Evaluation of simplified Frykman and AO classifications of fractures of the distal radius

Assessment of interobserver and intraobserver agreement

A. Illarramendi, A. González Della Valle, E. Segal, P. De Carli, G. Maignon, G. Gallucci

The Institute of Orthopaedics "Carlos E. Ottolenghi", Italian Hospital, Buenos Aires, Argentina

Accepted: 23 October 1997

Summary. *In order to evaluate interobserver and intraobserver agreement of the Frykman and AO classifications and their variations between assessors with different levels of experience, three hand specialists, a fellow and two senior residents classified radiographs of 200 fractures of the distal radius in anteroposterior and lateral views. Reproducibility was assessed by the use of the proportion of agreement and kappa coefficient between pairs of observers. The Frykman classification showed moderate interobserver reproducibility ($\kappa = 0.43$) and good intraobserver reproducibility ($\kappa = 0.61$). The experience of the reviewers did not significantly affect either of these. The AO system showed regular interobserver reproducibility ($\kappa = 0.37$) and moderate intraobserver reproducibility ($\kappa = 0.57$). The younger group obtained higher intraobserver agreement than the senior. Possible causes for the low reproducibility of both classifications are discussed together with a review of the literature. We do not recommend the Frykman or AO classifications for clinical application because of their questionable reproducibility.*

Résumé. *Les classification des fractures de l'extrémité distale du radius est un thème de controverse. Les objectifs du travail sont de déterminer la reproductibilité intra et interpersonnelle des classifications de Frykman et AO et évaluer ses variations avec l'expérience des observateurs. Trois*

spécialistes en chirurgie de la main et trois chirurgiens orthopédiques en formation ont évalué deux cents paires radiographiques en face et en profil. La reproductibilité a été calculée en utilisant la proportion d'accord et le coefficient "kappa". La classification de Frykman a obtenu une reproductibilité interpersonnelle modérée ($\kappa = 0.43$) et une bonne reproductibilité intrapersonnelle ($\kappa = 0.61$). L'expérience des évaluateurs n'intervient pas. La classification AO a montré une reproductibilité interpersonnelle régulière ($\kappa = 0.37$) et une reproductibilité intrapersonnelle modérée ($\kappa = 0.57$). Ce sont les non spécialistes qui ont obtenu les meilleurs valeurs de reproductibilité intrapersonnelle. Cette différence a été statistiquement significative. On discute les possibles causes de la baisse de reproductibilité dans ce groupe d'observateurs avec la littérature. Nous ne recommandons pas l'usage de ces classifications dans la pratique quotidienne.

Introduction

The use of numerous classification systems for distal fractures of the radius prevents comparable assessment of treatment modalities and outcomes from different series. Any classification system should provide guidelines to determine both choice of treatment and the prognosis. Good interobserver and intraobserver reproducibility are desirable features.

Interobserver reproducibility refers to the level of agreement between different observers in the classification of a lesion. Intraobserver reproducibility describes the level of agreement in one ob-

Reprint requests to: A. Illarramendi, Institute of Orthopaedics "Carlos E. Ottolenghi", Italian Hospital of Buenos Aires, Potosí 4215, CP 1199, Buenos Aires, Argentina

Table 1. Guidelines for kappa interpretation (Landis and Koch)

Kappa	Reproducibility
Less than 0	Bad
0–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1	Excellent

server for the classification of a lesion on separate occasions.

The Frykman and AO classifications for distal radial fractures have been widely used for clinical research [2, 3, 6, 11–13], and some authors suggest they are useful in evaluating treatment and prognosis [6].

This study was designed to assess the degree of interobserver and intraobserver consistency of these two classification systems in anteroposterior and lateral radiographs, and to determine their variations between observers of different levels of experience.

Material and methods

Anteroposterior and lateral radiographs of 200 acute fractures of the distal radius in 198 adult patients seen in the emergency room at this hospital were chosen for study. We excluded those with previous fractures of the ipsilateral wrist, those with bad quality radiographs, with rheumatoid arthritis and with articular involvement by other systemic conditions. Selection of radiographs was performed by 2 orthopaedic surgeons who were not assessors. The identification and date on the films were hidden in all studies and each was identified with a random number.

Six observers with different levels of experience were selected: 3 hand surgeons (O1, O2, O3) and 3 orthopaedic surgeons in training: a first-year hand surgery fellow (O4), a chief of residents (O5) and a third-year orthopaedic resident (O6). Although they were familiar with both classification systems, they were provided with detailed illustrations and the original descriptions [5, 10].

