Poor reproducibility of classification of proximal humeral fractures

Additional CT of minor value

Göran O J Sjödén[1], Tomas Movin[1], Peter Güntner[2], Peter Aspelin[2], Leif Ahrengart[1], Hans Ersmark[1] and Anders Sperber[1]

Fractures of the proximal humerus can be described using the Neer and AO fracture classifications. To assess the reproducibility and reliability of these classifications, we investigated 26 proximal humeral fractures with both plain radiographs and CT. 5 specialists in orthopedic surgery and 5 specialists in radiology independently classified all radiographs on 2 occasions. There was a moderate agreement between the observers when using the Neer classification, but only a fair agreement with the AO classification. The Neer system had a kappa value of 0.42 and the AO had a value of 0.31 in the first assessment. In the second assessment the kappa values were 0.45 and 0.30, respectively. Intraobserver reproducibility was slight to almost perfect agreement with Neer (kappa range 0.20–0.85) and slight to moderate agreement with AO (kappa range 0.16–0.60). The observers most familiar with shoulder fracture radiographs and shoulder fracture treatment were more consistent in their classifications.

We conclude that even with CT, the fracture classifications of Neer and AO have a low consistency. Neither classification system is reproducible enough to allow comparisons of different studies.

Departments of [1]Orthopedics and [2]Radiology, Karolinska Institute, Huddinge University Hospital, S-141 86 Huddinge, Sweden. Tel +46 8-5858 0000. Fax -711 4292
Submitted 96-09-06. Accepted 97-03-06

Most 2-part fractures of the proximal humerus are not operated on (Neer 1970a). The outcome of closed treatment of 3-part fractures varies (Leyshon 1984), but the outcome is generally poor for 4-part fractures (Stableforth 1984, Neer 1970b), although this has been disputed by some (Kristiansen and Christensen 1987, Rasmussen et al. 1992, Zyto et al. 1995). Avascular necrosis of the humeral head after 4-part fractures could be a contributing factor to the inferior results (Lee and Hansen 1981). There is a specific type of displaced 4-part fracture with a valgus impaction of the humeral head, where the incidence of avascular necrosis is reduced and the outcome is more favorable (Jakob et al. 1991).

Therefore, fractures should be evaluated in detail to plan the treatment and evaluate the outcome. There are two classification systems: the Neer (1970a) and the AO (Müller et al. 1990). It has been shown for both Neer and AO classifications that even shoulder surgeons find it difficult to categorize fractures with plain radiographs (Kristiansen et al. 1988, Sieberock and Gerber 1993, Sidor et al. 1993, Brien et al. 1995).

We evaluated whether CT could improve the reproducibility of classification of proximal humeral fractures.

Patients and methods

26 consecutive patients admitted to our hospital for proximal humeral fractures were examined by CT in addition to conventional radiography. The radiographs comprised one 45° angulated anteroposterior (AP) view and one scapular (Y) view. The CT examinations were performed with a Philips helical SR 7000 scan or a Siemens DRH scanner. Multiple CT scans were printed on plain films.

5 specialists in orthopedics and 5 specialists in radiology independently reviewed all cases. Each examiner was given a written description and drawings of the Neer (as modified by Sidor et al. 1993) and AO (Müller et al. 1990) classifications. Each examiner had access to a ruler and a goniometer. The conventional radiographs were analyzed together with the CT scans in each case.

With use of the Neer system, each fracture was classified as type 1 through 16. For statistical evaluation, the 16 categories in the Neer classification were reduced to 6: one-part fractures (type 1), two-part fractures (types 2–5), three-part fractures (types 8 and 9), four-part fractures (type 12), fracture-dislocations (types 6, 7, 10, 11, 13, and 14) and articular surface

Table 1. Intraobserver reproducibility after two reviews of 26 fractures, classified according to Neer and AO 2 months apart

	Kappa value	
	Neer	AO
Orthopedic surgeon		
Reconstructive surgeon	0.20	0.16
Traumatologist	0.59	0.33
Shoulder specialist	0.65	0.39
Shoulder specialist	0.85	0.42
Shoulder specialist	0.71	0.54
Radiologist		
General	0.23	0.37
General	0.32	0.20
General	0.46	0.36
Skeletal	0.45	0.28
Skeletal	0.60	0.60

fractures (types 15 and 16).

