Development of a new radiographic score for the follow-up of calcific tendinopathy of the rotator cuff

R. Dalla-Torre[1], A. Fouasson-Chailloux[2,3], B. Le Goff[1,3], C. Darrieutort-Laffite[1,3]

[1]Service de Rhumatologie, CHU Nantes; [2]Service de Médecine Physique et Réadaptation Locomotrice et Respiratoire, CHU Nantes; [3]Nantes Université, Oniris, CHU Nantes, INSERM, Regenerative Medicine and Skeleton, RMeS, UMR 1229, Nantes, France.

Abstract

Objective

The aim of this study was to evaluate the psychometric properties of a new x-ray scoring system for calcific tendinopathy of the rotator cuff (CTRC).

Methods

This is a post-hoc analysis of the CALCECHO trial. All patients received an ultrasound-guided puncture and lavage of their calcification. Clinical data and x-rays from baseline and follow-up visits at 7 days (D7), 3 months (M3) and 12 months (M12) were used. The scoring system was based on the reduction in size and density of the calcification compared to the initial x-ray (0 = no change; 1 = decrease of less than 50%; 2 = decrease of between 50 and 90%; 3 = decrease of more than 90%; 4 = complete disappearance). Inter-observer and intra-observer reliability were established between 3 independent investigators (2 experts and one junior) using weighted Kappa calculation. Construct validity was assessed as well as predictive validity and sensitivity to change.

Results

Between the two experts, inter-reader reliability was at 0.677, 0.744 and 0.656 at D7, M3 and M12 respectively. Intra-reader reliability was between 0.577 and 0.836 for the two expert readers and between 0.519 and 0.697 for the junior reader. Our score was correlated with shoulder pain and function at M3 and M12 and the score at M3 was predictive of the clinical outcome at M12. Finally, sensitivity to change was 0.8.

Conclusion

Our new score presented good psychometric properties and was correlated with clinical data. It could be useful in the follow-up of patients treated for CTRC.

Key words

rotator cuff, calcific tendinopathy, treatment, x-rays

Romain Dalla-Torre, MD
Alban Fouasson-Chailloux, MD, PhD
Benoit Le Goff, MD, PhD
Christelle Darrieutort-Laffite, MD, PhD

Please address correspondence to:
Christelle Darrieutort-Laffite
Department of Rheumatology,
CHU Nantes,
1 place Alexis Ricordeau,
44000 Nantes, France
E-mail:
christelle.darrieutort@univ-nantes.fr
ORCID iD: 0000-0002-1871-0073

Received on May 22, 2023; accepted in revised form on September 1, 2023.

Competing interests: none declared.

Introduction

Calcific tendinopathy of the rotator cuff (CTRC) is one of the most common causes of shoulder pain, affecting 10 to 42% of patients referred for a painful shoulder (1-4). CTRC affects patients between 30 and 60 years of age, with a female predominance, and the supraspinatus tendon is the most frequently affected (80% of deposits). The aetiology of the disease is still poorly known, and only little data is available on the mechanisms leading to calcium deposits (5). According to Uhthoff *et al*. (6), calcific deposits evolve in 4 consecutive phases. During the first two phases, the calcification appears "chalky" and well defined. It starts to become symptomatic when its volume increases to the point of causing tendon deformation and subacromial impingement. The resorption phase then occurs. The calcification becomes less dense, more fragmented, and with irregular boundaries. This resorption phase leads to the disappearance of the deposit followed by a final repair phase.

On x-ray, calcific deposits appear as a dense opacity around the humeral head. Various radiographic classifications have been developed to characterise rotator cuff calcific deposits (7-9). The two most widely used were provided by Molé (7) and Gärtner (8). However, previous data suggested a lack of inter-reader reliability with Kappa values at 0.47 and 0.25 for Gärtner and Molé, respectively (10-11). In clinical trials, these classifications are mainly used to characteryse the deposits at baseline but not during post-treatment follow-up (12-14). For the follow-up, x-ray changes are commonly reported using the size of the deposit in mm (maximum size of the longest axis) (12, 15, 16) or changes are defined as complete resolution, partial resolution (decrease in size) or no change (12, 17, 18). During long-term follow-up (≥1 year), previous publications have shown a positive correlation between radiological improvement (decrease in size of the calcification) and clinical results (19-21). However, it is not clear if early radiographical changes are associated with clinical improvement and whether they could be used to predict the final outcome and guide the treatment.

