Comparison of the Genant-modified Sharp and van der Heijde-modified Sharp scoring methods for radiographic assessment in rheumatoid arthritis

Objectives: To compare the Genant-modified Sharp and van der Heijde-modified Sharp composite scoring methods for evaluating radiographs of the hands, wrists and feet in patients with rheumatoid arthritis. **Materials & methods:** Standardized radiographs of the hands, wrists and feet of 28 patients with established rheumatoid arthritis taken 1 year apart were scored independently with each of the two methods by two radiologists blinded to each other's scores. One of the radiologists also re-read each case with each method to determine intrareader agreement. The smallest detectable change (SDC), normalized to the range of scores observed with each method, and the intra- and inter-reader reproducibilities expressed as intraclass correlation coefficients, were determined for each method. **Results:** The Genant–Sharp scores and van der Heijde–Sharp scores correlated strongly with each other (r^2 ranged from 0.95 to 0.99) and demonstrated high intra- and inter-reader agreement (intraclass correlation coefficients ranged from 0.84 to 0.98). The Genant–Sharp method demonstrated smaller normalized SDCs (4.7–7.9) than did the van der Heijde–Sharp method (6.4–10.5), but the difference was not statistically significant except for the joint space narrowing score read by the single radiologist. Furthermore, the proportion of patients showing change greater than 0 or greater than SDC were not significantly different between the two methods. **Conclusion:** The Genant–Sharp and van der Heijde–Sharp methods demonstrated relatively similar performances for scoring erosion and joint space narrowing in the hands, wrists and feet.

KEYWORDS: clinical trials ▪ imaging ▪ radiography ▪ rheumatoid arthritis ▪ scoring

Charles G Peterfy†1, Chun Wu2, Jan Szechinski2, Julie C DiCarlo1, Ying Lu3, Mark Genovese4, Vibeke Strand4 & Harry K Genant2,3

1Spire Sciences, LLC, 72 Rock Road, Kentfield, CA 94904, USA
2Synarc, Inc., 575 Market Street, 17th floor, San Francisco, CA 94105, USA
3Department of Radiology, University of California San Francisco, San Francisco, CA, USA
4Department of Rheumatology, Stanford University, 1000 Welch Road, Suite 203, Palo Alto, CA 94304, USA
†Author for correspondence:
Tel.: +1 415 233 1642
Fax: +1 415 461 1508
charles.peterfy@spiresciences.com

Radiographic evaluation of disease severity and progression in clinical trials of putative structure-modifying therapies for rheumatoid arthritis (RA) currently relies on semiquantitative scoring of bone erosion and joint space narrowing (JSN). Several scoring methods have been developed; however, the most widely used is a composite score of aggregated ordinal scales originally described in 1971 by John Sharp, which separately grades bone erosion and JSN in a number of locations in the hands and wrists [1]. Modifications of this scoring method, including extension to the feet, have been described by Genant *et al.* and van der Heijde *et al.* [2–6]. The two modified methods differ primarily in the scales used to grade erosions, but there are also minor differences in their JSN scales and the locations scored in the hands and wrists. Both methods have been accepted by regulatory agencies as valid and have been used successfully in randomized controlled trials to gain regulatory approval of structure-modifying therapies [4,7–12]. However, the relative performance of the Genant–Sharp and van der Heijde–Sharp methods is not known, as they have never been compared directly in the same cohort of patients, and since their scales differ numerically, comparisons across different trials are difficult to interpret. In this study we contrast the two modified-Sharp scoring methods with respect to the locations scored in the hands/wrists and feet, and the scales used to assess erosion and JSN, and compare the cross-sectional and longitudinal performances of the two methods using a common set of serially acquired radiographs of the hands, wrists and feet from a single cohort of RA patients.

Materials & methods

Radiographs of bilateral hands, wrists and feet were acquired from 28 patients with established RA of 2–10 years duration. Subjects at baseline had elevated C-reactive protein and rheumatoid factor seropositivity within the preceding 6 months, swollen joint count of four or more, and at least one definite radiographic erosion attributable to RA. Subjects with any indication of generalized osteoarthritis or significant vasculitis were excluded. The study underwent institutional board review, in which all subjects provided informed consent.

Radiographs were acquired at baseline and 1 year using single-emulsion, fine-grain film and single-screen cassettes. A separate radiograph was acquired for each extremity using standardized positioning, x-ray beam centering, focus-to-film distance and exposure settings. Radiographs were digitized at 12-bit pixel depth (4096 shades of gray) and a pixel size of 100 µm, and transferred to a Sun workstation for reading. Images selected for this study had sufficient technical quality and anatomical coverage to allow all joints included in the Genant–Sharp and van der Heijde–Sharp scoring methods to be read. Images were also selected to provide a broad range of severity and change in erosions and JSN, including cases demonstrating mild disease and those demonstrating moderately severe changes, as judged by several experienced readers other than those used in this study.

