Inter and Intraobserver Variability in DePalma's Classification of Shoulder Calcific Tendinitis

MARKUS MAIER, TAMARA MAIER-BOSSE, CHRISTOPH U. SCHULZ, MARC STEINBORN, CLEMENS A. SCHMITT, MARTIN KLEEN, and AXEL STÄBLER

ABSTRACT. ***Objective.*** To determine the intraobserver reproducibility and interobserver reliability of DePalma and Kruper's classification of calcific tendinitis of the rotator cuff. The result of this classification influences the choice of therapeutic procedures in patients with symptomatic calcific tendinitis.

Methods. Plain anteroposterior radiographs of shoulders from 100 patients with symptomatic calcified deposits of the rotator cuff were classified according to the criteria of DePalma and Kruper by 6 independent observers at 2 different time points within 4 months. The kappa values of intraobserver reproducibility and interobserver reliability were calculated.

Results. Kappa values of intraobserver reproducibility had a mean of 0.487 (SD 0.094); kappa values of interobserver reliability were 0.234 for the first test and 0.273 for the second test.

Conclusion. Determination of intraobserver reproducibility gave satisfactory to sufficient results and interobserver reliability was satisfactory for both tests indicating that studies based on the classification of DePalma and Kruper should be interpreted cautiously. (J Rheumatol 2003;30:1029–31)

Key Indexing Terms:
CALCIFIC TENDINITIS SHOULDER KAPPA STATISTICS
MORPHOLOGY CLASSIFICATION DEPALMA AND KRUPER

Calcified deposits in patients demonstrating pain and dysfunction of the shoulder (i.e., calcific tendinitis of the shoulder) are classified according to their morphological aspect on radiographs[1,2]. The morphology of the calcified deposit should be taken into clinical consideration when deciding how to treat the patient with calcific tendinitis of the shoulder[2-4]. The classification should therefore give unique and reproducible results[5]. We tested intraobserver reproducibility and interobserver reliability of DePalma and Kruper's classification[1] in plain anteroposterior radiographs of calcified deposits of the rotator cuff taken of patients with calcific tendinitis of the shoulder.

From the Department of Orthopedic Surgery, the Institute of Surgical Research, the Institute of Clinical Radiology, and the Institute of Anesthesia, Ludwig-Maximilians-University Munich, Klinikum Grosshadern, Munich; and the Max-Delbrueck-Center for Molecular Medicine, Charité/Campus Virchow-Klinikum, Berlin, Germany.

M. Maier, MD, Consultant Orthopedic Surgery, Institute of Surgical Research and Institute of Clinical Radiology; T. Maier-Bosse, MD, Registrar Orthopedic Surgery; C.U. Schulz, MD, Senior Consultant Orthopedic Surgery, Department of Orthopedic Surgery; M. Steinborn, MD, Registrar Radiology, Institute of Clinical Radiology; M. Kleen, MD, Assistant Professor Anesthesiology, Institute of Surgical Research and the Institute of Anesthesia; A. Stäbler, MD, Assistant Professor Radiology, Institute of Clinical Radiology, Ludwig-Maximilians-University Munich; C.A. Schmitt, Academic Clinician and Principal Investigator, Max-Delbrueck-Center for Molecular Medicine, Charité/Campus Virchow-Klinikum.

Address reprint requests to Dr. M. Maier, Department of Orthopedic Surgery, Ludwig-Maximilians University Munich, Klinikum Grosshadern, Marchioninistr. 15, D-81377 Munich, Germany.
E-mail: markus.maier@ort.med.uni-muenchen.de

Submitted April 9, 2002; revision accepted October 14, 2002.

MATERIALS AND METHODS

Patient selection. Out of a total of 124 anteroposterior digital radiographic images showing a single calcified deposit of the supraspinatus tendon of the rotator-cuff, 100 were randomly selected to be included in the study. The location of the calcified deposit in the supraspinatus tendon was confirmed by ultrasound. No calcified deposit displayed overlay with bone structures and all deposits were larger than 5 mm in diameter. The patients included 61 women and 39 men (mean-age 54 years, range 40 to 75 years, SD 7.5 years). Forty-nine calcified deposits were located in left shoulders and 51 in right shoulders. Deposits were classified according to DePalma and Kruper[1] to distinguish between 2 morphologic types: type I, fluffy and amorphous; type II, defined and homogenous.

Classification. Two weeks before classification, all 6 observers (orthopedic surgeons and radiologists with different levels of clinical expertise and familiar with this classification) received a copy of the original article[1] and were also allowed to use this article during the classification procedure. After removing all personal data, films were randomly numbered and presented to individual observers alone. Examination time per film was not restricted. Once a decision was made it could not be changed and the next film was presented. This rule was followed until the evaluation of the 100 deposits was completed by all observers. All deposits were classified for a second time after a latency of 16 to 17 weeks. In the meantime films were not available to the observers, and no results of the first classification were released. The second classification was performed in an identical manner after films were randomized and renumbered.

