3D-radiographic analysis does not improve the Neer and AO classifications of proximal humeral fractures

Göran O J Sjödén[1], Tomas Movin[1], Peter Aspelin[2], Peter Güntner[2] and Adel Shalabi[2]

The Neer and AO fracture classifications for fractures of the proximal humerus have shown poor reproducibility based on plain radiography. We wanted to investigate whether the addition of 3-dimensional (3D) reconstructions would increase the reproducibility of classification. 7 observers independently classified 24 fractures of the proximal humerus using both plain radiographs, CT and 3D and the classification was repeated 2 months later.

There was a moderate interobserver agreement when using the Neer classification, but only a fair agreement with the AO classification. The Neer system had a mean kappa value of 0.44 and the AO had a value of 0.32 for the first assessment. In the second assessment, the mean kappa values were 0.49 and 0.34, respectively. Intraobserver reproducibility was fair to substantial agreement for Neer (kappa range 0.27–0.73) and for AO (kappa range 0.29–0.74).

In conclusion, the addition of CT and 3D to plain radiographs did not improve the reproducibility of the classifications of Neer and AO of the proximal humerus.

Departments of [1]Orthopedics and [2]Radiology, Karolinska Institute, Huddinge University Hospital, SE-141 86 Huddinge, Sweden. Tel +46 8-5858 0000. Fax -711 4292. E-mail: goran.sjoden@karo.kl.se
Submitted 98-11-17. Accepted 99-04-18

A reproducible classification for fractures is a prerequisite for a systematic approach to treatment and evaluation of outcome. 2 classification systems are used for fractures of the proximal humerus, the Neer (1970) and the AO (Müller et al. 1990). It has previously been shown that classification remains difficult with both systems (Kristiansen et al. 1988, Brien et al. 1995). When only plain radiographs are available, there is low reproducibility of the classification of proximal humeral fractures (Sidor et al. 1993, Siebenrock and Gerber 1993). The addition of CT to plain radiographs does not enhance reproducibility and reliability (Bernstein et al. 1996, Sjödén et al. 1997).

The use of computer programs to produce 3-dimensional (3D) images from sets of axial computed tomograms is another tool to improve image interpretation. With 3D-reconstructions, spatial information is obtained in a form readily recognizable to the surgeon. 3D-reconstructions can also be of help in evaluating fractures involving the scapula and coracoid process, since overlapping lines are absent on conventional radiograms (Billet 1992). Even when using 3D-imaging, the description of fracture patterns is not easy (Sallay et al. 1997). We evaluated whether CT and 3D can improve the reproducibility and reliability of classification of proximal humeral fractures.

Patients and methods

24 patients admitted to our hospital for proximal humeral fractures were examined by CT, in addition to conventional radiography. The radiographs comprised one 15° angulated anteroposterior (AP) view and one scapular view. The CT examinations were performed with a Philips helical SR 7000 or a Siemens DRH scanner. Total scans were 30–40, performed with 3 mm slice thickness and a reconstruction interval of 2 mm. Multiple CT scans and multiplanar reconstructions, producing shaded surface 3D-images, were printed on plain films that were used for classification.

4 specialists in orthopedics and 3 specialists in radiology independently classified 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 and 3D images in every case.

With the Neer system, each fracture was classified as type 1 through 16. For statistical evaluation, the 16 categories were reduced to 6: one-part fractures (type 1), two-part fractures (type 2 through 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 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 to 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, 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 indicate poor agreement, 0.00–0.20 is slight agreement, 0.21–0.40 is fair agreement, 0.41–0.60 is moderate agreement, 0.61–0.80 is substantial agreement, and 0.81–1.00 is almost perfect agreement (Landis and Koch 1977).

Results

The mean kappa coefficient for pairwise agreement (interobserver reliability) for the Neer system was 0.44 for the first viewing and 0.49 for the second viewing. For the AO system, the kappa value was 0.32 for the first viewing and 0.34 for the second.

Table 1. Intraobserver reproducibility of 24 humeral fractures classified according to Neer and AO at 2 reviews 2 months apart

	Kappa value	
	Neer	AO
Orthopedist		
Traumatologist	0.57	0.74
Shoulder specialist	0.67	0.74
Shoulder specialist	0.73	0.46
Shoulder specialist	0.66	0.46
Radiologist		
General	0.56	0.72
General	0.27	0.29
General	0.48	0.28

Intraobserver reproducibility ranged from 0.27–0.73 for Neer and 0.29–0.74 for AO. The individual kappa values were in the same range among the orthopedists but somewhat lower among the radiologists (Tables 1 and 2).

