

## Scoring Methods



Quantification of radiographically detectable joint destruction is a prerequisite to measure damage progression in rheumatoid arthritis (RA). With the use of aggressive treatment and the availability of more potent drugs exact measurement becomes even more important. The best method to quantify damage, measuring erosion volume<sup>1</sup>, is still impossible from 2 dimensional radiographs, although computerized methods are being developed<sup>2</sup>.

Several scoring methods currently in use assess joint destruction in a semiquantitative way. All are derived from two original systems that have been frequently modified: the Larsen method<sup>3</sup> grades the global aspect of joint destruction from 0 to 5 with emphasis on (erosive) bone destruction. In contrast, the Sharp system<sup>4</sup> consists of an erosion score as an estimate of bone destruction and a separate joint space narrowing (JSN) score as a surrogate for cartilage destruction.

Erosions are scored in very different ways: the Sharp method counts the number of erosions from 0 to 5 irrespective of their size. Van der Heijde<sup>5</sup> also counts erosions but additionally takes into account their size in relation to joint surface. Larsen<sup>3</sup> and Genant<sup>6</sup> define grades less precisely using general terms (small to large erosions, questionable to severe changes, respectively) and standard reference films. In the Rau method (Ratingen Score)<sup>7</sup> a score increase of 1 represents 20% joint surface destruction (1–20%, 21–40%, 41–60%, etc.). Although extension into the bone is not measured, a correlation with erosion volume can be assumed. This system therefore is linear, clearly defined, and easy to learn. In practice it is often much easier to estimate how much of the cortical plate is destroyed or how much is preserved than to distinguish the number of erosions, which often merge into each other.

Unique for the Larsen method is the definition of grade 1 as soft tissue swelling and juxtaarticular osteoporosis. Both are very difficult to identify reliably on radiographs, decreasing agreement between raters. They indicate disease activity rather than joint destruction. Improvement of these features interferes with progression of destruction, thereby impairing the sensitivity to change.

RA affects both bone and cartilage. Measuring cartilage

destruction would give important additional information, especially as cartilage may be affected earlier than bone. Moreover, different treatments may have different protective effects on bone or cartilage. Since cartilage is not visible on radiographs, narrowing of the joint space is assumed to indicate cartilage destruction. Measuring the “real” joint space representing cartilage may be impaired by incorrect positioning due to swelling, contracture, or laxity of the capsule.

Other reasons for false projection are subluxation and luxation, which are not scored by Sharp<sup>4</sup>. Van der Heijde<sup>5</sup> includes both changes and assigns them the highest grades of JSN (3 + 4), thereby mixing in the same score cartilage destruction and malalignment. The score may even be dominated by malalignment, since luxation of all metatarsophalangeal joints can develop due to mechanical factors without any cartilage destruction. In order to distinguish it from other characteristics a separate malalignment score has been proposed<sup>8</sup>. Unfortunately, in his modification of Larsen’s method, Scott<sup>9</sup> also mixed different pathologies when defining grade 4 as “severe destructive abnormality or subluxation.”

The power of JSN to discriminate between more or less effective drugs has been questioned. Several recent studies described significant differences in the erosion score but not in JSN<sup>10,11</sup>. Further, several studies failed to demonstrate significant advantages of more detailed methods, with separate evaluation of JSN compared to global assessment of joints relying mostly on bone destruction<sup>12,13</sup>. This also raises the question of whether the double time expense for reading JSN separately is reasonable.

