REPORT

Psoriatic arthritis imaging: a review of scoring methods

D van der Heijde, J Sharp, S Wassenberg, D D Gladman

Ann Rheum Dis 2005;**64**(Suppl II):ii61–ii64. doi: 10.1136/ard.2004.030809

Structural damage assessed on conventional radiographs is an important outcome measure in psoriatic arthritis. This article reviews the available scoring methods. A full description of the methods is given as well as information on various aspects of validity.

Psoriatic arthritis (PsA) is characterised by arthritis, often leading to structural damage. This structural damage can be assessed on conventional radiographs and is an important outcome measure to judge the efficacy of treatment. Structural changes in PsA follow the pattern of joint involvement. The most frequently involved joints are those in the hands and wrists, followed by the feet, ankles, knees, and shoulders. Involvement of the distal interphalangeal joints (DIPs) and an asymmetrical pattern are characteristic of PsA. Several patterns of distribution occur in the hands,[1] with involvement of:

(1) DIPs and proximal interphalangeal joints (PIPs). There is usually an asymmetrical distribution and this may occur in a ray pattern, involving all the joints of one digit as opposed to all the joints at the same level in both hands, which tends to occur in rheumatoid arthritis (RA)
(2) wrist and an isolated ray
(3) multiple joints as in RA.

Entheseal involvement is a frequent manifestation in PsA.

In general, the radiographic features can be grouped into destructive and proliferative changes. Erosions are a typical destructive feature, frequently starting at the margins and then progressing towards the centre. Typical in PsA is an erosion with accompanying bone production. Erosions may become so extensive as to give an appearance of a widened joint, rather than a narrowed joint space. Widespread erosive changes may lead to the characteristic pencil in cup phenomenon: a blunt osseous surface on the proximal bone of a joint, which may protrude into an expanded surface of the distal bone of the joint. Marked osteolysis may be observed in severely destroyed joints, such that the whole phalanx may be destroyed. Proliferative new bone formation can occur along the shaft of the metacarpal and metatarsal bones and adjacent to the joints. At the same time as some joints demonstrate osteolysis, others demonstrate total ankylosis. Thus, in the same hand, and even in the same finger, one may detect the pencil in cup change in one joint and ankylosis in an adjacent joint.

The importance of radiographic evaluation in the assessment of patients with arthritis has been highlighted by recent studies in RA. Progression of radiographic changes occurs over time in RA, and there is a relation between the level of disease activity and subsequent radiological damage.[2–4] Radiographic changes have been used as an outcome measure in several drug trials and demonstrate superiority of some drug regimens over others.[5 6] It is also well known that radiographic damage at baseline is a strong predictor for progression of structural damage in RA.[7] Similar observations of predictive validity of baseline structural damage have been made in observational cohort studies in PsA where radiological damage at baseline was found to be predictive of increased mortality.[8] Thus, radiological evaluation is essential in the assessment of patients with PsA both as an outcome measure and as a predictor of other outcomes.

SCORING METHODS

In this review we focus on the changes in peripheral joints in patients with PsA. The spine and sacroiliac joints can also be involved in patients with PsA. However, the scoring methods developed for use in ankylosing spondylitis (AS) can be applied to assess these abnormalities in PsA as there are no distinguishing features between AS and PsA in the spine and sacroiliac joints with the exception of asymmetry. Validated methods are the Bath Ankylosing Spondylitis Radiology Index (BASRI), the Stoke Ankylosing Spondylitis Spine Score (SASSS), and a modification of the SASSS.[9–11]

Several scoring methods for the assessment of structural damage in peripheral joints in PsA have been proposed. All these are based on existing scoring systems for RA and have been adapted for use in PsA. In the following sections these methods are described in detail, including what is known about the validity of these methods for use in PsA.

MODIFIED STEINBROCKER SCORING METHOD FOR PsA

At the PsA Clinic at the University of Toronto, the focus has been on longitudinal observation of patients with PsA in an attempt to describe the clinical course and prognosis of the disease. Radiological progression in the peripheral joints of these patients is assessed by a modification of the Steinbrocker technique.[12] The original Steinbrocker classification scored a patient according to their worst joint, but the modified technique scores each joint on a 0–4 scale where:

- 0 is normal
- 1 reflects juxta-articular osteopenia or soft tissue swelling
- 2 is the presence of erosion
- 3 is presence of erosion and joint space narrowing
- 4 is total joint destruction, either lysis or ankylosis.

