Classification systems for distal radius fractures

Does the reliability improve using additional computed tomography?

Ydo V KLEINLUGTENBELT [1,2,6], Sylvester R GROEN [3], S John HAM [1], Peter KLOEN [4], Robert HAVERLAG [5], Maarten P SIMONS [5], Vanessa A B SCHOLTES [1], Mohit BHANDARI [6], J Carel GOSLINGS [7], and Rudolf W POOLMAN [1]

[1] Department of Orthopaedic and Trauma Surgery, JointResearch Onze Lieve Vrouwe Gasthuis, Amsterdam; [2] Department of Orthopaedic and Trauma Surgery, Deventer Ziekenhuis, Deventer; [3] University of Amsterdam; [4] Department of Orthopaedic and Trauma Surgery, Academic Medical Centre, Amsterdam; [5] Department of Orthopaedic and Trauma Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam; [6] Division of Orthopaedic Surgery, McMaster University, Hamilton, Ontario, Canada; [7] Trauma Unit, Department of Surgery, Academic Medical Center, University of Amsterdam
Correspondence: ijdoklb@hotmail.com
Submitted 2016-04-12. Accepted 2017-04-25.

Background and purpose — The reliability of conventional radiography when classifying distal radius fractures (DRF) is fair to moderate. We investigated whether reliability increases when additional computed tomography scans (CT) are used.

Patients and methods — In this prospective study, we performed pre- and postreduction posterior–anterior and lateral radiographs of 51 patients presenting with a displaced DRF. The case was included when there was a (questionable) indication for surgical treatment and an additional CT was conducted within 5 days. 4 observers assessed the cases using the Frykman, Fernández, Universal, and AO classification systems. The first 2 assessments were performed using conventional radiography alone; the following 2 assessments were performed with an additional CT. We used the intraclass correlation coefficient (ICC) to evaluate reliability. The CT was used as a reference standard to determine the accuracy.

Results — The intraobserver ICC for conventional radiography alone versus radiography and an additional CT was: Frykman 0.57 vs. 0.51; Fernández 0.53 vs. 0.66; Universal 0.57 vs. 0.64; AO 0.59 vs. 0.71. The interobserver ICC was: Frykman: 0.45 vs. 0.28; Fernández: 0.38 vs. 0.44; Universal: 0.32 vs. 0.43; AO: 0.46 vs. 0.40.

Interpretation — The intraobserver reliability of the classification systems was fair but improved when an additional CT was used, except for the Frykman classification. The interobserver reliability ranged from poor to fair and did not improve when using an additional CT. Additional CT scanning has implications for the accuracy of scoring the fracture types, especially for simple fracture types.

Distal radius fractures were initially, since 1814, called "Pouteau" fractures and later renamed "Colles" fractures, after the Irish surgeon Abraham Colles (Colles 1970). At that time, no further distinctions were made into various subtypes of distal radius fractures. After the introduction of the roentgen and the growing awareness of the diversity of fracture features, the number of subtypes along with fracture eponyms increased. The first classification system (Colles 1970) was originally based on clinical features only, but additional classification systems have been developed since through the use of conventional radiographs (CR).

The most common classification systems for distal radius fractures include Frykman (1967), Fernández (2001), Universal (Cooney 1993), and AO classification (Marsh et al. 2007). An overview of reliability studies evaluating these 4 classification systems is presented in Tables 1 and 2 (see Supplementary data). Using these full classification systems, the interobserver reliability was fair to moderate. Good to excellent agreement was found only when using the AO classification of only 3 types (Type A: extra-articular, Type B: partial articular and Type C: complete articular). This is comparable for the intraobserver reliability. However, 2 studies found substantial reliability for the Universal classification (Belloti et al. 2008, Kural et al. 2010). Currently, there is no gold standard for classifying distal radius fractures.

