Acta Orthop. Belg., 2020, **86**, 525-531

ORIGINAL STUDY

Radiographic assessment of calcifying tendinitis of the rotator cuff : an inter- and intraobserver study

Jan K.G. Louwerens, Femke M.A.P. Claessen, Inger N. Sierevelt, Denise Eygendaal, Arthur van Noort, Michel P.J. van den Bekerom

From the Department of orthopaedic surgery, Spaarne Gasthuis Hospital, Hoofddorp, The Netherlands & Shoulder & Elbow Platform collaborators

The radiographic appearance of calcific tendinitis of the rotator cuff varies according to the stage of the disease. We compared currently used classification systems in a large group of observers to identify the most reliable classification system. Thirty-seven orthopaedic surgeons evaluated shoulder radiographs of 25 patients to classify the stage of the calcific tendinitis according to the classifications by (1) Gärtner and (2) Molé on a Web-based study platform. Inter and intraobserver agreement among observers was measured using the Siegel and Castellan multirater κ. Both classification systems had fair interobserver agreement : κ was 0.25 for the Molé classification and 0.34 for the Gärtner classification. The Gärtner classification was significantly more reliable than the Molé classification. Currently there is no radiographic classification that can serve the purpose of guiding the treatment in a reliable way.

Keywords : Rotator cuff ; calcific tendinitis ; classifications ; radiographs ; agreement ; reliability.

INTRODUCTION

Calcific tendinitis is a frequently encountered cause of subacromial pain syndrome and its pathogenesis is still under debate *(14)*. The prevalence of calcific deposits in the rotator cuff tendons in either the general population (2.7% to 7.8%) as well as in a population with a painful shoulder (8% to 40%) is high *(7)* and the supraspinatus tendon is most frequently affected. It is postulated by Uhthoff et al. that calcific tendinitis can be divided into three

- Jan K.G. Louwerens[1],
- Femke M.A.P. Claessen[2],
- Inger N. Sierevelt[1],
- Denise Eygendaal[2],
- Arthur van Noort[1],
- Michel P.J. van den Bekerom[3,4]
- Shoulder & Elbow Platform & Science of Variation Group: J.A. Abboud, R. Blokzijl, J.P. Braman, J. Cheung, D.F. van Deurzen, J.S. Dines, S.F. Duncan, G.E. Garrigues, J.A. Hall, D.J. Hofstee, S.G. Kaar, P. Mansat, J.O. van der Meer, R. Page, P. Paladini, N. Parnes, M. Patel, J.M. Patino, L.I.F. Penning, G. Porcellini, M.V. Rademakers, S. Rowinski, C.P. Schonhuth, M. Shafi, A. Shafritz, C. Visser, A.J. Vochteloo, J.C. Vroemen, M.L. Wagener, M. Waseem, T.W. Wright, E.W. Zwitser, K. Zyto.

[1]*Department of orthopaedic surgery, Spaarne Gasthuis Hospital, Spaarnepoort 1, 2134TM, Hoofddorp, The Netherlands.*

[2]*Department of orthopaedic surgery, Amsterdam University Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.*

[3]*Department of orthopaedic surgery, Onze Lieve Vrouwe Gasthuis (OLVG), Jan Tooropstraat 164, 1061 AE Amsterdam, The Netherlands.*

[4]*Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, The Netherlands.*

Correspondence : Jan Louwerens, Department of orthopaedic surgery, Spaarne Gasthuis Hospital, Spaarnepoort 1, 2134TM, Hoofddorp, The Netherlands.

E-mail : jlouwerens@spaarnegasthuis.nl

No benefits or funds were received in support of this study. The authors report no conflict of interests.

