

The reliability and reproducibility of the Hertel classification for comminuted proximal humeral fractures compared with the Neer classification

Gijs I.T. Iordens, MD PhD¹, Kiran C. Mahabier, MD¹, Florian E. Buisman, MD¹, Niels W.L. Schep, MD PhD², Galied S.R. Muradin, MD PhD³, Ludo F.M. Beenen, MD PhD⁴, Peter Patka, MD PhD⁵, Esther M.M. Van Lieshout, MSc PhD¹, Dennis Den Hartog, MD PhD^{1*}

¹Trauma Research Unit Department of Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

²Trauma Unit Department of Surgery, Academic Medical Center, P.O. Box 22660, 1100 DD Amsterdam, The Netherlands

³Department of Radiology, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

⁴Department of Radiology, Academic Medical Center, P.O. Box 22660, 1100 DD Amsterdam, The Netherlands

⁵Department of Emergency Medicine, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

*Corresponding author: D. Den Hartog, MD PhD,

Erasmus MC, University Medical Center Rotterdam, Trauma Research Unit Dept. of Surgery
P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Phone: +31.10.7031050, Fax: +31.10.7032396

Mail: d.denhartog@erasmusmc.nl

ABSTRACT

Introduction: The Neer classification is the most commonly used fracture classification system for proximal humeral fractures. Inter- and intra-observer agreement is limited, especially for comminuted fractures. A possibly more straightforward and reliable classification system is the Hertel classification. The aim of this study was to compare the inter- and intra-observer variability of the Hertel with the Neer classification in comminuted proximal humeral fractures.

Materials and Methods: Four observers evaluated blinded radiographic images (X-rays, CT-scans, and CT-scans with 3D-reconstructions) of 60 patients. After at least two months classification was repeated.

Results: Inter-observer agreement on plain X-rays was fair for both Hertel ($\kappa=0.39$; 95% CI 0.23-0.62) and Neer ($\kappa=0.29$; 0.09-0.42). Inter-observer agreement on CT-scans was substantial ($\kappa=0.63$; 0.56-0.72) for Hertel and moderate for Neer ($\kappa=0.51$; 0.29-0.68). Inter-observer agreement on 3D-reconstructions was moderate for both Hertel ($\kappa=0.60$; 0.53-0.72) and Neer ($\kappa=0.51$; 0.39-0.58).

Intra-observer agreement on plain X-rays was fair for both Hertel ($\kappa=0.38$; 0.27–0.59) and Neer ($\kappa=0.40$; 0.15-0.52). Intra-observer agreement on CT-scans was moderate for both Hertel ($\kappa=0.50$; 0.38-0.66) and Neer ($\kappa=0.42$; 0.35-0.52). Intra-observer agreement on 3D-reconstructions was moderate for Hertel ($\kappa=0.55$; 0.45-0.64) and substantial for Neer ($\kappa=0.63$; 0.48-0.79).

Conclusions: The Hertel and Neer classifications showed a fair to substantial inter- and intra-observer agreement on the three diagnostic modalities used. Although inter-observer agreement was highest for Hertel classification on CT-scans, Neer classification had the highest intra-observer agreement on 3D-reconstructions. Data of this study do not confirm

superiority of either classification system for the classification of comminuted proximal humeral fractures.

INTRODUCTION

The incidence of fractures of the proximal humerus is 106 per 100,000 person years and a triplication of this number is expected by the year 2030.¹ Besides the impact of these fractures on health and quality of life, they also impose an economic burden on the society.²⁻⁶ The most important determinants for treatment choice include age, co-morbidities, functional demand, surgical expertise, and the personality of the fracture.⁷ Approximately eighty percent of the proximal humeral fractures are minimal or non-displaced fractures which can be treated conservatively. However, comminuted fractures (*i.e.*, three-part, four-part, and head-split fractures) often require surgical treatment.^{8,9}

Since most clinical studies include only specific fracture classes, a reliable and reproducible classification is needed for adequate patient selection. The most frequently used classification systems for the proximal humeral fractures are the Neer and the AO (Arbeitsgemeinschaft für Osteosynthesefragen) classifications.¹⁰ Unfortunately both classifications showed disappointing inter- and intra-observer agreement with kappa values below 0.40 for classifying comminuted proximal humeral fractures.¹¹⁻¹⁴ The Neer classification defines fracture displacement as a 1-cm distance and/or a 45° angle between fragments.¹⁵ Exact measurements of the displacement and angulation make this system difficult to apply in clinical practice. Therefore, a classification system with better reliability and reproducibility for comminuted proximal humeral fractures is warranted. Such a classification system may guide treatment and evaluation of results.^{16,17} A classification system that potentially meets these criteria is the Hertel classification. This classification is also known as the Lego-system, which is based upon the four-part concept of Codman.^{18,19}

The aim of this study was to compare the inter-observer reliability and intra-observer reproducibility of the Hertel with the Neer classification for comminuted proximal humeral

fractures.