In this study, we proposed a new scoring system based on evolution of the deposits over time. We hypothesised that a score evaluating the radiological changes in the calcific deposits beyond merely its initial aspect could be useful when managing patients treated for CTRC. We assumed that a score assessing the changes in size over time would be correlated with clinical symptoms and therapeutic success, and could help managing the patients. The objective of the study was to assess its reliability, sensitivity to change, correlation with the clinical data, and predictive value.

Methods

Study design

This is a *post-hoc* analysis of a double-blind multicentric non-inferiority randomised controlled study (CALCECHO trial (NCT02403856) [22]). To study our new scoring system, we used clinical data and x-rays of a sample of patients. The research protocol was approved by the local ethics committee and the French National Agency for Medicines and Health Products Safety (ANSM RC15_0019).

Patients

Inclusion criteria were age >18 years, chronic pain for more than 3 months, clinical subacromial impingement, and a calcification larger than 5 mm on standard anteroposterior x-ray. In the original study, all patients (n=132) underwent an ultrasound-guided puncture and lavage (UGPL) of the calcification and were randomised (1:1) to receive a steroid or saline injection in the subacromial bursitis at the end of the procedure.

Data collection

A sample of 69 patients was used for the present study. A total of 76 patients were included in our centre while the 56 others were included in the two other centres that participated in the study. However, x-ray data were incomplete for 7 patients (lost to follow-up). Including patients from a single centre allowed us to easily access x-rays on the Picture Archiving and Communica-

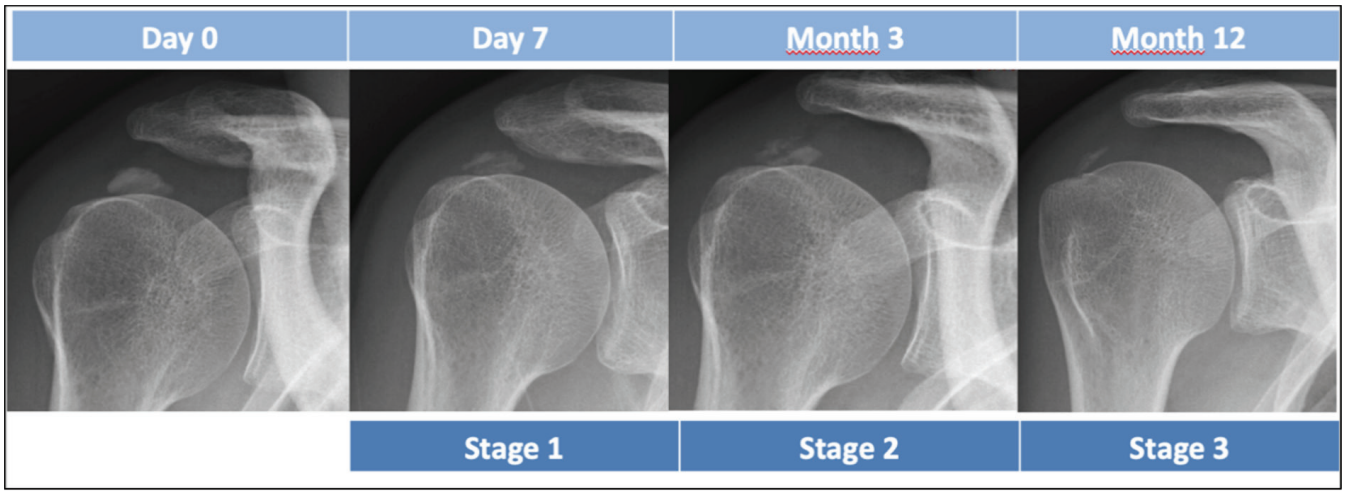

Fig. 1. Examples of x-rays taken at baseline (Day 0), day 7 (D7), month 3 (M3) and month 12 (M12) after an ultrasound-guided puncture and lavage of calcific tendonitis of the supraspinatus. In our score, the image at D7 corresponds to score 1 (decrease of the calcification of less than 50%); the image at M3 to score 2 (decrease of the calcification of between 50 and 90%), and the image at M12 to score 3 (decrease of more than 90%).

tion System of the hospital and to analyse all the images in the same system. We analysed baseline x-rays and those obtained during follow-up at 7 days (D7), 3 months (M3), and 12 months (M12). Shoulder anteroposterior view with the arm in a neutral position, internal rotation, external rotation, and scapular "Y" views were taken at each timepoint. A total of 276 exams were analysed, this number providing enough power to assess the reproducibility of the score. We collected the following information for the 69 patients at baseline: age, sex, duration of symptoms, presence of nocturnal pain, visual analogic scale (VAS) pain at rest and during activities and Disabilities of the Arm, Shoulder, and Hand (DASH) score. We also noted the tendon involved and the maximum long axis size of the calcific deposit.