Table I. — Radiographic classifications for calcific tendinitis of the rotator cuff

Title	*Definitions*	*Inter- / intraobserver agreement: κ (range)*
Gärtner *(5)*	(I) well circumscribed, dense (II) soft contour/dense or sharp/ transparent. (III) translucent and cloudy appearance without clear circumscription	0.33 - 0.48 / 0.36 - 0.42 *(8,9,21)*
Molé *(12)*	(A) dense, homogeneous, sharp contours (B) dense segmented, sharp contours (C) heterogeneous, soft contours (D) dystrophic calcification at the insertion	0.18 – 0.22 / 0.34 – 0.40 *(8,10)*
	(I) fluffy, amorphous and ill defined. (II) defined and homogeneous	0.25 - 0.34 / 0.24 - 0.49 *(8,10)*
DePalma *(15)* Patte and Goutallier *(16)*	(I) localised and homogeneous (II) diffuse, disseminated, heterogeneous	0.24 - 0.38 / 0.28 - 0.46 *(8,11)*

κ = Fleiss kappa.

main stages ; the pre-calcific stage, the calcific or formative phase and the resorption stage *(20)*. These stages are characterized by differences in size, shape and appearance on imaging techniques *(5)*. Different treatment options are advised depending on the stage of the disease and previous studies have suggested that there is a relationship between the size, location and morphology of calcifications and clinical outcome *(4,13)*. Imaging techniques can help the physician to localize and classify the calcific deposits and guide treatment by combining this information with clinical parameters *(13,19)*. Among these imaging techniques, radiography of the shoulder is widely available, cheap, fast and in most cases sufficient to diagnose calcific deposits in the rotator cuff. Standard radiographs include anterior-posterior, outlet and axillary view*(19)*. These views allow multidirectional assessment of the location and morphology of the deposits. Additional anterior-posterior views in external rotation and internal rotation could help in differentiating between a supraspinatus and infraspinatus located deposit. Various classifications systems exist to categorize radiographic signs of calcific tendinitis of which the Gärtner *(5)* and Molé *(12)* classification are most frequently used. (table I).

There is some knowledge on the psychometric properties of the classification systems in smaller observer groups and these studies did not show satisfactory inter- and intraobserver agreement *(9,21)*. The primary objective of this study was therefore to investigate whether the interobserver agreement of the Gärtner classification *(5)* and the Molé classification *(12)*, could be improved by using a large group of observers. The second objective was to assess whether the observers were able to correctly locate the deposit on the radiographs and the third objective was to re-assess the intraobserver agreement. The primary hypothesis was that both the Gärtner classification and the Molé classification systems would have a low inter- and intraobserver agreement. Our secondary hypothesis was that the interobserver agreement with respect to deposit location in the rotator cuff would be high.

MATERIALS AND METHODS

This study was approved by the institutional research board at the principal investigator's hospital and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its amendments.

Case selection

Radiographs of calcific tendinitis patients were selected from a database of patients treated at the senior investigator's hospital for shoulder pain. This database was composed for an earlier epidemiological study on the prevalence of calcific deposits in symptomatic and asymptomatic patients *(7)*. Each series contained three radiographs (anterior-posterior (AP), supraspinatus outlet view and axillary view). The AP radiographs were taken with the arm in neutral rotation and the scapula positioned parallel to the film (True AP / Grasney

Figure 1a. — Axial view of a subscapular deposit.

Figure 1b — AP view of a supraspinatus deposit.

view). In total 154 radiographs with correlative ultrasound examinations were available in the database *(7)*. Inclusion criteria for this study were cases with a single calcific deposit of at least 10 mm in size in a rotator cuff tendon and clinical signs of non-traumatic subacromial pain syndrome.

Figure 1c. — Outlet view of an infraspinatus deposit.

One of the authors (J.L., research fellow) selected twenty-five non-consecutive cases with calcific deposits of different size, morphology and location, representing a wide spectrum of radiographic presentations in a clinical setting. The distribution among the supraspinatus, the infraspinatus and subscapularis tendon was 17:5:3 (fig 1a-c). The standard of reference with regard to the location of the deposit was the ultrasound examination which was available in every case. Radiographs were anonymized, converted to DICOM files and uploaded to the research group's web-based survey platform.

