Are classifications of proximal radius fractures reproducible?

Fabio T Matsunaga§, Marcel J S Tamaoki, Eduardo F Cordeiro, Anderson Uehara, Marcos H Ikawa, Marcelo H Matsumoto, João B G dos Santos, João C Belloti.

1Second-year resident in the Department of Orthopedics and Traumatology, Universidade Federal de São Paulo — Escola Paulista de Medicina (Unifesp-EPM), São Paulo, Brazil.

2Attending physician in Shoulder and Elbow Surgery Sector in the Department of Orthopedics and Traumatology, Universidade Federal de São Paulo — Escola Paulista de Medicina (Unifesp-EPM), São Paulo, Brazil.

3Second-year resident in the Department of Orthopedics and Traumatology, Universidade Federal de São Paulo — Escola Paulista de Medicina (Unifesp-EPM), São Paulo, Brazil.

4Attending physician in Shoulder and Elbow Surgery Sector in the Department of Orthopedics and Traumatology, Universidade Federal de São Paulo — Escola Paulista de Medicina (Unifesp-EPM), São Paulo, Brazil.

5Attending physician in the Department Radiology, Universidade Federal de São Paulo — Escola Paulista de Medicina (Unifesp-EPM), São Paulo, Brazil.

6Head of the Shoulder and Elbow Surgery Sector, Department of Orthopedics and Traumatology, Universidade Federal de São Paulo — Escola Paulista de Medicina (Unifesp-EPM), São Paulo, Brazil.

7Adjunct professor and head of Hand Surgery Clinic, Department of Orthopedics and Traumatology, Universidade Federal de São Paulo — Escola Paulista de Medicina (Unifesp-EPM), São Paulo, Brazil.

8Attending physician in Hand Surgery Clinic, Department of Orthopedics and Traumatology, Universidade Federal de São Paulo — Escola Paulista de Medicina (Unifesp-EPM), São Paulo, Brazil.

§Corresponding author

Fabio Teruo Matsunaga
Departamento de Ortopedia e Traumatologia - UNIFESP
Rua Borges de Lagoa, 783 — 5o andar
São Paulo (SP) — Brasil —CEP 04038-032
Tel./Fax (+55 11) 5571-6621
E-mail:fteruo@gmail.com
Email addresses:

FTM: fteruo@gmail.com
MJST: marceltamaoki@terra.com.br
EFC: eduardo68@gmail.com
AU: andersonuehara@hotmail.com
MHI: mhikawa@gmail.com
MHM: marcelohmatsumoto@terra.com.br
JBGS: jbgsantos@ig.com.br
JCB: jbelloti@terra.com.br
Abstract

Background: Fractures of the proximal radius need to classify need to be classified in an appropriate and reproducible manner; the aim of this study was to assess the reproducibility of the three most widely used classification systems.

Methods: Fifty nine plain elbow radiographs in antero-posterior and lateral incidences of patients with proximal radius fractures were analyzed and rated with Mason, Morrey, and Arbeitsgemeinschaft für osteosynthesefragen/Association for the Study of Internal Fixation (AO/ASIF) classifications by four observers with different experience with this subject to verify their intra- and inter-observer reproducibility. Each observer analyzed the images on three different occasions on a computer with numerical sequence randomly altered. The intra- and inter-observers agreement was obtained calculating the kappa coefficient, proposed by Fleiss.

Results: We found that intra-observer agreement of Mason and Morrey classifications were satisfactory ($\kappa=0.582$ and $0.554$, respectively), while the AO/ASIF classification had poor intra-observer agreement ($\kappa=0.483$). Inter-observer agreement was higher in the Mason ($\kappa=0.429–0.560$) and Morrey ($\kappa=0.319–0.487$) classifications than in the AO/ASIF classification ($\kappa=0.250–0.478$), which poor concordance and the worst reproducibility.
Background

Fractures of the proximal radius are relatively common injuries, accounting for approximately 1/3 of all elbow fractures and about 1.7 to 5.4% of all fractures in adults\(^1,2\). Most of these (85%) occur in adults between 20 and 60 years of age (average of 30 to 40 years) with a male:female ratio of about 2:3\(^1,2\). One third of the lesions are associated with others upper limb injuries, such as carpal bone fractures\(^3\), distal radioulnar joint injuries\(^4,5\), interosseous membrane injuries\(^6,7\), capitellar fractures\(^8\), and damage of the medial collateral ligament.

