of the examiners maintained their treatment option (→ Table 2).

The overall average of interobserver agreement was considered moderate for the Fernandez classification and low for

Table 1 Landis et al. interpretation for kappa values<sup>23</sup>

Landis et al. interpretation for kappa values	
Kappa values	Interpretation
< 0	No agreement
0–0.19	Bad agreement
0.2–0.39	Low agreement
0.4–0.59	Moderate agreement
0.6–0.79	Substantial/good agreement

the AO/OTA 2018 classification. The treatment option after the evaluators had classified according to the classification Fernandez obtained low agreement, and the maintenance of treatment after the AO/OTA 2018 classification was considered excellent (→ Table 3). The result of the mean intra- and interobserver agreement with all participants is found in → Appendix 3 (supplementary material).

**Table 2** Mean intra- and interobserver agreement according to each evaluated item

Percentage of Average Agreement						
	Fernandez classification	Treatment option	AO/OTA Rating 2018 radius segment	AO/OTA Rating 2018 ulna segment	AO/OTA Rating 2018 segment quantity of modifiers	Maintenance of the treatment option
INTEROBSERVER	76.40%	62.00%	59.20%	81.20%	52.60%	95.40%
INTRAOBSERVER	77.33%	61.77%	56.66%	80.44%	49.55%	94.17%

► **Table 4** shows the results of the intraobserver interpretation. Kappa values remained similar to those from time *t0*. Consistency was observed between the values of general percentage agreement (percent overall agreement) in *t0* and *t1*.

Regarding the AO/OTA 2018 ulna segment classification, that is, the association of ulnar fracture, kappa indexes of 0.64 and 0.61 were obtained in time *t0* and *t1*, respectively, as observed in ► **Table 5**.

Regarding the modifying segment of the AO/OTA 2018 classification, which obtained kappa values of 0.18 in time *t0* and 0.17 in time *t1* (► **Table 5**), the agreement was considered poor in relation to the number of modifiers used to classify each fracture. When we individually evaluated the use of modifiers, it was noticed that the most selected data was option "2 - Diverted", followed by option "5b - Subsequent Diverted", as detailed in ► **Table 6**.

**Discussion**

The data obtained in the present study showed an average intraobserver percentage agreement of 77.3% with the Fernandez classification and of 56.6% with the AO/OTA 2018, and an interobserver agreement of 76.4 and 59.2%, respectively. This comparison was made with the part of the AO/OTA 2018 classification that evaluates the radius segment because this system separates into segments for isolated evaluation of the radius and the ulna. Analyzing the kappa index for inter- and intraobserver reproducibility, the Fernandez classification resulted in moderate agreement (0.57

**Table 3** Interpretation of the interobserver kappa values

Interpretation of Interobserver Kappa Values		
Online Kappa Calculator		
	Kappa Index	Agreement
Fernandez classification	0.57	MODERATE
Choice of treatment	0.33	LOW
AO/OTA Rating 2018 - Radius Segment	0.34	LOW
Maintenance of the treatment option	0.83	EXCELLENT

and 0.59, respectively), and the AO/OTA 2018 classification resulted in low agreement (0.34 and 0.31, respectively). The same results were found by Naqbi et al.<sup>4</sup> regarding the Fernandez classification, demonstrating moderate agreement in the evaluation of 25 radiographs by specialists, and in the work by Van Leerdam et al.<sup>25</sup> regarding the AO/OTA classification, which detected a low agreement.

Corroborating our study, Yinjie et al.<sup>21</sup> published an intra- and interobserver comparison between the Fernandez and the AO/OTA 2018 classifications in which 5 experienced surgeons evaluated 160 radiographic images. Their results were comparable to ours: moderate intraobserver reproducibility with the Fernandez classification, and low with the AO. They established that the reproducibility of the AO classification decreases with the increase of subgroups, modifiers, and qualifiers. When comparing the classification proposed by Waever et al., they proved that there was no

**Table 4** Interpretation of intraobserver kappa values

Interpretation of Intraobserver Kappa Values (General Percentage Agreement)					
Online Kappa Calculator					
	Kappa index in <i>t0</i>	Kappa index in <i>t1</i>	Agreement	General percentage agreement in <i>t0</i>	General percentage agreement in <i>t1</i>
Fernandez classification	0.57	0.55	Moderate	65.28%	63.69%
Choice of treatment	0.33	0.36	Low	44.22%	46.27%
AO/OTA Rating 2018- Radius Segment	0.34	0.31	Low	41.02%	38.61%
Maintenance of the treatment option	0.83	0.81	Excellent	91.47%	92.89%

**Table 5** Interpretation of kappa p values for associated ulna fractures

Interpretation of the Agreement Values of the AO/OTA Classification 2018 - Ulnar Segment and Segment Quantity of Modifiers					
Online Kappa Calculator					
	Kappa index in t0	Kappa index in t1	Agreement	General Percentage Agreement at t0	General Percentage Agreement in t1
AO/OTA Classification 2018 - Ulna Segment	0.64	0.61	Substantial/Good	69.82%	67.67%
Segment Quantity modifiers	0.18	0.17	Bad	34.53%	33.56%

