Inter-rater reliability of radiographic scoring results in rheumatoid arthritis patients

Poster No.:	C-1497
Congress:	ECR 2013
Type:	Scientific Exhibit
Authors:	G. Spirov, E. Jordanoska, L. Damjanovska, V. Vasilevska; Skopje/ US
Keywords:	Musculoskeletal system, Musculoskeletal joint, Conventional radiography, Observer performance, Inflammation
DOI:	10.1594/ecr2013/C-1497

Purpose

Rheumatoid arthritis (RA) is a chronic and progressive systemic autoimmune inflammatory disease the cause of which remains unknown. The disease may present at any age, but the peak onset is in the forth to the sixth decade affecting women more commonly, 3:1. Synovium is the primary site of the pathologic process. In the course of the disease adjacent structures such as the bone, tendons, capsule and ligaments typically are involved. The small joints of the hands and feet are among the first to be affected and the larger joints are caught up later, too.

The conventional radiography is widely accepted to evaluate the disease as a result of the low cost, high availability, possibility of standardization and blinded centralized reading, reasonable reproducibility and existence of validated assessment methods.

Evaluating structural joint damage using the scored radiographs method as an outcome measure helps to evaluate the severity of RA. The van der Heijde modification of Sharp technique provides separate scores for erosion and for joint space narrowing (JSN). The value of any scoring radiographs method depends on its inter-rater reliability or the degree of agreement among raters. It is a well known fact that inter-rater reliability is being strengthened by establishing clear guidelines and through experience. The inter-rater reliability in scoring same radiographs by double blind reading is estimated in this paper to help to understand the need of experienced raters in evaluating structural joint damage using the guidelines of van der Heijde modification of Sharp's grading system.

Methods and Materials

29 patients (23 female, 6 male) fulfilling the classification criteria for RA according to American College of Rheumatology that have been randomly chosen took part in the study. Single films of each hand and foot were performed by standardized positioning. Radiographs were made in posteroanterior view of both hands and anteroposterior view of both feet. The x-ray beam was centred on the third metacarpal bone in the hand or metatarzal-phalangeal joint in the foot.

Totally 116 radiographs, four radiographs (two hands and two feet) of each of the 29 patients were scored blindly by two readers, both of them radiology residents in their second and third year without any previous experience in radiographic assessment of RA.

1) Erosion score for hands and feet:

The erosion score per joint of the hands can range from 0 to 5. Erosions are scored 1 if they are discrete but clearly present. A score of 3 is given if the erosion is large and extends over the imaginary middle of the bone. A score of 5 is given if a complete collapse of the joint is present or if the full surface of the joint is affected. Score 2 and 4 are in between. In each joint, individual erosions are summed up to a maximum of 5. The maximal erosion score for each hand is thus 80, considering the 16 reviewed areas for erosions per hand.

The erosion score of the feet is different according the number of the evaluated joints (only six joints). Each joint can range from 0 to 10, because each side of the joint is independently scored from 0 to 5 (total score per joint is 10). The maximal erosion score per foot is 60.

Maximal total erosion score (hands and feet) is 280 (Figure 1).

2) Joint space narrowing and joint subluxation or luxation score for hands and feet:

Joint space narrowing and joint subluxation or luxation are combined in a single score with a range of 0 to 4. A normal joint space is scored 0. The score 1 is used when there is a suspicion of joint space narrowing. A generalized narrowing leaving more than 50% of the original joint space present or focal narrowing of the joint is scored 2. A generalized narrowing leaving less than 50% of the original joint space present, or subluxation is scored 3. A bony ankylosis or a complete luxation of the joint is scored 4. Fifteen joints of the hand are evaluated thus the maximal narrowing/(sub)luxation score for each hand is 60.

The same score criteria are applicable for the feet and only six joints are evaluated with maximum score of 24.

Maximal total narrowing/(sub)luxation score (hands and feet) is 168 (Figure 2).

Maximal total Sharp/van der Heijde score is 448.

The absolute scores were compared using ***the percent of agreement between the raters*** as a crude measure. We defined the inter-rater reliability agreement as

- *matching* - when less than 10% variation of absolute score was found,
- *acceptable differences* - when variation between 10 and 30% of absolute score was calculated, and
- *unacceptable differences* - when over 30% variation of absolute score was figured out.