Fractures of the proximal radius play an important role in injuries of the elbow, not so much by how frequently they occur, but mainly by the potential difficulties in treatment and the complications that can arise, sometimes with serious impairment of function due to pain and loss of mobility. Therefore, classification is an essential part of understanding the before treatment.

The most common mechanism of injury in these fractures is a fall onto an outstretched hand with an axial load on the radius. In such cases, the radial head and/or neck fracture when they collide with the capitellum, usually with forearm in pronation and the elbow in partial flexion\(^9,10,11\).

The classification systems that have been developed for the fractures are intended to allow surgeons to classify them into clinically useful groups.

In 1954, Mason based the first classification system of these fractures on deviation,
angle, and comminution of the radial head, dividing them into three types. Type I fractures were nondisplaced or minimally displaced fractures of the head or neck; type II were displaced fractures (more than 2 mm) of the head or neck; and type III were severely comminuted fractures of the proximal radius. In 1962, Johnston expanded the classification of Mason adding type IV, a fracture associated with dislocation of the elbow.

In 1997, Hotchkiss expanded the Mason classification, emphasizing the need and type of intervention according to the type of fracture. In 2008, van Riet and Morrey published a revision of the Mason classification, distinguishing between injuries associated with coronoid fractures, the olecranon fractures, and ligamentous injuries.

The AO/ASIF classification was created in 1986 and revised in 1996. It considers the seriousness of the bone injury and serves as a basis for treatment and evaluation of results. The AO/ASIF system specifies three basic types: extra-articular, articular of the radius or ulna, and articular of the radius and ulna. With each group, the fractures are organized in increasing order of severity with regard to morphological complexity, difficulty in treatment, and prognosis. This system is one of the most complete classifications, but its intra- and inter-observer reproducibility has been a problem when the groups and subgroups have been evaluated.

The classification systems should guide clinical evaluation. A good system needs to be valid, reliable, and reproducible. The perfect classification system should also standardize the language used to describe the fractures, offer guidelines for treatment, indicate the possibility of complications, and help determine the prognosis. The ideal
system should also provide a mechanism to evaluate and compare the results with treatment of similar fractures treated at various centres and reported at different times in the literature\textsuperscript{19}.

Considering the need to classify the fractures of the proximal radius in an appropriate and reproducible manner, we sought to assess the reproducibility of the three most widely used classification systems. Thus we evaluated the intra- and inter-observer agreement of the Mason modified by Hotchkiss, Morrey and AO/ASIF classifications of proximal radius fractures.
Methods

We analyzed 65 consecutive elbow radiographs performed on patients with fractures of the proximal radius. The patients were treated in the same hospital. Each of the 65 radiographs consisted of two views, anterior-posterior and lateral, and they were numbered, with the patients’ names and ages concealed. Radiographs were excluded if the patient had incomplete skeletal development, pathologic fractures or previous elbow surgery.

The image quality was determined by two orthopaedic surgeons, not evaluators of concordances. The radiography was accepted only when both of these surgeons considered the radiographs acceptable.

Four observers familiar with the classification systems were selected for analysis. These observers were a second-year resident of orthopaedics (R2), a general orthopaedist (GO), a shoulder and elbow surgery specialist (SES), and a radiologist (RD).

To standardize the information for all observers, each were given self-explanatory diagrams with the classification systems. Each observer classified each fracture at three different times, according to the three systems (Mason modified by Hotchkiss, Morrey and AO/ASIF). In the first evaluation (T1), the digitized images of radiographs were viewed on a computer in numerical sequence. Three weeks later, in the second assessment (T2), the sequence of radiographs was randomly altered, as it was in the third assessment (T3), three weeks after T2. This sequence of randomization was known only by a person uninvolved in the assessment of the
images.

The data were collected in spreadsheets and kappa coefficients were calculated for analysis according to the method proposed by Fleiss et al. This method not only calculates the agreement expected by chance, as described earlier in the method of Scott and Cohen, but also the correlation among more than two observers in the evaluation of nominal variables. The kappa coefficient of agreement indicates the proportion of agreement among observers. The kappa values range from -1 to +1: values between -1 and 0 indicate that observed agreement was lower than that expected by chance, 0 indicates a level of agreement equal to that expected by chance, and +1 indicates total agreement. Overall, kappa values below 0.5 are considered unsatisfactory; values between 0.5 and 0.75 are considered satisfactory and appropriate, and values over 0.75 are considered excellent\textsuperscript{20,21,22,23}.

This project was approved by the Research Ethics Committee under No. 363/08, on April 4, 2008.
Results

From the 65 initial radiographs, six were excluded because of poor technical quality, leaving a sample size to 59 radiographs.