With the AO system, each fracture was classified into 1 of 3 types (A, B and C) and each type was divided into 3 groups (1–3). Thus a fracture could be classified into one of 9 patterns (A1–C3). After 2 months, the same radiographs were presented in a different order and were classified again by the same observers.

Statistics

We used computer-calculated kappa statistics (PC-Agree 1.4 data program, Svanholm et al. 1989) to analyze interobserver reliability and intraobserver reproducibility. This analysis involves adjustment for agreement among observers with that occurring by chance. If all observers agree in all cases, the kappa value is +1. A kappa value of 0 indicates that the agreement is equal to that occurring by chance. Kappa values of less than 0.00 indicate poor agreement; 0.00–0.20 = slight agreement; 0.21–0.40 fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = substantial agreement; 0.81–1.00 = almost perfect agreement (Landis and Koch 1977). Since kappa also depends on the prevalence of the various categories used (Kraemer 1979), care should be used when comparing different kappa values.

Results

The mean kappa coefficient for pairwise agreement (interobserver reliability) with the Neer system was 0.42 in the first viewing and 0.43 in the second viewing. With the AO system, the kappa value was 0.31 in the first viewing and 0.26 in the second.

Intraobserver reproducibility ranged from 0.20–0.85 with Neer and 0.16–0.60 with AO. The individual kappa values were in the same range in the orthopedic and the radiologist groups (Tables 1 and 2).

Discussion

Like Bernstein et al. (1996), we found that CT, together with plain radiographs, did not make fracture classification more consistent. There was a moderate agreement between the observers when using the Neer classification, but only a fair agreement with the AO classification.

Others have found that classification based on plain radiographs is extremely difficult, even for shoulder surgeons (Siebenrock and Gerber 1993, Sidor et al. 1993). It has been questioned if the classification systems are useful at all (Burstein 1993). Recently, Neer (1997) stated that even the most experienced surgeon occasionally is in doubt and has to make the final classification at surgery. It is therefore likely that the results of treatment of various categories of fractures reported in the literature may be inaccurate because of the difficulties of making a consistent classification (Cowell 1994).

Our hypothesis was that a better imaging technique would increase the reproducibility of fracture classification. CT is useful in patients where the degree of displacement of the greater and lesser tuberosity or rotation of fragments is difficult to determine on conventional radiographs (Kilcoyne et al. 1990, Billet et al. 1992). As the delineation of the fracture lines is better visualized by CT, the risk of avascular necrosis of the humeral head can better be estimated (Jurik and Albrechsen 1994). However, the use of CT in our study did not improve the classification, when compared with previous studies using plain radiographs (Siebenrock and Gerber 1993, Sidor et al. 1993). Bernstein et al. (1996) found that CT, added to plain radiographs, increased intraobserver reliability slightly, but did not improve interobserver reproducibility.

Classification systems for fractures of the proximal femur (Frandsen et al. 1988, Andersen et al. 1990), distal radius (Kreder et al. 1996) and the ankle (Nielsen et al. 1990, Thomsen et al. 1991) have also shown a low degree of interobserver reliability. This can be due to a lack of diagnostic ability, but also to lack of ability to translate the image information into the classification patterns. The interpretation of visual information is difficult and varies between observers. Knowledge gained from training and experience in radiology improves the ability to interpret radiographs (Kinnunen et al. 1988, Bass and Chiles 1990). In our study, the observers most familiar with radiographs and treatment of shoulder fractures were more

Table 2. Classification of 26 proximal humeral fractures by 5 orthopedic surgeons and 5 radiologists