This method has face and content validity in that it reflects the biological changes thought to occur in the arthritic joint, from soft tissue swelling to total joint destruction.

Abbreviations: DIP, distal interphalangeal (joint); DS, destruction score; ICC, intraclass correlation; IP, interphalangeal (joint); MCP, metacarpophalangeal (joint); MTP, metatarsophalangeal (joint); OMERACT, Outcome Measures in Rheumatology; PIP, proximal interphalangeal (joint); PS, proliferation score; PsA, psoriatic arthritis; RA, rheumatoid arthritis; TNF, tumour necrosis factor, TS, total score

In this method, the Toronto group led by Gladman score all the joints of the hands (with the wrist considered one joint), all metatarsophalangeal joints (MTPs), and the interphalangeal joint (IP) of the big toe. This includes a total of 28 joints in the hands and 12 joints in the feet, thus 40 joints altogether. The maximum score possible is 160, if all joints had a score of 4.

The method has been tested for interobserver and intraobserver reproducibility, as well as sensitivity to change, and compared with the Larsen method.[13] It proved to be reliable when a musculoskeletal radiologist and a rheumatologist examined repeat films from 68 patients in a blinded fashion. The interobserver intraclass correlation coefficient (ICC) for the original Steinbrocker, modified Steinbrocker, and Larsen techniques was 0.86, 0.86, and 0.87, respectively. The intraobserver ICCs for the original Steinbrocker method were 0.90 and 0.86; for the modified Steinbrocker 0.80 and 0.81; and for Larsen method 0.84 and 0.85. The original Steinbrocker method was not sensitive to change over time. However, the modified Steinbrocker and the Larsen were comparatively sensitive to change for both observers.

In a study using the modified Steinbrocker method, patients with PsA were compared with patients with RA, matched for sex, age, and disease duration. There was a similar degree of severity in the two groups.[13] Based on this radiological scoring system, radiographic damage at the first visit has been found to be a predictor for mortality in patients with PsA, adding information to the validity of the method.[8] In addition to the changes in the peripheral joints, the presence of atlantoaxial subluxation, presence of both classic and paramarginal syndesmophytes, presence of spurs at the calcaneus (both at Achilles' tendon and plantar fascia insertion), enthesitis at the pelvis area, periostitis, and tuft resorption are recorded separately. The presence of sacroiliitis is recorded according to the New York criteria.

The modification of the Steinbrocker method by Gladman is feasible for assessing radiographic changes in the peripheral joints in the office/clinic. It has not been used systematically in a randomised controlled trial, but has been used in nested case–control studies of several drugs in PsA showing that traditional disease modifying antirheumatic drugs (DMARDs) have not been able to prevent progression of joint damage in PsA.[14–17]

PsA SCORING METHOD BASED ON THE SHARP SCORING METHOD FOR RA

In preparation for the radiographic assessment included in recent trials of antitumour necrosis factor (anti-TNF) agents in PsA, two groups of rheumatologists and radiologists developed a scoring method for radiographic abnormalities based on the Sharp scoring method for RA (Ory P, unpublished observations).[18] In developing the scoring method it was agreed that erosions and joint space narrowing should be scored for the same joints as for RA.[18] The following joints in the hands are scored for erosions: the second through fifth DIPs, all five metacarpophalangeal joints (MCPs), the interphalangeal IP joint of the thumb, seven bones in the wrist including the first metacarpal base, the multangulars as a unit, the navicular, the lunate, the triquetrum and pisiform as a unit, the radius, and the ulna. In the feet, the five MTPs and the IP of the big toe are scored for erosions. For joint space narrowing the following joints are scored in the hands: the second through fifth DIPs, all five MCPs and six joints in the wrist including the fourth, fifth, and sixth carpometacarpal joints, the multangular-navicular, capitate-navicular and capitate-lunate as a unit, and the radiocarpal joints; in the feet the five MTPs are scored. In RA, the erosion scale has a range of 5 per joint. Scores are applied as:

- 0 = no erosion
- 1 = one discrete erosion or involvement of less than 21% of the joint area by erosion
- 2 = two discrete erosions or involvement of 21–40% of the joint
- 3 = three discrete erosions or involvement of 41–60% of the joint
- 4 = four discrete erosions or involvement of 61–80% of the joint
- 5 = extensive destruction involving more than 80% of the joint.

