Observer variabilities of radiological classifications of calcified deposits in calcifying tendinitis of the shoulder

Markus MAIER[1,2], Tamara MAIER-BOSSE[1], Andreas VEIHELMANN[1,2], Christoph PELLENGAHR[1], Marc STEINBORN[3], Martin KLEEN[2,4], Christoph U. SCHULZ[1]

The radiological morphology of calcified deposits in calcifying tendinitis of the shoulder is classified according to Patte and Goutallier and according to Mole *et al*. The results of these classifications influence the choice of therapeutic procedures. In this study, the intraoberserver reproducibility and interobserver reliability of these classifications were determined. Plain anteroposterior radiographs of shoulders from 100 patients with symptomatic calcified deposits of the rotator cuff were classified according to the criteria of Patte and Goutallier as well as to the criteria of Mole *et al*, by six independent observers, twice within four months. The kappa values of intraoberserver reproducibility and interobserver reliability were calculated. Classification of Patte and Goutallier : intraoberserver reproducibility, mean kappa value 0.458 (standard deviation 0.098) ; interobserver reliability, mean kappa values 0.4 (first test) and 0.354 (second test). Classification of Mole *et al* : intraoberserver reproducibility, mean kappa value 0.402 (standard deviation 0.092) ; interobserver reliability, mean kappa values 0.239 (first test) and 0.191 (second test). Both classifications demonstrated a satisfactory to sufficient intraobserver reproducibility. The classification of Patte and Goutallier showed a satisfactory interobserver reliability, whereas the classification of Mole *et al* had a satisfactory to insufficient interobserver reliability. Studies dealing with both classifications should therefore be interpreted carefully.

INTRODUCTION

Calcified deposits in patients with calcifying tendinitis of the shoulder can be classified according to their radiological aspect (10, 13). Besides the clinical symptoms and the course of the disease, the morphology of the calcified deposit as noted on radiographs plays an important role to determine whether a patient with symptomatic calcifying tendinitis of the shoulder is treated or not and which type of treatment is applied (7, 9, 15, 16). Such classifications should therefore give reproducible results (17). Therefore, we tested the intraoberserver reproducibility and the interobserver reliabiliy of two different radiological classifications widely used in Europe (4, 10, 13). The study was carried out on anteroposterior radiographs of the shoulder

From the University Hospital (Klinikum Grosshadern), Ludwig-Maximilians-University Munich, Germany

[1] Department of Orthopaedic Surgery.

[2] Institute of Surgical Research.

[3] Institute of Clinical Radiology.

[4] Institute of Anaesthesia.

Markus Maier, Assistant Professor Orthopaedic Surgery.

Tamara Maier-Bosse, Registrar Orthopaedic Surgery.

Andreas Veihelmann, Assistant Professor Orthopaedic Surgery.

Christoph Pellengahr, Consultant Orthopaedic Surgery.

Marc Steinborn, Consultant Radiology.

Martin Kleen, Assistant Professor Anaesthesia.

Christoph U. Schulz, Consultant Orthopaedic Surgery.

Correspondence and reprints : Markus Maier, Department. of Orthopaedic Surgery, Ludwig-Maximilians University Munich, Klinikum Grosshadern, Marchioninistr. 15, D-81377 Munich, Germany. E-mail : markus.maier@ort.med.uni-muenchen.de.

obtained from patients with calcifying tendinitis of the shoulder.

MATERIALS AND METHODS

One hundred randomly selected digital anteroposterior radiographs demonstrating a single calcified deposit of the rotator cuff not overlaid or underlaid by bone were included in the study. Deposits were classified according to Patte and Goutallier (13) (type I : localised and homogeneous ; type II : diffuse, disseminated, heterogeneous) and according to the classification of the French Society of Arthroscopy (SFA) (Mole *et al* (10)) : type A : dense, homogeneous, sharp contours ; type B : dense, cloudy, sharp contours ; type C : inhomogeneous, soft contours, type D : dystrophic calcifications of the insertion zone of the rotator cuff tendons. Two weeks before classification all six observers (orthopaedic surgeons and radiologists with different levels of clinical expertise and familiar with the classifications) received a copy of the original articles (10, 13) and were also allowed to use these articles during the classification procedure.

After removing all personal data, the films were randomly numerized and presented to each observer alone without limitation of time. Once a decision was taken it could not be changed and not until the decision was taken, was the next film presented. This procedure went on until all 100 deposits were classified by each observer. All deposits were classified twice within an interval of 16 to 17 weeks. In the meantime films were not available to the observers and no results of the first classification were given. The second classification was done identically after films were re-randomised and renumbered.