Images for this section:

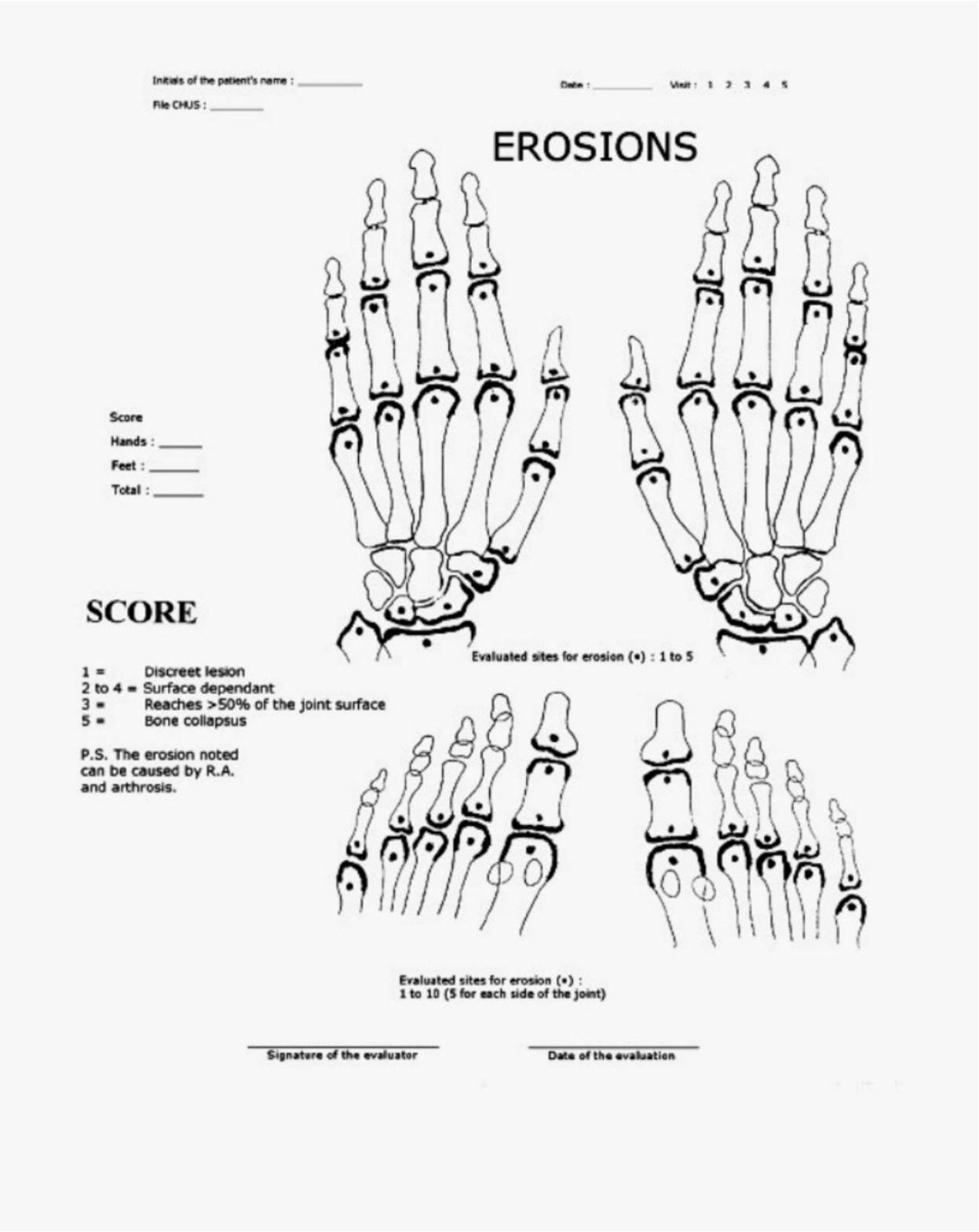

Fig. 1: Scoring sheet for erosion score of hands and feet

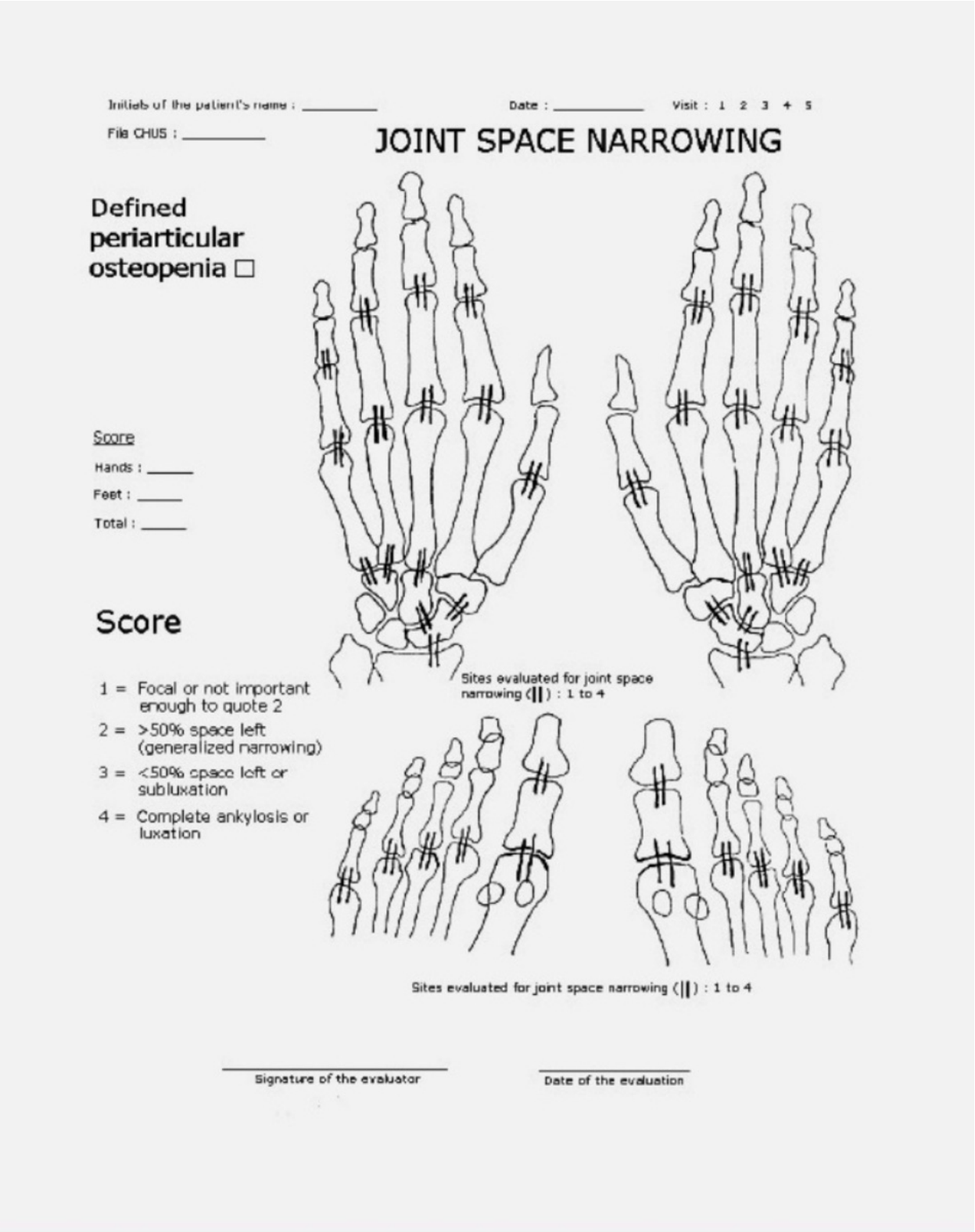

Fig. 2: Scoring sheet for joint space narrowing and joint subluxation or luxation score of hands and feet

Results

The acquired scoring results (total 58) were assessed by analyzing the total score for erosion (29 cases) and JSN (29 cases) separately.

- *matching* - only 20.7% or 12 total scores were accounted as matching results (Figure 3 and 4).
- *acceptable differences* - roughly 27.6% or 16 total scores were in a wider range assumed as acceptable differences in results.
- *unacceptable differences* - over 50% (51.7%) or 30 total scores had unacceptably high differences (Figure 5 and 6).

Standard statistical evaluation:

The single measure intraclass correlation coefficient (evaluates the reliability of individual) for single rater is 0.57 which is defined as moderate (0.40-.59). Cronbach's # coefficient (coefficient of internal consistency) is 0.764, defined as acceptable.

Images for this section:

Fig. 3: An example of matching scores for joint space narrowing

Fig. 4: An example of matching scores for joint space narrowing

Fig. 5: An example of scores with unacceptable differences for erosions

Fig. 6: An example of scores with unacceptable differences for erosions

Conclusion

