Our purpose was to verify if the Neer and AO-ASIF classifications for fractures of the proximal humerus satisfy the requisites of simplicity and reproducibility and if the parameters that they consider to establish the severity of the fracture are similar. Two of the authors classified the proximal humeral fractures of 227 patients based on plain radiographs, and they repeated the classification five years later. The reliability, reproducibility and coherence of the classifications were investigated. Inter-observer reliability was $K = 0.77$ (Neer) and $K = 0.65$ (AO-ASIF) while intra-observer reproducibility was $K = 0.68$ (examiner I) and $K = 0.63$ (examiner II). In 1/5 of the cases, disagreement led to a different classification of the same fracture. Furthermore, neither classification establishes a linear scale of gravity able to provide an indication for treatment.

The Neer and AO-ASIF classifications have a low reproducibility and reliability when fractures, especially those with 3 or 4 parts, are assessed by means of plain radiographs. Therefore, patients with complex fractures should be submitted to CT to have a correct pre-operative diagnosis.

Keywords: humeral head fractures; Neer classification; AO-ASIF classification; reproducibility, reliability.

INTRODUCTION

The purpose of a fracture classification is to establish, following conventional parameters such as morphologic characteristics and fracture level, a simple and memorisable scale of fracture severity and outcome and to guide fracture treatment. A classification should have a high reproducibility and reliability and allow a meaningful comparison of results between different studies.

Several classifications of proximal humeral head fractures have been suggested. In 1896 Kocher (19) proposed a classification based on the anatomical level of the fracture. Codman (8) distinguished four main anatomical fragments of the fracture: greater tuberosity, lesser tuberosity, articular segment and surgical neck. In 1945 Dehen (11) presented a new classification based on actiopathogenetic mechanisms. These classifications did not have a wide...
application, but represented the basis of the success-

sive systems, together with the studies of Laing (22)
on the vascularisation of the humeral head. In 1970
Neer (29) proposed a new classification, which is
currently the most widely used; this was based on
the anatomy of the humeral head, on the bio-

mechanics of the injury and on displacement of the
fragments. This classification distinguishes four
main parts (humeral head, lesser and greater
tuberosity and shaft), 6 Groups (I-VI) and 16 Sub-
groups, according to the level of the fracture or
fracture-dislocation and to the importance of the
displacement of one or more parts (30). The AO-
ASIF classification (34), which is not so widely
used, distinguishes 27 subgroups on the basis of
location, articular involvement, degree of com-

mination and associated shoulder dislocation, with
special emphasis on the integrity of the vascular
supply. This system distinguishes valgus impacted
four-part proximal humerus fractures from other
four-part injuries with partial preservation of vascu-
lar inflow to the articular segment through the
medial capsule.

The validity of the Neer and AO-ASIF classifica-
tion has been scarcely analysed. Court-Brown et
al (10) found the AO-ASIF classification more
comprehensive than the Neer system, while
Naranja and Iannotti (28) emphasised the usefulness
of Neer’s criteria in intraoperative decision making.
Nevertheless it emerged that both classifications
have a low reproducibility and reliability (0.25 < K
coefficient of Cohen < 0.75) (20,35).

The purpose of this study was to verify if the two
most commonly used classifications of humeral
head fractures satisfy the requisites of simplicity,
clinical usability and reproducibility and if their
differing principles of codification may lead to a dif-
ferent treatment approach of the same fracture.

MATERIALS AND METHODS

Between 1993 and 2001, 642 patients with a proximal
humeral fracture were conservatively treated at our
department. At the time of injury, the mean age of the
patients (205 males and 437 females) was 73 years
(range: 18-97 years). In 2004 all the fractures were
catalogued by two authors using the Neer system. Six
years later, 250 of the 642 medical records were ran-
domly selected and 23 were eliminated because of poor
radiographic quality. For each of the 227 cases included,
least two preoperative radiographic views were avail-
able (true AP + axillary views in 99 cases and true AP +
axillary and outlet view in 128). The same two authors
separately catalogued again the 227 fractures according
to the Neer and AO and ASIF classification. They were not
involved in the random selection of the cases. Both
expert examiners with a different clinical experience (the
first examiner more specifically competent than the
second one), were provided with a goniometer, a milli-

meter-scale and the description of the two classification
systems.

Intra-observer reliability was analysed for both
classifications while intra-observer reproducibility, after
6 years, was analysed for each of the two authors.
Reasons for disagreement were assessed as well as the
comparison between reproducibility and degree of
clinical experience of the two authors. Finally, the clas-
sifications were compared in order to assess the coher-
ence between the two systems.