All x-rays were independently and blindly analysed by 3 rheumatologists (2 were considered "experts" thanks to their experience in the management of CTRC and the third was a junior doctor. At each timepoint, they rated the x-rays according to the proposed score and Molé and Gärtner classifications. Our score was based on the assessment of changes in size and density of the calcification compared to pre-treatment x-ray. It was established as follows: 0 = no or minimal change from the initial x-ray; 1 = decrease of the calcification of less than 50%; 2 = decrease of between 50 and 90%; 3 = decrease of more than 90%; 4 = complete disappearance of the deposit (Fig. 1). They performed two readings three months apart to measure intra-observer reliability. Finally, the two expert readers (1 and 2) reviewed the discordant X-rays to obtain a consensus. This consensus was used to assess the construct validity and the predictive value of the score. To assess the correlation with clinical data, we used follow-up data obtained at D7, M3, and M12: VAS pain at rest and during activities and DASH score (except at D7). The DASH score is an auto-questionnaire used to assess the function of the upper limb. A DASH score <15 corresponds to a clinical state where patients have no remaining daily disability related to their shoulder condition and are able to work. A DASH score <40 corresponds to patients who can work despite the presence of pain (23). We also analysed patients reaching the minimal clinically important difference (MCID) of the DASH score estimated at 16 points (20).

Data analysis

Inter-observer reliability was assessed using x-rays from the entire cohort (n=69), while intra-observer reliability was assessed in 30 randomly selected patients. Weighted Kappa values were calculated. The Kappa was calculated using both the observed accuracy and the random accuracy. Kappa was calculated as: (observed accuracy - random accuracy) / (1 - random accuracy) (24). When the number of categories increases, the value of κ decreases because there is more room for disagreement so we used the weighted Kappa to take the degree of disagreement into account. The reading was done while aware of the order of the x-rays. The readers were blinded to the patient allocation group in the CALCECHO trial and blinded to clinical and ultrasound data. Construct validity was assessed through the cross-sectional relationship between x-ray stages and pain (VAS pain at rest and during activities) or function (DASH score), using a Spearman test. Predictive validity, *i.e.* the ability of the radiographic score to predict clinical outcome (DASH <15, DASH <40, and MCID DASH), was assessed using a χ^2 test. Sensitivity to change was assessed by standardised response mean (SRM): mean (M3-D7 and M12-M3) scoring change/standard deviation of change (25-27). Statistical analysis was performed using SPSS statistical software. Statistical significance was set at 0.05.

Results

Baseline characteristics of the patients

Within the cohort, there was a predominance of women (n=46, 66.7%), and the mean age was 49.8±15 years. Mean VAS pain at rest and during activities

Table I. Patients' characteristics at baseline.

Characteristics	Total (n=69)
Age, mean (SD)	49.8 (15)
Female sex, n (%)	46 (66.7%)
Duration of symptoms in months, mean (SD)	32.4 (40.8)
VAS at rest/100 mm, mean (SD)	25 (22)
VAS at motion/100 mm, mean (SD)	68 (15)
Nocturnal pain, n (%)	52 (75.4%)
DASH Score, mean (SD)	43.5 (15.2)
Tendon affected, n (%)	
Supraspinatus	60 (87%)
Infraspinatus	8 (11.5%)
Subscapularis	1 (1.5%)
Maximum long axis of the calcification in mm, mean (SD)	16.5 (4.9)

VAS: visual analogue scale; DASH: disabilities of the arm, shoulder and hand; SD: standard deviation.

was 25±22/100 and 68±15/100, respectively. On x-ray, the mean long axis of the calcification was 16.5±4.9 mm and the calcification was located within the supraspinatus in most cases (n=60, 87%) (Table I).

The baseline characteristics of the patients included in this study (n=69) were not different from those of the CALCECHO entire cohort (n=132) (Supplementary Table S1).

Inter and intra-reader reliability

At D7, inter-reader reliability between the two experts was substantial for our score (weighted Kappa values at 0.677) (Table II). It was fair for the Molé and Gärtner classifications (weighted Kappa values at 0.356, and 0.389 for Molé score and Gärtner score, respectively) (Suppl. Table S2) (28). At M3 and M12, weighted Kappa values were moderate to substantial for our score (0.744 and 0.656) (Table II).

