Skeletal Radiol (1988) 17:420–422

Skeletal Radiology

The Neer classification of fractures of the proximal humerus

An assessment of interobserver variation

Bjarne Kristiansen, M.D., Ulrich L. S. Andersen, M.D., Claus A. Olsen, M.D., and Jens-Erik Varmarken, M.D.

Department of Orthopaedic Surgery, University Hospital of Copenhagen, Rigshospitalet, Copenhagen, Denmark

Abstract. The reliability of the Neer classification of proximal fractures of humerus was examined by determining the agreement between pairs of observers using weighted kappa statistics. Anteroposterior and lateral radiographs of 100 surgical neck fractures were grouped independently by four observers. A low degree of agreement was found, especially between the most inexperienced observer and the rest. Considering the therapeutic consequences of a correct classification, these fractures should be assessed by experienced orthopedic surgeons or radiologists.

Key words: Classification, fractures – Proximal humerus, fractures – Shoulder, injuries – Weighted kappa statistics

In 1970, Neer published a study of proximal humeral fractures [7] in which he included a classification based on the recognition that these fractures essentially involve one or more of four anatomic segments: the articular component, the greater tuberosity, the lesser tuberosity, and the shaft [1]. Combining vascular and functional considerations, in addition Neer defined significant segment displacement as occurring with angulation greater than 45 degrees or more than 10 mm of separation between segments. The Neer system has gained wide acceptance in comparative studies of these fractures as well as in clinical practice because of its proven value in management and prognosis [3, 5, 6]. In no previous study, however, has the degree of interobserver using the system error been reported.

Address reprint requests to: Bjarne Kristiansen, M.D., Department of Orthopaedic Surgery U, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark

The present study was therefore designed to determine the reliability of the Neer classification, based on the radiologic picture in anteroposterior as well as lateral projections, by examining the agreement between pairs of observers.

Materials and methods

Radiographs of 100 shoulders showing fractures of the surgical neck of the humerus were studied. Of the four observers taking part, one (A) was a specialist in orthopaedic surgery, two had more than two years (C and D) and the fourth, less than one years' experience in orthopaedic surgery (B). All four had studied the paper by Neer [7] and used the classification routinely. The radiographs were studied independently by each observer and were grouped into one of 5 groups defined in Fig. 1.

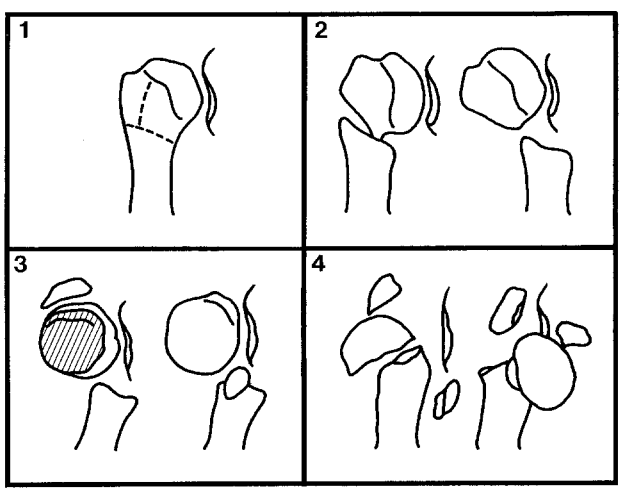

Fig. 1. The Neer classification of fractures of the surgical neck of the humerus. *Group 1:* Non – or minimally-displaced fractures. *Group 2:* Displaced 2-part fractures, only involving the surgical neck. *Group 3:* Displaced 3-part fractures involving one of the tuberosities. *Group 4:* Displaced 4-part fractures involving both tuberosities. In addition all the other types of proximal humeral fractures including anatomical neck fractures, isolated tuberosity fractures, and fractures-dislocations are combined in *group 5*

Statistical methods. The reliability of the Neer grouping was expressed in terms of interobserver agreement in pairs and calculated using weighted kappa statistics [2]. Weighted kappa values can vary from −1 (complete disagreement) through 0 (chance agreement) to +1 (complete agreement). The degree of disagreement between pairs of observers in assigning the group of each case was given a disagreement weighting v_x, which was 0, 1, 2, or 3. Weighted kappa was then calculated from the formula

$$k_w = 1 \div \frac{\sum v_x \times p_{ox}}{\sum v_x \times p_{cx}}$$

where p_{ox} is the observed and p_{cx} the chance agreement in each group.

Results

The total number of cases assigned to each group by each of the observers is outlined in Table 1; examples of assessments of the radiographs are shown in Figs. 2 and 3. The assessments of observer A and C are set out in full in Table 2 as an example of disagreements, showing for example that 28 fractures were placed in group 1 by both observers, and that 10 fractures were placed in group 2 by observer A and in group 3 by observer C. The percentage of interobserver agreement varied between 24% and 59%, whereas kappa values varied from 0.07 to 0.48 (Table 3). All the lowest values were found in pairs involving observer B. Considering each group separately, we found that the agreement observed in relation to the agreement expected by chance was higher in group 1. Combining groups 2, 3, and 4 produced a higher degree of agreement and raised the percentage values, but the kappa statistics were unchanged (Table 4).