**Table 6** Percentage of modifiers selected by examiners in t0 and t1

Modifier	time t0 (%)*	time t1 (%)*	Modifier	time t0 (%)*	time t1 (%)*
(0) Does not apply	0.45%	0.28%	(5d) Radial diverted	6.00%	6.32%
(1) Not diverted	1.29%	1.19%	(5e) Diverted multidirectional	2.51%	1.97%
(2) Diverted	25.43%	23.33%	(6a) Subluxation - Volar Ligament Instability	1.16%	1.48%
(3a) Joint impaction	12.07%	11.24%	(6b) Subluxation - Dorsal Ligament Instability	1.74%	2.53%
(3b) Metaphyseal impaction	13.23%	13.00%	(6c) Subluxation - Ulnar Ligament Instability	0.96%	0.49%
(4) Not impacted	2.32%	1.76%	(6d) Subluxation - Radial Ligament Instability	0.38%	0.28%
(5a) Previous diverted (Volar)	3.80%	4.08%	(6e) Subluxation - Multidirectional Ligament Instability	0.58%	0.35%
(5b) Posterior diverted (Dorsal)	19.04%	21.71%	(7) Diaphysarian Extension	1.87%	2.25%
(5c) Diverted ulnar	1.03%	1.41%	(8) Low Bone Quality	6.06%	6.32%

superiority in the results of reproducibility. These authors studied 573 radiographs of patients with DRF seeking to point out a unified and universal classification system.<sup>2</sup>

It is important to highlight the increase in the intra-observer agreement in the analysis comparing times *t0* and *t1*. After the classification is assimilated by the evaluator, it tends to be used in a more reproducible way. It was noticed that the fewer the options to choose in the classification, the higher will be the agreement in its use. In the evaluation of the ulnar segment of the AO/OTA 2018 classification, ~ 80% of agreement with only 6 options was obtained. Compared with the result of the use of modifiers, which comprise > 15 options, it was observed that the agreement fell to ~ 50%. Therefore, the agreement tends to decrease with more choices in the classification, as has already been pointed out.

The classification does not seem to interfere with the choice of treatment. After classifying according to the Fernandez classification, the evaluators opted for one conduct for each presented DRF. This conduct was maintained in almost all cases (94%) after the evaluation according to the AO/OTA 2018. At this point, the question if knowledge and personal experi-

ence tend to be preferable to the use of classification systems in the prediction of prognosis and decision-making of DRF treatment arises.<sup>26</sup> In a multicenter study, Mulders et al. emphasized that there it is unlikely that a consensus on the treatment of these fractures if guided specifically by classification systems will be reached, since surgeons will always tend toward strategies based on their experiences.<sup>9,27,28</sup>

Another point to be discussed is the familiarity with the proposed classification. The Fernandez system has been used since the early 2000s, which may explain the most reproducible results. Perhaps, the proposed AO/OTA 2018 system, due to its greater richness of detail and more options for the identification of DRFs, will increase its reproducibility over time and its use will become disseminated. It is worth mentioning that none of the various classification systems for DRF have high reproducibility.<sup>21</sup>

We found several limitations in the present study. We highlight the free application of the questionnaire, which may evidence a response bias, as well as that unlimited time was granted to answer the questionnaire, which allowed the examiners to pause the survey and resume it after a deadline of one week. Also, the

only quantitative and nonqualitative evaluation of the modifiers of the AO/OTA 2018 classification, which leads to more details and difficulty to classify. In addition, we evaluated hand surgeons at only one level of training, although they were experienced. Another possible limitation was to not have evaluated which classification the examiner was accustomed to, because the higher the familiarity, the greater the reproducibility in its use. Ideally, we recommend further research on a larger scale, with evaluators of different levels of experience, with a higher sample number analyzed in the same period, and also with grouping of multiple subtypes, to be performed in order to possibly increase the consistency of the results obtained.

## Conclusion

In the present study, the classifications studied did not present high agreement in inter- and intraobserver reproducibility. It is suggested that the complexity and detail of the new AO/OTA 2018 classification is the cause of its low reproducibility when compared with that of the system proposed by Fernandez.

### Financial Support

The present survey has not received any specific funding from public, commercial, or not-for-profit funding agencies.

### Conflict of Interests

The authors have no conflict of interests to declare.