Validated trauma classification systems offer a structured framework to communicate effectively about clinical cases, and support the treatment decision process (i.e., nonsurgical vs. surgical management, type of surgical intervention). In addition, classifying the severity of fractures is important in clinical research, as the classified grade or type can be used as

part of a study's eligibility criteria. However, in order to apply the treatment recommendations arising from these trials, the applicable classification systems must also be used in daily practice. Given the low degree of reliability using conventional radiography (CR) alone for fracture classifications, supplemental information may be warranted for more accurate and reproducible evaluations.

Prior studies have used computed tomography (CT) to investigate the AO, Fernandez and Universal classification of distal radius fractures. However, these studies have been limited by lack of standardization on expertise of the reviewers (Yunes Filho et al. 2009) and focus on simplified versions of the original classifications (Flikkilä et al. 1998). Currently, no evaluation of the utility of an additional CT scan on any original full classification system with experienced reviewers has been published. Additionally, prior studies have not evaluated radiographs versus radiographs plus CT scans.

To address this current lack of knowledge, we aimed to determine the intra- and interobserver reliability of the most commonly used fracture classification systems, using both CR and CR with the addition of a CT scan in a representative clinical setting of cases with a questionable indication for surgery. We evaluated the most commonly used classification systems that have been developed to classify any type of distal radius fracture; the Frykman (1967), Fernández (2001), Universal (Cooney 1993), and AO classification (Marsh et al. 2007). By using experienced observers, we hypothesized that the intraobserver and interobserver reliability is higher when using conventional radiography with additional CT. In addition, we determined the accuracy of the classification systems using the CT scan as a reference standard.

Patients and methods

Study design
A prospective database was established between January 1, 2007 and March 2, 2011 of patients with a displaced distal radius fracture seen at the emergency rooms in a hospital in Amsterdam (Onze Lieve Vrouwe Gasthuis).

Experience of the observers
The observers consisted of 4 experienced Dutch surgeons, of whom 2 were trauma surgeons (MS, RH) and 2 were orthopedic surgeons (JH, PK). Each had over 10 years of experience in fracture treatment. All were responsible for the (distal radius) fracture care within their department.

Study patients
Patients were eligible for inclusion if they (1) were 18 years of age or older presenting with a displaced distal radius fracture in the emergency department, (2) had pre- and postreduction conventional posterior–anterior and lateral radiographs of the wrist, and (3) had an additional CT within 5 days in cases of

a (questionable) indication for surgery. Questionable indication for surgery was defined as an inadequate reduction of the fracture as described by the AAOS guidelines (Lichtman et al. 2010) or in case of a presumably unstable fracture (Mackenney et al. 2006). Patients were excluded if they had a prior fracture or pathology of the distal radius.

Scoring procedure
The 4 observers independently classified the radiographic and/or CT images at 4 different time points. Each scoring round was performed with an interval of at least 4 weeks. All images were digitalized and anonymized.

Although increasing the number in either group would yield a more precise reliability estimate, the number of fractures has a greater impact on the precision than the number of observers (Steiner and Norman 2013). For this reason we chose a relatively low, but clinically representative, number of 4 observers.

At time points 1 and 2, the pre- and postreduction CR were used to classify the fracture according to the Frykman, Fernández, Universal and AO classification systems. At time points 3 and 4, both the CR and all the 2D CT scan images were used (axial, sagittal and coronal planes). The order of the images was randomized at each time point. A short description of the 4 classification systems with additional illustrations was available for each observer.

Classification systems (Table 3)
The subgroups of the AO classification were not used in this study, to simplify the evaluation and keep the number of grading criteria comparable to the other classification systems.

Sample size
Based on the methodology proposed by Giraudeau and Mary (2001), we used the expected value of the ICC, along with the number of raters and the desired confidence interval and confidence level, to determine the number of subjects to be evaluated in this study. When using an additional CT, we expected a higher ICC than is shown in previous literature when using CR. We therefore estimated an ICC between 0.6 and 0.8. To obtain a 95% confidence interval (CI) with a confidence level of ±0.10 we needed between 30 and 81 patients.