The series was studied individually and without supervision by each practitioner. Only one classification was applied at each session. Two special conditions were recorded: (a) that a fracture could not be included in any group, and (b) that a

Table 3. Frykman classification – intraobserver agreement

Observer	PA (%)	Kappa
O1	61	0.48
O2	61	0.52
O3	69	0.77
O4	85	0.65
O5	77	0.62
O6	74	0.61
Total	73	0.61

O1, 2, 3, 4, 5, 6 = Observer No. 1, 2, 3, 4, 5, 6

PA = Proportion of agreement

fracture could be included in more than one group. The fractures included in either of these conditions by any reviewer were defined as inadequately classified and excluded from the reproducibility analysis.

We simplified both classifications for the purpose of the statistical analysis. We eliminated fractures of the ulnar styloid to reduce the Frykman classification to 4 possible choices: type 1 – extra-articular fractures; type 3 – intra-articular radiocarpal fractures; type 5 – fractures with distal radioulnar joint involvement, and type 7 – fractures with radiocarpal and radioulnar involvement.

The AO classification system is complex and includes 3 types, 9 groups and 27 subgroups. We simplified it by regrouping them into 5 categories as suggested by Safar in 1993 (personal communication). Type I includes all extra-articular fractures (9 subgroups of AO type A). Type II includes all partial articular fractures (9 subgroups of AO type B). Type III includes non-communited intra-articular fractures (AO group C1). Type IV includes intra-articular fractures with metaphyseal comminution (AO group C2) and type V includes fractures with articular and metaphyseal comminution (AO group C3).

In order to calculate intraobserver agreement, each observer analysed the series for each classification on 2 occasions at an interval of 2 months. In the second test, the series was presented in reverse order.

Statistical analysis

Assessment of inter- and intraobserver consistency was accomplished by the use of two parameters: the proportion of agreement and the kappa coefficient as proposed by Fleiss [4].

The observed proportion of agreement is the percentage of instances in which the observers agreed. The kappa coefficient involves adjustment of the observed proportion of agreement

Table 2. Frykman classification – interobserver agreement. Proportion of agreement and kappa coefficient in pairs of observers

Observer	O2 PA	Kappa	O3 PA	Kappa	O4 PA	Kappa	O5 PA	Kappa	O6 PA	Kappa
O1	57	0.43	51	0.36	52	0.37	55	0.40	55	0.40
O2	–		59	0.39	58	0.38	57	0.40	57	0.39
O3	–		–		90	0.84	53	0.36	59	0.42
O4	–		–		–		55	0.38	62	0.45
O5	–		–		–		–		67	0.54

O1, 2, 3, 4, 5, 6 = Observer No. 1, 2, 3, 4, 5, 6

PA = Proportion of agreement

Table 4. AO classification – interobserver agreement. Proportion of agreement and kappa coefficient in pairs of observers

Observer	O2 PA	Kappa	O3 PA	Kappa	O4 PA	Kappa	O5 PA	Kappa	O6 PA	Kappa
O1	53	0.25	58	0.36	63	0.36	59	0.38	53	0.29
O2	–		57	0.33	64	0.40	54	0.33	57	0.36
O3	–		–		68	0.48	51	0.31	62	0.45
O4	–		–		–		57	0.35	63	0.43
O5	–		–		–		–		58	0.41

O1, 2, 3, 4, 5, 6 = Observer No. 1, 2, 3, 4, 5, 6

PA = Proportion of agreement

Table 5. AO classification – intraobserver agreement

Observer	PA (%)	Kappa
O1	71	0.52
O2	64	0.41
O3	70	0.57
O4	85	0.72
O5	75	0.62
O6	70	0.56
Total	73	0.57

O1, 2, 3, 4, 5, 6 = Observer No. 1, 2, 3, 4, 5, 6

PA = Proportion of agreement

by correction for the proportion of agreement which arises due to chance. The kappa coefficient values can vary from 1 to –1. Kappa is equal to 1 when there is perfect agreement, is equal to 0 when agreement corresponds to chance agreement. In situations where there is less than chance agreement, kappa values are less than 0, and in complete disagreement kappa is equal to –1. We used the guidelines proposed by Landis and Koch for interpretation of kappa values [8] (Table 1).

Kappa coefficients were calculated with the aid of computer software Kappa Statistics (Epidemiological Research Unit, University of Manchester, UK). Interobserver agreement was assessed by calculating kappa coefficients for every possible pair of observers (Table 2). Intraobserver agreement was determined by individual comparison of the two tests (Tables 3 and 5). The observed proportion of agreement between specialists and non-specialists was compared with the chi-square test. The level of significance was $P < 0.05$.

Results

Frykman classification – interobserver agreement

Five cases (2.5%) could not be adequately classified by one or more assessors and were excluded from the intraobserver consistency test.