Discussion

Our hypothesis that a better imaging technique would increase the reproducibility of fracture classification failed, since the use of 3D did not improve the classification, as compared with previous studies using CT and conventional radiography (Sjödén et al. 1997). There was a moderate agreement between the observers when using the Neer classification, but only a fair agreement with the AO classification. Since kappa also depends on the prevalence of the different categories used (Kraemer 1979), a comparison between different kappa values must be made cautiously. It can be inferred from this study that experience of treatment is of value when classifying fractures, since the most experienced shoulder surgeons showed a more consistent classification in comparison with the radiologists. However, others have found that a classification based on plain radiographs is also difficult for shoulder surgeons (Siebenrock and Gerber 1993, Sidor et al. 1993). When simplifying the AO or Neer classification by reducing the number of categories, the inter- and intraobserver variations did not improve (Siebenrock and Gerber 1993, Sidor et al. 1993). In our study as well as that of others (Siebenrock and Gerber 1993), agreement tended to be better for

Table 2. Classification of 24 proximal humeral fractures by 4 orthopedists and 3 radiologists

		Neer classification							AO classification						
Pat no.	Exam no.	Orth 1	Orth 2	Orth 3	Orth 4	Radio 1	Radio 2	Radio 3	Orth 1	Orth 2	Orth 3	Orth 4	Radio 1	Radio 2	Radio 3
1	1	4	3	4	4	4	5	5	c2	b2	b2	c2	c2	c3	c3
1	2	4	4	3	4	4	4	4	c2	b2	b2	c2	c2	b2	c2
2	1	4	2	3	3	2	4	2	c1	a2	b1	c2	b1	b1	a3
2	2	3	3	2	2	3	2	2	b3	b1	a2	a2	b1	a2	a2
3	1	4	3	2	3	2	3	2	c2	b1	b1	b1	b1	b1	b1
3	2	2	4	3	3	2	2	2	b1	b1	b1	b2	a2	b1	b1
4	1	4	3	3	4	2	5	3	c2	c2	b2	c2	c2	b2	b2
4	2	3	3	3	3	3	3	4	b2	c2	b1	b2	b1	b1	b2
5	1	5	5	5	5	5	5	5	c3	c3	b3	c3	c3	c3	c3
5	2	5	5	5	5	5	5	5	c3	b3	b3	c3	c3	c3	b3
6	1	2	2	2	2	2	4	4	a3	b1	a3	a3	a3	b3	b3
6	2	2	2	2	2	2	3	4	b2	a3	a3	a3	a3	b2	b2
7	1	1	1	1	2	1	2	2	a1	a1	a1	a1	a1	a1	a1
7	2	1	1	2	2	2	1	1	a1	a1	a1	a1	a1	a1	a1
8	1	5	5	5	5	6	5	5	c3	b3	b3	c3	c3	c3	c3
8	2	5	5	5	5	5	6	5	c3	c3	b3	b3	c3	c2	b3
9	1	4	3	4	4	2	4	4	c2	c1	b2	c2	c2	c2	b1
9	2	4	4	4	4	4	3	4	c2	b2	c2	c2	c2	c2	b2
10	1	2	2	2	3	2	2	2	b1	b1	a2	b1	b1	b1	a2
10	2	2	2	2	2	2	2	3	b1	a2	a2	a2	b1	c1	b2
11	1	2	2	2	2	2	2	2	a2	a2	a2	a2	a2	b2	a3
11	2	2	2	2	2	2	2	2	a2	a2	a2	a2	b1	a2	a1
12	1	1	1	1	3	2	1	–	a2	b1	b1	b1	b1	c1	–
12	2	2	2	3	4	2	1	2	a2	b1	b1	b1	b1	a2	a2
13	1	3	3	4	3	2	4	4	b2	c2	b2	b1	b1	c2	c2
13	2	3	3	3	3	3	3	4	c2	c2	b2	b1	c2	b2	b2
14	1	4	4	4	4	4	5	5	c2	c2	b2	b2	c2	c2	c3
14	2	4	4	4	4	4	4	5	c2	c2	b2	c2	c2	c2	c3
15	1	5	5	5	5	5	5	5	c3	c3	b3	b3	c3	c3	c3
15	2	5	5	5	5	5	6	6	c3	b3	b3	c3	c3	c3	c3
16	1	5	5	5	5	5	5	5	c3	c3	b3	b3	b3	b3	b3
16	2	5	5	5	5	5	5	5	c3	c3	b3	b3	c3	c3	b3
17	1	5	3	4	2	5	4	5	c3	b2	b2	a3	c3	c2	c3
17	2	5	3	4	5	5	6	5	c3	b2	b2	b3	c3	c3	c3
18	1	5	5	5	5	5	5	5	c3	b3	b3	b3	c3	c3	c3
18	2	5	5	5	5	5	6	5	c3	b3	b3	c3	c3	c3	c3
19	1	2	2	2	3	2	2	2	a1	b2	a1	b1	a1	a1	b1
19	2	2	2	5	3	2	2	1	a1	a1	a1	b1	a1	a1	a1
20	1	4	4	3	4	4	4	3	c2	c2	b2	b2	c2	c2	a3
20	2	4	4	6	3	6	4	4	c2	c2	c2	c2	c2	b2	b2
21	1	5	5	5	5	5	5	5	c3	c3	b3	b3	c3	c3	b3
21	2	5	5	5	5	5	5	5	c3	b3	b3	b3	c3	c3	c3
22	1	4	3	3	3	3	3	5	c2	c2	b2	b2	c2	c2	c3
22	2	3	3	3	3	4	3	4	c2	b2	b2	b2	c2	b3	c2
23	1	2	6	5	5	6	5	5	a1	c3	b3	b3	c3	c3	c3
23	2	2	6	5	5	6	6	5	a1	c3	b3	c3	c3	c3	c3
24	1	4	3	4	3	4	4	4	c2	c2	b2	b2	c2	c2	b2
24	2	3	3	4	3	4	4	4	c2	b2	c2	b2	c2	b3	b3