Statistical analysis. Observer variability was determined by kappa statistics[6], established to determine observer variability in the interpretation of morphologic findings[5,7,8]. For calculation, a proportion of possible agreements between observers that occurred by chance (p_c) is calculated from observed data. The difference between the observed proportion of agreements (p_o) and p_c is divided by 1 minus p_c (Formula 1).

For classification schemes using more than 2 levels, we used a quadratic weighted variation of kappa. For this, observations were weighted according to the extent of disagreement before calculation of kappa: the larger the disagreement, the more weight was assigned to a given

observation[6]. A kappa index of + 1.0 shows complete agreement and an index of 0.0 shows incidental agreement. Interpretation of kappa values was according to Landis and Koch[9]: 0.0 to 0.2, insufficient; 0.21 to 0.4, satisfactory; 0.41 to 0.6, sufficient; 0.61 to 0.8, good; and 0.81 to 1.0 excellent. The mathematical precision of the classification could not be determined since the theoretically true classification was unknown. Therefore, the level of agreement between the classifications of the same observer at 2 different time-points (intraobserver reproducibility) and between the 6 different observers (interobserver reliability) were determined.

RESULTS

Intraobserver reproducibility. In the first evaluation, 600 deposits were classified according to their morphology as type I (n = 324) and type II (n = 276). In the second evaluation 241 of 324 deposits were classified as type I (74.4%) and 198 of 276 deposits as type II (71.7%). Kappa values of the intraobserver reproducibility had a mean of 0.487 (SD 0.094) (Table 1). Consultants (mean 0.530; SD 0.031) achieved higher mean kappa values than registrars (mean 0.444; SD 0.124), and surgeons (mean 0.5083; SD 0.053) achieved higher mean kappa values than radiologists (mean 0.444; SD 0.1725).

Interobserver reliability. In the first evaluation, 16 deposits were classified in agreement by 6 observers, 28 deposits by 5 observers, 34 deposits by 4 observers, and 22 deposits by 3 observers. In the second round, 26 deposits achieved identical classification by 6 observers, 18 deposits by 5 observers, 41 deposits by 4 observers, and 15 deposits by 3 observers. In the first round, 10 deposits were classified unequivocal as type I, and 6 deposits as type II. In the second evaluation, agreement was reached in over 19 deposits regarding type I, and in over 7 deposits regarding type II. The kappa value of interobserver reliability in the first round was 0.234 and in the second series was 0.273.

DISCUSSION

Calcific tendinitis of the shoulder is a self-limiting disease showing different stages that can be classified by clinical and radiological criteria. The classification of DePalma and Kruper[1] correlates the morphology of the calcified deposit on the rotator cuff tendons with both the patient's clinical situation and the stage of the disease. Type I deposits are usually found in acute cases and indicate the stage of resorption, whereas type II deposits are usually associated with subacute and chronic cases and indicate the stage of formation[10]. A prerequisite of rational therapeutic planning is to attribute the morphologic aspect and the clinical symptoms to the stage of the disease. During the formation stage, oral non-steroid medication or subacromial injections with local anesthetics with or without steroids are considered the standard therapies. If these modalities fail, needling, extracorporeal shockwave application, or surgical removal of the deposit become therapeutic alternatives. In contrast, during the stage of resorption only pain reduction and needling of the deposit are recommended[11].

Table 1. Kappa values of intraobserver reproducibility. A, B, Registrar Orthopedic Surgery; C, Consultant Radiology; D, Registrar Radiology; E, F, Consultant Orthopedic Surgery.

Observer	Kappa Values
A	0.440
B	0.570
C	0.566
D	0.322
E	0.512
F	0.511

However, our results indicate that the attribution of the deposits according to their morphologic aspect to the stage of the disease is not fully accomplished by the classification system proposed by DePalma and Kruper[1]. According to Landis and Koch[9] this classification only produced a satisfactory to sufficient intraobserver reproducibility and a satisfactory interobserver reliability with respect to both tests.

Due to the artificial test setting in our present study, the decision making process of the observers might differ from an actual clinical situation. Therefore, the present study has a number of limitations, even if the study design we used was established to test observer variability[12-14]: (1) the observers were aware that the interpretation of the radiography images had no direct clinical consequence for a patient' s treatment; (2) in a real clinical situation observers know about the clinical symptoms of each individual patient; (3) observers might have interpreted radiography images more freely and easily than in a clinical situation, and as a result, the study may have been biased towards a higher observer variability; or (4) conversely the observers might have assessed the radiographs more accurately than in a real clinical situation. This could denote that observer variability would be better under more regular clinical conditions.

