Reproducibility of three classifications of proximal humeral fractures
Reprodutibilidade de três classificações de fraturas na região proximal do úmero

Eduardo da Frota Carrera¹, Andre Wajnsztejn², Mario Lenza¹, Nicola Archetti Netto³

ABSTRACT
Objective: To propose a new system for classifying proximal humeral neck fractures, and to evaluate intra- and interobserver agreement using the Neer system that is the most commonly used in the area and the Arbeit Gemeinschaft für Osteosynthesefragen system created by an European group, and a new classification system proposed by the authors of this study. Methods: A total of 56 patients with proximal humeral fractures were selected, and submitted to digitized simple radiography in antero-posterior shoulder and scapular profile. Radiographs were analyzed by three observers at time one, and then three and six weeks later. The kappa coefficient modified by Fleiss was used for the analysis. Results: The mean intra-observer Kappa agreement index (k=0.687) of the new classification, was higher than both the Neer classification (k=0.362) and the Arbeit Gemeinschaft für Osteosynthesefragen system (k=0.46). The mean interobserver Kappa agreement index (0.446) of the new classification, also had better results than both the Neer classification (k=0.063) and the Arbeit Gemeinschaft für Osteosynthesefragen system (k=0.028). Conclusion: The new classification considering bone compression had higher results for intra- and interobserver compared to the Neer system, and the Arbeit Gemeinschaft für Osteosynthesefragen system.

Keywords: Shoulder fractures/classification; Observer variation

INTRODUCTION
Proximal humerus fractures are the seventh most common fracture in frequency and correspond to approximately 80% of all humeral fractures¹. The incidence of fractures at this site may vary from 63 to 73% per 100,000 person-years, being more predominant among the elderly population. This type of fracture often occurs in approximately 75% of the population over 60 years old, and is more common among women. The proportion of women to men is three to one¹². The classification system proposed by Neer in 1970 is widely used as a way to assess and determine guidelines to treat fractures of the proximal humerus³. Recently, an European group⁴ the Arbeit Gemeinschaft

Study carried out at Universidade Federal de São Paulo – UNIFESP, São Paulo (SP), Brazil.
¹ Hospital Israelita Albert Einstein – HAE, São Paulo (SP), Brazil.
² Universidade Federal de São Paulo – UNIFESP, São Paulo (SP), Brazil.
Corresponding author: Eduardo da Frota Carrera – Avenida Albert Einstein, 627, Building A1, office #322 – Zip code: 05652-000 – São Paulo (SP), Brazil – Phone: (55 11) 2151 2417 – E-mail: e.carrera@me.com
Received on: Jun 29, 2012 – Accepted on: Nov 5, 2012
Conflicts of interest: none.
Für Osteosynthesefragen (AO/ASIF)\(^4\) proposed a system that also became acknowledged as a classification for these fractures.

Currently classification systems used for fractures of the proximal extremity of the humerus have low agreement and reproducibility in radiographies and also in computed tomography. Such difference towards classifications’ agreement could be due to multiple variables, for example, studies of low quality because of difficulties in placing patients with a fracture of proximal humerus, and also the surgeon’s lack of experience to analyze these fractures\(^5\)\(^-\)\(^10\).

In spite of improvements demonstrated by some authors in the agreement of intra- and interobserver results using more complex tests like simple tomography and tridimensional (3D) reconstructions, these results are inconclusive when the relationship between classifications and suggestions of treatment is evaluated towards physician’s experience\(^7\)\(^-\)\(^9\).

**OBJECTIVE**

This study aimed to evaluate intra- and interobserver agreement of classification systems proposed by Neer and by the AO/ASIF group, and to compare their results with our classification using radiographic exams of patients with fracture of the proximal extremity of the humerus.

**METHODS**

This study was carried out at the Orthopedic and Traumatology Department at Universidade Federal de São Paulo (UNIFESP), and was approved by the Ethical and Research Committee of the same institution (process #0234/08). From January 2002 to January 2008 antero-posterior and lateral shoulder radiographic views at the scapular plane of adult patients with isolated fracture or fracture-dislocation of proximal humerus were selected.

Radiographs were performed with the patient sitting down or standing upright. For the antero-posterior radiography the posterior face of the affected shoulder was placed close to the film and the patient’s shoulder was externally rotated in about 40°. To perform the lateral radiographs, the anterior face of the affected shoulder was placed close to the film and to the patient’s opposite shoulder also rotated in about 40°. The axillary view was not performed in all patients of the study mainly for the difficulty in placing them. Hence, because not all patients had this view, it was not included in this analysis.