Table 1. Number of cases assigned to each group by each of the four observers

Groups	Observers			
	A	B	C	D
1	38	5	32	34
2	20	6	5	7
3	17	42	29	35
4	13	36	19	20
5	12	11	15	4

Table 2. Comparison of the assessments of observers A and C

Group according to observer A	Group according to observer C				
	1	2	3	4	5
1	28	3	3	2	2
2	0	2	10	3	5
3	3	0	9	4	1
4	0	0	4	6	3
5	1	0	3	4	4

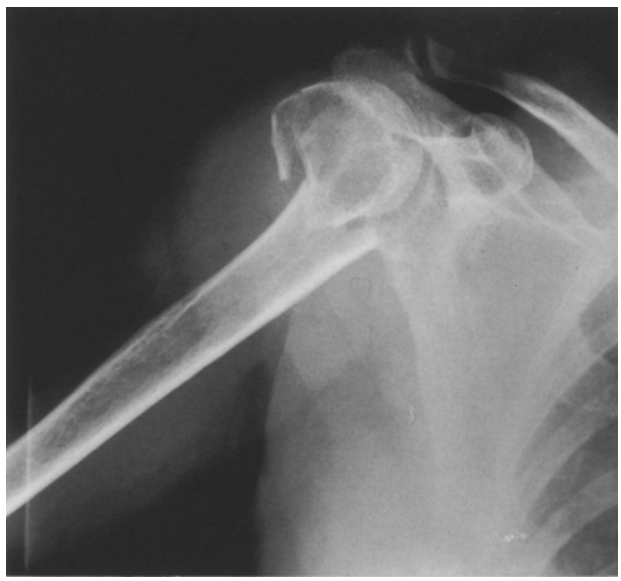

Fig. 2. *Case 1.* Surgical neck fracture assessed as a 2-part fracture by one observer, as a 3-part fracture by two and as a 4-part fracture by the fourth observer

Fig. 3. *Case 90.* Surgical neck fracture assessed as a 3-part fracture by one observer, as a 4-part fracture by another, and as a fracture-dislocation group 5 by two observes

Table 3. Agreement between pairs of observers. The figures in brackets are the numbers of agreements which would be expected if selection was made by chance alone

Pairs of observers	Agreement						
	Numbers in each group					Percentage correct	Kappa value
	1	2	3	4	5		
A and B	4 (2)	2 (1)	10 (7)	7 (5)	1 (1)	24	0.07
A and C	28 (12)	2 (1)	9 (5)	6 (2)	4 (2)	49	0.38
A and D	27 (13)	4 (1)	12 (6)	5 (3)	2 (0)	50	0.40
B and C	4 (2)	2 (0)	15 (12)	11 (7)	6 (2)	38	0.24
B and D	4 (2)	2 (0)	20 (15)	14 (7)	2 (0)	42	0.20
C and D	25 (11)	3 (0)	16 (10)	11 (7)	4 (1)	59	0.48

Table 4. Agreement between pairs of observers when Neer groups 2, 3, and 4 are combined. The figures in the brackets are the numbers of agreement which would be expected if selection was made by chance alone

Pairs of observers	Agreement				
	Numbers in each group			Percentage correct	Kappa value
	1	2+3+4	5		
A and B	4 (2)	44 (42)	1 (1)	49	0.03
A and C	28 (12)	38 (27)	4 (2)	70	0.33
A and D	27 (13)	45 (31)	2 (0)	74	0.43
B and C	4 (2)	50 (45)	6 (2)	60	0.30
B and D	4 (2)	55 (52)	2 (0)	61	0.17
C and D	25 (11)	46 (33)	4 (1)	75	0.47

Discussion

It is generally agreed that the stable, minimally displaced proximal humerus fracture has an excellent prognosis following conservative treatment consisting in short-term immobilization in a sling with early functional exercises. The displacement type of fracture, on the contrary, has a poor prognosis, and the treatment of choice is still disputed. In order to institute the optimal treatment it is thus of great importance to classify these fractures correctly from the radiographs. The reliability of the Neer grouping in this study has been tested using weighted kappa statistics. This method, compared to the more commonly used chi-square method, has the advantages that the degree of disagreement is taken into account, and allowance is made for chance agreement [2, 4]. The weighted kappa coefficient is usually used to compare the judgement of two different observers in diagnosing or evaluating the same cases, i.e., to determine if different levels of education are of importance [4]. No significant or acceptable level of the coefficient is suggested in the original work by Cohen [2] and must be discussed in any given series on this subject.

We found the highest interobserver agreement in the minimal displaced group of fractures. However the level of agreement expressed as kappa-coefficient is unchanged, when 2-, 3-, and 4-segment fractures are considered as one group, indicating serious difficulties in classifying the displaced types of fractures. Kappa values below 0.5 are, in our opinion, unacceptable and it must be emphasized that if the Neer classification is to form the basis for decisions regarding treatment, doubtful cases must be assessed by senior registrars or consultants in orthopaedic surgery or skeletal radiology.

References

1. Codman EA (1934) The shoulder: rupture of the supraspinatus tendon and other lesions in or about the subacromial bursa. Miller, Brooklyn
2. Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 70:213
3. Clifford PC (1981) Fractures of the neck of the humerus: a review of the late results. Injury 12:91
4. Gjørup T, Jensen AM (1986) The role of kappa coefficient in evaluating reproducibility of test results. Nord Med 101:90
5. Kristiansen B, Christensen SW (1987) Proximal humeral fractures. Late results in relation to classification and treatment. Acta Orthop Scand 58:124
6. Mills HJ, Horne G (1985) Fractures of the proximal humerus in adults. J Trauma 25:801
7. Neer CS (1970) Displaced proximal humeral fractures. Classification and evaluation. J Bone Joint Surg [Am] 52:1077