Statistics
Classifications at time points 1 and 2 were used to determine the intraobserver reliability for the CR for each observer separately. Classifications at time points 3 and 4 were used to determine the intraobserver reliability for the CR with added CT scans for each observer separately.
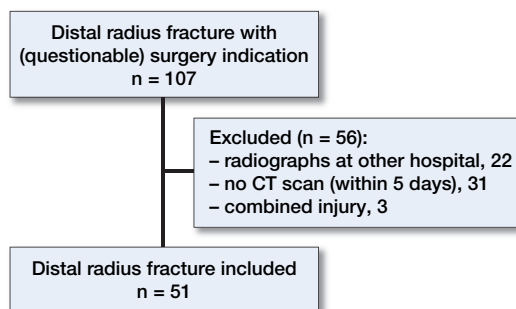
Classifications at time points 1 and 3 were used to determine the interobserver reliability for the CR for each pair of observers (observer 1–2, 1–3, 1–4; 2–3, 2–4; 3–4) and we report the mean of these results with the associated CI.

We present descriptive statistics of the study patients, including means (SD) for continuous data. Intra- and interob-

Table 3. Overview of included classification systems

Classification, types (n), and description of types
Frykman (1967), 8 (I–VIII)
I — Extra-articular
II — Extra-articular with ulnar fracture
III — Intra-articular into radiocarpal joint
IV — Intra-articular into radiocarpal joint with ulnar fracture
V — Intra-articular into radioulnar joint
VI — Intra-articular into radioulnar joint with ulnar fracture
VII — Intra-articular into radiocarpal + radioulnar joints
VIII — Intra-articular into radiocarpal + radioulnar joints with ulnar fracture
Fernández (2001), 5 types, based on trauma mechanism
Type 1 — Bending fracture of metaphysis
Type 2 — Shearing fracture of joint surface
Type 3 — Compression fracture of joint surface
Type 4 — Avulsion fractures or radiocarpal fracture-dislocation
Type 5 — Combined fractures associated with high high-velocity injuries
Universal (1993), 4 types, subdivision in 2 x 3 groups
Type 1 — Extra-articular fracture, without deviation
Type 2 — Extra-articular fracture, with deviation
2A — Reducible and stable
2B — Reducible and unstable
2C — Irreducible
Type 3 — Intra-articular fracture, without deviation
Type 4 — Intra-articular fracture, with deviation
4A — Reducible and stable
4B — Reducible and unstable
4C — Irreducible
AO/ASIF (2007), 3 types, 9 groups
A — Extra-articular fractures
A1 — Ulnar fracture, radius intact
A2 — Radius fracture, simple and impacted
A3 — Radius fracture, multifragmentary
B — Partial articular fractures
B1 — Radius fracture, sagittal
B2 — Radius fracture, frontal, dorsal rim
B3 — Radius fracture, frontal, volar rim
C — Complete articular fractures
C1 — Articular simple + metaphyseal simple
C2 — Articular simple, metaphyseal multifragmentary
C3 — Articular multifragmentary

Flow chart of patients in the study.

volar and dorsal approach, which are used in the treatment of the majority of distal radius fractures, do not provide an adequate view of the dorsal, volar and intra-articular comminuted fracture. We used the CT scan as a reference standard instead of the "gold standard" to more accurately classify the fracture (Knottnerus and Muris 2003). With regard to the distribution of fracture types, absolute and percentile frequencies were calculated and differentiated according to CR (round 1) and CR with an additional CT (round 3). In addition, the percentage of change per fracture type was determined for all 4 classification systems. We compared the distribution of fracture classifications using CR only and CR with added CT scans for each classification system using chi-square tests. We corrected for multiple testing using a Bonferroni correction.

Ethics, funding and potential conflicts of interest

Ethics approval was obtained from the medical ethical committee at our hospital (WO 10.086). We conducted this study according to the Collaboration for Outcome Assessment in Surgical Trials (COAST) guidelines (Karanicolas et al. 2009).