Exclusion criteria: (1) radiographs of a pathologic fracture in the proximal humerus region, or any other abnormal change or tumor that could interfere in the joint normal anatomy; (2) skeletally immature patients’ radiographs; (3) a previous fracture; (4) a tuberosity isolated fracture; (5) low quality X-ray exams.

To calculate the sample size the kappa index modified by Fleiss et al. was used as the main variable\(^11\). Type I error was pre-stated as 5% (confidence of 95%) and type II error as 20% (test power of 80%). A population standard deviation of 0.40 of kappa value and a minimal difference to be detected of 0.30 of Kappa value were used. A total sample of 14 radiographs per observer was calculated after the definition of such values.

Simple shoulder radiographs in antero-posterior and lateral position of the scapula were digitized by a radiologist not involved in the study. To take the radiographs a Nikon® Coolpix 4500 model camera was used. Images were analyzed by four observers after digitalization. In the analysis they considered the Neer classification\(^3\), the AO/ASIF European group system\(^4\) and the new classification suggested in this study.

The observers were: a sixth year medical student (1), an orthopedist specialized in shoulder (2), a radiologist specialized in musculoskeletal diseases (3) and one orthopedist specialized in trauma\(^4\).

Classifications were illustrated and explained on a pamphlet describing each classification (Figures 1 to 3). To each observer a ruler and goniometer were given to be used during the fractures’ assessment. The participants’ personal identifications in the radiographs were protected and then they were randomly numbered. Images were assessed and classified by all observers at time one (T1); then they were randomly numbered and the procedure redone after three (T2) and six (T3) weeks. The four observers assessed the same radiography on three different times (T1, T2, and T3).

According to the Neer classification\(^3\), the fractures were divided into six groups (Figure 1):

- **Group I**: minimally displaced fracture, with displacement of <1 cm or angle <45°;
- **Group II**: fracture with displacement of the proximal humerus anatomical neck;
- **Group III**: fracture with displacement of the proximal humerus surgical neck;
- **Group IV**: fracture with displacement of the greater humerus tuberosity dividing into two parts with no surgical neck displacement; into three parts with surgical neck displacement; or into four parts with fracture and displacement of the lesser tuberosity;
- **Group V**: fracture with displacement of the lesser tuberosity presenting the same features of group IV.
Reproducibility of three classifications of proximal humeral fractures

In the AO/ASIF system(4), fractures were classified in three types, structured in three groups and totalizing nine types of fractures (Figure 2).

In the classification system proposed by the present study, fractures were divided as follow: compression fractures and non-compression fractures (Figure 3).

Fractures were considered as compression fractures when a plastic or permanent deformity of the spongious metaphyseal bone occurred by compression or shearing.
among the fractured fragments no matter the number of fragments involved. Radiographic image of fractured fragments are usually not well identified or defined in compression fractures. Non-compression fractures are often identified by well-defined fragments on radiographic images. In this type of fracture no loss of bone tissue is seen due to compression mechanism or shearing among fragments.

**Statistical analysis**

The kappa coefficient agreement modified by Fleiss et al. was used for intra- and interobserver analyses\(^{(1)}\).

The kappa coefficient provided an agreement proportion matched among observers. Values of kappa coefficient varied from -1 to +1. When the value was -1 it meant total disagreement, if +1 meant total agreement, and when zero meant disagreement. The kappa coefficient values may also be attributed arbitrarily to subdivisions. If values between 0.00 and 0.20 are found, they suggest unsatisfactory agreement; between 0.21 and 0.40 little agreement; between 0.41 and 0.60 moderate agreement; between 0.61 and 0.80 satisfactory and adequate agreement; values over 0.80 indicated almost perfect agreement\(^{(12-14)}\).

A significance level of 5% was adopted, and there was rejection for a null coefficient for descriptive levels \(p < 0.05\). It is important to note that a coefficient different from zero does not indicated high agreement.

Agreement percentage was also calculated among several measures that had easier interpretation.

Calculations were done using the R statistical package.

**RESULTS**

A total of 174 patients with fracture of proximal extremity of the humerus were assisted from January 2002 to January 2008. From this total, only 71 patients met the inclusion criteria. Radiographs of the 71 patients were evaluated by an orthopedist and a radiologist not involved in the study, who selected 56 images in anteroposterior view of the shoulder and lateral of the scapula. The demographic data of patients are presented on table 1.