		Neer classification										AO classification									
A	B	C	D	E	F	G	H	I	J	K	L	C	D	E	F	G	H	I	J	K	L
1	1	8	8	8	8	8	16	8	8	12	12	B2	B1	B2	B2	B1	C2	B1	C2	C2	B1
1	2	8	8	8	8	9	16	8	8	11	8	C2	B1	B2	B1	B2	C3	B1	C2	C3	C1
2	1	8	14	8	12	13	14	14	14	12	14	C2	C3	B2	C2	C3	C3	B3	C3	C2	C2
2	2	8	12	12	14	13	11	14	14	14	16	C2	C2	C2	C3	C3	C3	C3	C3	C3	C3
3	1	8	8	8	8	8	8	8	8	9	12	B1	B1	B2	B2	B2	B1	B1	B1	C2	B2
3	2	8	8	8	8	11	8	8	8	14	10	C2	C2	B2	B1	B3	C2	B1	B2	C2	C2
4	1	12	12	12	12	9	14	12	12	12	14	C2	C2	C2	C2	B2	C3	C2	C2	C2	C3
4	2	12	12	12	14	13	12	14	12	13	14	B2	C2	B2	C3	C3	C2	C3	C2	C3	C3
5	1	8	12	12	8	4	12	8	8	8	12	C2	C2	B1	C2	A2	C2	B1	B2	C2	C1
5	2	8	12	12	12	10	12	12	12	12	11	C2	C2	B1	C2	B1	C3	C2	B2	C2	C3
6	1	12	12	12	12	4	12	12	12	12	13	C2	C2	B1	C2	A1	C2	B1	B1	C2	B1
6	2	12	12	12	12	15	12	12	12	12	12	C2	C2	B1	C2	C3	C2	B1	B2	C2	C2
7	1	16	16	3	8	1	15	3	3	9	4	C2	C1	A2	B1	A2	C2	A2	B1	C2	A1
7	2	16	15	16	3	2	16	2	16	11	9	C2	C3	C2	A2	A2	C1	C1	C3	C2	A1
8	1	10	13	10	13	13	13	13	10	13	10	C3	C3	C3	C3	C3	C3	C3	B3	C3	B1
8	2	10	13	13	13	13	10	14	10	14	6	C3	C3	B3	C3	B3	C3	C3	C3	C3	B3
9	1	14	14	13	14	14	14	14	14	14	14	C3	C3	C3	B3	C3	C3	C3	C3	C3	C3
9	2	11	14	14	14	15	14	14	14	11	14	B3	C3	B3	C3	B2	C3	C3	C3	C3	B3
10	1	12	12	15	12	15	14	11	12	13	10	C2	C2	C2	C2	C3	C3	B3	C2	C3	B2
10	2	12	16	12	16	10	13	14	12	13	10	C2	C3	B2	C3	C3	C3	C3	B3	C3	B2
11	1	8	8	8	8	3	12	8	8	12	8	B2	B1	B1	B1	C1	C2	B1	B2	C2	B2
11	2	8	8	8	4	8	8	8	12	12	8	C1	B1	B1	B1	B2	C2	B1	C2	C2	B2
12	1	12	12	12	12	3	12	8	12	12	12	C2	C2	B2	C2	C1	C2	C1	B2	C2	B2
12	2	12	12	12	12	8	12	8	12	12	12	C2	C2	B2	C2	C3	C2	B2	C2	B3	B1
13	1	15	6	6	6	6	15	10	6	10	6	A1	A1	B3	A1	C3	C3	B3	B3	B3	A1
13	2	15	6	6	6	4	16	13	6	10	6	C3	B3	B3	B3	B2	C3	B3	C3	A2	B3
14	1	4	4	1	4	4	1	4	1	4	4	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1
14	2	4	4	4	4	4	4	3	1	4	4	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1
15	1	4	4	1	4	4	1	4	4	4	4	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1
15	2	4	1	4	4	2	1	4	1	4	4	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1
16	1	9	12	12	12	12	12	12	15	12	13	C2	C2	C2	B2	B3	C2	C1	C2	C2	C1
16	2	3	9	12	12	12	3	12	8	12	12	A3	B1	B1	C2	C2	C2	C1	C2	C2	B1
17	1	16	16	15	13	8	14	14	11	12	10	C3	C3	C3	C3	A2	C3	C3	C3	C2	C1
17	2	16	11	9	11	13	16	11	16	7	16	C2	C3	C2	B3	C3	C3	C3	C3	C2	C3
18	1	16	16	16	2	4	16	2	11	9	4	C3	C3	C3	C2	C1	C1	C1	C2	C2	A1
18	2	16	16	16	16	4	15	2	16	2	7	C3	C2	B3	C1	A1	C3	C1	C3	C1	C3
19	1	12	13	13	13	13	13	13	13	12	13	C3	C3	C3	C3	C3	C3	C3	C3	C2	C3
19	2	10	13	13	4	14	13	10	13	11	10	C3	C3	C3	C3	C3	C3	B3	B3	C3	C3
20	1	16	7	11	11	12	14	11	16	14	10	C3	A1	C3	B3	A3	C3	B3	C3	C3	C3
20	2	16	7	11	11	14	15	11	11	14	16	C3	B3	B3	B3	B2	C3	B3	C3	C3	C3
21	1	8	12	12	12	10	12	8	12	12	12	C2	C2	B2	C2	C3	C2	B1	C2	C2	B2
21	2	8	12	12	12	12	12	8	8	13	12	B2	C2	B2	C2	C3	C2	B1	C2	C3	C3
22	1	10	13	10	10	14	13	8	13	13	13	C3	C3	B3	B3	C3	C3	B1	C3	C3	C2
22	2	10	13	10	13	10	13	10	13	11	15	C3	C3	B3	C3	B2	C3	C3	B3	C3	B2
23	1	5	5	4	4	4	4	4	4	12	4	B2	A1	A1	A1	A1	A1	A1	A1	B2	A1
23	2	5	5	4	4	4	4	4	4	9	4	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1
24	1	15	16	8	15	10	15	8	15	10	10	C2	C2	B2	C2	C3	C2	B1	C2	C3	B1
24	2	15	16	8	15	8	16	8	10	11	14	C2	C2	C2	C2	C3	C2	B1	C3	C2	C3
25	1	11	11	14	14	10	14	11	16	14	14	C3	C3	C3	C3	C3	C3	B3	C3	C3	C3
25	2	11	14	14	11	14	11	11	14	12	14	C3	C3	B3	B3	B2	C3	B3	C3	C3	C2
26	1	8	8	8	8	10	12	8	8	12	14	C2	C2	B2	B2	C3	B3	B1	B1	B2	C3
26	2	8	8	12	12	9	8	11	8	14	8	B2	B2	B2	C2	B2	B3	B3	C2	C3	B2