MATERIALS AND METHODS

Radiographs

Radiographic images were selected from hospital records and from the radiology system PACS (Picture Archiving and Communication System) of two Level 1 trauma centers using a unique identifying code for diagnosis and treatment of all consecutive proximal humerus and humeral shaft fractures (Diagnose Behandel Combinatie, DBC, 207), or based upon the Abbreviated Injury Score (AIS90; 752600.2 Humerus - fracture NFS, 752602.2 Humerus - fracture - closed/undisplaced, 752604.3 Humerus - fracture - open/displaced/comminuted, 752606.3 Humerus - fracture - with radial nerve involvement).

All consecutive adult patients diagnosed with a comminuted (three- and four part and head-split fractures) fracture of the proximal humerus between January 1 2003 and October 15 2010 of whom plain X-rays and CT-scans were available, were found eligible.

Pathological and recurrent fractures were excluded. The principal investigator (GITI) was adequately trained and had sufficient experience to select the radiographic images meeting the criterion of representing a comminuted fracture. The first eligible 60 patients were selected. Comminuted fractures were defined as three-part, four-part, and head-split fractures according to Hertel.

Radiographs obtained from the standard trauma series were used. This series at least had to include anteroposterior and lateral views. Radiographs accepted for clinical decision making were regarded of sufficient quality for inclusion. Two X-rays were available for 50 patients, three for seven patients and four for three patients. The 3D-volume rendering reconstructions were made using an open-source program (OsiriX version 3.9.2, Geneva, Switzerland).²⁰ The reconstructions could be rotated over both X- and Y-axis and consisted of 40 images per axis. All X-rays and CT-scans were collected and blinded by the principal

investigator, who did not participate in the classification of the images. In order to guarantee identical viewing conditions all observers evaluated all images with the same open-source viewer (ClearCanvas Workstation version 2.0, Toronto, Canada). For every observer the cases were presented in a different, random order.

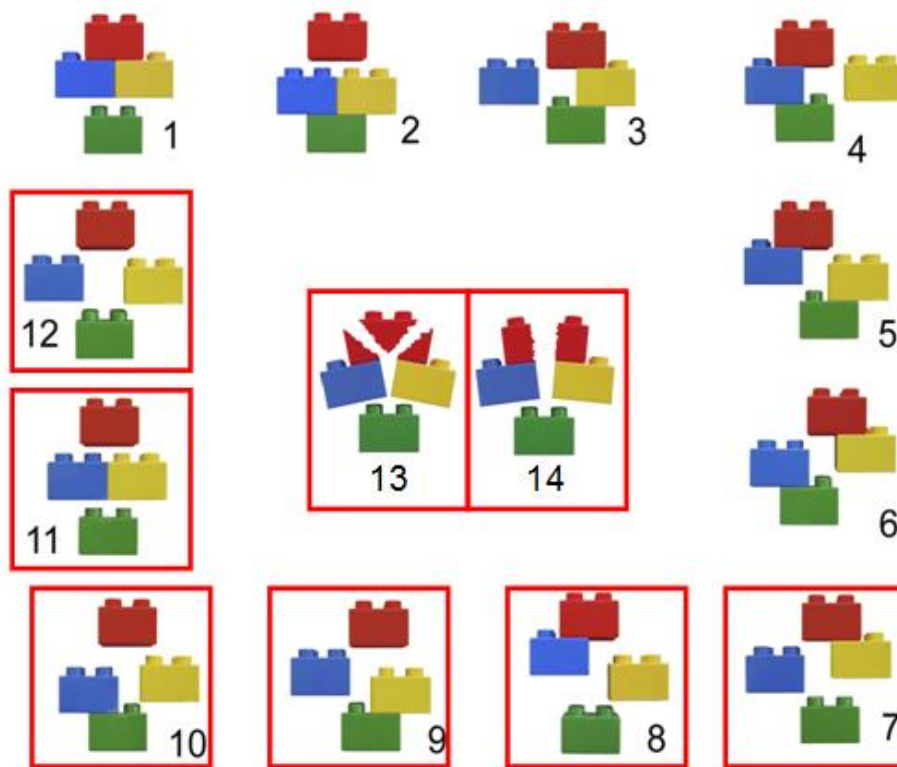
Hertel classification

The Hertel classification is based upon Codman's traditional four-part concept (Figure 1).²¹ It provides a precise description of the fracture pattern by means of five basic fracture planes. These fracture planes lie between the greater tuberosity and the humeral head, the greater tuberosity and the shaft, the lesser tuberosity and the head, the lesser tuberosity and the shaft, and the lesser tuberosity and the greater tuberosity. There are six possible fractures dividing the humerus into two fragments, five possible fractures dividing the humerus into three fragments, and a single fracture dividing the humerus into four fragments.^{18, 19} The comminuted (*i.e.*, 3- and 4-part and head-split fractures) are marked with red boxes.