Evaluating radiographs of RA patients using the same method by untrained raters produces great differences in the final score for each patient separately for erosion and for JSN as well. The significant difference in absolute scores most obviously results from the reader's insufficient reading experience level despite the very clearly defined scoring method.

References

1. van der Heijde D, Dankert T, Nieman F, Rau R, Boers M. Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis. Rheumatology 1999; 38:941-947.

2. Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. Ann Rheum Dis 2001; 60:817-827.

3. Sokka T. Radiographic scoring in rheumatoid arthritis: a short introduction to the methods. Bulletin of the NYU Hospital for Joint Diseases 2008; 66(2):166-8.

4. Østergaard M, Pedersen SJ, Døhn UM. Imaging in rheumatoid arthritis - status and recent advances for magnetic resonance imaging, ultrasonography, computed tomography and conventional radiography. Best Pract Res Clin Rheumatol. 2008 Dec; 22(6):1019-44.

5. Pope TL, et al.: Imaging of the musculoskeletal system. Philadelphia, Saunders Elsevier, 2008.

6. Sharp van der Heijde Score. University of Sherbrooke, Faculty of medicine and health sciences, Department of medicine, Division of rheumatology: **http:// rheumatology.usherbrooke.ca/?q=scoresharp** [Accessed 20.06.2012]

Personal Information