Table 1 summarizes the kappa values for intra-observer comparisons for each observer, at the three time points. The concordance was higher among the different with the Mason and Morrey classifications (mean $\kappa=0.582$ and 0.554, respectively) than with the AO/ASIF classification (mean $\kappa=0.483$).

Tables 2 and 3 show the kappa coefficients for intra-observer comparisons between times T1 and T2 and between T2 and T3, respectively. The intra-observer agreement was higher with the Mason and Morrey systems than the AO/ASIF system; and higher between T2 and T3, than between T1 and T2.

Table 4 shows the inter-observer kappa coefficients at each of the times of assessment. Inter-observer agreement was higher in the Mason ($\kappa=0.429–0.560$) and Morrey ($\kappa=0.319–0.487$) classifications than in the AO/ASIF classification ($\kappa=0.250–0.478$). The AO/ASIF and Morrey classifications had greater kappa coefficients at T3 than at T1 and T2. For the Mason classification, the coefficients were almost the same at T2 and T3, and these were larger than for the other classifications and considered satisfactory.
Discussion

The classification systems for this study were selected because they are the most commonly used and studied for fractures of the proximal radius\textsuperscript{15}. Classification systems are of great importance in orthopaedic practice because they are used to describe fractures, guide treatment, and compare treatment outcome within and between studies in the literature. As a result, intra- and inter-observer concordances are essential for any classification system.

In the analysis of intra-observer agreement between three time points, the average kappa coefficient for the AO/ASIF classification was unsatisfactory ($\kappa = 0.483$), ranging from 0.305 (for the SES) to 0.676 (for the R2). These results are probably due to the complexity of the classification, and accord with the results of other studies that evaluated the classification for fractures of other bones\textsuperscript{24,25}. Professional experience had no effect on intra-observer agreement with this system, as indicated by the highest kappa coefficient for the R2 and the lowest for the SES.

For the Mason and Morrey classifications, intra-observer concordances were satisfactory and similar ($\kappa$ of 0.692 and 0.644, respectively). The similarity of these coefficients was expected because the Morrey classification is derived from and more complex than the of Mason classification\textsuperscript{15}.

The intra-observer agreement was higher between T2 and T3 than between T1 and T2. This difference was probably due to the evaluators’ conditioning in examining and classifying the fractures\textsuperscript{26}. 

- 10 -
Regarding inter-observer concordances, the AO/ASIF rating was unsatisfactory at all times, although it improved with time, probably due to conditioning. A similar pattern occurred with the Morrey classification, but the values were slightly higher but still unsatisfactory. The Mason classification had an unsatisfactory mean kappa coefficient at T1 (0.429). However, the means at T2 and T3 were satisfactory (0.560 and 0.551, respectively). These values may also be explained by the complexity of the classifications. These results are similar to those reported by Sheps et al.\textsuperscript{27}, which indicated that the correlation was unsatisfactory for the AO classification and better for the system adapted by Hotchkiss from the Mason system. However, even this adapted Mason system was unsatisfactory when considering the lower limit of 95% of the confidence interval.

It is important to note that this study was limited to assessing the correlation between the observers, and it was not possible to measure the accuracy of these reviews. To do so would require a diagnostic study in which each observer’s classification was compared with the results of an evaluation or standard procedure with high diagnostic sensitivity and specificity.
Conclusions

Inter- and intra-observer concordances of the Mason classification are satisfactory and better than those of the AO/ASIF and Morrey classification systems. The AO/ASIF system has poor concordance and the worst reproducibility.
Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

FTM participated in the design of the study, collected, digitalized and randomized the radiographs, helped to draft the manuscript, revised the bibliography for references and performed the statistical analysis; MJST conceived the study design, analyzed the radiographs and classified the fractures, helped to draft the manuscript and revised the bibliography for references; EFC: participated in the design of the study, analyzed the radiographs and classified the fractures and translated the manuscript to English; AU: participated in the design of the study, analyzed the radiographs and classified the fractures; MHI participated in the design of the study, analyzed the radiographs and classified the fractures; MHM conceived the study design, revised the article and gave the final approval of the version to be published; JBGS participated in the design of the study, revised the article and gave the final approval of the version to be published; JCB participated in the design of the study, revised the article. All authors read and approved the final manuscript.
References


2. Kaas L, Van Riet R, Vroemen JPAM, Eyyendaal D. **The incidence of associated fractures of the upper limb in fractures of the radial head.** *Strat Trau Limb Recon* 2008 Online First