No external funding was received for this study. No competing interests were declared.

server reliability was evaluated using the intraclass correlation coefficient (ICC). While other reliability studies have chosen a Kappa statistic, the ICC is able to take into account skewed data as well as to give credit for partial agreement. Kappa statistics are less accurate if responses are skewed and only appropriate for categorical data (Karanicolas et al. 2009). Fleiss and Cohen (1973) showed that weighted kappa and ICC are equivalent in general cases when interval scales are used. To compare our results with previous literature, Cohen's Kappa was determined as well. The values were interpreted as described by Cicchetti (1994); ICC values less than 0.40 indicate poor agreement, values between 0.40 and 0.59 indicate fair agreement, values between 0.60 and 0.74 indicate good agreement, and values ranging from 0.75 to 1.00 indicate excellent agreement. To determine the accuracy, direct visualization through operative intervention would theoretically be the gold standard, but practically this is unrealistic. Both the

Results

Study participants

From the 107 patients who entered the emergency room during the study period with a distal radius fracture with a (questionable) indication for surgery, 51 patients met the inclusion criteria (Figure). The included patients had a mean age of 50 (14) years. 38 patients (75%) were female. The postreduction CT scan was performed after a mean of 2.5 (2.2) days. The number of cases selected for surgical treatment ranged widely (31%, 53%, 82% and 96%) between the 4 observers.

Reliability of classification systems

All ICCs for the intraobserver reliability with the range of the 4 observers are presented in Table 4. All ICCs for the interobserver reliability and their respective 95% confidence intervals (CI) are presented in Table 5. The calculated Kappa values, to

Table 4. Intraobserver reliability

| Classification | CR | | CR + CT scan | | |
	ICC	Agreement	ICC	Agreement	p
Frykman	0.57 (0.34–0.77)	Fair	0.51 (0.33–0.80)	Fair	0.6
Fernandez	0.53 (0.32–0.62)	Fair	0.66 (0.53–0.90)	Good	0.1
Universal	0.57 (0.43–0.71)	Fair	0.64 (0.50–0.78)	Good	0.5
AO groups	0.59 (0.51–0.66)	Fair	0.71 (0.56–0.91)	Good	0.3

CR: Conventional radiographs
ICC: Mean intra-class correlation coefficient of the Intraobserver reliability with the range of the 4 observers in parentheses

Table 5. Interobserver reliability

| Classification | CR | | CR + CT scan | | |
	ICC	Agreement	ICC	Agreement	p
Frykman	0.45 (0.31–0.60)	Fair	0.28 (0.14–0.44)	Poor	0.03
Fernandez	0.38 (0.21–0.55)	Poor	0.44 (0.30–0.59)	Fair	0.4
Universal	0.32 (0.18–0.48)	Poor	0.43 (0.20–0.51)	Fair	0.01
AO groups	0.46 (0.31–0.60)	Fair	0.40 (0.26–0.53)	Fair	0.4

CR: Conventional radiographs
ICC: Mean intraclass coefficient of the interobserver reliability with 95% CI in parentheses.

compare with previous literature, are presented in Tables 8 and 9 (see Supplementary data).

Frykman classification

Intraobserver reliability: The mean ICC of the Frykman classification was 0.57 when using CR, representing fair reliability. The addition of CT showed no statistically significant improvement. The mean reliability was also fair (mean ICC = 0.51).

Interobserver reliability: The mean ICC of the Frykman classification was 0.45 when using CR, representing fair reliability. The addition of CT scanning was less reliable (p = 0.03). The mean reliability was poor (mean ICC = 0.28).

Fernández classification

Intraobserver reliability: The mean ICC of the Fernández classification was 0.53 when using CR, representing fair reliability. The addition of CT showed a trend toward improvement but this was not statistically significant. The mean reliability was good (mean ICC = 0.66).