The erosion scale was expanded giving a range of 0–7 points. The scores of 6 and 7 were added to accommodate more extensive bone destruction seen in many cases of PsA. Initially the erosion score of 6 was used for pencil in cup appearance in patients with marked bone destruction, and the score of 7 was defined as gross osteolysis. The score of 5 was retained for erosion involving more than 80% of the joint but not associated with gross osteolysis or a pencil in cup lesion. In a subsequent study, 6 and 7 were defined as progressively more extensive osteolysis, and pencil in cup abnormality was scored as present or absent for all lesions given an erosion score of either 6 or 7. However, the scores of 6 and 7 are not added to get the total erosion score; these features are kept separate. The maximum possible score for erosion is 210 for hands and 60 for feet. Joint space narrowing was scored on a scale of 0–4, as for RA, but with the addition of widening (score 5), which was automatically scored when gross osteolysis was present. The joint space narrowing scores are:

- 0 = normal joint
- 1 = asymmetrical and or minimal narrowing
- 2 = definite narrowing with loss of up to 50% of the normal space
- 3 = definite narrowing with loss of 51–99% of the normal space
- 4 = absence of a joint space, presumptive evidence of ankylosis
- 5 = widening.

Again, the score for widening is not included in the total narrowing score but analysed as a separate feature. The joint space narrowing score has a maximum possible score of 160 for hands and 40 for feet.

In addition to erosions and joint space narrowing, a number of radiographic features seen in PsA were scored: shaft periostitis, juxta-articular periostitis, periostitis in the wrist, and tuft resorption. Initially shaft periostitis was scored as present or absent in the proximal and middle phalanges of the fingers. In a later study, which included foot radiographs, this feature was scored for the metatarsals and proximal phalanges of the toes, and the scale was expanded to 0–3 (absent, mild, moderate, and severe). Juxta-articular periostitis was scored in the immediate area of the joint as the count of the number of quadrants involved giving a range of 0–4. In addition, six areas in the wrist were scored for periostitis on a scale of 0–3 (none, mild, moderate, and severe): first metacarpal base and the multangulars scored as one unit; the navicular; the radius; the fifth metacarpal base and the hamate score as one unit; the triquetrum and pisiform as one unit; and the ulna. Finally, tuft resorption was scored according to the extent of bone loss by quintiles for the five fingers and the big toe. Thus 0 = none, 1 = 1–20% resorption, etc.

The scoring method has been applied in two clinical trials, of which only one has been analysed to date. In the

completed trial, which included only hand radiographs, there was a significant difference between the treatment arms, showing reduced progression of damage in patients treated with etanercept compared with patients given placebo, thus indicating sensitivity to change of this method.[19] The difference was based on the erosion and joint space narrowing score. Inclusion or exclusion of the DIPs in the scores did not change the result. Scoring of the additional features did not give extra information for differentiating between the treatment arms. Based on the limited data available, the scoring of these additional features gave no supplementary information. The interreader ICCs on status scores evaluated by four readers ranged from 0.81 to 0.88. For the annualised progression rate the interreader ICC was 0.63. The moderate interreader ICC for change scores was more a reflection of the limited amount of progression observed in the patients than differences in scores between the readers.

SHARP–VAN DER HEIJDE MODIFIED SCORING METHOD FOR PsA

This method is based on the Sharp–van der Heijde method for assessing erosions and joint space narrowing of joints of hands and feet in RA.[20] The proposed adapted scoring method for PsA is a detailed scoring method evaluating erosions, joint space narrowing, (sub)luxation, ankylosis, gross osteolysis, and pencil in cup phenomena. In addition to the joints evaluated for RA, the DIPs of the hands are assessed. Each of the following joints of the hands are scored for erosions: ten DIPs/IPs, ten MCPs, two first metacarpal bones, two radius and ulnar bones, two multangular units (trapezium and trapezoid combined); in the feet, ten MTPs and two IPs of the big toes are scored. Joint space narrowing, (sub)luxation, ankylosis, gross osteolysis and pencil in cup are assessed in the hands in ten DIPs/IPs, ten MCPs, two third, fourth, and fifth carpometacarpal joints, two multangular-navicular joints, two capitate-navicular-lunate joints, two radiocarpal joints, and in the feet in ten MTPs and two IPs of the big toes.