Independent members of the Shoulderelbow Platform *(22)* were invited by an invitation e-mail that included a short study description. They were asked to evaluate 25 shoulder radiograph series from patients with calcific rotator cuff tendinitis on a web-based study platform. Other than an acknowledgment as part of the author collaborative in the paper, no incentives were provided. The goal of the Shoulderelbow Platform is to facilitate online

Table II. — Interobserver / intraobserver agreement

Classification	Categorical kappa	κ* (inter / intra)	abs ǂ (inter / intra)
Gartner & Simons	fair / substantial	0.34 / 0.70	0.58 / 0.79
Molé	fair / substantial	0.25 / 0.65	0.46 / 0.75
Location			
Supraspinatus	moderate / substantial	0.47 / 0.79	0.75 / 0.93
Infraspinatus / teres minor	fair / substantial	0.38 / 0.60	0.72 / 0.82
Subscapularis	moderate / substantial	0.53 / 0.76	0.90 / 0.95

* = Fleiss kappa. ǂ = absolute agreement

interobserver agreement and diagnostic accuracy studies in the field of orthopedic shoulder and elbow injuries. Members of the Shoulderelbow Platform are fully trained, actively practicing surgeons from different countries. The observers independently logged in to the website. After logging in they received an instruction on the use of the classification systems and were asked to provide the following demographic and professional data : (1) observer's gender (2) location of practice (3) years of practice (4) observer's clinical specialty and (5) number of treated calcific tendinitis patients a year. For the interobserver agreement observers were asked to classify the deposit according to the Gärtner *(5)* and the Molé classification *(12)* (Table 1).The observers were then asked to answer a multiple option question on the location of the deposit. Options were (1) supraspinatus tendon (2) infraspinatus / teres minor tendon (3) subscapularis tendon. Observers evaluated radiographs using a built-in Digital Imaging and Communications in Medicine viewer (MedDream, Softneta, Kaunas, Lithuania) and were able to zoom and adjust brightness, contrast, and window levelling. A case had to be completed to continue with the next case. Observers completed the study at their own pace, in their own time on various computers if necessary. Six months later randomly selected senior surgeons were contacted until six agreed to re-asses the previous cases to determine the intraobserver agreement.

Statistical analysis was performed by use of SPSS 21 software (IBM Corp, Armonk, NY, USA). Observer characteristics are described as frequencies with accompanying percentages. Agreement among observers was determined using absolute agreement and the Fleiss kappa measure described by Siegel and Castellan *(18)*. The Fleiss kappa measure is a frequently used statistics measure to describe chance- corrected agreement between ratings made by multiple observers *(3,6,17)*. The generated kappa values were interpreted according to the guidelines by Landis and Koch *(6)* : values of 0.01 to 0.20 indicate poor agreement ; 0.21 to 0.40 fair agreement ; 0.41 to 0.60, moderate agreement ; 0.61 to 0.80, substantial agreement ; and more than 0.81, almost perfect agreement. Kappa values were compared by use of a two sample Z-test. A p-value <0.05 was considered statistically significant. A subgroup analysis was performed on the inter-observer data to assess whether differences in observer characteristics (years in practice, number of treated cases a year or continent of residence) influenced the Fleiss kappa measure.

Post-hoc power analysis revealed that a minimum sample of 25 patients evaluated by a minimum of 37 observers would provide 97% power ($\alpha = 0.05$, $ß = 0.20$) in a two-sample Z-test to detect a clinically significant difference of one categorical rating of kappa ($\kappa = 0.10$).

RESULTS

In total 150 invitations were send. Fifty-seven surgeons logged in to the Shoulder Elbow Platform. Thirty-seven observers (25%) completed the survey. The majority of the observers worked in Continental Europe (59%), were in practice 5 years or more (68%) and treated >25 cases of calcific tendinitis patients a year.

Surgeons had fair interobserver agreement for both the Molé classification (κ=0.25) and the Gärtner classification (κ=0.34). The Gärtner classification

was more reliable than the Molé classification system (p=0.014*)*. A moderate interobserver agreement was found for the presence of calcific deposits in the supraspinatus tendon (κ =0.47) and the subscapularis tendon (κ=0.53) whereas a fair interobserver agreement was found for the presence of a calcific deposit in the infraspinatus tendon (κ =0.38) on radiographs. Table II shows the kappa and the absolute agreement values.

There was substantial intraobserver agreement for both the Molé (κ =0.65) and the Gärtner classification (κ =0.70) ranging from moderate (κ = 0.55) to almost perfect (κ = 0.85). With regard to the localization of the deposits on the radiograph, substantial intraobserver agreement was achieved for the supraspinatus (κ = 0,79), infraspinatus (κ =0,65) and subscapularis (κ =0,76).