Interobserver reliability: The mean ICC of the Fernández classification was 0.38 when using CR, representing poor reliability. The addition of CT showed no statistically significant improvement. The mean reliability was fair (mean ICC = 0.44).

Universal classification

Intraobserver reliability: The mean ICC of the Universal classification was 0.57 when using CR, representing fair reliability. The addition of CT showed no statistically significant improvement. The mean reliability was good (mean ICC = 0.64).

Interobserver reliability: The mean ICC of the Universal classification was 0.32 when using CR, representing poor reliability. The addition of CT showed significant improvement (p = 0.01). The mean reliability was fair (mean ICC = 0.43).

AO classification

Intraobserver reliability: The mean ICC of the AO classification was 0.59 when using CR, representing fair reliability. The addition of CT showed no statistically significant improvement. The mean reliability was good (mean ICC = 0.71).

Interobserver reliability: The mean ICC of the AO classification was 0.46 when using CR, representing fair reliability. The addition of CT showed no statistically significant change. The mean reliability was only fair (mean ICC = 0.40).

Distribution of fracture types with and without CT scan

The overall distribution of fracture types changed after adding a CT scan using the AO and Fernandez classification systems (p < 0.001 and p = 0.006 respectively). The overall distribution of fracture types did not significantly change using the Universal and Frykman classification systems (p = 0.09 and p = 0.06 respectively).

In general, in each classification system approximately half of the extra-articular fractures were classified as an intra-articular fracture when adding CT scanning. For example, in the AO classification the ratio of intra-articular to extra-articular fractures increased from 171:33 (= 5) in round 1 (based on conventional radiography), to 189:15 (= 13) in round 3 (based on CT scanning). The other 3 classifications showed a similar increase in the number of intra-articular fractures. In addition, when adding CT scanning the extra-articular fracture types were classified differently between 60% (Universal type 1) and 100% (Universal: 2B and AO: A3), as the scoring of the intra-articular fracture types changed between 17% (AO: C3) and 53% (Frykman: III/IV). Besides these features, the other statistically significant changes for each classification system are described in Tables 6 and 7.

Frykman: Only the number of fractures classified as type VIII (intra-articular radio-ulnar and radio-carpal joint with an ulnar fracture) increased. The other changes were not statistically significant.

Fernández: The number of fractures classified as type 2 (shearing fracture of joint surface) increased, while the number of fractures classified as type 1 (bending fractures of metaphysis) decreased.

Universal: None of the changes in distribution were statistically significant.

AO: The number of fractures classified as type C3 (intra-articular multifragmentary) increased, while the number of fractures classified as type A2 (extra-articular simple) and type C1 (articular simple + metaphyseal simple) decreased.

Table 6. Distribution of fracture types in round 1 (CR: conventional radiography) and round 3 (CR + CT: CR + additional CT scan) of all 4 observers given in percentages

Frykman	I	II	III	IV	V	VI	VII	VIII
CR	9	5	20	25	3	1	18	19 [a]
CR + CT	5	2	18	16	3	0	24	31 [a]
Fernandez	1	2	3	4	5			
CR	18 [a]	11 [a]	50	1	20			
CR + CT	7 [a]	21 [a]	59	0	13			
Universal	1	2A	2B	2C	3	4A	4B	4C
CR	2	8	4	0	4	25	39	17
CR + CT	1	5	1	0	9	27	32	24
AO groups	A2	A3	B1	B2	B3	C1	C2	C3
CR	12 [a]	4	1	3	12	31 [a]	30	6 [a]
CR + CT	2 [a]	5	3	5	12	19 [a]	32	22 [a]

[a] Significant changes in fracture distribution in that category.