At M3 and M12, weighted Kappa values for intra-reader reliability were 0.792 and 0.836 for reader 1 and 0.801 and 0.809 for reader 2 (Table II). Considering the junior, intra-reader reliability was better for our score compared to the two other classifications at each timepoint, with weighted Kappa values closer to those of the experts (Table II and Suppl. Table S3).

Construct validity (cross-sectional association between radiographic stages and symptoms)

At D7, pain level at rest or during activities was not correlated with our score (Fig. 2a). However, at M3, there was a significant correlation between the score and clinical outcomes considering VAS pain during activities (r =- 0.420, $p<0.001$) and DASH score (r = -0.449, $p<0.001$) (Fig. 2b). Thus, decreased pain and improved function were associated with increasing x-ray score, *i.e.* with the reduction of the calcification (Fig. 2). At M12, increasing X-ray score was correlated with a decrease in VAS pain at rest (r = -0.486, $p<0.001$) and an increase in DASH score (r = -0.327, $p<0.007$) (Fig. 2c).

Predictive validity

The score at D7 were not significantly associated with the clinical improvement at M3. However, a significant association was found between the x-ray score at M3 and the clinical data at M12 using DASH <15 ($p=0.016$), DASH <40 ($p=0.045$) and DASH Minimal Clinically Important Difference ($p<0.001$) (Table IV).

Sensitivity to change

The results for sensitivity to change were expressed as SRMs: 0.81 and 0.80 for sensitivity to change between D7 and M3 and between M3 and M12 respectively.

Discussion

In this *post-hoc* analysis, we investigated the reliability of a new radiographic scoring system for calcific tendinopathy assessing its inter-reader and intra-reader reliability, as well as its correlation with clinical data and its predictive value in a cohort of 69 patient using a total of 276 exams. Our score presented substantial inter-observer reliability (weighted Kappa values at 0.677, 0.744 and 0.656 at D7, M3 and M12, respectively) and substantial to almost perfect intra-reader reliability. The score was correlated with pain level and functional status at M3 and M12 and the score at M3 was capable of predicting clinical outcome at M12.

Various studies have investigated the reliability of radiographic classifications in CTRC, especially the Molé

Table II. Inter-reader reliability and intra-reader reliability for our score.

Inter-observer reliability		Our score			Intra-observer reliability		Our score		
		Reader 1/2	Reader 2/3	Reader 1/3		Reader 1	Reader 2	Reader 3	
D0	Weighted kappa (95% CI)	-	-	-	D0	Weighted kappa (95% CI)	**-**	**-**	**-**
D7	Weighted kappa (95% CI)	**0.677** (0.557-0.798)	**0.486** (0.17-0.654)	**0.548** (0.385-0.712)	D7	Weighted kappa (95% CI)	**0.615** (0.390-0.841)	**0.577** (0.34-0.772)	**0.519** (0.271-0.767)
M3	Weighted kappa (95% CI)	**0.744** (0.637-0.852)	**0.517** (0.381-0.653)	**0.545** (0.418-0.671)	M3	Weighted kappa (95% CI)	**0.792** (0.652-0.932)	**0.801** (0.687-0.915)	**0.697** (0.502-0.893)
M12	Weighted kappa (95% CI)	**0.656** (0.509-0.802)	**0.539** (0.397-0.682)	**0.630** (0.522-0.739)	M12	Weighted kappa (95% CI)	**0.836** (0.690-0.982)	**0.809** (0.711- 0.907)	**0.671** (0.495-0.846)

D0: Day 0 (baseline); D7: day 7; M3: Month 3; M12: Month 12 after ultrasound-guided percutaneous lavage of the calcification.

