Original article

Evaluation of intra- and interobserver reliability of the AO classification for wrist fractures ☆

Check for updates

Pedro Henrique de Magalhães Tenório, Marcelo Marques Vieira, Abner Alberti,
Marcos Felipe Marcatto de Abreu, João Carlos Nakamoto, Alberto Cliquet Júnior*

Departamento de Ortopedia e Traumatologia, Hospital de Clínicas, Universidade Estadual de Campinas, Campinas, SP, Brazil

A B S T R A C T

Objective: This study evaluated the intraobserver and interobserver reliability of the AO classification for standard radiographs of wrist fractures.

Methods: Thirty observers, divided into three groups (orthopedic surgery senior residents, orthopedic surgeons, and hand surgeons) classified 52 wrist fractures, using only simple radiographs. After a period of four weeks, the same observers evaluated the initial 52 radiographs, in a randomized order. The agreement among the observers, the groups, and intraobserver was obtained using the Kappa index. Kappa-values were interpreted as proposed by Landis and Koch.

Results: The global interobserver agreement level of the AO classification was considered fair (0.30). The three groups presented fair global interobserver agreement (residents, 0.27; orthopedic surgeons, 0.30; hand surgeons, 0.33). The global intraobserver agreement level was moderated. The hand surgeon group obtained the higher intraobserver agreement level, although only moderate (0.50). The residents group obtained fair levels (0.30), as did the orthopedics surgeon group (0.33).

Conclusion: The data obtained suggests fair levels of interobserver agreement and moderate levels of intraobserver agreement for the AO classification for wrist fractures.

© 2017 Sociedade Brasileira de Ortopedia e Traumatologia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Avaliação da reprodutibilidade intra e interobservadores da classificação AO para fratura do punho

R E S U M O

Objetivo: Este estudo avaliou a confiabilidade interobservador e intraobservador da classificação AO para radiografias simples em fraturas do terço distal do punho.

Métodos: Trinta observadores, divididos em três grupos (residentes de ortopedia e traumatologia, ortopedistas e cirurgiões de mão), classificaram 52 fraturas do terço distal do antebraço com radiografias simples. Após quatro semanas, os mesmos observadores avaliaram as mesmas 52 fraturas em ordem aleatória. O índice *kappa* foi usado

para estabelecer o nível de concordância entre os observadores individualmente e entre os grupos de residentes, ortopedistas e cirurgiões da mão, bem como para avaliar a concordância intraobservador. O índice de *kappa* foi interpretado conforme proposto por Landis e Koch.

Resultados: A confiabilidade interobservador global da classificação AO foi considerada baixa (0,30). Os três grupos apresentaram índices globais de concordância considerados baixos (residentes, 0,27; ortopedistas, 0,30 e cirurgiões da mão, 0,33). A concordância intraobservador global obteve índice moderado (0,41), foi maior no grupo dos cirurgiões da mão, no qual foi considerada moderada (0,50). No grupo dos residentes e ortopedistas foi considerada baixa, com valores de 0,30 e 0,33, respectivamente.

Conclusão: A partir desses dados, concluímos que a classificação AO para fraturas do punho apresenta baixa reprodutibilidade interobservador e moderada reprodutibilidade intraobservador.

Introduction

A public health problem, the incidence of wrist fractures has increased, a fact attributed to the increase in the elderly people of the population, as well as to the increase of high-energy traumas. In a 2001 American study, it was observed that these fractures are the most commonly observed in emergency rooms, representing 3% of all upper limb fractures, with 640,000 cases per year in the United States alone.[1] In the Brazilian population, it is estimated that these fractures account for 10–12% of all fractures.[2]

The distribution of these fractures is bimodal; the most prevalent fracture patterns are associated with high-energy trauma in young people, while the elderly present fractures related to bone fragility.[3] Most fractures (57–66%) are extra-articular; between 9% and 16% are classified as partial articular and 25–30%, as total articular fractures.[4]

Since their first description[5] in 1814 by Abraham Colles, several classification systems have been proposed, in an attempt to find patterns that could indicate the energy of the trauma, the fracture stability, and the prognosis. Ideally, a classification system should be anatomically reproducible, diagnostic, prognostic, and able to evaluate associated lesions and indicate treatment. Such a classification does not yet exist; currently, the most widely used classification is that proposed by the AO group[3] (*Arbeitsgemeinschaft für Osteosynthesefragen* – Association for the Study of Internal Fixation).