Neer classification

The Neer classification is based on the existence of displacements of one or more of the major segments of the proximal humerus: the articular surface, the greater and the lesser tuberosity, and the shaft. Displacement is defined as an at least 1-cm distance and/or a 45° angle between fragments.^{22, 23}

Figure 1 Hertel classification



Combining the fracture planes between the head (red), the greater (blue) and lesser (yellow) tuberosity and the shaft (green) results in 12 possible fracture patterns. Eight fracture patterns were considered as comminuted and were included this study (7, 8, 9, 10, 11, 12, 13 and 14). (Reprinted with permission from Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. *J Shoulder Elbow Surg* 2004;13:427-33.)^{18, 19}

Classification

All images were classified independently by two senior shoulder expert trauma surgeons (DDH and NWLS) and by two senior radiologists with primary orthopedic trauma focus (GSRM and LFMB). All images were provided in random order, and the observers were given as much time as needed for accurate assessment. The observers were blinded to clinical information and treatment strategies of the patients, and were not allowed to discuss their

observations with other investigators. All observers were familiar with the Neer and Hertel classification. In order to ensure unambiguous application of both fracture classification systems, a clarification of both classification systems was provided with each questionnaire, along with a standard evaluation form.

The images were classified three times, and were randomly provided in different order each time. The first evaluation was used for determining the inter-observer agreement on X-rays, CT-scans, and 3D-reconstructions separately. In order to determine the intra-observer agreement, the images were re-evaluated again after at least two months..

Statistical analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 16.0 (SPSS, Chicago, Ill., USA). Kappa statistics were calculated using MedCalc version 12.4.0. Normality of continuous data was tested by inspecting frequency histograms (Q-Q plots), and homogeneity of variances was tested using the Levene's test.

Data were analyzed using kappa statistics, as described by Cohen.²⁴ The kappa coefficient represents the agreement between two sets of observations compared with the likelihood of agreement based on chance alone. The kappa coefficient ranges from 1 (perfect agreement) to <0 (systematic disagreement, or no more agreement than would be expected by chance alone). The kappa values for inter-observer agreement were calculated for each possible pair of observers in the first round before calculating the mean kappa value. The kappa values for intra-observer agreement were calculated for each of the four individual observers before calculation the mean kappa value.²⁵ Interpretation of the values was carried out according to the guidelines of Landis and Koch which suggest that values <0 represent poor reliability; 0.00-0.20 slight agreement; 0.21-0.40 fair agreement; 0.41-0.60, moderate

agreement; 0.61-0.80, substantial agreement; and 0.81-1.00 almost perfect agreement.²⁶

Kappa values are reported with a 95% confidence interval.

The inter- and intra-observer kappa-values of both classifications and for comparing the radiographic modalities (*i.e.*, X-ray versus CT-scan, X-ray versus 3D-reconstruction, and CT-scan versus 3D-reconstruction) were compared using the Student's t-test. The Levene's test was used for assessing equality of variance. The corresponding p-value of the Student's t-test was used accordingly. A P-value of < 0.05 was considered statistically significant.

Ethical approval

The study was exempted by the local Medical Research Ethics Committee Erasmus MC (No. MEC-2011-151). For this type of study formal consent is not required.

RESULTS

Inter-observer agreement

An overview of the inter-observer agreement between the Hertel and Neer classifications for comminuted proximal humeral fractures on plain radiographs, CT-scans and CT-scans with 3D-reconstructions is shown in Table 1. The inter-observer agreement on plain radiographs was fair for both the Hertel classification ($\kappa=0.39$; 95% CI 0.23-0.62) and the Neer classification ($\kappa=0.29$; 95% CI 0.09-0.42). The inter-observer agreement on CT-scans was substantial for the Hertel classification ($\kappa=0.63$; 95% CI 0.56-0.72) and moderate for the Neer classification ($\kappa=0.51$; 95% CI 0.29-0.68). The inter-observer agreement on CT-scans with 3D-reconstructions was moderate for both the Hertel classification ($\kappa=0.60$; 95% CI 0.52-0.72) and the Neer classification ($\kappa=0.51$; 95% CI 0.39-0.58). Despite the kappa being consistently approximately 0.1 point higher for the Hertel classification on X-ray, CT-scans, and CT-scans with 3D-reconstructions., no statistically significant differences were found between the Hertel and Neer classification for these three modalities.