The maximum score for erosions is 5 in the joints of the hands and 10 in the joints of the feet. Scores for erosions are as follows

- 0 = no erosions
- 1 = discrete erosion
- 2 = large erosion not passing the mid-line
- 3 = large erosion passing the mid-line.

A combination of the above scores may lead to a maximum of 5 per entire joint in the hands, and 5 at each site of the joint (for the entire joint a maximum of 10) in the feet.

The so called joint space narrowing score is based on the following features:

- 0 = normal
- 1 = asymmetrical or minimal narrowing up to a maximum of 25%
- 2 = definite narrowing with loss of up to 50% of the normal space
- 3 = definite narrowing with loss of 50–99% of the normal space or subluxation
- 4 = absence of a joint space, presumptive evidence of ankylosis, or complete luxation.

Gross osteolysis and pencil in cup is scored separately. In the final summary score, joints with one of these abnormalities get the maximum score assigned for both erosions and for joint space narrowing. The maximum possible score for erosions is 200 for the hands and 120 for the feet; the maximum possible score for joint space narrowing is 160 for the hands and 48 for the feet. Thus, the maximum possible scores are 320 for erosions, 208 for joint space narrowing, and 528 for the total score.

The method is being applied by two readers in two placebo controlled clinical trials evaluating efficacy of anti-TNF treatment in PsA. These data will give insight on intrareader and interreader agreement, discrimination among patients with different disease status, and sensitivity to change. Also, from these trials, one may be able to deduce which joints are giving the most information and how gross osteolysis and pencil in cup should be included in the total score, thus addressing many of the aspects of the Outcome Measures in Rheumatology (OMERACT) filter.

PSORIATIC ARTHRITIS RATINGEN SCORE

Psoriatic Arthritis Ratingen Score (PARS) was developed specifically for the radiographic assessment of patients with PsA. It includes 40 joints of the hands and feet (eight DIPs, two IPs of the thumbs, eight PIPs, ten MCPs, both wrists, both IPs of the great toes, and second to fifth MTPs).[21] All joints are scored separately for destruction and proliferation. The destruction score (DS) is based on the amount of joint surface destruction on a 0–5 scale:

- 0 = normal
- 1 = one or more definite erosions with an interruption of the cortical plate of >1 mm but destruction of less than 10% of the total joint surface
- 2 = destruction of 11–25%
- 3 = destruction of 26–50%
- 4 = destruction of 51–75%
- 5 = destruction of more than 75% of joint surface.

The proliferation score (PS) considers any kind of bony proliferation typical for PsA on a 0–4 scale:

- 0 = normal
- 1 = bony proliferation measured from the original bone surface of 1–2 mm, or, if the margins of the proliferation cannot be distinguished from the original bone surface, clearly identifiable bone growth not exceeding 25% of the original diameter of the bone
- 2 = bony proliferation of 2–3 mm or bone growth between 25% and 50%
- 3 = bony proliferation >3 mm or bone growth >50%
- 4 = bony ankylosis.

The DS (0–200) and the PS (0–160) are added to give the total score (TS) (0–360) for each patient.

The method has been validated using complete sets of *x* rays of 20 patients with active PsA. Radiographs were taken at a mean time interval of three years and were read twice several weeks apart in known chronological order by two readers blinded towards patient identity. The data were analysed with a hierarchical analysis of variance model using the variance of the change over time (progression) and the variance of intrareader and interreader reliability. The more the ratio of the intrapatient SD (change) and intrareader or interreader SD (measurement error) exceeds 1, the more likely it is that the readers describe real progression.[22] The ratios of change scores to reader variation were 3.3 (reader 1), 2.0 (reader 2), and 3.8 (both readers) for the DS; 2.2, 4.2, and 2.7 for the PS; and 3.6, 2.8, and 3.9 for the TS. Agreement between the readers was 3.9 for the DS, 2.8 for the PS, and 4.1 for the TS. Thus the reliability of the method was good compared with similar data from methodological studies with established scoring systems in RA patients.[23] The same was true for the minimal detectable change (interrater

MDC)—5.8% (DS), 5.0% (PS), and 4.6% (TS) of the maximum possible scores.