A Patient no.
B Examination no.
C–E Shoulder specialists
F Trauma surgeon
G Reconstruvtive surgeon
H–I Skeletal radiologists
J–L General radiologists

consistent in their classifications, while the observer with less experience in this area had a greater tendency to change the classifications from one time to time.

Failure to detect abnormal findings is known as "underreading". It may happen when another finding in the same radiograph captures the attention. Once the clinical question is answered the examination is terminated before other lesions are detected. This "satisfaction of search" effect could occur even if observers continue to inspect films after the discovery of the first lesion (Berbaum et al. 1990).

Another possibility is related to perceptual set effects. Perceptual set is defined as a readiness to inter-

pret images in particular ways. The conclusion that certain image features are indicative of one diagnostic category might make it more difficult to detect features of another diagnostic category (Haber 1966). Lesions of the same type as the lesion already found may then become more detectable, whereas lesions of a different type may become less detectable.

It seems obvious that problems in translating information in the radiographic image into an arbitrary classification system, make the evaluation of results of shoulder fracture treatment from several centers difficult to compare.

References

Andersen E, Jorgensen L G, Hededam L T. Evans' classification of trochanteric fractures: an assessment of the interobserver and intraobserver reliability. Injury 1990; 21: 377-8.

Bass J C, Chiles C. Visual skill. Correlation with detection of solitary pulmonary nodules. Invest Radiol 1990; 9: 994-8.

Berbaum K S, Franken E A, Dorfman D D, et al. Satisfaction of search in diagnostic radiology. Invest Radiol 1990; 25: 133-40.

Bernstein J, Adler L M, Blank J E, Dalsey R M, Williams G R, Iannotti J P. Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. J Bone Joint Surg (Am) 1996; 78: 1371-5.

Billet F P J, Schmitt W G H, Gay B. Computed tomography in traumatology with special regard to the advances of three-dimensional display. Arch Orthop Trauma Surg 1992; 111: 131-7.

Brien H, Noftall F, MacMaster S, Cummings T, Landells C, Rockwood P. Neer's classification system: a critical appraisal. J Trauma 1995; 38: 257-60.

Burstein A H. Editorial. Fracture classification systems: do they work and are they useful? J Bone Joint Surg (Am) 1993; 75: 1743-4.

Cowell H R. Editorial. Patient care and scientific freedom. J Bone Joint Surg (Am) 1994; 76: 640-1.