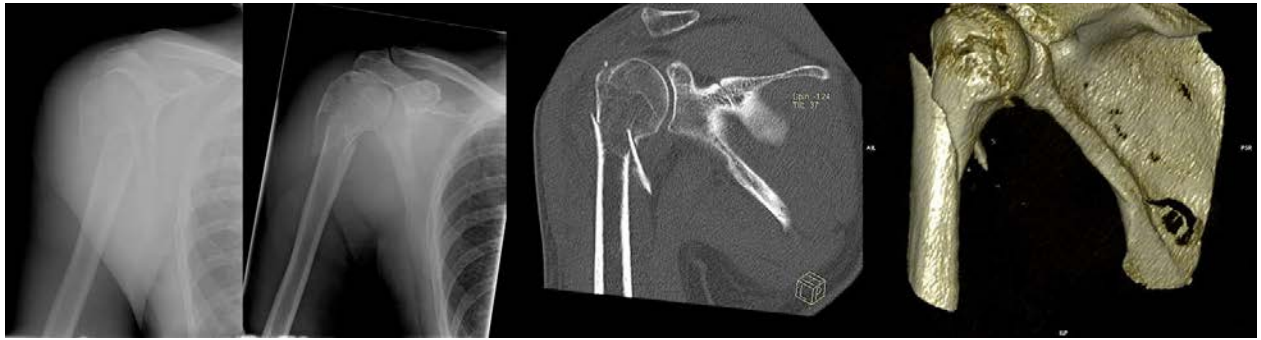
Inter-observer agreement was lowest for fractures between the head and the lesser tuberosity in radiographs. All four investigators consistently classified the radiographs as Hertel type 7 (n=4; see a typical example in Figure 2a) or Hertel type 12 (n=3; Figure 2b). However, in 18 other patients, disagreement was noted (*i.e.*, at least one investigator scored different from the others; Figure 2c) and were classified as Hertel type 7 or 12. For Neer classification only one fracture was classified unanimously (Figure 2d). As opposed to the Hertel classification, no consistent disagreement was identified.

Table 1. Inter-observer agreement of the Hertel and Neer classification on X-rays, CT-scans and 3D-reconstructions

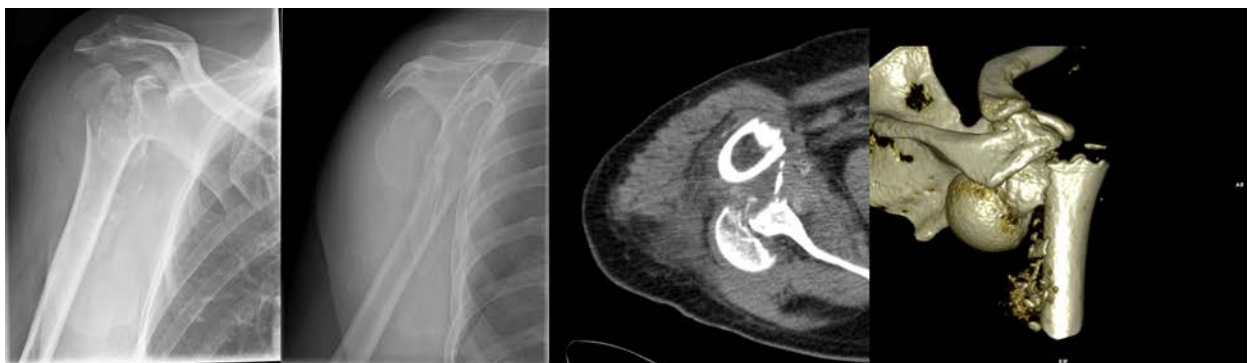
Observer	X-ray		CT-scan		3D-reconstruction	
	Hertel	Neer	Hertel	Neer	Hertel	Neer
1 + 2	0.23 (0.00- 0.46)	0.31 (0.07- 0.56)	0.59 (0.39- 0.78)	0.68 (0.54- 0.83)	0.53 (0.32- 0.75)	0.58 (0.38- 0.77)
1 + 3	0.34 (0.11- 0.56)	0.39 (0.17- 0.61)	0.60 (0.40- 0.79)	0.49 (0.31- 0.68)	0.52 (0.32- 0.72)	0.53 (0.31- 0.75)
1 + 4	0.31 (0.07- 0.54)	0.38 (0.12- 0.65)	0.64 (0.45- 0.83)	0.57 (0.32- 0.82)	0.71 (0.55- 0.86)	0.58 (0.35- 0.80)
2 + 3	0.62 (0.46- 0.78)	0.09 (- 0.13- 0.31)	0.56 (0.35- 0.77)	0.29 (0.17- 0.40)	0.57 (0.40- 0.74)	0.52 (0.28- 0.75)
2 + 4	0.37 (0.15- 0.59)	0.15 (- 0.12- 0.43)	0.66 (0.47- 0.85)	0.52 (0.33- 0.72)	0.57 (0.37- 0.77)	0.45 (0.22- 0.68)
3 + 4	0.46 (0.23- 0.70)	0.42 (0.20- 0.64)	0.72 (0.57- 0.88)	0.49 (0.28- 0.69)	0.71 (0.54- 0.89)	0.39 (0.16- 0.63)
Total	Fair 0.39 (0.23- 0.62)	Fair 0.29 (0.09- 0.42)	Substantial 0.63 (0.56- 0.72)	Moderate 0.51 (0.29- 0.68)	Moderate 0.60 (0.52- 0.72)	Moderate 0.51 (0.39- 0.58)
	P = 0.249		P = 0.067		P = 0.065	