A comparison of the change over time of the DS with the change of the PS revealed that there was only a weak correlation between both features, suggesting that proliferation develops independently from destruction. Measuring both features separately, therefore, adds significant information compared with measuring just one.

Applying the OMERACT filter to this method, one can state that with respect to truth the PARS instrument has proved to measure two separate features of radiographic change that have been described as the typical radiographic signs of PsA. It is reliable both in terms of intrarater and interrater reliability comparable with standard radiographic scoring methods used in RA. Scoring is easily performed and can be done as it is usually done in clinical trials. As in every trial the sensitivity to change in the given population must be determined for the readers. There is no hint that the feasibility of the method is poorer than using other detailed composite radiographic scores in RA but this has to be proved.

DISCUSSION

Further evaluation of the various methods is needed, including performance in clinical trials as well as comparisons of the proposed scoring methods. A working group has been formed to accomplish this task, and the first study on the comparison of the methods is underway. The ultimate purpose is to select the most appropriate method for evaluation in clinical trials and the most appropriate method for evaluating disease severity and in long-term (observational) studies. It may well be possible that not the same method is suitable for both purposes. For example, in clinical trials, the ability to pick up small changes or small differences among treatments over a relatively short period of time is of great importance. Feasibility and the amount of time consumed by a method are of lesser significance in such trials; these issues are of major importance in large cohort studies. In this type of study it is often also more important to be able to discriminate among patients with different disease status than to detect small changes over time. It is expected that in the near future a lot of progress will be made in this field.

Authors' affiliations

D van der Heijde, Department of Rheumatology, University Hospital Maastricht, and Caphri Research Institute, University Maastricht, Maastricht, the Netherlands
J Sharp, University of Washington School of Medicine, Seattle, WA, USA
S Wassenberg, Fachkrankenhaus Ratingen, Ratingen, Germany
D D Gladman, University of Toronto; Toronto Western Research Institute; Psoriatic Arthritis Program, University Health Network; Centre for Prognosis Studies in the Rheumatic Diseases, Toronto Western Hospital, Toronto, Ontario, Canada

Correspondence to: D van der Heijde, Department of Rheumatology, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, the Netherlands; dhe@sint.azm.nl