RESULTS

NEER classification

In 2004, the examiners’ classifications were in
agreement for 71.8% of the fractures (Concordance
K = 0.77). In 48 of the 227 fractures, the degree of
disagreement between the examiners was apprecia-
table. The same fracture was allotted to different
Groups in 20.7% of the cases or subgroups in 7.4%
of the cases (Fig. 1); in 82% of these cases there
was a fracture of the surgical neck with the partici-
patation of one or both tuberosities. After 6 years,
examiner II classified the same fracture similarly as
before in 73.1% of the cases (reproducibility: K =
0.63) and examiner I in 68.7% of the cases (repro-

Statistical Analysis

The kappa coefficient of Cohen was used, as a statis-
tical measure of inter-rater agreement, to assess the inter-
observer reliability and reproducibility of the classifica-
tion (9). This coefficient (K) can have values from -1 to 1
(-1 = highest disagreement; 0 = random agreement; 1 =
highest agreement). K values < 0.8 stand for a non
optimal reproducibility or reliability (23).

RESULTS

NEER classification

In 2004, the examiners’ classifications were in
agreement for 71.8% of the fractures (Concordance
K = 0.77). In 48 of the 227 fractures, the degree of
disagreement between the examiners was apprecia-
table. The same fracture was allotted to different
groups in 20.7% of the cases or subgroups in 7.4%
of the cases (Fig. 1); in 82% of these cases there
was a fracture of the surgical neck with the partici-
patation of one or both tuberosities. After 6 years,
examiner II classified the same fracture similarly as
before in 73.1% of the cases (reproducibility: K =
0.63) and examiner I in 68.7% of the cases (repro-

Acta Orthopaedica Belgica, Vol. 77 - 6 - 2011
ducibility: K = 0.68). The different experience of the examiners did not influence the reproducibility of this classification (K = 0.68 and K = 0.63). Table I shows the classifications of the 227 fractures carried out by the two examiners.

**AO-ASIF classification**

Concordance between the two examiners was observed in 62.9% of the cases (K = 0.65). In 45 of the 227 fractures, the degree of disagreement between the examiners was considerable and led to a different classification of the same fracture; 20.2%, 4.8% and 11.9% of the fractures examined were catalogued, respectively, into different Types, Groups or Subgroups (Fig. 1). Disagreement occurred above all for extra-articular fractures. Table II shows the classifications of the 227 fractures according to the two examiners.

**Coherence between the two classifications**

A single fracture belonging to a specific group of the Neer classification could be catalogued into different Groups of the AO-ASIF classification, which includes fractures with satisfactory (type A) or unsatisfactory (type C) prognosis. This discordance was seen especially for the following fractures: Groups: I, II, IV 2 part (2p), V 2p and VI 2p. At the same time, it emerged that different AO-ASIF Groups corresponded to different Neer Groups: e.g., B1.1 and B2.1 fractures (AO-ASIF) corresponded, respectively, to Group I, III, IV2p, IV3p and I, III, V 2p, V 3p (Neer). Furthermore, Groups B2.1 and B2.3 (AO-ASIF) corresponded to Neer Group I, III, IV 2p, V 2p and IV 3p. Table III shows the correspondence between the two systems.

**DISCUSSION**

Both classifications (Neer and AO-ASIF) were difficult to memorise, compromising their practical use during ordinary clinical practice. After 6 years, the reproducibility of the Neer classification was low for both examiners and not significantly dependent of the level of expertise, contrary to what has been reported by other studies \(^{(20,39)}\). Moreover, a recent study demonstrated that training improves agreement among both experts and non-experts \(^{(74)}\).

The highest percentage of disagreement was found with fractures belonging to Group III, IV 2p and 4p of Neer and Type B of AO-ASIF classification, or with those pluri-fragmented fractures in which it was hard to establish the degree and type of displacement. Neither classification establishes a clear linear scale of the severity of the fracture, able to provide an immediate indication for treatment; besides, the prognosis of the fracture is not always deducible from the group it belongs to, and does not worsen with the progression of the groups. For
example, the non or slightly displaced fractures of the anatomical neck may be classified in opposite groups between the two classifications – Group I (Neer) and Group C (AO-ASIF) – as well as displaced fractures of the greater tuberosity combined with gleno-humeral dislocation – Group A (AO-ASIF) and Group VI (Neer). The fracture of the greater tuberosity with proximal migration of the fragment causing narrowing of the subacromial space (generally with a negative prognosis) could be catalogued in Group I of Neer and in Group A1.2 of AO-ASIF, while the same fracture with a distal migration of the tuberosity (generally with a good prognosis) is catalogued in Group IV of Neer and A2.2 of AO-ASIF (32). Furthermore, a fracture of the surgical neck with a 30° varus displacement leading to poor results in 80% of cases (32), is classified in the first group of both classifications (Group I of Neer and Group A2.2 of AO-ASIF), while fractures of the surgical neck with a marked valgus displacement of the humeral head and an inferior or posterior displacement of the greater tuberosity (11% of unsatisfactory results in our experience) are catalogued, respectively, in Group III and B2.3.