Kappa values, with the 95% confidence interval between brackets, are shown. For the total scores the strength of agreement according to the guidelines of Landis and Koch is also shown.²⁶

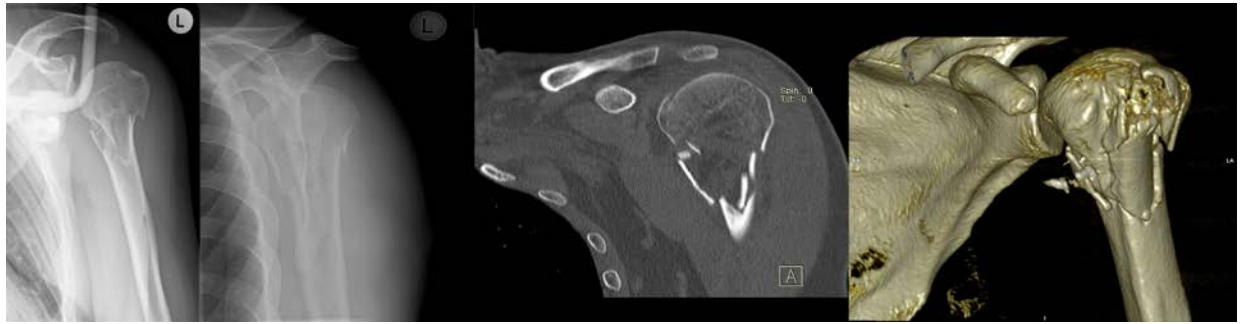
Fig. 2 Radiographs of proximal humeral fractures



- a. Fracture pattern classified by all observers as a Hertel type 7. This fracture was classified as a Neer two-part surgical neck fracture and a two-part greater tuberosity fracture by the observers.



- b. Fracture pattern classified by all observers as a Hertel type 12. This fracture was classified as Neer three-part greater tuberosity fracture, three-part anterior fracture dislocation, four-part anterior fracture dislocation and four-part posterior fracture dislocation by the observers



- c. Fracture pattern classified as a Hertel type 7 by two observers and Hertel type 12 by the other two observers. This fracture was classified as Neer two-part surgical neck fracture by three and as a Neer two-part greater tuberosity fracture by one observer.



- d. Fracture pattern classified by all observers as a Neer three-part anterior fracture dislocation. This fracture was classified as Hertel type 7 by two and Hertel type 12 by the other two observers.

Intra-observer agreement

An overview of the intra-observer agreement comparing the Hertel with the Neer classification is shown in Table 2. On plain radiographs, the intra-observer agreement was fair for the Hertel classification ($\kappa=0.38$; 95% CI 0.27-0.59) as well as for the Neer classification ($\kappa=0.40$; 95% CI 0.15-0.52). On CT-scans, it was moderate for the Hertel classification ($\kappa=0.50$; 95% CI 0.38-0.66) as well as the Neer classification ($\kappa=0.42$; 95% CI 0.35-0.52). 3D-reconstructions showed the highest agreement. It was moderate for the Hertel classification ($\kappa=0.55$; 95% CI 0.45-0.64) and even substantial for the Neer classification ($\kappa=0.63$; 95% CI 0.48-0.79). No statistically significant differences were found. No clear trend towards specific fracture lines causing disagreement was found for either classification.

When comparing the agreement between different radiographic modalities, the agreement between X-rays and either CT-scans or 3D-reconstructions was fair for Hertel and poor for Neer (Table 3). Agreement was moderate when comparing CT-scans with 3D-reconstructions for both classification systems.