Factor associated with interobserver agreement

Various subgroup analyses were performed based on demographic parameters of the observers but no significant differences were observed. (supplementary table III-V)

DISCUSSION

In the current study we showed that the radiographic classification systems as developed by Gärtner and Molé lack interobserver agreement. Therefore these classifications are not reliable enough to classify calcific tendinitis of the rotator cuff. The intraobserver agreement was acceptable with a substantial agreement (surgeons tend to agree with themselves more than with each other) but the interobserver agreement for the Gärtner classification was only fair according to the criteria by Landis and Koch. The Gärtner classification showed a higher (p = 0.014*)* interobserver reliability (κ = 0.34) than the classification by Molé (κ =0.24). The highest agreement among observers was seen among observers who treat >50 patients a year (κ = 0.47 / moderate agreement). Increasing the amount of observers to thirty seven in relation to previous studies with lower numbers did not improve interobserver agreement. In these studies, with four and six observers respectively, a fair to moderate agreement was found (table I) *(9,21)*. These results are consistent with previous studies in which alternative classification systems for calcific tendinitis have been tested (table I). While the results between these studies differ, the interobserver data never exceeds moderate agreement. Patte and de Palma both suggested a system with only two options but this did not result in a more reliable classification. It seems that observers can't agree on vague terms such as 'ill-defined, cloudy, inhomogeneous and localised or diffuse'.

In this study agreement between observers with regard to the location of the calcific deposit on radiographs was fair to substantial with absolute agreement ranging from 0.72 to 0.90. It appeared easier for observers to differentiate if a deposit was in the supraspinatus or the subscapularis than in the infraspinatus. This could be due to the fact that both tendons have their insertion on the greater tuberosity. Additional AP external and internal rotation views, supplementary to the outlet view, could have made the differentiation more easy but these radiograph were not routinely available in current database. Furthermore, deposits in the subscapularis tendon can be identified on the anterior aspect of the humeral head on the axillary-view radiographs, making it easier to determine the location.

A limitation of most studies of observer variation is the use of only a few observers. Strengths of this study include the use of a web-based study platform and it provides insight in how interobserver agreement studies can benefit from a web-based international study platform. While in this study the results between the small and large observer groups were similar, a web-based platform has the potential to gather a large amount of data from an international collaboration of surgeons, makes it easier to recruit observers and provides the observers the tools to assess the radiographs, or other imaging modalities, in an uniform way. In comparison to smaller observer studies this provides the researchers the possibility to perform additional subgroup analyses.

The study should however be interpreted in light of several limitations. First, the results would be more generalizable if the observers included a range of clinical staff including residents and reporting radiologists who may also be responsible

for diagnosis, triage and delivery of appropriate care. Second, observers received an explanation on the classification systems prior to the survey but did not receive any specific training. Third there may be a difference in quality between the web interface that was utilized and the usual way in which physicians view radiographs. However, the DICOM viewer provides all the usual tools that are required for an appropriate assessment. Fourth, for practical purposes we chose to limit the study to two classification systems although other, less frequently used, systems are available (table I). Fifth, additional AP in internal rotation views might have increased the detection rate of infraspinatus deposits. Finally, the data may be subject to the so-called ‘‘kappa paradox’’ because the kappa measure was considerably lower than the overall percentage of agreement (table II). If the prevalence of an outcome is low, it causes an imbalance in the marginal totals, generating a lower kappa than one might expect *(1,2)*.

Based on this study we can conclude that inter-observer radiographic classifications for calcific tendinitis of the rotator cuff are not reliable enough and would need more precise and simplified criteria to improve reliability. This would be of importance since tools, whether imaging or clinical, are needed to guide physicians in their treatment algorithm for patients with symptomatic calcific tendinitis. Development of these tools however is difficult, because there are no clear clinical or radiographic cut-off points between the different phases of the disease and patients may even have multiple calcific deposits in different phases. Currently there is no classification that can serve this purpose and physicians remain largely dependent on the development of symptoms over time and a combination of screening examinations to determine what phase of the disease a patient is in.