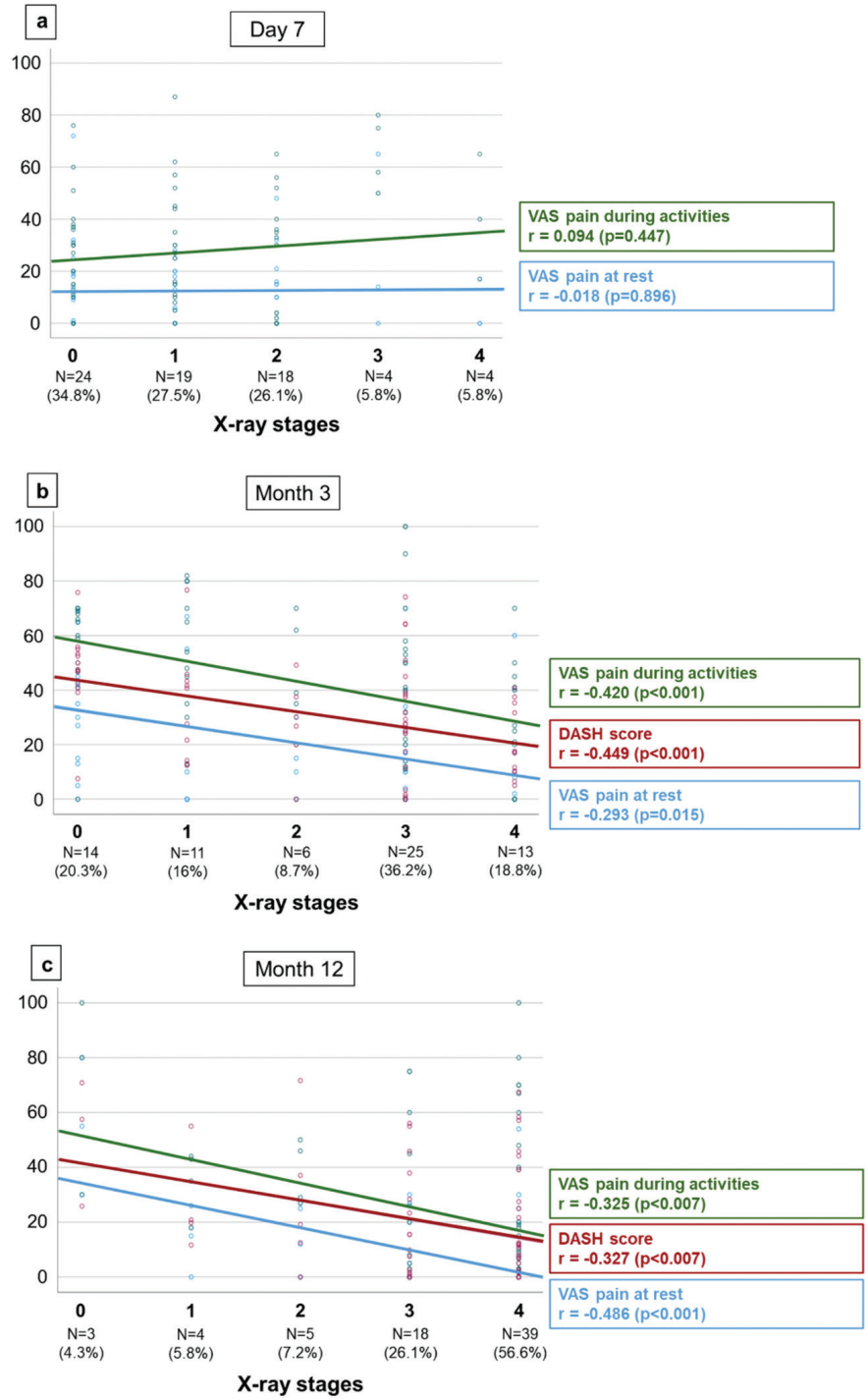

Fig. 2. Correlation between our radiographic score and pain level or function over time. At day 7, there was no correlation between the radiographic score and VAS pain. However, at month 3 and month 12, we found a correlation between symptoms (VAS pain and DASH score for function) and radiographical scores. Clinical improvement correlated with reduction in calcific deposits (*i.e.* increase in radiographic score).

and Gärtner classifications, which are the most widely used. Kappa values for Molé and Gärtner classifications in previous studies were slight to fair: 0.18 to 0.25 and 0.33 to 0.47 respectively for inter-observer reliability (9-11, 29). However, the intra-reader reliability was better (0.34 to 0.65 for Molé and 0.36 to 0.70 for Gärtner) (9, 11). Here, we reported higher inter-reader reproducibility, especially between the 2 experts (0.677, 0.744 and 0.656 at D7, M3 and M12 respectively). Regarding intra-reader reliability, we found Kappa values close to those recently published for Gärtner classifications when looking at the expert readings (11). The intra-reader agreement looked better for our junior reader compared to Gärtner classifications, suggesting a reduced impact of the clinician's experience on correct classification of the x-rays. Moreover, in our study, the inter-observer reproducibility of Gärtner scores tended to decrease over time, while it remained stable for our score.