This is an alphanumeric binary classification, subdivided into three types, nine groups, and 27 subgroups. Due to its great detail, the intra and interobserver agreement presents divergent results in previous studies assessing its types, groups, and subgroups.[6]

This study is aimed at assessing the intra and interobserver reliability of the AO classification with only the use of simple radiographs in patients with wrist fractures.

Material and methods

This study was approved by the institution's research ethics committee under the number CAAE 69671317.0.0000.5404.

Fifty-two images made in 2017, of patients of both genders with fractures of the distal third of the forearm, were retrieved by PACS (picture archiving and communication system). Only the initial radiographs of skeletally mature patients with an acute fracture, without previous treatment and without splints, fixators, casts, and any other objects that could cover or distort the radiographic image, were selected. Only the posteroanterior and lateral views were included in the study. The images were identified using only numbers, for future reference.

The images were initially analyzed by 30 physicians, divided into groups that progressively had greater contact with wrist fractures (ten orthopedic and traumatology residents, ten orthopedists, and ten hand surgeons), in random order and with no patient identification, with the aid of a descriptive table of the classification (Fig. 1). Participants were asked to classify the fractures as types A (extra-articular), B (partial articular), and C (total articular). After the type classification, the volunteers classified them into the nine groups (from A1 to C3) and the subgroups (from A1.1 to C3.3).

After four weeks, the same participants again classified the same 52 radiographs, in a randomly determined new order, without patient identification. The participants had no access to the results of their initial assessments, or to those of the other volunteers.

Statistical analysis

The data were analyzed using the kappa statistical method. The kappa coefficient is used to evaluate intraobserver agreement, removing the agreement that would be attributed to chance. The values were interpreted using the classification proposed by Landis and Koch[7] (Table 1) that has been traditionally adopted in studies that use the kappa coefficient. Kappa values above 0.8 indicate excellent agreement; between 0.61 and 0.8, good; between 0.41 and 0.6, moderate; between 0.21 and 0.4, low; and between zero and 0.2, poor. Negative values indicate disagreement.

The AO classification was assessed at three different levels of detail. Interobserver agreement was assessed between the participants of a given group (residents, orthopedists and

```
A1 – Extra-articular ulnar fracture with intact radius
A1.1 Styloid process
A1.2 Ulnar metaphysis
A1.3 Fragmented ulnar metaphysis


A2 – Extra-articular ulnar fracture, simple or impacted
A2.1 Without deviation
A2.2 With dorsal deviation (Pouteau-Colles)
A2.3 With volar deviation (Goyran-Smith)


A3 – Extra-articular multifragmentary fracture
A3.1 Impacted, with axial shortening
A3.2 Impacted, with wedge
A3.3 Complex (metaphyseal comminution)


B1 – Partial articular fracture, with sagittal fracture line
B1.1 Lateral, simple (chauffeur)
B1.2 Lateral, multifragmentary
B1.3 Medial


B2 – Partial articular fracture, dorsal fragment (Barton's)
B2.1 Simple
B2.2 With lateral fracture, sagittal fracture line
B2.3 With dorsal carpal displacement


B3 – Partial articular fracture, volar fragment (reverse Barton's)
B3.1 Simple, with small fragment
B3.2 Simple, with large fragment
B3.3. Multifragmentary


C1 – Complete articular fracture (simple articular fracture line and
simple metaphysial fracture line)
C1.1 Postero-medial articular fragment
C1.2 Sagittal fracture line
C1.3 Frontal fracture line
```

Fig. 1 – AO classification for wrist fractures.

Table 1 – Landis and Koch interpretation for *kappa* values.

kappa value	Interpretation
<0	No agreement
0–0.19	Poor agreement
0.20–0.39	Low agreement
0.40–0.59	Moderate agreement
0.60–0.79	Substantial agreement
0.80–1.0	Excellent agreement

hand surgeons) in relation to types A, B, and C. This correlation was then assessed considering the types that varied from A1 to C3. Finally, the level ranging from A1.1 to C3.3.8 was assessed.[8]

After four weeks, new assessments were made and, when comparing these with the baseline, the intraobserver agreement was calculated.

Results

The overall mean interobserver agreement of the AO classification, without distinction of group and for all levels, was considered low (kappa index of 0.30). This result was repeated for all levels, regardless of the group of examiners, from 0.40 for the first and most general level, 0.30 for the second, and 0.20 for the more detailed. When the groups of examiners were taken into account, low levels of agreement were obtained for residents (0.27), orthopedists (0.30), and hand surgeons (0.33).