less complex fractures, and as the complexity of the fracture increased, there was greater disagreement. An experienced surgeon is often uncertain which segment is fractured when consulting CT scans and trauma radiographs (Bernstein et al. 1996). Neer (1997) stated that even the most experienced shoulder surgeon occasionally is in doubt, and has to make the final classification at surgery. It has therefore been questioned whether the classification systems are useful at all (Burstein 1993).

The reason for the poor interobserver agreement in fracture classifications could be the difficulties in translating the image information into the classification patterns. A complex classification system may result in a low reproducibility. The classifications of AO and Neer are based on conventional radiographs in the frontal plane. To translate the information of CT and 3D images into plain 2-dimensional frontal images introduces a new difficulty. It can be assumed that any fracture classification system dependent on images will have a low reproducibility if fracture patterns are not easily recognized. Thus a complex classification not easily translated from images irrespective of 2D, plain or CT, or 3D images is of limited clinical value.

The problems of translating the information in the radiographic image into an arbitrary classification makes the evaluation of treatment difficult. We should seek a classification that is easily extracted from the images—be it plain radiographs, CT or MRI—and in which the various classes have a prognostic value. Hopefully there will emerge a classification that is based on the functional rather than the anatomical features of the fracture (Bernstein et al. 1997, Colton 1997).

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.

Bernstein J, Monaghan B A, Silber J S, DeLong W G. Taxonomy and treatment—a classification of fracture classifications. J Bone Joint Surg (Br) 1997; 79: 706-7.

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.

Colton C. Fracture classification—a response to Bernstein et al. J Bone Joint Surg (Br) 1997: 79: 708-9.

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

Kristiansen B, Andersen U L S, Olsen C A, Varmarken J E. The Neer classification 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.

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) 1970; 52: 1077-89.

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

Sallay P I, Pedowitz R A, Mallon W J, Vandemark R M, Dalton J D, Speer K P. Reliability and reproducibility of radiographic interpretation of proximal humeral fracture pathoanatomy. J Shoulder Elbow Surg 1997; 6: 60-9.

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.

Sjödén G, Movin T, Güntner P, Aspelin P, Ahrengart L, Ersmark H, Sperber A. Poor reproducibility of classification of proximal humeral fractures. Additional CT of minor value. Acta Orthop Scand 1997; 68: 239-42.

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