We observed a significant correlation between our score and shoulder pain and function at M3 and M12. At D7, the correlation with pain was not found, nor was the predictive value. This could be explained by a low level of pain overall at D7. During the week after the UGPL, most patients had significant improvement in their pain (22) while the x-rays did not yet show any obvious changes. This immediate effect on pain could be related to the work interruption during the week after UGPL, to the systematic prescription of non-steroidal anti-inflammatory drugs and analgesics during the first days following the procedure, or to the corticosteroid infection for those randomised in the steroid group. Considering the absence of radio-clinical correlation and the absence of predictive value at this time, performing an x-ray at D7 is probably not useful in the follow-up of UGPL. However, at M3, the x-ray score could be used to guide treatment. If patients have no changes in their calcification at 3 months, they only have a 20% probability of having improvement at 12 months whereas they have a 75 to 100% chance of improvement in case of reduction by more than 90% or disappearance. A follow-up visit at 3 months with a new x-ray would allow clinicians to determine whether additional treatment is necessary or not. Finally, the scoring system showed a sensitivity to change considered "significant" according to Cohen's threshold (30) with an SRM of 0.8. This suggests that our score is capable of detecting clinically significant changes. It would be interesting to use it as an imaging outcome in future clinical trials on calcific tendinopathy.

Table III. Predictive value of our score at D7 and M3.

	Score at D7 (n=66)					
	0 (minimal changes)	**1** (decrease in size <50%)	**2** (decrease in size 50-90%)	**3** (decrease in size >90%)	**4** (complete disappearance)	*p*
DASH M3 <15	21.7% (n=5/23)	27.8% (n=5/18)	35.3% (n=6/17)	0%(n=0/4)	25% (n=1/4)	0.657
DASH M3 <40	43.5% (n=10/23)	72.2% (n=13/18)	76.5% (n=13/17)	75%(n=3/4)	100%(n=4/4)	0.077
DASH MCID M3(n=63)	21.7% (n=5/23)	38.9% (n=7/18)	66.7% (n=10/15)	33.3% (n=1/3)	50% (n=2/4)	0.095
	Score at M3 (n=67)					
	0 (minimal changes)	**1** (decrease in size <50%)	**2** (decrease in size 50-90%)	**3** (decrease in size >90%)	**4** (complete disappearance)	*p*
DASH M12 <15	21.4% (n=3/14)	72.7% (n=8/11)	50% (n=3/6)	62.5% (n=15/24)	83.3% (n=10/12)	**0.016**
DASH M12 <40	57.1% (n=8/14)	90.9% (n=10/11)	66.7% (n=4/6)	91.7% (n=22/24)	91.7% (n=11/12)	**0.045**
DASH MCID M12(n=64)	21.4% (n=3/14)	90% (n=9/10)	66.7% (n=4/6)	75% (n=18/24)	100% (n=10/10)	**<0.001**

VAS: visual analogue scale; DASH: disabilities of the arm, shoulder and hand; MCID: mean clinical important difference.
(3 missing DASH scores at baseline (no calculation of the MCID), 3 others missing at M3 and 2 at M12).

Limitations

Our *post-hoc* analysis has certain limitations. First, we only used part of the initial cohort, which involved patients from a single centre. Although this group of patients was similar to the overall cohort, this may have induced a selection bias. Second, the readers were aware of the order of the x-rays. To use our score, referring to baseline x-ray was required but knowing the order of follow-up x-rays could have represented a bias. Third, our score, which is composed of 5 options, might have induced more differences between readers and readings and could have altered the inter- and intra-reader reliability compared to the other classifications which only include 3 and 4 options. In addition, Molé and Gärtner classifications did not consider disappearance of the calcification. For this reason, part of the x-rays could not be evaluated, especially at M12 when part of the calcifications had disappeared. This may have induced bias in the reliability analyses for these two classifications. Furthermore, the observers only received an explanation on the classification systems prior to the readings but did not receive any specific training. Finally, the data may be subject to the so-called "Kappa paradox". In some cases, when the prevalence of an outcome is low, it causes an imbalance in the marginal totals, generating a lower Kappa than expected (31, 32). Afterwards, it will be necessary to validate this new scoring system in other patient cohorts.

Conclusion

This study showed that our new score based on the percentage of reduction of the deposit had good psychometric parameters, was correlated to shoulder pain and function at M3 and M12, and was capable of predicting the long-term outcome in a cohort of patients undergoing treatment for CTRC. We believe that it could help clinicians to manage patients with CTRC.