REFERENCES

1 **Hochberg MC**, Silman AJ, Smolen JS, Weinblatt ME, Weisman MH, eds. *Practical Rheumatology*, 3rd edn., Edinburgh: Mosby, 2003:1198–200, 1247–49.
2 **Wolfe F**, Sharp JT. Radiographic outcome of recent-onset rheumatoid arthritis: a 19-year study of radiographic progression. *Arthritis Rheum* 1998;**41**:1571–82.
3 **Welsing P**, Landewé R, van Riel PL, Boers M, van Gestel AM, van der Linden S, *et al*. The relationship between disease activity and radiological progression in patients with rheumatoid arthritis: a longitudinal analysis. *Arthritis Rheum* 2004;**50**:2082–93.
4 **Boers M**, Kostense PJ, Verhoeven AC, van der Linden S; COBRA Trial Group. COBRA Trial Group. Combinatietherapie Bij Reumatoide Artritis. Inflammation and damage in an individual joint predict further damage in that joint in patients with early rheumatoid arthritis. *Arthritis Rheum* 2001;**44**:2242–6.
5 **Lipsky PE**, van der Heijde DM, St. Clair EW, Furst DE, Breedveld FC, Kalden JR, *et al*. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. Infliximab and methotrexate in the treatment of rheumatoid arthritis. *New Engl J Med* 2000;**343**:1594–602.
6 **Klareskog L**, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, *et al*. TEMPO (Trial of Etanercept and Methotrexate with Radiographic Patient Outcomes) study investigators. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *Lancet* 2004;**363**:675–81.
7 **Guillemin F**, Gerard N, van Leeuwen M, Smedstad LM, Kvien TK, van den Heuvel W; EURIDISS Group. Prognostic factors for joint destruction in rheumatoid arthritis: a prospective longitudinal study of 318 patients. *J Rheumatol* 2003;**30**:2585–9.
8 **Gladman DD**, Farewell VT, Wong K, Husted J. Mortality studies in psoriatic arthritis: results from a single outpatient centre. II. Prognostic indicators for mortality. *Arthritis Rheum* 1998;**41**:1103–10.
9 **Pincus T**, Ferraccioli G, Sokka T, Larsen A, Rau R, Kushner I, Wolfe F. Evidence from clinical trials and long-term observational studies that disease-modifying anti-rheumatic drugs slow radiographic progression in rheumatoid arthritis: updating a 1983 review. *Rheumatology (Oxford)* 2002;**41**:1346–56.
10 **Averns HL**, Oxtoby J, Taylor HG, Jones PW, Dziedzic K, Dawes PT. Radiological outcome in ankylosing spondylitis: use of the Stoke Ankylosing Spondylitis Spine Score (SASSS). *Br J Rheumatol* 1996;**35**:373–6.
11 **Creemers M**, Franssen M, Hof Mv M, Gribnau F, Van De Putte L, Van Riel P. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. *Ann Rheum Dis* 29 March 2004 [Epub ahead of print].
12 **Rahman P**, Gladman DD, Cook RJ, Zhou Y, Young G, Salonen D. Radiological assessment in psoriatic arthritis. *Br J Rheumatol* 1998;**37**:760–65.
13 **Rahman P**, Nguyen E, Cheung C, Schentag CT, Gladman DD. Comparison of radiological severity in psoriatic arthritis and rheumatoid arthritis. *J Rheumatol* 2001;**28**:1041–4.
14 **Mader R**, Gladman DD, Long J, Gough J, Farewell VT. Does injectable gold retard radiologic evidence of joint damage in psoriatic arthritis? *Clin Invest Med* 1995;**18**:139–43.
15 **Abu-Shakra M**, Gladman DD, Thorne JC, Long J, Gough J, Farewell VT. Longterm methotrexate therapy in psoriatic arthritis: clinical and radiologic outcome. *J Rheumatol* 1995;**22**:241–5.
16 **Rahman P**, Gladman DD, Cook RJ, Zhou Y, Young G. The use of sulfasalazine in psoriatic arthritis: a clinic experience. *J Rheumatol* 1998;**25**:1957–61.
17 **Lee JC**, Gladman DD, Schentag CT, Cook RJ. The long-term use of azathioprine in patients with psoriatic arthritis. *J Clin Rheumatol* 2001;**7**:160–5.
18 **Sharp JT**, Bluhm GB, Brook A, Brower AC, Corbett M, Decker JL, *et al*. Reproducibility of multiple-observer scoring of radiologic abnormalities in the hands and wrists of patients with rheumatoid arthritis. *Arthritis Rheum* 1985;**28**:6–24.
19 **Mease PJ**, Kivitz AJ, Burch FX, Siegel EL, Cohen SB, Ory P, *et al*. Etanercept treatment of psoriatic arthritis: safety, efficacy, and effect on disease progression. *Arthritis Rheum* 2004;**50**:2264–72.
20 **van der Heijde D**. How to read radiographs according to the Sharp/van der Heijde method. *J Rheumatol* 2000;**27**:261–3.
21 **Wassenberg S**, Fischer-Kahle V, Herborn G, Rau R. A method to score radiographic change in psoriatic arthritis. *Z Rheumatol* 2001;**60**:156–66.
22 **Rückmann A**, Ehle B, Trampisch HJ. How to evaluate measuring methods in the case of non-defined external validity. *J Rheumatol* 1995;**22**:1998–2000.
23 **Wassenberg S**, Herborn G, Fischer S, Rau R, Ehle B, Trampisch HJ. Comparison of Larsen's and Sharp's method of scoring radiographs in rheumatoid arthritis [abstract]. *Arthritis Rheum* 1994;**37**:S250.