Both classifications have an excessive subdivision into Groups, which is not always useful during ordinary clinical practice: for example, fractures 1p and 2p of Neer classification and those belonging to Group A1, A2, A3, B1 are similar in prognosis and therapeutic indications in most cases (16). Despite this, some types of fractures associated with poor results, for example those of the surgical neck with

<table>
<thead>
<tr>
<th>Groups</th>
<th>Examiner I</th>
<th>Examiner II</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 1.1</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>A 1.2</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>A 1.3</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>A 2.1</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>A 2.2</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>A 2.3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>A 3.1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>A 3.2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>A 3.3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>B 1.1</td>
<td>71</td>
<td>83</td>
</tr>
<tr>
<td>B 1.2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>B 1.3</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>B 2.1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>B 2.2</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>B 2.3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>B 3.2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C 1.1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C 1.2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>C 1.3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C 2.1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>C 2.2</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>C 2.3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>C 3.2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C 3.3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>TOTAL</td>
<td>227</td>
<td>227</td>
</tr>
</tbody>
</table>

Table III. — The Neer and AO-ASIF classification of each fracture compared: a fracture catalogued in a particular Neer Group may be catalogued in different Groups of the AO-ASIF system

<table>
<thead>
<tr>
<th>Group I</th>
<th>A1.1, A1.2, A2.1, A2.2, (A2.3), (A3.1), (A3.3), B1.1, (B1.2), B1.3, (B2.1), (B2.3), C1.1, C1.2, (C1.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group II</td>
<td>C1.3, (C2.1), (C2.2)</td>
</tr>
<tr>
<td>Group IV 2p</td>
<td>A1.2, B1.1, B1.3, (B2.1), B2.3, (C1.1), (C1.2)</td>
</tr>
<tr>
<td>Group V 2p</td>
<td>A1.1, B1.2, (B2.1), (B2.3), (C1.1), (C1.2)</td>
</tr>
<tr>
<td>Group V 3p</td>
<td>B1.2, (B2.2)</td>
</tr>
<tr>
<td>4 part fractures</td>
<td>C1.1, C2.1, C2.2, C2.3, C3.3</td>
</tr>
<tr>
<td>Group VI 2p</td>
<td>A1.3, (B3.1), (B3.2), (C3.1), C3.2</td>
</tr>
<tr>
<td>Group VI 3p</td>
<td>B3.2, (B3.3), C3.2</td>
</tr>
<tr>
<td>Group VI 4p</td>
<td>C3.2, C3.3</td>
</tr>
<tr>
<td>Fractures of the articular segment</td>
<td>(C2.3), (C3.3)</td>
</tr>
</tbody>
</table>
slight varus displacement or those with an upward migrated greater tuberosity, are not well identified in either classification (2,32).

Our study highlighted the scarce coherence between the two classifications: the same fracture may be catalogued differently by the two systems. For example, among extra-articular bifocal fractures, those of Group B1.1 (AO-ASIF), are included in different Groups of Neer system (Group I, III, IV 2p and IV 3p). Furthermore, Group I of Neer (generally with a good prognosis) corresponds, in the AO-ASIF system, to many Groups with satisfactory and unsatisfactory prognosis (A1.1, A1.2, A2.1, A2.2, A2.3, A3.1, A3.3, B1.1, B1.2, B1.3, B2.1, B2.3, C1.1, C1.2, C1.3). We found many contrasting points between the two classifications and a low reproducibility and reliability. These data may explain the great disagreement concerning therapeutic indications and results presented in literature with humeral head fractures (14,15,17,21,24,31,33, 36,37,39,42,43,48).

Finally, these two systems are based on the accurate determination of the degree of angular displacement of the four main segments; nevertheless, the radiographical examination did not allow an accurate evaluation, especially in the case of complex fractures. This fact raised doubts concerning the classification of a single fracture using the two systems; such doubts can be blamed on the disagreement and low reproducibility which resulted from this study and from others (3,4,6,20,26,35,38, 39,41).

Neer in 2002 stated that reliable use of the 4-segment system requires exacting roentgen studies and knowledgeable interpretation of the films. Previous studies revealed that lateral scapular projections do not improve the reliability and reproducibility of the Neer and AO-ASIF classifications (35,40,44). CT has been shown to be useful when conventional radiography is not sufficient due to low quality imaging and presence of osseous overlap (1). Furthermore, CT reveals fractures not clearly seen on plain radiographs and contributes to a better understanding of displaced three and four-part fractures (5,46); it is also valuable in delineating the configuration of the fracture, helping to plan surgical reconstruction (13,47). For this reason, further studies have tried to incorporate computed tomographies (CTs and 3D reconstructions) in delineating the fracture patterns and subsequently applying the Neer classification, reporting contrasting results (12,45).

In conclusion, our study shows that the Neer and AO-ASIF classifications are not compatible with each other and are time-consuming. They may lead to different treatment approaches for the same fracture, depending on the classification used.

The two classifications have a low reproducibility and reliability, above all when they refer to 3 and 4 parts fractures and when the classification is made based on standard radiographic views. Therefore, we suggest submitting patients with complex fractures to CT evaluation.

REFERENCES