REFERENCES

1. **Bexkens R, Claessens F, Kodde I, et al**. The Kappa Paradox. *Shoulder Elb* 2018 ; 10 : 308.
2. **Bruinsma WE, Guitton TG, Warner JJ, et al**. Interobserver Reliability of Classification and Characterization of Proximal Humeral Fractures. *J Bone Jt Surg* 2013 ; 95 : 1600-1604.
3. **Cohen J**. Weighted kappa : Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 1968 ; 70 : 213-220.
4. **Gärtner J, Heyer A**. Calcific tendinitis of the shoulder. *Orthopade* 1995 ; 24 : 284-302.
5. **Gartner J, Simons B**. Analysis of calcific deposits in calcifying tendinitis. *ClinOrthopRelat Res* 1990 ; May : 111-120.
6. **Landis RJ, Koch GG**. The measurement of observer agreement for categorical data. *Biometrics* 1977 ; 33 : 159-174.
7. **Louwerens JKG, Sierevelt IN, van Hove RP, et al**. Prevalence of calcific deposits within the rotator cuff tendons in adults with and without subacromial pain syndrome : clinical and radiologic analysis of 1219 patients. *J Shoulder Elbow Surg.* 2015 ; 24 : 1588-93.
8. **Maier M, Amsin JSC, Laser CG, et al**. Intra- and interobserver reliability of classification scores in calcific tendinitis using plain radiographs and CT scans. *Acta Orthop. Belg* 2008; 74: 590-595.
9. **Maier M, Maier-Bosse T, Eckermann M, et al**. Die rontgenmorphologische Gestalt symptomatischer Verkalkungen bei Tendinosis calcarea der Schulter. *Unfallchirurg* 2003 ; 106 : 185-189.
10. **Maier M, Maier-bosse T, Schulz CU, et al**. Inter and intraobserver variability in DePalma ’ s classification of shoulder calcific tendinitis. *J Rheumatol* 2003 ; 30: 4-7.
11. **Markus M, Osse T, Eihelmann A V, et al**. Observer variabilities of radiological classifications of calcified deposits in calcifying tendinitis of the shoulder. *Acta Orthop Belg* 2003; 69: 222-225.
12. **Mole D, Kempf JF, Gleyze P, et al**. Results of endoscopic treatment of non-broken tendinopathies of the rotator cuff. *RevChir OrthopReparatrice ApparMot* 1993 : 532-541.
13. **Ogon P, Suedkamp NP, Jaeger M, et al**. Prognostic factors in nonoperative therapy for chronic symptomatic calcific tendinitis of the shoulder. *Arthritis Rheum* 2009 ; 60 : 2978-84.
14. **Oliva F, Via AG, Maffulli N**. Physiopathology of intratendinous calcific deposition. *BMC Med* 2012 ; 10 : 95.
15. **de Palma A, Kruper J**. Long-term study of shoulder joints afflicted with and treated for calcific tendinitis. *Clin Orthop* 1961; 20 : 61-72.
16. **Patte D, Goutallier D**. Periarthritis of the shoulder. Calcifications. *Rev Chir Orthop Reparatrice Appar Mot* 1988 ; 74 : 277-8.
17. **Posner KL, Sampson PD, Caplan RA, et al**. Measuring interrater reliability among multiple raters : An example of methods for nominal data. *Stat Med* 1990 ; 9 : 1103-1115.
18. **Siegel, Sidney and N John Castellan J**. Nonparametric Statistics for the Behavioral Sciences. 2nd editio. New York, McGraw-Hill, 1988.
19. **Suzuki K, Potts A, Anakwenze O, et al**. Calcific tendinitis of the rotator cuff : management options. *J Am Acad Orthop Surg* 2014 ; 22 : 707-717.

20. **Uhthoff HK, Sarkar K**. Calcifying tendinitis. *Baillieres Clin Rheumatol* 1989 ; 3 : 567-581.

21. **de Witte PB, van Adrichem RA, Selten JW, et al**. Radiological and clinical predictors of long-term outcome in rotator cuff calcific tendinitis. *Eur Radiol* 2016 ; 26 : 3401-3411.

22. Shoulder Elbow Platform. *www.shoulderelbowplatform.com*