In this issue of *The Journal*, Tanaka, *et al*<sup>14</sup> compare Larsen’s with the Rau method (Ratingen Score) in patients with a symptom duration of only 3.5 months fulfilling the American College of Rheumatology criteria for RA during followup. Table 1 of the article indicates that the Larsen score increased to 12% of the maximum possible score during the first 2 years compared to only 6% of the Rau score. In this very early stage of the disease the number of swollen joints increases, scored 1 by Larsen but ignored by

---

See Comparison of the Rau and the Larsen methods in the evaluation of radiographic progression in early RA, page 682

---

Rau. Therefore the increase in the Larsen Score in this phase of the disease reflects enhanced disease activity more than damage. Moreover, every new eroded joint is scored 2 in the Larsen and 1 in the Ratingen score. Also, one small erosion in the wrist (multiplied by 5) adds 10 points to the Larsen score compared to 1 point in the Ratingen score. During the next 4 years, the yearly progression remained constant with the Rau method and decreased to 50% with the Larsen method, indicating less sensitivity to change of the latter in established disease. This is confirmed by the results of the standardized response means<sup>14</sup>. Another comparison of the two methods<sup>15</sup> [Wassenberg S, Herborn G, Sharp JT, van der Heijde D, Larsen A, Rau R. Comparison of 4 different scoring methods in the evaluation of radiographic progression in RA (in preparation)] in patients with early disease entering a clinical trial demonstrated a higher baseline score, less inter-rater agreement, and less sensitivity to change with Larsen's method.

The linearity of progression in the Ratingen Score is likely to be caused by the equal intervals between grades (20% each). Other scoring methods fail to meet this prerequisite for ordinal scales. For example, in the Larsen system one small erosion is graded 2, already representing 40% of the maximum score; in the Sharp system 4 small erosions (grade 4) count for 80% of the maximum score, while one large erosion standing for the same amount of damage is only graded 1 (20%). Therefore, van der Heijde considers the size of the erosions in her modification. Nevertheless, both methods reach the maximum possible grade when only 50% of the joint surface is destroyed, inducing a substantial ceiling effect. The overestimation of early changes and the ceiling effect contribute to the often cited hypothesis that progression is greater in early than in later disease.

Knowledge about the chronological sequence of the films is an important factor influencing the results of radiographic evaluation. Based on the assumption that radiographic damage is irreversible, traditional scoring methods do not allow a reduction of the score indicating improvement (4–6). An erosion is continuously counted during followup, even if it is no longer visible. Since the reader tends to score more deterioration with known sequence, this way of reading is more sensitive to change<sup>15</sup>, but overrates progression and neglects improvement. Allowing change and variation in only one direction is scientifically questionable.

To exclude the prejudice of steady progression, radiographs were read with unknown time sequence in recent drug trials, e.g., of biologic therapies. Surprisingly, a score reduction was found in many patients, resulting in very little or even negative median progression. If the time sequence is not known, readers tend to score much more conservatively (“no change”) to avoid mistakes. This is mainly true in advanced disease, where minor changes in severely damaged joints are difficult to detect. The favorable results

of these trials, therefore, cannot be compared with the results of older trials reporting greater progression. The difference may mostly be based on methodological differences. The same is true for the doctrine of a steadily progressive longterm course of RA.

The score reduction observed in recent studies may reflect the uncertainty of the reader. However, real improvement (“healing,” “repair”) has been observed<sup>16</sup>, and can be recognized by reading radiographs in unknown sequence<sup>17</sup>. These phenomena can be documented separately<sup>17</sup>, but agreement should be reached on whether and how joints with improvement should be scored in the “regular” score: should the score be reduced in the case of recortication, of partial filling in, or only with complete restoration? During longterm observation of successfully treated patients the number of joints with “active” erosions decreased, while the number of joints with (inactive) “secondary” osteoarthritis increased<sup>18</sup>. At present, both are scored equally. A scoring system acknowledging substantial difference between active arthritis and its consequence would fundamentally change our view on RA as an irreversibly progressive disease.

The case mix in clinical trials is another crucial aspect of radiographic evaluation. The results of many trials are flawed because many patients with low potential for progression were included. If only patients with really active, seropositive, already erosive disease were selected, trials evaluating drug effects on progression would provide stronger results with fewer patients. It has to be stated that radiographic evaluation is easier, less time consuming, and more reliable in early compared to advanced disease.