## References

- Tenório PHM, Vieira MM, Alberti A, Abreu MFM, Nakamoto JC, Cliquet A. Evaluation of intra- and interobserver reliability of the AO classification for wrist fractures. *Rev Bras Ortop* 2018;53(06):703–706
- Wæver D, Madsen ML, Rölfing JHD, et al. Distal radius fractures are difficult to classify. *Injury* 2018;49(Suppl 1):S29–S32
- Plant CE, Hickson C, Hedley H, Parsons NR, Costa ML. Is it time to revisit the AO classification of fractures of the distal radius? Inter- and intra-observer reliability of the AO classification. *Bone Joint J* 2015;97-B(06):818–823
- Naqvi SG, Reynolds T, Kitsis C. Interobserver reliability and intra-observer reproducibility of the Fernandez classification for distal radius fractures. *J Hand Surg Eur Vol* 2009;34(04):483–485
- Alffram PA, Bauer GC. Epidemiology of fractures of the forearm. A biomechanical investigation of bone strength. *J Bone Joint Surg Am* 1962;44-A:105–114
- Brogan DM, Richard MJ, Ruch D, Kakar S. Management of Severely Comminuted Distal Radius Fractures. *J Hand Surg Am* 2015;40(09):1905–1914
- Shehovych A, Salar O, Meyer C, Ford DJ. Adult distal radius fractures classification systems: essential clinical knowledge or abstract memory testing? *Ann R Coll Surg Engl* 2016;98(08):525–531
- Arora R, Gabl M, Gschwentner M, Deml C, Krappinger D, Lutz M. A comparative study of clinical and radiologic outcomes of unstable colles type distal radius fractures in patients older than 70 years: nonoperative treatment versus volar locking plating. *J Orthop Trauma* 2009;23(04):237–242
- Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Fracture and Dislocation Classification Compendium-2018. *J Orthop Trauma* 2018;32(Suppl 1):S1–S170
- Kanakaris NK, Lasanianos NG. Distal Radial Fractures. In: Lasanianos NG, Kanakaris NK, Giannoudis PV, editors. *Trauma and Orthopaedic Classifications*. New York: Springer; 2014:95–105
- Jayakumar P, Teunis T, Giménez BB, Verstreken F, Di Mascio L, Jupiter JB. AO Distal Radius Fracture Classification: Global Perspective on Observer Agreement. *J Wrist Surg* 2017;6(01):46–53
- Lee DY, Park YJ, Park JS. A Meta-analysis of Studies of Volar Locking Plate Fixation of Distal Radius Fractures: Conventional versus Minimally Invasive Plate Osteosynthesis. *Clin Orthop Surg* 2019;11(02):208–219
- Melone CP Jr. Articular fractures of the distal radius. *Orthop Clin North Am* 1984;15(02):217–236
- Fernandez DL. Distal radius fracture: the rationale of a classification. *Chir Main* 2001;20(06):411–425
- Cooney WP. Fractures of the distal radius. A modern treatment-based classification. *Orthop Clin North Am* 1993;24(02):211–216
- Marsh JL, Slongo TF, Agel J, et al. Fracture and dislocation classification compendium - 2007: Orthopaedic Trauma Association classification, database and outcomes committee. *J Orthop Trauma* 2007;21(10, Suppl):S1–S133
- Illarramendi A, González Della Valle A, Segal E, De Carli P, Maignon G, Gallucci G. Evaluation of simplified Frykman and AO classifications of fractures of the distal radius. Assessment of interobserver and intraobserver agreement. *Int Orthop* 1998;22(02):111–115
- Koval K, Haidukewych GJ, Service B, Zirgibel BJ. Controversies in the management of distal radius fractures. *J Am Acad Orthop Surg* 2014;22(09):566–575
- Thurston AJ. 'Ao' or eponyms: the classification of wrist fractures. *ANZ J Surg* 2005;75(05):347–355
- Porrino JA Jr, Maloney E, Scherer K, Mulcahy H, Ha AS, Allan C. Fracture of the distal radius: epidemiology and premanagement radiographic characterization. *AJR Am J Roentgenol* 2014;203(03):551–559
- Yinjie Y, Gen W, Hongbo W, et al. A retrospective evaluation of reliability and reproducibility of Arbeitsgemeinschaft für Osteosynthesefragen classification and Fernandez classification for distal radius fracture. *Medicine (Baltimore)* 2020;99(02):e18508
- 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
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(01):159–174
- Randolph JJ. Online Kappa Calculator [Computer software]. 2008. Available from: <http://justus.randolph.name/kappa>
- van Leerdam RH, Souer JS, Lindenhovius AL, Ring DC. Agreement between Initial Classification and Subsequent Reclassification of Fractures of the Distal Radius in a Prospective Cohort Study. *Hand (N Y)* 2010;5(01):68–71
- Belloti JC, Tamaoki MJ, Franciozi CE, et al. Are distal radius fracture classifications reproducible? Intra and interobserver agreement. *Sao Paulo Med J* 2008;126(03):180–185
- Kleinlugtenbelt YV, Groen SR, Ham SJ, et al. Classification systems for distal radius fractures. *Acta Orthop* 2017;88(06):681–687
- Mulders MA, Rikli D, Goslings JC, Schep NW. Classification and treatment of distal radius fractures: a survey among orthopaedic trauma surgeons and residents. *Eur J Trauma Emerg Surg* 2017;43(02):239–248