Frandsen P A, Andersen E, Madsen F, Skjødt T. Garden's classification of femoral neck fractures. An assessment of the interobserver variation. J Bone Joint Surg (Br) 1988; 70: 588-90.

Haber R N. Nature of the effect of set on perception. Psychol Rev 1966; 73: 335-51.

Jakob R J, Miniaci A, Anson P S, Jaberg H, Osterwalder A, Ganz R. Four-part valgus impacted fractures of the proximal humerus. J Bone Joint Surg (Br) 1991; 73: 295-8.

Jurik A G, Albrechsen J. The use of computed tomography with two- and three-dimensional reconstructions in the diagnosis of three- and four-part fractures of the proximal humerus. Clin Radiol 1994; 49: 800-4.

Kilcoyne R F, Shuman W P, Matsen F A, Morris M, Rockwood C A. The Neer classification of displaced proximal humeral fractures: spectrum of findings on plain radiographs and CT scans. Am J Roentgenol 1990; 154: 1029-33.

Kinnunen J, Gothlin JH, Totterman S. Effect of training and experience on radiologic diagnostic performance in midfacial trauma. Acta Radiol 1988; 28: 83-7.

Kraemer H C. Ramifications of a population model for k as a coefficient of reliability. Psychometrika 1979; 44: 461-72.

Kreder H J, Hanel D P, McKee M, Jupiter J, McGillivary G, Swiontkowski M F. Consistency of AO fracture classification for the distal radius. J Bone Joint Surg (Br) 1996; 78: 726-31.

Kristiansen B, Christensen S W. Proximal humeral fractures. Late results in relation to classification and treatment. Acta Orthop Scand 1987; 58: 124-7.

Kristiansen B, Andersen U L S, Olsen C A, Varmarken J E. The Neer classfication of fractures of the proximal humerus. Skeletal Radiol 1988; 17: 420-2.

Landis J R, Koch G G. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-74.

Lee C K, Hansen H R. Posttraumatic avascular necrosis of the humeral head in displaced proximal humeral fractures. J Trauma 1981; 21: 788-91.

Leyshon R L. Closed treatment of fractures of the proximal humerus. Acta Orthop Scand 1984; 55: 48-51.

Müller M E, Nazarian S, Koch P, Schatzker J. The comprehensive classification of fractures of long bones. Springer, New York 1990: 54-63.

Neer C S. Displaced proximal humeral fractures. Part I. Classification and evaluation. J Bone Joint Surg (Am) 1970a; 52: 1077-89.

Neer C S. Displaced proximal humeral fractures. Part II. Treatment of three-part and four-part displacement. J Bone Joint Surg (Am) 1970b; 52: 1090-103.

Neer C S. Four-segment classification of proximal humeral fractures. In: Shoulder surgery (ed. Copeland S). W B Saunders Company Ltd, London 1997: 321-31.

Nielsen J O, Dons-Jensen H, Sørensen H T. Lauge-Hansen classification of malleolar fractures. An assessment of the reproducibility in 118 cases. Acta Orthop Scand 1990; 61: 385-7.

Rasmussen S, Hvass I, Dalsgaard J, Christensen B S, Holstad E. Displaced proximal humeral fractures: results of conservative treatment. Injury 1992; 23: 41-3.

Sidor M L, Zuckerman J D, Lyon T, Koval K, Cuomo F, Schoenberg N. The Neer classification system for proximal humerus fractures. J Bone Joint Surg (Am) 1993; 75: 1745-50.

Siebenrock K A, Gerber C. The reproducibility of the classification of the fractures of the proximal end of the humerus. J Bone Joint Surg (Am)1993; 75: 1751-5.

Stableforth P G. Four-part fractures of the neck of the humerus. J Bone Joint Surg (Br) 1984; 66: 104-8.

Svanholm H, Starklint H, Gundersen H J G, Fabricius J, Barlebo H, Olsen S. Reproducibility of histomorphologic diagnoses, with special reference to kappa statistics. APMIS 1989; 97: 689-98.

Thomsen N O B, Overgaard S, Olsen L H, Hansen H, Nielsen S T. Observer variation in the radiographic classification of ankle fractures. J Bone Joint Surg (Br) 1991; 73: 676-8.

Zyto K, Kronberg M, Broström L Å. Shoulder function after displaced fractures of the proximal humerus. J Shoulder & Elbow Surg 1995; 4: 331-6.