Table 7. Percentage of changes in classification after adding a CT scan (round 1 versus 3)

Frykman	I+II	III+IV	V+VI	VII+VIII			
	69	48	78	0			
Fernandez	1	2	3	4	5		
	69	41	7	n/a	0		
Universal	1	2A	2B	3	4A	4B	4C
	60	71	75	25	29	23	0
AO groups	A2	A3	C1	C2	C3		
	88	89	51	25	0		

Frykman: The fracture types with and without an ulnar fracture are added together.
Fernandez: type 4 and Universal type 2c were not taken into account, because these were not classified.

Discussion

In contrast to our hypothesis, the results of this study revealed that the increase in reliability when using additional CT scanning was seen only in the intraobserver reliability, with the exception of the Frykman classification. The Frykman classification distinguishes between intra-articular and extra-articular fractures of the distal radio-ulnar joint. On CR, the distal radio-ulnar joint fracture line is not always clearly imaged and therefore generally not taken into account in the classification evaluation. However, on a CT scan a small fracture line is often seen in the region of the distal radio-ulnar joint, allowing room for interpretation and potentially discrepant results. This could explain the decrease in reliability of the Frykman classification when using additional CT scanning.

3 prior studies also used CT scanning to investigate the interobserver reliability of the AO classification for distal radius fractures (Flikkilä et al. 1998, Yunes Filho et al. 2009, Arealis et al. 2014). In our study, the interobserver reliability (kappa values) was found to be comparable, both using CR alone and with an additional CT scan. One would expect a higher reliability, since determining the three-dimensional morphology of the fracture might be more difficult when CT scans are not combined with radiographs. Surprisingly, using an additional CT scan and only experienced observers in our study the reliability did not improve in comparison with some other studies that used a CT scan alone and observers of all levels of experience (Yunes Filho et al. 2009, Arealis et al. 2014).

Using the CT scan as the reference standard, we can state that simple fracture types are less accurately classified when using only a radiograph than more severe types. This is contradictory to the clinical practice in which CT scans are especially used in the more severe fracture types to plan treatment. Similar to other published reports (Johnston et al. 1992, Dahlen et al. 2004), we found a systematic decrease of about 50% in the ratio of extra- to intra-articular fractures when the CT scan was added to CR.

Furthermore, our results confirm earlier statements that the severity of a distal radius fracture may be underestimated in standard radiographs. For instance, Cole et al. (1997) reported an improved reliability in assessing specific displacement features, in particular the measurement of gapping or stepping-off, based on CT compared with CR. This is best shown by the AO classification as the number of type C3 (articular multifragmentary) fractures, which increased after adding CT to the evaluation.

Rozental et al. (2001) and Heo et al. (2012) reported that sigmoid notch involvement is underestimated when using only CR. This feature is also seen in our study. The Frykman classification distinguishes between intra-articular and extra-articular fractures of the distal radio-ulnar joint. As shown in Table 3, the number of type VIII (intra-articular distal radio-ulnar joint and distal radio-carpal joint with ulnar fracture) increased when using the additional CT scan as also shown by Goldwyn et al. (2012).

We suggest use of the AO classification, as this is currently the classification system most frequently used and the reliability is comparable to the other classification systems. Preferably, a new classification system also based on CT instead of CR alone should be developed. Such a new classification system should focus mainly on giving direction to the type of treatment.

A limitation of our study is that the sample size was underestimated for the interobserver reliability. The pre-specified estimation of ICC for intraobserver reliability (as described in the methods) was comparable to our estimation so we can be confident that our number of raters is sufficient. The range of intraobserver ICCs was relatively large for some classifications; however, replacing either the best or worst observer would be unlikely to change the conclusion that an additional CT scan improves reliability. It would likely only affect the absolute ICC, not the difference between CR alone and CR with additional CT scan. By choosing a group of patients with a (questionable) indication for surgery we introduced a selec-

tion bias, which possibly influenced our results and therefore risks a lack of generalizability to other patients. However, the optimal treatment of this group of patients lacks consensus and therefore these patients will likely benefit most from additional evaluation criteria for accurate classification. Although the intraobserver reliability improved from fair to good, the p-values were not significant. However, these t-tests were likely to be underpowered and had a high rate of Type II error. Previous studies have shown better reliability for younger patients when classifying DRFs (Wadsten et al. 2009). The relatively low mean age in this study may affect the outcome and could give higher reliability. Additionally, it is important to note that there is some inferential uncertainty with these results. It may be difficult to apply our results to broader populations, although we took precautions to ensure a representative sample and as close to a real-world clinical setting as possible.