7. Taylor TKF, O’Connor BT. **The effect upon the inferior radio-ulnar joint and excision of the head of the radius in adults.** *J Bone Joint Surg.* 1964;46[B]:83-84


10. Bucholz RW, Heckman JD, Court-Brown CM. Rockwood and Green’s Fractures in Adults. Vol 1, Sixth ed. Philadelphia: Lippincot Williams & Wilkins 2006:1011-12


   Consistency of AO fracture classification for the distal radius. J Bone Joint 
   Surg [Br]. 1996;78-B:726-31


20. Fleiss JL. Measuring nominal scale agreement among many raters. 
   Psychological Bulletin 1971;76(5):378-82


22. Martin JS, Marsh JL, Bonar SK, DeCoster TA, Found EM, Brandser EA. 
    Assessment of the AO/ASIF Fracture Classification for distal tibia. J Orthop 
    Trauma. 1997;11:477-483

    Reproducibility of histomorphological diagnosis with special reference to the 
    kappa statistic. APMIS. 1989;97:689-698

24. Flinkkila T, Nikkola-Sihto A, Kaarela O, Paakko E, Raatikainen T. Poor 
    interobserver reliability of AO classification of fractures of distal radius. J 


26. Belloti JC, Tamaoki MJS, Franciozi CES, dos Santos JBG, Balbachevsky D, 
    Chap EC, Albertoni WM, Faloppa F. As classificações das fraturas do rádio
**Tables**

Table 1. Kappa coefficients for intra-observer comparisons between the three time points (T1, T2 and T3).

<table>
<thead>
<tr>
<th>Observer</th>
<th>AO/ASIF</th>
<th>MORREY</th>
<th>MASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>0.676</td>
<td>0.666</td>
<td>0.710</td>
</tr>
<tr>
<td>GO</td>
<td>0.439</td>
<td>0.548</td>
<td>0.523</td>
</tr>
<tr>
<td>SES</td>
<td>0.305</td>
<td>0.356</td>
<td>0.409</td>
</tr>
<tr>
<td>RD</td>
<td>0.513</td>
<td>0.645</td>
<td>0.685</td>
</tr>
<tr>
<td>Mean</td>
<td>0.483</td>
<td>0.554</td>
<td>0.582</td>
</tr>
</tbody>
</table>

Table 2. Kappa coefficients for intra-observer comparisons between time points T1 and T2.

<table>
<thead>
<tr>
<th>Observer</th>
<th>AO/ASIF</th>
<th>MORREY</th>
<th>MASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>0.674</td>
<td>0.641</td>
<td>0.793</td>
</tr>
<tr>
<td>GO</td>
<td>0.500</td>
<td>0.643</td>
<td>0.676</td>
</tr>
<tr>
<td>SES</td>
<td>0.287</td>
<td>0.330</td>
<td>0.353</td>
</tr>
<tr>
<td>RD</td>
<td>0.496</td>
<td>0.551</td>
<td>0.540</td>
</tr>
<tr>
<td>Mean</td>
<td>0.489</td>
<td>0.541</td>
<td>0.591</td>
</tr>
</tbody>
</table>
Table 3. Kappa coefficients for intra-observer comparisons between moments T2 and T3.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Classification</th>
<th>AO/ASIF</th>
<th>MORREY</th>
<th>MASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td></td>
<td>0.657</td>
<td>0.749</td>
<td>0.689</td>
</tr>
<tr>
<td>GO</td>
<td></td>
<td>0.524</td>
<td>0.604</td>
<td>0.554</td>
</tr>
<tr>
<td>SES</td>
<td></td>
<td>0.485</td>
<td>0.457</td>
<td>0.696</td>
</tr>
<tr>
<td>RD</td>
<td></td>
<td>0.616</td>
<td>0.766</td>
<td>0.829</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.570</td>
<td>0.644</td>
<td>0.692</td>
</tr>
</tbody>
</table>

Table 4. Kappa coefficients for inter-observer comparisons at each time point.

<table>
<thead>
<tr>
<th>Time Point</th>
<th>Classification</th>
<th>AO/ASIF</th>
<th>MORREY</th>
<th>MASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td></td>
<td>0.250</td>
<td>0.319</td>
<td>0.429</td>
</tr>
<tr>
<td>T2</td>
<td></td>
<td>0.386</td>
<td>0.443</td>
<td>0.560</td>
</tr>
<tr>
<td>T3</td>
<td></td>
<td>0.478</td>
<td>0.487</td>
<td>0.551</td>
</tr>
</tbody>
</table>