J Shoulder Elbow Surg (2016) ■■, ■■–■■

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

ORIGINAL ARTICLE

Reliability testing of the Larsen and Sharp classifications for rheumatoid arthritis of the elbow

Nicholas B. Jew, MD[a], Anthony M. Hollins, MD[b], Benjamin M. Mauck, MD[a], Richard A. Smith, PhD[a], Frederick M. Azar, MD[a], Robert H. Miller III, MD[a], Thomas W. Throckmorton, MD[a,*]

[a]*University of Tennessee-Campbell Clinic, Department of Orthopaedic Surgery and Biomedical Engineering, Memphis, TN, USA*
[b]*Memphis Shoulder & Orthopedic Surgery, Memphis, TN, USA*

Background: Two popular systems for classifying rheumatoid arthritis affecting the elbow are the Larsen and Sharp schemes. To our knowledge, no study has investigated the reliability of these 2 systems. We compared the intraobserver and interobserver agreement of the 2 systems to determine whether one is more reliable than the other.

Methods: The radiographs of 45 patients diagnosed with rheumatoid arthritis affecting the elbow were evaluated. Anteroposterior and lateral radiographs were deidentified and distributed to 6 evaluators (4 fellowship-trained upper extremity surgeons and 2 orthopedic trainees). Each evaluator graded all 45 radiographs according to the Larsen and Sharp scoring methods on 2 occasions, at least 2 weeks apart.

Results: Overall intraobserver reliability was 0.93 (95% confidence interval [CI], 0.90-0.95) for the Larsen system and 0.92 (95% CI, 0.86-0.96) for the Sharp classification, both indicating substantial agreement. Overall interobserver reliability was 0.70 (95% CI, 0.60-0.80) for the Larsen classification and 0.68 (95% CI, 0.54-0.81) for the Sharp system, both indicating good agreement. There were no significant differences in the intraobserver or interobserver reliability of the systems overall and no significant differences in reliability between attending surgeons and trainees for either classification system.

Conclusion: The Larsen and Sharp systems both show substantial intraobserver reliability and good interobserver agreement for the radiographic classification of rheumatoid arthritis affecting the elbow. Differences in training level did not result in substantial variances in reliability for either system. We conclude that both systems can be reliably used to evaluate rheumatoid arthritis of the elbow by observers of varying training levels.

Level of evidence: Basic Science Study; Validation of Classification System

Keywords: Elbow; rheumatoid arthritis; classification; radiographic; reliability; severity; progression

Approximately 50% of patients with rheumatoid arthritis eventually develop involvement of the elbow, which can have a significant effect on the function of the upper extremity.[4,9,10] Plain radiographs are an important part of the

The University of Tennessee Health Science Center Institutional Review Board approved this study (# 14-03143-XP).

*Reprint requests: Thomas W. Throckmorton, MD, 1211 Union Ave, Ste 510, Memphis TN 38104, USA.

E-mail address: tthrockmorton@campbellclinic.com

1058-2746/$ - see front matter © 2016 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.
http://dx.doi.org/10.1016/j.jse.2016.07.074

initial evaluation of these patients, as well as a means by which to monitor the progression of the disease over time.[14,19] They also are used to monitor the effectiveness of therapy and are an important part of preoperative planning.[6,17,18]

Common radiographic findings in joints affected by rheumatoid arthritis include periarticular osteopenia, erosions, joint space narrowing, soft-tissue swelling, osteoporosis, subluxation and malalignment, ankylosis, and osteophyte formation. Because erosions and joint space narrowing generally are agreed to be the 2 most important findings, they form the basis for most scoring systems.[3,19] Numerous radiographic classification methods have been developed to grade the severity of joint involvement. Among these, the Larsen[8] and Sharp[14,15] methods are most commonly used. The Sharp method evaluates erosions and joint space narrowing separately, whereas the Larsen scheme grades the global aspect of destruction and includes erosions and joint space narrowing in a single score.[8,15,16,19] Both systems have good interobserver and intraobserver reliability in detecting overall severity and disease progression of rheumatoid arthritis.[2,7,12,13]

Although these systems are commonly used, there are scant data to suggest that one is more reliable than the other for radiographic assessment of the rheumatoid elbow. We proposed to evaluate the intraobserver and interobserver reliability of the Larsen and Sharp classification systems for rheumatoid arthritis of the elbow and to determine whether the level of training had an effect on the reliability. We hypothesized that both systems would have acceptable reliability in grading rheumatoid arthritis of the elbow and that higher levels of training would have a positive effect on reliability.

Materials and methods

We identified all patients in our electronic medical record with a diagnosis of rheumatoid arthritis affecting the elbow. Patients with a diagnosis of fracture or those with previous surgical intervention involving implants were excluded, leaving 45 elbows for evaluation. Anteroposterior and lateral radiographs of the affected elbow were obtained using a standardized institutional protocol at the time of initial evaluation. All images were deidentified and randomized into a single file. The images were then independently reviewed by 6 evaluators (4 attending orthopedic surgeons with fellowship training in elbow disorders and 2 orthopedic trainees). The methods of scoring as developed by Larsen[8] (Table I) and Sharp[14,15] (Table II) were included at the beginning of the file to serve as a reference for the evaluators. Each evaluator then scored the radiographs according to both methods on 2 separate occasions at least 2 weeks apart.

Statistical analysis was performed using AgreeStat 2013.3 software (Advanced Analytics, LLC. Gaithersburg, MD, USA) and SPSS 22 software (IBM Corp., Armonk, NY, USA). Intraobserver reliability was calculated using Spearman rank correlation coefficients, and interobserver reliability was calculated using the weighted Conger κ. The 95% confidence intervals (CI) were calculated for intraobserver and interobserver agreement as well. Correlation coefficients and κ scores >0.8 were considered to indicate substantial agreement: 0.6 to 0.8, good agreement; 0.4 to 0.6, moderate agreement; and <0.4, fair agreement. Two-tailed *t* tests were used to evaluate average reliability figures between scoring systems and between attending surgeons and trainees. Statistical significance was set at $P < .05$.

Results

Overall average intraobserver reliability was 0.93 (95% CI, 0.90-0.95) for the Larsen system and 0.92 (95% CI, 0.86-0.96) for the Sharp classification (Table III), both indicating substantial agreement. When attending surgeons and trainees were compared, the average Larsen intraobserver reliability was 0.93 for staff and 0.94 for trainees, and the average Sharp intraobserver reliability was 0.92 for staff and 0.95 for trainees. There were no statistically significant differences in intraobserver reliability between the systems overall or based on training level ($P > .05$).

Table I Larsen classification[8]

- Grade 0: Intact bony outlines and normal joint space
- Grade 1: Slight abnormality with periarticular soft-tissue swelling, periarticular osteoporosis, or slight joint space narrowing
- Grade 2: Definite abnormality. Erosion is obligatory.
- Grade 3: Medium destructive abnormality. Erosion is obligatory.
- Grade 4: Severe abnormality where there is usually no joint space left, and the original bony outlines are partly preserved
- Grade 5: Mutilating changes, where the original articular surfaces have disappeared

Table II Sharp classification[14]

Joint Space Narrowing Score:	Erosion Score:
• 0 = normal	• 0 = normal
• 1 = focal joint narrowing	• 1 = discrete erosions
• 2 = narrowing less than 50% of original joint space	• 2 to 3 = larger erosions according to surface area involved
• 3 = narrowing of more than 50% of original joint space	• 4 = erosions extending over middle of the bone
• 4 = amyloses	• 5 = complete collapse

Total score = Joint Space Narrowing Score + Erosion Score.