Digitized images were read independently by two radiologists using the Genant–Sharp and the van der Heijde–Sharp composite scoring methods [2,4,5,10]. Both radiologists were specialized in and dedicated to image analysis for clinical trials. One radiologist had many years of experience using the Genant–Sharp scoring method and had read more than 1000 radiographs using the van der Heijde–Sharp method in previous clinical trials. The second radiologist had less experience with either method, but was a dedicated clinical trials radiologist with experience scoring images using other methods and had extensive training in Genant–Sharp and van der Heijde–Sharp scoring using hand, wrist and foot images from a different cohort of RA patients.

Serial images from individual patients were displayed side-by-side on a pair of 2K, high-resolution monitors with 256-shade grayscale and matrix size of 1728 × 2304 pixels, and were read simultaneously in random subject order with random shuffling of the two time points and masking of visit dates to blind the readers to chronological sequence. Scores were entered on an electronic report form that was directly linked to a central database in a manner fully compliant with US FDA regulations [101]. One of the radiologists read all of the images twice with each of the methods to allow determination of intrareader reproducibility. This was performed by initially reading the images twice using the Genant–Sharp method on two separate occasions approximately 4 months apart, and then reading the same images twice with the van der Heijde–Sharp method on two separate occasions 3 months apart. Each of these readings was carried out independently without knowledge of previous reading results or the other reader's results.

Genant-modified Sharp scoring method

Erosion score

A total of 14 locations in each hand and wrist (proximal interphalangeal joints of digits 2–5, the interphalangeal joint of the thumb and five metacarpophalangeal joints, the carpometacarpal [CMC] joint of the thumb, the scaphoid bone, the distal radius and the distal ulna); (Figure 1A) and six joints in each foot (five metatarsophalangeal joints and the interphalangeal joint of digit 1 [i.e., the big toe] (Figure 2A) were scored individually using an 8-point scale from 0 to 3.5 based on the amount of articular bone eroded:

- 0: normal
- 0.5: subtle loss of cortical continuity or equivocal findings of bone erosion
- 1.0: mild: definite but small erosions involving less than 25% of the articular bone(s)
- 1.5: mild-to-moderate: small–medium erosions involving less than 25% of the articular bone(s)
- 2.0: moderate: medium–large erosions involving approximately 26–50% of the articular bone(s)
- 2.5: moderate-to-severe: erosion of approximately 51–75% of the articular bone(s)
- 3.0: severe: erosion of approximately 76–90% of the articular bone(s)
- 3.5: very severe: erosion of 100% of the articular bone(s)

JSN score

A total of 13 locations in each wrist and hand (proximal interphalangeal joints of digits 2–5, the interphalangeal joint of the thumb, metacarpophalangeal joints 1–5, CMC 3–5 as a single unit, the pericapitate space [scaphoid-capitate and lunate-capitate combined] and the radiocarpal joint [scaphoid-radius and lunate-radius articulations]) and six locations in each foot (metatarsophalangeal joints 1–5 and the interphalangeal joint of the big toe) were scored using a 9-point scale from 0 to 4:

- 0: normal
- 0.5: subtle or equivocal narrowing
- 1.0: focal or mild (<25%) narrowing
- 1.5: mild-to-moderate (26–50%) narrowing