Table 2. Intra-observer agreement of the Hertel and Neer classification for X-rays, CT-scans and 3D-reconstructions

Observer	X-ray		CT-scan		3D-reconstruction	
	Hertel	Neer	Hertel	Neer	Hertel	Neer
1	0.59 (0.42- 0.76)	0.52 (0.27- 0.77)	0.38 (0.23- 0.53)	0.43 (0.29- 0.58)	0.55 (0.36-0.75)	0.55 (0.38-0.72)
2	0.27 (0.10- 0.43)	0.15 (0.03- 0.28)	0.47 (0.31- 0.62)	0.39 (0.24- 0.54)	0.49 (0.30-0.69)	0.74 (0.59-0.90)
3	0.41 (0.32- 0.59)	0.52 (0.37- 0.66)	0.66 (0.50- 0.82)	0.35 (0.20- 0.49)	0.63 (0.45-0.81)	0.55 (0.34-0.76)
4	0.27 (0.10- 0.45)	0.41 (0.52- 0.57)	0.52 (0.35- 0.69)	0.52 (0.37- 0.66)	0.52 (0.32-0.71)	0.69 (0.53-0.86)
Total	Fair 0.38 (0.27- 0.59)	Fair 0.40 (0.15- 0.52)	Moderate 0.50 (0.38- 0.66)	Moderate 0.42 (0.35- 0.52)	Moderate 0.55 (0.45-0.64)	Substantial 0.63 (0.48-0.79)
	P = 0.902		P = 0.288		P = 0.188	

Kappa values, with the 95% confidence interval between brackets, are shown. For the total scores the strength of agreement according to the guidelines of Landis and Koch is also shown. [ENREF 1](#)²⁶

Table 3. Intra-observer agreement of the Hertel and Neer classification on X-rays vs. CT-scans, X-rays vs. 3D-reconstructions and CT-scans vs. 3D-reconstructions

Observer	X-ray vs. CT-scan		X-ray vs. 3D-reconstruction		CT-scan vs. 3D-reconstruction	
	Hertel	Neer	Hertel	Neer	Hertel	Neer
1	0.04 (-0.11-0.19)	0.17 (0.5-0.30)	0.01 (-0.14-0.17)	0.13 (0.02-0.24)	0.35 (0.18-0.51)	0.34 (0.20-0.49)
2	0.15 (-0.03-0.34)	0.01 (-0.10-0.12)	0.11 (-0.07-0.29)	0.10 (-0.01-0.21)	0.39 (0.23-0.56)	0.30 (0.17-0.44)
3	0.28 (0.11-0.45)	0.17 (0.03-0.30)	0.25 (0.08-0.42)	0.11 (0.00-0.22)	0.48 (0.31-0.66)	0.33 (0.19-0.48)
4	0.43 (0.28-0.57)	0.37 (0.22-0.52)	0.45 (0.29-0.61)	0.35 (0.20-0.49)	0.59 (0.44-0.75)	0.64 (0.51-0.78)
Total	Fair 0.23 (-0.04-0.49)	Poor 0.18 (-0.05-0.41)	Fair 0.21 (-0.01-0.51)	Poor 0.17 (-0.02-0.36)	Moderate 0.46 (0.28-0.63)	Moderate 0.41 (0.15-0.66)
	P = 0.692		P = 0.756		P = 0.628	

Kappa values, with the 95% confidence interval between brackets, are shown. For the total scores the strength of agreement according to the guidelines of Landis and Koch is also shown.²⁶

DISCUSSION

The results of this study showed that for classification of comminuted proximal humeral fractures both the Neer and the Hertel classification had a fair to substantial inter- and intra-observer agreement. There was no statistically significant difference between the inter-observer agreement for both classification systems, nor when comparing the different radiographic modalities. Overall, the Hertel classification showed a trend towards being a more reliable classification system. The Hertel classification showed a 35, 24 and 18% higher mean kappa value for inter-observer agreement than the Neer classification when applied to plain radiographs, CT-scans, and CT-scans with 3D-reconstructions, respectively. In previous studies, both the inter-observer agreement (kappa 0.27-0.64) as well as the intra-observer agreement (kappa 0.19-0.66) for the Neer classification on plain radiographs were generally higher than the agreement observed in the current study ($\kappa=0.29$ and $\kappa=0.40$, respectively).¹¹,

²⁷ This difference could be explained by the fact that we selected only patients with comminuted fractures. Classification of these types of fractures is known to have poorer inter- and intra-observer agreement.¹⁴ One study used 3D-printed models of proximal humeral fractures instead of radiographic images. They demonstrated a higher inter-observer agreement for the Hertel classification compared with the Neer and AO classification ($\kappa=0.44$ versus $\kappa=0.33$ and $\kappa=0.11$, respectively).¹⁴ which is in line with the present study results.