All 6 observers agreed on the classification of 48 fractures (24.6%), 5 observers for 43 fractures (22%), 4 for 50 fractures (25.6%), 3 for 51 fractures (26.1%), and only 2 for 3 fractures (1.5%).

The average proportion of agreement between pairs of assessors was 59% (range 51% to 90%) and the average kappa coefficient was 0.43 (range 0.36 to 0.84) which indicates moderate reproducibility (Table 2).

Fig. 1a, b. Case 19. Anteroposterior (a) and lateral (b) radiographs of a distal fracture of the radius classed as Frykman type 7 by 3 observers and type 3 by 3 others. Two changed their classification at the second test. When using the simplified AO system, 4 classed the fracture as type III, and 2 as type IV. They changed their classification at the second test

Fig. 2a, b. Case 152. Anteroposterior (a) and lateral (b) radiographs of a distal fracture of the radius classed as Frykman type 5 by 3 observers, types 1, 3 and 7 by 3 others. Two changed their classification at the second test. When using the simplified AO system, 4 classed the fracture as type III, one as type I and one as type IV. Three changed their classification at the second test

The group of specialised observers achieved an average proportion of agreement of 56% and average kappa coefficient of 0.39 (fair reproducibility). The non-specialised observers obtained an average proportion of agreement of 61% and an average kappa coefficient of 0.46 (moderate reproducibility). The difference between the average proportion of agreement obtained by both groups was not significant ($P = 0.075$).

Frykman classification – intraobserver agreement

Eight observations in 5 fractures classified by O1 were excluded from his consistency assessment due to inadequate classification.

The average proportion of agreement between first and second grading sessions for all 6 observers was 73% (range 61% to 85%), and the average kappa coefficient was 0.61 (range 0.48 to 0.77). This represents good reproducibility (Table 3).

For experienced observers, the proportion of agreement was 72% and the kappa coefficient was 0.59 (moderate reproducibility). The orthopaedic surgeons in training obtained an average proportion of agreement of 75% and a kappa coefficient of 0.63 (good reproducibility). There was no significant difference between the proportion of agreement of both groups ($0.20 < P < 0.30$) (Table 3).

AO classification – interobserver agreement

Nine cases (4.5%) could not be adequately grouped by one or more observers and were excluded.

All 6 observers agreed on the classification for 40 fractures (21%), 5 for 49 fractures (26%), 4 for 55 fractures (29%), 3 for 39 fractures (20%) and only 2 for 8 fractures (4%).

The average proportion of agreement between pairs of observers was 58% (range 51% to 68%) and the average kappa coefficient was 0.37 (range 0.25 to 0.48) which corresponds to a fair reproducibility (Table 4).

Experienced observers achieved an average proportion of agreement of 56% and an average kappa coefficient of 0.31 (fair reproducibility). The non-specialists obtained an average proportion of agreement of 59% and an average kappa coefficient of 0.40 (fair reproducibility).

The difference between the proportion of agreement obtained by both groups was not significant ($P > 0.30$).

AO classification – intraobserver agreement

Three observers could not adequately classify one or more fractures which were excluded from their intraobserver consistency analysis.

The average proportion of agreement between first and second tests for all 6 observers was 73% (range 64 to 85%) and the average kappa coefficient was 0.57 (range 0.41 to 0.72), representing moderate reproducibility (Table 5).

For experienced observers, the proportion of agreement was 68% and the kappa coefficient was 0.50 (moderate reproducibility). The non-specialists obtained an average proportion of agreement of 77% and the kappa coefficient of 0.63 (good reproducibility). There was a significant difference between the proportion of agreement of both groups ($P < 0.03$).

Discussion

Any classification system should have adequate intraobserver and interobserver consistency in order to be widely accepted as a common tool for comparing different series. Andersen et al. found good intra- and interobserver reproducibility for the Older classification system in 185 fractures evaluated by 4 hand surgeons [1].

The Frykman and AO classifications have been used to group patients, select treatment and estimate outcome [2, 3]. Moran et al. evaluated interobserver variations for the Frykman classification and reported kappa values of less than 0.5, but they did not analyse intraobserver reproducibility or variations in agreement between observers with different levels of experience [9]. Kreder et al. evaluated the consistency of the AO classification in 36 physician and non-physician observers in a series of 30 fractures, and reported diminished interobserver agreement in comminuted fractures and in non-physicians [7].