Observer variabilities were determined by kappa statistics (11), established to determine observer variabilities in the interpretation of radiomorphologic findings (5, 12, 17). For calculation, a proportion of possible agreements between observers that occurred by chance (p_C) was calculated from observed data. The difference between the observed proportion of agreements (p_O) and p_C was divided by 1 minus p_C

For classification schemes using more than two levels, we used a quadratic weighted variation of kappa. For this, observations were weighted according to the degree of disagreement before calculation of kappa : the larger the disagreement, the more weight was assigned to a given observation (11). A kappa value of +1.0 is accepted to show complete agreement and of 0.0 to show full agreement. Interpretation of kappa values was done according to Landis and Koch (6) : 0.0 – 0.2, insufficient ; 0.21 – 0.4, satisfactory ; 0.41 – 0.6, sufficient ; 0.61 – 0.8, good ; 0.81 – 1.0 excellent. Precision of the classification could not be determined as the theoreticaly "true" classification was unknown. We therefore determined the level of agreement between the classifications of the same observer at two different moments (intraobserver reproducibility) and between the six different observers (interobserver reliability).

Table I. — Kappa values of intraobserver reproducibility. A, B, Registrar Orthopaedic Surgery ; C, Consultant Radiology ; D, Registrar Radiology ; E, F, Consultant Orthopaedic Surgery

Observer	Kappa values Patte and Goutallier (13)	Kappa values Mole *et al* (10)
A	0.575	0.519
B	0.425	0.318
C	0.387	0.401
D	0.479	0.498
E	0.556	0.384
F	0.324	0.293

RESULTS

Intraobserver reproducibility

Classification of Patte and Goutallier (13) : in the first test all six observers classified 6 × 100 deposits or 600 deposits according to their radiological aspect as type I (n = 336) and type II (n = 264) In the second test 238 of 336 deposits were classified as type I (70.8%) and 207 of 264 deposits as type II (78.4%). Kappa values of the intraobserver reproducibility showed a mean of 0.458 (SD 0.098) (table I). Classification of Mole *et al* (10) : in the first test all six observers classified 600 deposits as type I (n = 72), type II (n = 270), type III (n = 144) and type IV (n = 114). In the second test 69 of 72 deposits were classified as type I (95.8%), 252 of 270 as type II (93.3%), 140 of 144 as type III (97.2%) and 105 of 114 as type IV (92.1%). Kappa values of the intraobserver reproducibility showed a mean of 0.402 (SD 0.092) (table I).

Interobserver reliability

Classification of Patte and Goutallier (13) : in the first test 33 deposits were classified equally by

six observers, 25 deposits by five observers, 28 deposits by four observers and 14 deposits by three observers. In the second test 28 deposits were classified equally by six observers, 24 deposits by five observers, 38 deposits by four observers and 10 deposits by three observers. In the first test 18 deposits were classified in unison as type I and 15 deposits in unison as type II. In the second test each 14 deposits were classified in unison as type I and type II. Kappa value of interobserver reliability in the first test was 0.4 and in the second test 0.354.

Classification of Mole *et al* (10) : in the first test 14 deposits were classified equally by six observers, 19 deposits by five observers, 27 deposits by four observers and 37 deposits by three observers. In the second test 15 deposits were classified equally by six observers, 15 deposits by five observers, 35 deposits by four observers and 31 deposits by three observers. In the first test no deposits were classified in unison as type I, six deposits in unison as type II, seven deposits as type III and one deposit as type IV. In the second test no deposits were classified in unison as type I, five deposits in unison as type II, six deposits as type III and four deposits as type IV. Kappa value of interobserver reliability in the first test was 0.239 and in the second test 0.191.

DISCUSSION

According to the criteria of Landis and Koch (6) both classifications demonstrated satisfactory to sufficient intraobserver reproducibility. The classification of Patte and Goutallier (13) showed a satisfactory interobserver reliability for both tests. In contrast the classification of Mole *et al* (10) had a satisfactory (first test) and insufficient (second test) interobserver reliability.

Radiological morphologic classifications of calcified deposits used in Orthopaedic Surgery must fulfill certain prerequisites : first, one observer should classify identically the radiological aspect of a deposit at various moments (intraobserver reproducibility), second, different observers should classify identically the radiological aspect of a deposit (interobserver reliability) (1). These prerequisites are not fulfilled by the tested classifications (10, 13) whereas former studies dealing with radiological classifications already pointed to that problem of orthopaedic classifications. Interobserver reliabilities were only sufficient testing the classification of ankle fractures according to Lauge-Hansen (kappa value 0.50) and Weber (kappa value 0.57) (18). Neer‘s classification of proximal humeral fractures only had insufficient to satisfactory interobserver reliability (kappa values ranging from 0.26 to 0.50) (2, 3, 14).