Reporting of studies is also relevant. Publications should report not only mean or median values but also the percentage of patients showing real progression greater than the “smallest detectable change” (MDC)<sup>19</sup>. Since the MDC very much depends on the case mix and the quality of radiographs, this should be established for every individual study. Under this condition the advantage of multiple readers is questionable. Different readers scoring different films may reduce the significance of study results.

Finally, the most important factor in the evaluation of radiographs remains the reader. His or her concentration, attention, experience, and consistency will probably influence the quality of the results more than the method applied.

**ROLF RAU**, MD, PhD,  
Professor of Medicine,  
E-mail: rrau@uni-duesseldorf.de  
**SIEGFRIED WASENBERG**, MD,  
Department of Rheumatology,  
Evangelisches Fachkrankenhaus,  
Rosenstrasse 2, D-40882 Ratingen, Germany.

*Address reprint requests to Professor Rau.*

## REFERENCES

1. Sharp JT. Assessment of radiographic abnormalities in rheumatoid arthritis: What have we accomplished and where should we go from here? *J Rheumatol* 1995;22:1787-91.
2. Sharp JT, Gardner JC, Bennett EM. Computer-based methods for measuring joint space and estimating erosion volume in the finger and wrist joints of patients with rheumatoid arthritis. *Arthritis Rheum* 2000;43:1378-86.
3. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. *Acta Radiol* 1977;18:481-91.
4. Sharp JT, Young DY, Bluhm GB, et al. How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? *Arthritis Rheum* 1985;28:1326-35.
5. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. *J Rheumatol* 1999;26:743-45.
6. Genant HK. Methods of assessing radiographic change in rheumatoid arthritis. *Am J Med* 1983;48:35-47.
7. Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. *J Rheumatol* 1998;25:2094-107.
8. Kaye JJ, Nance PE Jr, Callahan LF, et al. Observer variation in quantitative assessment of rheumatoid arthritis. II. A simplified scoring system. *Invest Radiol* 1987;22:41-6.
9. Scott D, Houssien D, Laasonen L. Proposed modification of Larsen's scoring method for hand and wrist radiographs. *Br J Rheumatol* 1995;34:56.
10. Bathon JM, Martin RW, Fleischmann RM, et al. A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. *N Engl J Med* 2000;343:1586-93.
11. Boers M, Verhoeven AC, Markusse HM, et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. *Lancet* 1997;350:309-18.
12. Plant MJ, Saklatvala J, Borg AA, Jones PW, Dawes PT. Measurement and prediction of radiological progression in early rheumatoid arthritis. *J Rheumatol* 1994;21:1808-13.
13. Paimela L, Laasonen L, Helve T, Leirisalo-Repo M. Comparison of the original and the modified Larsen methods and the Sharp method in scoring radiographic progression in early rheumatoid arthritis. *J Rheumatol* 1998;25:1063-6.
14. Tanaka E, Yamanaka H, Matsuda Y, et al. Comparison of the Rau method and the Larsen method in the evaluation of radiographic progression in early rheumatoid arthritis. *J Rheumatol* 2002; 29:682-7.
15. van der Heijde D, Boonen A, Boers M, Kostense P, van der Linden S. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. *Rheumatology* 1999;38:1213-20.
16. Rau R, Herborn G. Healing phenomena of erosive changes in rheumatoid arthritis patients undergoing disease-modifying antirheumatic drug therapy. *Arthritis Rheum* 1996;39:162-8.
17. Rau R, Wassenberg S, Herborn G, Perschel WT, Freitag G. Identification of radiologic healing phenomena in patients with rheumatoid arthritis. *J Rheumatol* 2001;28:2608-15.
18. Rau R, Herborn G, Karger T, Werdier D. Retardation of radiographic progression in rheumatoid arthritis with methotrexate therapy: a controlled study. *Arthritis Rheum* 1991;34:1236-44.
19. Lassere M, Boers M, van der Heijde D, et al. Smallest detectable difference in radiological progression. *J Rheumatol* 1999;26:731-9.