The highest kappa intra-observer agreement index obtained in the three moments (M1, M2 and M3) was seen for the classification proposed in this study \((k=0.687)\) followed by the AO/ASIF classification \((k=0.460)\), and by the Neer classification \((k=0.362)\) (Table 2).

As for the kappa interobserver index the classification of this study had a high mean value in the three moments \((k=0.446)\) as well, followed by the Neer classification \((k=0.063)\), and by the AO/ASIF classification \((k=0.028)\) (Table 3).

**Figure 3.** Classification proposed by authors of this study
(scheme by: Professor Caio Nery)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (years/variation)</td>
<td>63.6 (20-89)</td>
</tr>
<tr>
<td>Men (%)</td>
<td>18 (32.1)</td>
</tr>
<tr>
<td>Women (%)</td>
<td>38 (67.9)</td>
</tr>
<tr>
<td>Affected side</td>
<td></td>
</tr>
<tr>
<td>Right (%)</td>
<td>15 (26.8)</td>
</tr>
<tr>
<td>Left (%)</td>
<td>41 (73.2)</td>
</tr>
<tr>
<td>Dominant side</td>
<td></td>
</tr>
<tr>
<td>Right (%)</td>
<td>48 (85.7)</td>
</tr>
<tr>
<td>Left (%)</td>
<td>8 (14.3)</td>
</tr>
<tr>
<td>Trauma mechanism</td>
<td></td>
</tr>
<tr>
<td>Fall without level difference (%)</td>
<td>43 (76.8)</td>
</tr>
<tr>
<td>Fall down stairs (%)</td>
<td>6 (10.7)</td>
</tr>
<tr>
<td>Motorcycle accident (%)</td>
<td>2 (3.6)</td>
</tr>
<tr>
<td>Car accident (%)</td>
<td>5 (8.9)</td>
</tr>
</tbody>
</table>
The less experienced observer (1) obtained the lowest values for the kappa intra-observers index in all classification systems. The orthopedist specialized in trauma had the highest values of the kappa index in the classification proposed in this study and in the AO/ASIF. The orthopedist specialized in shoulder presented the highest kappa index for the Neer classification (Table 2).

**DISCUSSION**

Classification systems should provide tools to support clinical evaluation. A good classification system has to be valid, reliable and reproducible. However, to be ideal, a system must have a standard language for safety communication, create guidelines for treatments, and help to determine the disease prognosis. In addition, it should be an instrument to evaluate and compare results achieved by the treatment of similar diseases at several research centers registered in the literature at different times (15).

The concept that the main systems for the classification of proximal extremity humeral fractures had low agreement and are of little reproducibility was demonstrated by the results of this study. The classification models proposed by Neer and the AO/ASIF group are widely accepted and used nowadays, however, there are many criticisms related to the difficulty to reproduce them (5, 10, 16, 17).

In a literature review by Brorson and Hrobjartsson (17) evaluating the Neer classification, involving 11 studies, 88 observers and 468 fractures, the kappa interobserver agreement varied between 0.17 and 0.52, that is, unsatisfactory to moderate. This study evaluated a heterogeneous group of observers and had low agreement index in the Neer classification, that is, unsatisfactory.

Table 2. Intraobserver agreement

<table>
<thead>
<tr>
<th>Observer</th>
<th>Kappa</th>
<th>p value</th>
<th>%</th>
<th>Observer</th>
<th>Kappa</th>
<th>p value</th>
<th>%</th>
<th>Observer</th>
<th>Kappa</th>
<th>p value</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.501</td>
<td>&lt;0.001</td>
<td>61.4</td>
<td>1</td>
<td>0.134</td>
<td>&lt;0.001</td>
<td>8.8</td>
<td>1</td>
<td>0.109</td>
<td>&lt;0.001</td>
<td>10.5</td>
</tr>
<tr>
<td>2</td>
<td>0.722</td>
<td>&lt;0.001</td>
<td>80.7</td>
<td>2</td>
<td>0.595</td>
<td>&lt;0.001</td>
<td>50.9</td>
<td>2</td>
<td>0.600</td>
<td>&lt;0.001</td>
<td>61.4</td>
</tr>
<tr>
<td>3</td>
<td>0.659</td>
<td>&lt;0.001</td>
<td>73.7</td>
<td>3</td>
<td>0.390</td>
<td>&lt;0.001</td>
<td>40.4</td>
<td>3</td>
<td>0.151</td>
<td>&lt;0.001</td>
<td>3.51</td>
</tr>
<tr>
<td>4</td>
<td>0.966</td>
<td>&lt;0.001</td>
<td>91.2</td>
<td>4</td>
<td>0.723</td>
<td>&lt;0.001</td>
<td>73.7</td>
<td>4</td>
<td>0.587</td>
<td>&lt;0.001</td>
<td>56.1</td>
</tr>
</tbody>
</table>

Kappa mean = 0.687

AO/ASIF: Arbeit Gemeinschaft für Osteosynthesefragen.