References

1. LOUWERENS JKG, SIEREVELT IN, VAN HOVE RP, VAN DEN BEKEROM MPJ, VAN NOORT A: 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(10): 1588-93. https://doi.org/10.1016/j.jse.2015.02.024
2. FRIEDMAN MS: Calcified tendinitis of the shoulder. *Am J Surg* 1957; 94(1): 56-61. https://doi.org/10.1016/0002-9610(57)90618-9
3. HARMON PH: Methods and results in the treatment of 2,580 painful shoulders: With special reference to calcific tendinitis and the frozen shoulder. *Am J Surg* 1958; 95(4): 527-44. https://doi.org/10.1016/0002-9610(58)90436-7
4. FARIN PU, JAROMA H: Sonographic findings of rotator cuff calcifications. *J Ultrasound Med* 1995;14(1): 7-14. https://doi.org/10.7863/jum.1995.14.1.7
5. DARRIEUTORT-LAFFITE C, BLANCHARD F, LE GOFF B: Calcific tendonitis of the rotator cuff: From formation to resorption. *Joint Bone Spine* 2018; 85(6): 687-92. https://doi.org/10.1016/j.jbspin.2017.10.004
6. UHTHOFF HK, SARKAR K, MAYNARD JA: Calcifying tendinitis: a new concept of its pathogenesis. Clin *Orthop Relat Res* 1976; 118: 164-8.
7. MOLÉ 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(7): 532-41.
8. GÄRTNER J, HEYER A: Calcific tendinitis of the shoulder. *Orthopade* 1995; 24(3): 284-302.
9. MAIER M, SCHMIDT-RAMSIN J, GLASER C, KUNZ A, KÜCHENHOFF H, TISCHER T: Intra- and interobserver reliability of classification scores in calcific tendinitis using plain radiographs and CT scans. *Acta Orthop Belg* 2008; 74(5): 590-5.
10. DE WITTE PB, VAN ADRICHEM RA, SELTEN JW, NAGELS J, REIJNIERSE M, NELISSEN RGHH: Radiological and clinical predictors of long-term outcome in rotator cuff calcific tendinitis. *Eur Radiol* 2016; 26(10): 3401-11. https://doi.org/10.1007/s00330-016-4224-7
11. LOUWERENS JK, CLAESSEN FM, SIEREVELT IN, EYGENDAAL D, VAN NOORT A, VAN DEN BEKEROM MP: Radiographic assessment of calcifying tendinitis of the rotator cuff : an inter- and intraobserver study. *Acta Orthop Belg* 2020; 86(3): 525-31.
12. DE WITTE PB, SELTEN JW, NAVAS A *et al*.: Calcific tendinitis of the rotator cuff: a randomized controlled trial of ultrasound-guided needling and lavage versus subacromial corticosteroids. *Am J Sports Med* 2013; 41(7): 1665-73. https://doi.org/10.1177/0363546513487066
13. VERSTRAELEN F, SCHOTANUS M, KLEMANN-HARINGS S, LAMBERS HEERSPINK O, JANSEN E: Comparison of clinical and radiological outcomes after three different surgical treatments for resistant calcifying tendinitis of the shoulder: a short-term randomized controlled trial. *J Orthop Surg Res* 2022; 17(1): 480. https://doi.org/10.1186/s13018-022-03373-1
14. OGON P, SUEDKAMP NP, JAEGER M, IZADPANAH K, KOESTLER W, MAIER D: Prognostic factors in nonoperative therapy for chronic symptomatic calcific tendinitis of the shoulder. *Arthritis Rheum* 2009; 60(10): 2978-84. https://doi.org/10.1002/art.24845