The three levels of classification were evaluated within the groups of examiners. For the group of residents, a low agreement was observed in the first level (0.34); the agreement in the second level was also low (0.27), while in the most detailed level, it was poor (0.19). In the group of orthopedists, a moderate agreement (0.42) was observed in the first level, low (0.30) in the second, and poor (0.18) in the most detailed level. In the group of hand surgeons, a moderate (0.44) agreement was observed in the first level; this agreement was low in the second (0.32) and third (0.23) levels.

The overall intraobserver agreement was considered moderate (0.41). The mean agreement observed in the group of residents was considered low (0.36). When the intraobserver agreement was stratified according to classification levels, a moderate agreement was observed for the first level (0.50), and low agreement for the second (0.34) and third (0.23) levels. The mean agreement observed in the group of orthopedists was considered low (0.39). In the first level, a moderate (0.51) agreement was observed. In turn, the agreement observed in the second (0.37) and third (0.29) levels was low. In the hand surgeons group, a moderate interobserver agreement (0.50) was observed; it was considered good (0.63) for the first level, moderate (0.49) for the second, and low (0.37) for the third.

Discussion

An ideal rating system should provide a means to report results, as well as to enable fast and straightforward communication among professionals. It should also provide information on trauma mechanism and energy, indicate anatomical patterns, allow a prompt diagnosis, estimate prognosis, assess the degree of soft tissue injuries, and guide treatment. Furthermore, it should be easy to use, widely accepted, intuitive, and reproducible.

In this study, it was observed that the greater the daily contact of the observers with wrist fractures, the greater the agreement, but it never exceeded moderate levels. It was also observed that the higher the level of detail of the classification, the lower the agreement in all groups.

When the intraobserver agreements were analyzed, a high frequency of moderate agreement rates was observed that indicates that after the classification is learned by the observer, it tends to be used coherently.

It can be concluded that although this classification is comprehensive, as its subtypes cover most of the existing fracture patterns, it has low levels of interobserver agreement, not being reproducible in daily clinical practice.

According to a 2015 study, there are 13,147 active registered orthopedists in Brazil.[9] In order to achieve a statistically significant sample size, 1067 volunteers would have had to be interviewed for a 95% confidence interval. Thus, although this study included a larger number of volunteers than other studies retrieved in the literature, a much larger number of participants would be necessary to refute the use of this classification.

Conclusion

The AO classification presents low levels of reproducibility among residents of orthopedics and traumatology, orthopedists, and hand surgeons. However, its intraobserver reproducibility is moderate.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

To all study participants who collaborated with the dedication of their limited time.

REFERENCES

1. Chung KC, Spilson SV. The frequency and epidemiology of hand and zorearm fractures in the United States. J Hand Surg Am. 2001;26(5):908–15.

2. Reis FB, Faloppa F, Saone RP, Boni JR, Corvelo MC. Fraturas do terço distal do rádio: classificação e tratamento. Rev Bras Ortop. 1994;29(5):326–30.

3. Wolfe SW. Distal radius fractures. In: Wolfe SW, Pederson WC, Hotchkiss RN, Kozin SH, Cohen MS, editors. Green's operative hand surgery. 7th ed. Philadelphia: Elsevier Churrchill Livingstone; 2017. p. 516–87.

4. McQueen MM. Fractures of the distal radius and ulna. In: Rockwood and Green's fractures in adults. 8th ed. Philadelphia: Wolters Kluwer Healt; 2015. p. 1057.

5. Colles A. On the fracture of the carpal extremity of the radius. N Engl J Med Surg. 1814;3:368–72.

6. Kreder HJ, Hanel DP, McKee M, Jupiter J, McGillivary G, Swiontkowski MF. Consistency of AO fracture classification for the distal radius. J Bone Joint Surg Br. 1996;78(5): 726–31.

7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1): 159–74.

8. Cooney WP, Agee JM, Hastings H, Melone CP, Rayback JM. Symposium: management of intrarticular fractures of distal radius. Contemp Orthop. 1990;21:71–104.

9. Scheffer M, Biancarelli A, Cassenote A. Demografia Médica no Brasil 2015. São Paulo: Departamento de Medicina Preventiva da Faculdade de Medicina da USP; 2015. Conselho Regional de Medicina do Estado de São Paulo; Conselho Federal de Medicina.