The results of the present study could lead to the conclusion that the two tested classifications (10, 13) should use more precise criteria for reproducible and reliable classification of calcified deposits of the rotator cuff tendons. This would be of great clinical impact as both classifications (10, 13) are in common clinical use and influence the decision to treat patients with symptomatic calcifying tendinitis of the shoulder as well as the type of procedure to be applied in these patients (8, 10, 13, 16). However, the potential benefit of such classification systems is limited if their reproducibility and reliability is limited. Therefore, studies dealing with these classifications should be interpreted carefully.

REFERENCES

1. **Bernstein J.** Fracture classification systems : do they work and are they useful ? *J Bone Joint Surg* 1994 ; 76-A : 792-793
2. **Bjorkenheim JM, Paavolainen P, Ahovuo J, Slatis P.** Subacromial impingement decompressed with anterior acromioplasty. *Clin Orthop* 1990 ; 252 : 150-155
3. **Brien H, Noftall F, MacMaster S, Cummings T, Landells C, Rockwood P.** Neer’s classification system : a critical appraisal. *J Trauma* 1995 ; 38 : 257-260
4. **Jerosch J, Strauss JM, Schmiel S.** Arthroscopic therapy of tendinitis calcarea—acromioplasty or removal of calcium ? *Unfallchirurg* 1996 ; 99 : 946-952
5. **Krummenauer F, Kalden P, Kreitner KF.** Cohen’s kappa or McNemar’s test ? A comparison of binary repeated measurements. *Röfo Fortschr Geb Röntgenstr Neuen Bildgeb Verfahr* 1999 ; 171 : 226-231
6. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics* 1977 ; 33 : 159-174
7. **Loew M, Daecke W, Kusnierczak D, Rahmanzadeh M, Ewerbeck V.** Shock-wave therapy is effective for chronic calcifying tendinitis of the shoulder. *J Bone Joint Surg* 1999 ; 81-B : 863-867
8. **Maier M, Stäbler A, Lienemann A, Köhler S, Feitenhansl A, Dürr HR, Pfahler M, Refior HJ.**

Shockwave application in calcifying tendinitis of the shoulder — prediction of outcome by imaging. *Arch Orthop Trauma Surg* 2000 ; 120 : 493-498

9. **Maier M, Stäbler A, Schmitz C, Lienemann A, Köhler S, Dürr HR, Pfahler M, Refior HJ.** On the impact of calcified deposits within the rotator cuff tendons in shoulders of patients with shoulder pain and dysfunction. *Arch Orthop Trauma Surg* 2001 ; 121 : 371-378
10. **Mole D, Kempf JF, Gleyze P, Rio B, Bonnomet F, Walch G.** [Results of endoscopic treatment of non-broken tendinopathies of the rotator cuff. 2. Calcifications of the rotator cuff]. *Rev Chir Orthop* 1993 ; 79 : 532-541
11. **Norman GR, Streiner DL.** *Biostatistics, The Bare Essentials.* B.C. Decker Inc. Hamilton, London 2000 ; pp 218-221
12. **Ostergaard M, Klarlund M, Lassere M, Conaghan P, Peterfy C, McQueen F, O'Connor P.** Interreader agreement in the assessment of magnetic resonance images of rheumatoid arthritis wrist and finger joints — an international multicenter study. *J Rheumatol* 2001 ; 28 : 1143-1150
13. **Patte D, Goutallier D.** [Periarthritis of the shoulder. Calcifications]. *Rev Chir Orthop* 1988 ; 74 : 277-278
14. **Peh WC, Farmer TH, Totty WG.** Acromial arch shape : assessment with MR imaging. *Radiology* 1995 ; 195 : 501-505
15. **Rompe JD, Zoellner J, Nafe B.** Shock wave therapy versus conventional surgery in the treatment of calcifying tendinitis of the shoulder. *Clin Orthop* 2001 ; 387 : 72-82
16. **Rupp S, Seil R, Kohn D.** Tendinosis calcarea of the rotator cuff. *Orthopaede* 2000 ; 29 : 852-867
17. **Smith SW, Meyer RA, Connor PM, Smith SE, Hanley EN JR.** Interobserver reliability and intraobserver reproducibility of the modified Ficat classification system of osteonecrosis of the femoral head. *J Bone Joint Surg* 1996 ; 78-A : 1702-1706
18. **Thomsen NO, Overgaard S, Olsen LH, Hansen H, Nielsen ST.** Observer variation in the radiographic classification of ankle fractures. *J Bone Joint Surg* 1991 ; 73-B : 676-678