The inter-observer reliability for both the Hertel and the Neer classification was higher when classified on CT-scans (with or without 3D-reconstructions) than when classified on X-rays. The 3D-volume rendering, however, did not improve the inter-observer agreement of the Neer classification. Although this may be due to the fact that the reviewers were more used to assessing fracture patterns on plain CT-scans, it is also in agreement with previous data.^{14, 28} This study showed that the same holds true for the Hertel classification.

Inter-observer agreement of the Hertel classification was substantial ($\kappa=0.63$) when applied to CT-scans alone and fair ($\kappa=0.60$) when applied to CT-scans with 3D-reconstructions. The intra-observer reliability for the Neer and Hertel classifications increased from fair on X-ray to moderate on CT. Reliability using 3D-reconstructions improved even further for the Neer classification, but not for the Hertel classification. All observers judged the Hertel classification as the simpler to use system. For the Hertel classification, the observers had difficulties discriminating type 7 from 12, implying that the fracture line separating these types requires specific attention. For the Neer classification, no specific disagreement was found. Most difficulties for the Neer classification were directly related to the measurements required to be able to use this classification appropriately. Especially the reference points for the degrees of dislocation and the measurement of the degrees of angulation proved difficult. This suggests that the Hertel classification is a more straightforward classification, although this was not supported by a significantly improved agreement.

This study had some limitations. The inter- and intra-observer agreement for the two classification systems was not studied when applied to a combination of plain radiographs and CT-scans in the same session. Although this would more closely reflect common practice, most previous studies used the same method. Moreover by this method it was possible to assess both classification systems for the different imaging modalities separately. Nevertheless, it could be an interesting topic for further research. Another limitation is the selection of the radiographic images by a single person. This person however was not an observer. All radiographic images were blinded and randomized for each observer during all of the evaluations. This minimized the chance that images would be memorized and made exchange of data between observers impossible. Also, in order to accurately reflect daily routine, the quality of the radiographic images was not used as an exclusion criterion. The radiographs used by the treating surgeons were considered as of good enough quality, since

the treatment strategies were based on them. So no additional quality aspects were added to the inclusion criteria. Although the strength of this study is the number of patients enrolled, the number of observers was relatively low. This may have contributed to not finding statistically significant results when comparing the Hertel and Neer classifications. As a final limitation, both classification systems share the inability to designate risk factors for a disrupted perfusion of the humeral head; an important determinant in the choice of treatment in comminuted fractures of the proximal humerus. These factors include the size of the calcar segment, the part of the metaphysis that remains attached to the head (metaphyseal extension), of less than 8 mm and disruption of the medial hinge of more than 2 mm, which is the pivot point of the head at the level of the posteromedial fracture line. An intra-operative perfusion study has proven that Hertel fracture types 2, 9, 10, 11 and 12 are prone to develop avascular necrosis.¹⁸

In conclusion, the results of the current study showed a moderate inter-observer agreement for both the Hertel and the Neer classifications for radiographs. When applied to CT-scans, the Hertel classification showed a trend towards a higher inter-observer agreement than the Neer classification, *i.e.*, substantial versus moderate, respectively, but this was not a significant difference. Although inter-observer agreement was highest for Hertel classification on CT-scans, Neer classification had the highest intra-observer agreement on 3D-reconstructions. Data of this study do not confirm superiority of either classification system for the classification of comminuted proximal humeral fractures.

REFERENCES

1. Kannus P, Palvanen M, Niemi S, Parkkari J, Jarvinen M, Vuori I. Increasing number and incidence of osteoporotic fractures of the proximal humerus in elderly people. *Brit Med J*. 1996 Oct 26;313(7064):1051-2.
2. Palvanen M, Kannus P, Niemi S, Parkkari J. Update in the epidemiology of proximal humeral fractures. *Clin Orthop Relat R*. 2006 Jan;442(442):87-92.
3. Den Hartog D, De Haan J, Schep NWL, Tuinebreijer WE. Primary shoulder arthroplasty versus conservative treatment for comminuted proximal humeral fractures: a systematic literature review. *Open Orthop J*. 2010;4:87-92.
4. Lanting B, MacDermid J, Drosdowech D, Faber KJ. Proximal humeral fractures: A systematic review of treatment modalities. *J Shoulder Elb Surg*. 2008 Jan-Feb;17(1):42-54.
5. Chu SP, Kelsey JL, Keegan THM, Sternfeld B, Prill M, Quesenberry CP, et al. Risk factors for proximal humerus fracture. *Am J Epidemiol*. 2004 Aug 15;160(4):360-7.
6. Bell JE, Leung BC, Spratt KF, Koval KJ, Weinstein JD, Goodman DC, et al. Trends and Variation in Incidence, Surgical Treatment, and Repeat Surgery of Proximal Humeral Fractures in the Elderly. *J Bone Joint Surg Am*. 2011 Jan 19;93A(2):121-31.
7. Guy P, Slobogean GP, McCormack RG. Treatment Preferences for Displaced Three- and Four-Part Proximal Humerus Fractures. *J Orthop Trauma*. 2010 Apr;24(4):250-4.
8. Court-Brown CM, Caesar B. Epidemiology of adult fractures: A review. *Injury*. 2006 Aug;37(8):691-7.
9. Petit CJ, Millett PJ, Endres NK, Diller D, Harris MB, Warner JJP. Management of proximal humeral fractures: Surgeons don't agree. *J Shoulder Elb Surg*. 2010 Apr;19(3):446-51.