15. OUDELAAR BW, HUIS IN 'T VELD R, OOMS EM, SCHEPERS-BOK R, NELISSEN RGHH, VOCHTELOO AJH: Efficacy of adjuvant application of platelet-rich plasma after needle aspiration of calcific deposits for the treatment of rotator cuff calcific tendinitis: a double-blinded, randomized controlled trial with 2-year follow-up. *Am J Sports Med* 2021; 49(4): 873-82. https://doi.org/10.1177/0363546520987579
16. LOUWERENS JKG, SIEREVELT IN, KRAMER ET *et al*.: Comparing ultrasound-guided needling combined with a subacromial corticosteroid injection versus high-energy extracorporeal shockwave therapy for calcific tendinitis of the rotator cuff: a randomized controlled trial. *Arthroscopy* 2020; 36(7): 1823-33.e1. https://doi.org/10.1016/j.arthro.2020.02.027
17. REBUZZI E, COLETTI N, SCHIAVETTI S, GIUSTO F: Arthroscopy surgery versus shock wave therapy for chronic calcifying tendinitis of the shoulder. J *Orthop Traumatol* 2008; 9(4): 179-85. https://doi.org/10.1007/s10195-008-0024-4
18. CHO NS, LEE BG, RHEE YG: Radiologic course of the calcific deposits in calcific tendinitis of the shoulder: does the initial radiologic aspect affect the final results? *J Shoulder Elbow Surg* 2010; 19(2): 267-72. https://doi.org/10.1016/j.jse.2009.07.008
19. COMPAGNONI R, MENON A, RADAELLI S *et al*.: Long-term evolution of calcific tendinitis of the rotator cuff: clinical and radiological evaluation 10 years after diagnosis. *J Orthop Traumatol* 2021; 22(1): 42. https://doi.org/10.1186/s10195-021-00604-9
20. DE WITTE PB, KOLK A, OVERES F, NELISSEN RGHH, REIJNIERSE M: Rotator cuff calcific tendinitis: ultrasound-guided needling and lavage versus subacromial corticosteroids: five-year outcomes of a randomized controlled trial. *Am J Sports Med* 2017; 45(14): 3305-14. https://doi.org/10.1177/0363546517721686
21. DUMOULIN N, CORMIER G, VARIN S *et al*.: Factors associated with clinical improvement and the disappearance of calcifications after ultrasound-guided percutaneous lavage of rotator cuff calcific tendinopathy: a post hoc analysis of a randomized controlled trial. *Am J Sports Med* 2021; 49(4): 883-91. https://doi.org/10.1177/0363546521992359
22. DARRIEUTORT-LAFFITE C, VARIN S, COIFFIER G *et al*.: Are corticosteroid injections needed after needling and lavage of calcific tendinitis? Randomised, double-blind, non-inferiority trial. *Ann Rheum Dis* 2019; 78(6): 837-43. https://doi.org/10.1136/annrheumdis-2018-214971
23. ANGST F, SCHWYZER HK, AESCHLIMANN A, SIMMEN BR, GOLDHAHN J: Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis Care Res* (Hoboken) 2011; 63 Suppl 11: S174-88. https://doi.org/10.1002/acr.20630
24. SIM J, WRIGHT CC: The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005; 85(3): 257-68.
25. AMMITZBØLL-DANIELSEN M, ØSTERGAARD M, NAREDO E, TERSLEV L: Validity and sensitivity to change of the semi-quantitative OMERACT ultrasound scoring system for tenosynovitis in patients with rheumatoid arthritis. *Rheumatology* (Oxford) 2016; 55(12): 2156-66. https://doi.org/10.1093/rheumatology/kew317
26. GOSSEC L, JORDAN JM, LAM MA *et al*.: Comparative evaluation of three semi-quantitative radiographic grading techniques for hip osteoarthritis in terms of validity and reproducibility in 1404 radiographs: report of the OARSI-OMERACT Task Force. *Osteoarthritis Cartilage* 2009; 17(2): 182-7. https://doi.org/10.1016/j.joca.2008.06.009
27. TAN YK, ALLEN JC, LYE WK, CONAGHAN PG, CHEW LC, THUMBOO J: Dichotomous versus semi-quantitative scoring of ultrasound joint inflammation in rheumatoid arthritis using novel individualized joint selection methods. *Clin Rheumatol* 2017; 36(5): 1137-41. https://doi.org/10.1007/s10067-016-3432-5
28. LANDIS JR, KOCH GG: The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1): 159-74.
29. MAIER M, MAIER-BOSSE T, ECKERMANN M *et al*.: The roentgenmorphologic aspects of symptomatic calcifications in patients with calcifying tendinitis of the shoulder: determination of intra- and interobserver variabilities of Gärtner's classification. *Unfallchirurg* 2003; 106(3): 185-9. https://doi.org/10.1007/s00113-002-0548-7
30. COHEN J: Statistical Power Analysis for the Behavioral Sciences (2nd ed.). New York: Routledge; 1988.
31. BEXKENS R, CLAESSEN FM, KODDE IF, OH LS, EYGENDAAL D, VAN DEN BEKEROM MP: The kappa paradox. *Shoulder Elbow* 2018; 10(4): 308. https://doi.org/10.1177/1758573218791813
32. BRUINSMA WE, GUITTON TG, WARNER JJP, RING D, SCIENCE OF VARIATION GROUP: Interobserver reliability of classification and characterization of proximal humeral fractures: a comparison of two and three-dimensional CT. *J Bone Joint Surg Am* 2013; 95(17): 1600-4. https://doi.org/10.2106/jbjs.l.00586