A strength of our study is that we used the COAST criteria to ensure we addressed all components of a reliability study. Another strength is that the number of patients selected for surgical treatment ranged widely between the 4 observers, showing that this group of patients is representative of the group of patients lacking consensus.

Our study results suggest that the additional value of CT scanning over CR is limited in regard to reliability. However, it has significant implications for accurate scoring of the fracture types. Using an additional CT scan changes how patients are classified into fracture types, therefore trials using CR alone to evaluate eligibility will have different patients included compared with trials using additional CT scans. This has implications for external validity (generalizability) and for comparing trials with each other.

Although previous literature showed that CT scans are more reliable than CR quantifying articular surface incongruencies, to our knowledge no previous studies have reported the impact on clinical outcome of intra-articular involvement without a step or gap. The outcomes of the current study are not necessarily related to better patient outcomes. Prospective randomized studies—comparing CR for patients who have displaced DRF with additional CT scans—should be conducted to confirm the additional value of a CT scan for patient outcomes. Also a cost-effectiveness analysis should be conducted as national care budgets are limited.

In summary, our study results suggest that the additional value of CT scanning over CR is limited with regard to reliability, but has significant implications for accurate scoring of the fracture types. The reliability of the classification system might be decreased due to the fact that the additional information concerning fracture morphology provided by the additional CT scan leaves increased room for interpretation when classifying a distal radius fracture.

Summary

To our knowledge this is the first reliability study on 4 classification systems that determines the intra- and interobserver reliability using CR alone and with additional CT scanning. The intraobserver reliability of the classification systems was found to be fair but improves to good agreement if an additional CT is used, with the exception of the Frykman classification. The interobserver reliability of the investigated classification systems for distal radius fractures was poor to fair and did not improve when using additional CT scanning. Additional CT scanning has significant implications for accurate scoring of the fracture types in AO and Fernandez classifications, especially for the less severe fractures.

Supplementary data

Tables 1, 2, 8, and 9 are available in the online version of this article, http://dx.doi.org/ 10.1080/17453674.2017.1338066

Arealis G, Galanopoulos I, Nikolaou V S, 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.

Belloti J C, Tamaoki M J, Franciozi C E, Santos J B, Balbachevsky D, Chap Chap E, Albertoni W M, Faloppa F. Are distal radius fracture classifications reproducible? Intra and interobserver agreement. Sao Paulo Med J 2008; 126 (3): 180-5.

Cicchetti D V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 1994; 6 (4): 284-90.

Colles A. Historical paper on the fracture of the carpal extremity of the radius (1814). Injury 1970; 2 (1): 48-50.

Cole R J, Bindra R R, Evanoff B A, Gilula L A, Yamaguchi K, Gelberman R H. Radiographic evaluation of osseous displacement following intra-articular fractures of the distal radius: Reliability of plain radiography versus computed tomography. J Hand Surg Am 1997; 22 (5): 792-800

Cooney W P. Fractures of the distal radius: A modern treatment-based classification. Orthop Clin North Am 1993; 24 (2): 211-16.

Dahlen H C, Franck W M, Sabauri G, Amlang M, Zwipp H. Incorrect classification of extra-articular distal radius fractures by conventional X-rays: Comparison between biplanar radiologic diagnostics and CT assessment of fracture morphology. Unfallchirurg 2004; 107 (6): 491-8.

Fernandez D L. Distal radius fracture: The rationale of a classification. Chir Main 2001; 20 (6): 411-25.