Table 3. Interobserver agreement

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Kappa</th>
<th>p value</th>
<th>%</th>
<th>Evaluation</th>
<th>Kappa</th>
<th>p value</th>
<th>%</th>
<th>Evaluation</th>
<th>Kappa</th>
<th>p value</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0.520</td>
<td>&lt;0.001</td>
<td>56.1</td>
<td>T1</td>
<td>0.0263</td>
<td>0.202</td>
<td>1.75</td>
<td>T1</td>
<td>0.0301</td>
<td>0.186</td>
<td>1.75</td>
</tr>
<tr>
<td>T2</td>
<td>0.331</td>
<td>&lt;0.001</td>
<td>38.6</td>
<td>T2</td>
<td>0.0326</td>
<td>0.131</td>
<td>1.75</td>
<td>T2</td>
<td>0.0677</td>
<td>0.005</td>
<td>1.75</td>
</tr>
<tr>
<td>T3</td>
<td>0.488</td>
<td>&lt;0.001</td>
<td>54.4</td>
<td>T3</td>
<td>0.0252</td>
<td>0.235</td>
<td>1.75</td>
<td>T3</td>
<td>0.0918</td>
<td>&lt;0.001</td>
<td>3.51</td>
</tr>
</tbody>
</table>

Kappa mean = 0.446

AO/ASIF: Arbeit Gemeinschaft für Osteosynthesefragen.
experienced group with a less experienced one did not find significant differences in the interobservers agreement indexes\(^{(15)}\).

So, it is believed that when an observer is comfortable using a specific classification the result is higher. However, some studies have been observing that if the same classification is used repeatedly in different times it does not change intra- and interobservers reproducibility\(^{(20)}\). This study included a less experienced observer (a medical student). In spite of the fact that a student participation could be a complication factor for the data analysis, his participation was on purpose in order to assess the validity and reproducibility of the classifications used.

A randomized clinical trial\(^{(21)}\) noted an improvement in the agreement using the Neer classification after training 14 observers, being the kappa index before the training 0.34 and after it 0.62, which was not seen in the control group. Although, a systematic review \(^{(16)}\) done in 1993 did not show that more experienced observers had less disagreement than those with less experience.

Because classifications most used by investigators present low agreement and reproducibility to identify of proximal humerus fractures, it is proposed a new concept to classify these fractures in order to improve interpretations and investigate new treatment alternatives.

This study has a number of limitations such as: (1) this is a retrospective study in which bias rate may affect results; (2) only two radiographs were used to evaluate the fractures not being possible to examine the three perpendicular plane radiographies therefore it was not possible to precisely evaluate fracture dislocations; (3) this study did not include more complex and widely imaging exams used in clinical practice, as simple tomography or with reconstruction, causing bias and misunderstandings. This approach reflects the reality of some developing countries.

Bone compression determines the great difference between the two types of the proximal humerus extremity fractures (the so called compression and non-compression fractures) enabling to identify complex fractures that often present poor prognosis. The metaphyseal spongious bone which is compressed between the fractured shaft fragment and the humeral head, constitutes a barrier to reduce and maintain the reduction during the intra- and post-operative periods, being these fractures classified as complex or of poor prognosis. However, if this bone abnormal situation is considered at reduction or fixation time, it will provide a better understanding for the surgeon regarding the fracture type and its reduction, therefore, enabling a better prognosis\(^{(22)}\).

The literature describes some unsatisfactory prognosis criteria for fractures of the proximal humerus, among them it is possible to highlight the metaphyseal comminution and the epiphyseal varus displacement, the elderly patients over 65 years old, the associated fractures, and the humeral head dislocation\(^{(23,24)}\). It is believed that one of the most important factors to define the fracture complexity is the displacement with shearing or compression among fragments that create a loss of the spongious metaphyseal bone support. To consider this feature in fractures, no matter the number of involved fragments, is the concept that was applied to the classification proposed by this study.

**CONCLUSION**

The classification considering bone compression had better results in intra- and interobserver interpretation than the Neer and the AO/ASIF classifications.

**REFERENCES**