More classifications concerning the morphological aspect of calcifications of the rotator cuff have been described[15-17]. However, we decided to test the classification of DePalma and Kruper[1] as all observers knew the original description of this classification and were familiar with its use in a routine clinical assessment. Those readers using the classification of DePalma and Kruper[1] for a long time (i.e., consultants) achieved a smaller observer variability than observers in training. However, there is no obvious explanation why radiologists showed poorer observer variability than surgeons.

A morphologic classification of calcified deposits commonly used in clinical rheumatology and orthopedic surgery must comply with certain prerequisites: first, any given observer should classify identically the morphologic aspect of a deposit at various time-points (intraobserver reproducibility); second, identical classifications should be

obtained from different observers (interobserver reliability)[18]. These prerequisites are not completely met by the classification of DePalma and Kruper[1], whereas former studies using morphologic classifications already pointed to that problem. Interobserver reliability was only sufficient when testing the classification of ankle fractures according to Lauge-Hansen (kappa index 0.50) and Weber (kappa index 0.57)[19]. Neer's classification of proximal humeral fractures achieved had only insufficient to satisfactory interobserver reliability (kappa indices ranging from 0.26 to 0.50)[20-22].

Our results could lead to the conclusion that the classification of DePalma and Kruper[1] needs more precise criteria for reproducible and reliable classification of calcified deposits of the rotator-cuff tendons. This would be of great clinical use since their classification is in common clinical use and contributes to the decision to treat or not to treat patients with symptomatic calcific tendinitis of the shoulder, and also suggests the type of therapeutic procedure that should be applied[3,4]. However, the value of a classification system must be questioned if its reproducibility and reliability are limited. Therefore, studies based on this classification should be interpreted cautiously.

REFERENCES

1. DePalma AF, Kruper JS. Long term study of shoulder joints afflicted with and treated for calcific tendinitis. Clin Orthop 1961;20:61-72.
2. McKendry RJ, Uhthoff HK, Sarkar K, Hyslop PS. Calcifying tendinitis of the shoulder: prognostic value of clinical, histologic, and radiologic features in 57 surgically treated cases. J Rheumatol 1982;9:75-80.
3. 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 Br 1999;81:863-7.
4. 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.
5. 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 Am 1996;78:1702-6.
6. Norman GR, Streiner DL, editors. Biostatistics, the bare essentials. London: B.C. Decker; 2000.
7. Krummenauer F, Kalden P, Kreitner KF. Cohen's kappa or McNemar's test? A comparison of binary repeated measurements. Rofo Fortschr Geb Rontgenstr Neuen Bildgeb Verfahr 1999;171:226-31.
8. Ostergaard M, Klarlund M, Lassere M, et al. Interreader agreement in the assessment of magnetic resonance images of rheumatoid arthritis wrist and finger joints — an international multicenter study. J Rheumatol 2001;28:1143-50.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
10. Uhthoff HK, Sarkar K. Calcifying tendinitis. Baillieres Clin Rheumatol 1989;3:567-81.
11. Rupp S, Seil R, Kohn D. Tendinosis calcarea of the rotator cuff. Orthopade 2000;29:852-67.
12. Morgan SJ, Groshen SL, Itamura JM, Shankwiler J, Brien WW, Kuschner SH. Reliability evaluation of classifying radial head fractures by the system of Mason. Bull Hosp Jt Dis 1997;56:95-8.
13. Rehan VK, Seshia MM, Johnston B, Reed M, Wilmot D, Cook V. Observer variability in interpretation of abdominal radiography of infants with suspected necrotizing enterocolitis. Clin Pediatr 1999;38:637-43.
14. Gregson JM, Leathley MJ, Moore AP, Smith TL, Sharma AK, Watkins CL. Reliability of measurements of muscle tone and muscle power in stroke patients. Age Ageing 2000;29:223-8.
15. Gartner J, Heyer A. Calcifying tendinitis of the shoulder. Orthopade 1995;24:284-302.
16. Patte D, Goutallier D. Periarthritis of the shoulder. Calcifications. Rev Chir Orthop Reparatrice Appar Mot 1988;74:277-8.
17. 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 Reparatrice Appar Mot 1993;79:532-41.
18. Bernstein J. Fracture classification systems: do they work and are they useful? J Bone Joint Surg Am 1994;76:792-3.
19. Thomsen NO, Overgaard S, Olsen LH, Hansen H, Nielsen ST. Observer variation in the radiographic classification of ankle fractures. J Bone Joint Surg Br 1991;73:676-8.
20. Bjorkenheim JM, Paavolainen P, Ahovuo J, Slatis P. Subacromial impingement decompressed with anterior acromioplasty. Clin Orthop 1990;252:150-5.
21. 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.
22. Peh WC, Farmer TH, Totty WG. Acromial arch shape: assessment with MR imaging. Radiology 1995;195:501-5.