We used radiographs of 200 distal radial fractures to determine inter- and intraobserver agreement of the Frykman and AO classification systems for 6 observers with varying expertise. With the simplified Frykman classification, in the first test, all 6 observers agreed for only 48 fractures (24.6%) (Figs. 1, 2). When both first and second viewings were considered, unanimous classification occurred only in 29 fractures (14.8%). We found moderate interobserver reproducibility for the simplified Frykman classification (kappa = 0.43), which is consistent with previous reports [9], but 10 pairs of observers showed only fair interobserver agreement. Intraobserver consistency

for this system was good. We believe that subjective interpretation of the parameters considered in this classification is partly the reason for this observation. In addition, radiographic assessment in complex fractures may be insufficient. Evaluation of such lesions by CT scan would probably provide more precise information and improve reproducibility rates.

For the simplified AO classification, in the first test, only 40 fractures (21%) were classified identically by all 6 observers (Fig. 1, 2). When both first and second viewings were considered, all 6 assessors agreed on the type in only 24 fractures (13%). We found fair interobserver reproducibility for this classification ($\kappa = 0.37$) which is consistent with previous reports [7, 9]. Intraobserver agreement was moderate. This can be attributed to several factors: firstly, the classification is complex and differences between subgroups can sometimes be subtle; secondly, there are no strict guidelines for assessment of the extent of metaphyseal and articular comminution, and thirdly, the determination of the limits of the distal radioulnar joint.

The simplified Frykman classification has proved more consistent than the simplified AO scheme with regard to interobserver reproducibility. There are two factors responsible for this; our observers regularly use the former, and the simplified AO system has more types than the Frykman scheme. We believe that full classifications would have shown less reproducibility.

In contrast to the results reported by Kreder et al. using the simplified AO classification [7], we found that the proportion of inter- and intraobserver agreement did not improve with increased experience.

For both classifications, the values of intraobserver reproducibility were higher than values of interobserver reproducibility.

These classifications have been frequently used to assist decisions regarding treatment, to determine the prognosis, to compare treatment modalities and for grouping patients. Unsatisfactory reproducibility could be responsible for the heterogeneous grouping and the difficulties in assessing comparability between series.

We suggest that these classifications should be no longer used; both properly describe the char-

acteristics of the lesions. We have identified their weakness in order to determine their application in daily practise. Our results contribute to the understanding of distal radial fractures and to the formulation of a universal classification. Such a system should serve to discriminate between treatment modalities and expectation of outcome, and should provide adequate interobserver and intraobserver consistency.

References

1. Andersen GR, Rasmussen JB, Dahl B, Solgaard S (1991) Older's classification of Colles' fractures. Good intraobserver and interobserver reproducibility in 185 cases. *Acta Orthop Scand* 62: 463–464
2. Bradway JK, Amadio PC, Cooney WP (1989) Open reduction and internal fixation of displaced, comminuted intra-articular fractures of the distal end of the radius. *J Bone Joint Surg [Am]* 71: 839–847
3. Edwards GS Jr (1991) Intra-articular fractures of the distal part of the radius treated with the small AO external fixator. *J Bone Joint Surg [Am]* 73: 1241–1250
4. Fleiss JL (1981) Statistical methods for rates and proportions, (2nd edn.) Wiley and Sons, New York, pp 217
5. Frykman G (1967) Fracture of the distal radius including sequelae shoulder-hand-finger syndrome, disturbance in the distal radio-ulnar joint and impairment of nerve function. A clinical and experimental study. *Acta Orthop Scand [Suppl]* 108
6. Havemann D, Busse FW (1990) Unfallmechanismen und Klassifikation bei der distalen Radiusfraktur. *Langenbecks Arch Chir [Suppl II]* pp 639–642
7. Kreder H, Hanel D, McKee M, Jupiter J, McGillivray G, Swiontkowski M (1996) Consistency of AO fracture classification for the distal radius. *J Bone Joint Surg [Br]* 78: 726–731
8. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174
9. Moran T, Gerwin M, Weiland A, McCormack R, Stern P, Hotchkiss R (1996) Interobserver reliability of distal radius fracture classifications. *Am Acad Orthop Surgeons Annual Meeting*. San Francisco, February 22, 1996, Paper No. 33
10. Muller ME, Nazarian S, Koch P (1987) Classification AO des fractures. *Les os longs*. Springer, Berlin
11. Weber SC, Szabo RM (1986) Severely comminuted distal radial fracture as an unsolved problem: complications associated with external fixation and pins and plaster techniques. *J Hand Surg [Am]* 11: 157–165
12. Wittner B, Holz U (1993) Allgemeine Klassifikation der distalen Radiusfrakturen und Behandlung der extra-articularen distalen Radiusfrakturen (Typ A2 und A3). *Chirurg* 64: 880–888
13. Wuster-Hofmann M, Hofmann AK, Kinzl L (1993) Behandlung der einfachen intraartikularen Radiusfrakturen (Typ B1–B3). *Chirurg* 64: 889–893