Fleiss J L, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement 1973; 33: 613-19.

Flikkilä T, Nikkola-Sihto A, Kaarela O, Paakko E, Raatikainen T. Poor interobserver reliability of AO classification of fractures of the distal radius: Additional computed tomography is of minor value. J Bone Joint Surg Br 1998; 80 (4): 670-2.

Frykman G. 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 1967; 108 (Suppl): 108-53.

Giraudeau B, Mary J Y. Planning a reproducibility study: How many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. Stat Med 2001; 20 (21): 3205-14.

Goldwyn E, Pensy R, O'Toole R V, Nascone J W, Sciadini M F, LeBrun C, Manson T, Hoolachan J, Castillo R C, Eglseder W A. Do traction radiographs of distal radial fractures influence fracture characterization and treatment? J Bone Joint Surg Am 2012; 94 (22): 2055-62.

Heo Y M, Roh J Y, Kim S B, Yi J W, Kim K K, Oh B H, Oh H T. Evaluation of the sigmoid notch involvement in the intra-articular distal radius fractures: The efficacy of computed tomography compared with plain X-ray. Clin Orthop Surg 2012; 4 (1): 83-90.

Johnston G H, Friedman L, Kriegler J C. Computerized tomographic evaluation of acute distal radial fractures. J Hand Surg Am 1992; 17 (4): 738-44.

Karanicolas P J, Bhandari M, Kreder H, Moroni A, Richardson M, Walter S D, Norman G R, Guyatt G H, Collaboration for Outcome Assessment in Surgical Trials (COAST) Musculoskeletal Group. Evaluating agreement: Conducting a reliability study. J Bone Joint Surg Am 2009; 91 (Suppl 3): 99-106.

Knottnerus J A, Muris J W. Assessment of the accuracy of diagnostic tests: The cross-sectional study. J Clin Epidemiol 2003; 56 (11): 1118-28.

Kural C, Sungur I, Kaya I, Ugras A, Erturk A, Cetinus E. Evaluation of the reliability of classification systems used for distal radius fractures. Orthopedics 2010; 33 (11): 801.

Lichtman D M, Bindra R R, Boyer M I, Putnam M D, Ring D, Slutsky D J, Taras J S, Watters W C, 3rd, Goldberg M J, Keith M, Turkelson C M, Wies J L, Haralson R H, 3rd, Boyer K M, Hitchcock K, Raymond L. Treatment of distal radius fractures. J Am Acad Orthop Surg 2010; 18 (3): 180-9.

Mackenney P J, McQueen M M, Elton R. Prediction of instability in distal radial fractures. J Bone Joint Surg Am 2006; 88 (9): 1944-51.

Marsh J L, Slongo T F, Agel J, Broderick J S, Creevey W, DeCoster T A, Prokuski L, Sirkin M S, Ziran B, Henley B, Audige L. Fracture and dislocation classification compendium—2007: Orthopaedic trauma association classification, database and outcomes committee. J Orthop Trauma 2007; 21 (10 Suppl): S1-133.

Rozental T D, Bozentka D J, Katz M A, Steinberg D R, Beredjiklian P K. Evaluation of the sigmoid notch with computed tomography following intra-articular distal radius fracture. J Hand Surg Am 2001; 26 (2): 244-51.

Steiner D L, Norman G R. Health measurement scales: A practical guide to their development and use. Oxford University Press, 2013.

Wadsten M A, Sayed-Noor A S, Sjoden G O, Svensson O, Buttazzoni G G. The Buttazzoni classification of distal radial fractures in adults: Interobserver and intraobserver reliability. Hand (NY) 2009; 4 (3): 283-8.

Yunes Filho P R M, Pereira Filho M V, Gomes D C P, Medeiros R S, De Paula E J L, Mattar Junior R. Classifying radius fractures with X-ray and tomography imaging. Acta Orthopedica Brasileira 2009; 17 (2): 9-13.