10. Marsh JL, Slongo TF, Agel J, Broderick JS, Creevey W, DeCoster TA, et al. Fracture and dislocation classification compendium-2007 - Orthopaedic Trauma Association classification, database and outcomes committee. *J Orthop Trauma*. 2007 Nov-Dec;21(10):S1-S133.
11. Bahrs C, Schmal H, Lingenfelter E, Rolauffs B, Weise K, Dietz K, et al. Inter- and intraobserver reliability of the MTM-classification for proximal humeral fractures: A prospective study. *Bmc Musculoskel Dis*. 2008 Feb 17;9:21.
12. Brorson S, Bagger J, Sylvest A, Hrobjartsson A. Diagnosing displaced four-part fractures of the proximal humerus: a review of observer studies. *Int Orthop*. 2009 Apr;33(2):323-7.
13. Mahadeva D, Dias RG, Deshpande SV, Datta A, Dhillon SS, Simons AW. The reliability and reproducibility of the Neer classification system - Digital radiography (PACS) improves agreement. *Injury*. 2011 Apr;42(4):339-42.
14. Majed A, Macleod I, Bull AMJ, Zyto K, Resch H, Hertel R, et al. Proximal humeral fracture classification systems revisited. *J Shoulder Elb Surg*. 2011 Oct;20(7):1125-32.
15. Neer CS, 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg Am*. 1970 Sep;52(6):1077-89.
16. Bernstein J, Monaghan BA, Silber JS, DeLong WG. Taxonomy and treatment - a classification of fracture classifications. *J Bone Joint Surg Br*. 1997 Sep;79B(5):706-7.
17. Schepers T, van Lieshout EMM, Ginai AZ, Mulder PGH, Heetveld MJ, Patka P. Calcaneal Fracture Classification: A Comparative Study. *J Foot Ankle Surg*. 2009 Mar-Apr;48(2):156-62.
18. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. *J Shoulder Elb Surg*. 2004 Jul-Aug;13(4):427-33.

19. Hertel R. Fractures of the proximal humerus in osteoporotic bone. *Osteoporosis Int.* 2005 Mar;16:S65-S72.
20. Rosset A, Spadola L, Ratib O. OsiriX: An open-source software for navigating in multidimensional DICOM images. *J Digit Imaging.* 2004 Sep;17(3):205-16.
21. Codman EA. Fractures in relation to the subacromial bursa. Codman EA, editor. Malabar, FL: Krieger Publishing; 1934.
22. Neer CS. Four-segment classification of proximal humeral fractures: Purpose and reliable use. *J Shoulder Elb Surg.* 2002 Jul-Aug;11(4):389-400.
23. Robinson BC, Athwal GS, Sanchez-Sotelo J, Rispoli DM. Classification and Imaging of Proximal Humerus Fractures. *Orthop Clin N Am.* 2008 Oct;39(4):393-403.
24. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull.* 1968 Oct;70(4):213-20.
25. Svanholm H, Starklint H, Gundersen HJG, Fabricius J, Barlebo H, Olsen S. Reproducibility of Histomorphologic Diagnoses with Special Reference to the Kappa-Statistic. *Apmis.* 1989 Aug;97(8):689-98.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977 Mar;33(1):159-74.
27. Brunner A, Honigmann P, Treumann T, Babst R. The impact of stereo-visualisation of three-dimensional CT datasets on the inter- and intraobserver reliability of the AO/OTA and Neer classifications in the assessment of fractures of the proximal humerus. *J Bone Joint Surg Br.* 2009 Jun;91B(6):766-71.
28. Sjoden GO, Movin T, Aspelin P, Guntner P, Shalabi A. 3D-radiographic analysis does not improve the Neer and AO classifications of proximal humeral fractures. *Acta Orthop Scand.* 1999 Aug;70(4):325-8.