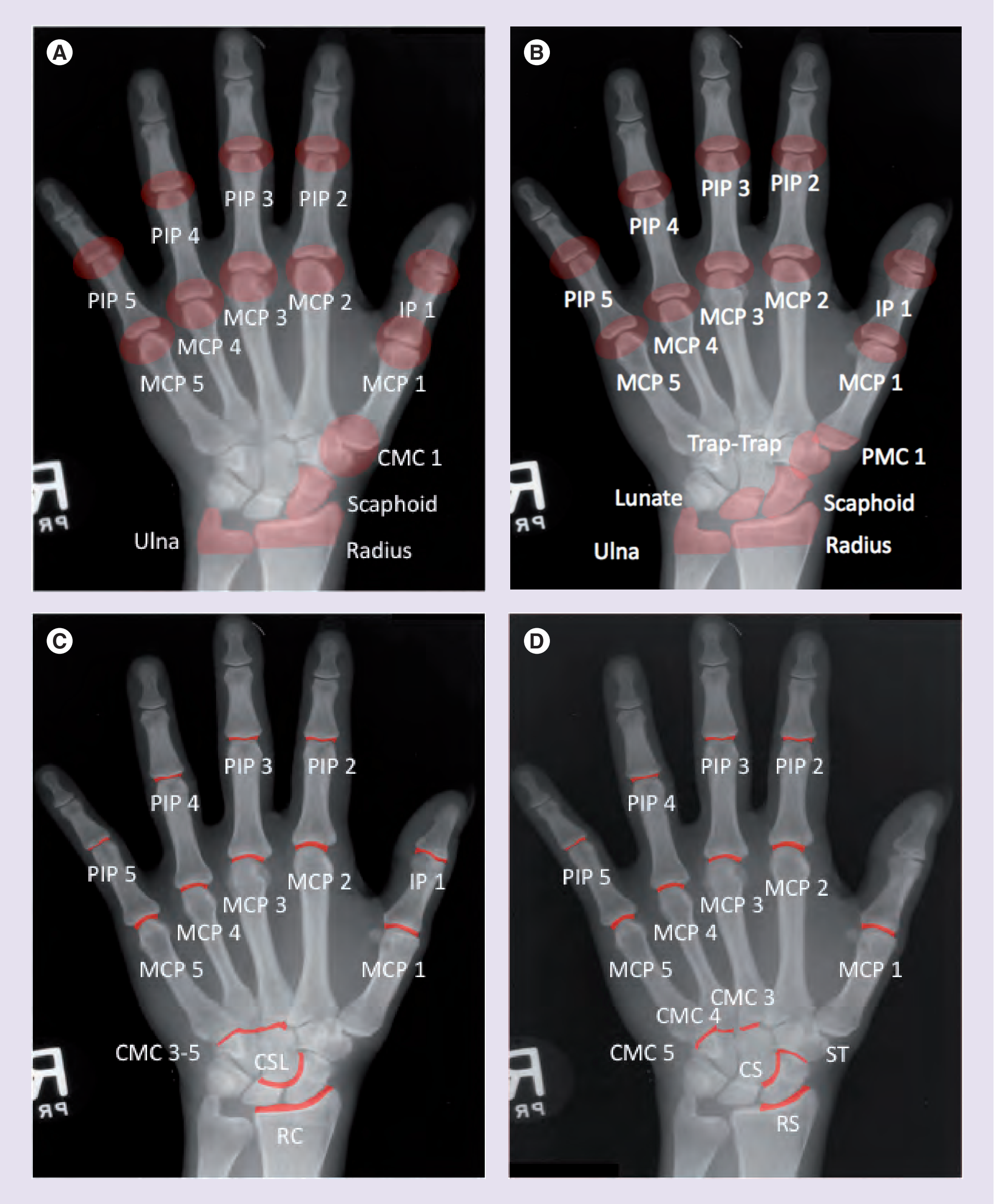

Figure 1. Locations in the hands and wrists for (A) Genant–Sharp erosion, (B) van der Heijde–Sharp erosion, (C) Genant–Sharp joint space narrowing and (D) and van der Heijde–Sharp joint space narrowing scoring. Note that the articular components of CMC-1 (PMC-1 and Trap–Trap) are scored independently for erosion with the van der Heijde–Sharp method, but are scored together with the Genant–Sharp method. The only bone scored for erosion with the van der Heijde–Sharp method but not the Genant–Sharp method is the lunate bone. The only joint scored for joint space narrowing with van der Heijde–Sharp but not Genant–Sharp method is the ST joint. However, the Genant–Sharp method includes the IP-1, scaphoid–lunate and radius–lunate joints in joint space narrowing evaluation, whereas the van der Heijde–Sharp method leaves these out.
CMC: Carpometacarpal; CS: Capitate–scaphoid; CSL: Capitate–scaphoid–lunate; IP: Interphalangeal joint; MCP: Metacarpophalangeal joint; PIP: Proximal interphalangeal joint; PMC: Proximal metacarpal; RC: Radiocarpal; RS: Radius–scaphoid; ST: Scaphoid–trapezium; Trap–Trap: Trapezium–trapezoid.

- 2.0: moderate (51–75%) narrowing or dislocation in the absence of erosions
- 2.5: moderate-to-severe (76–95%) narrowing
- 3.0: complete loss of joint space (bone on bone) or dislocation in the presence of erosion
- 3.5: partial or equivocal ankylosis
- 4.0: definite ankylosis

The individual location scores were summed to create a total erosion score (E score) and a total JSN score for each patient. E scores and

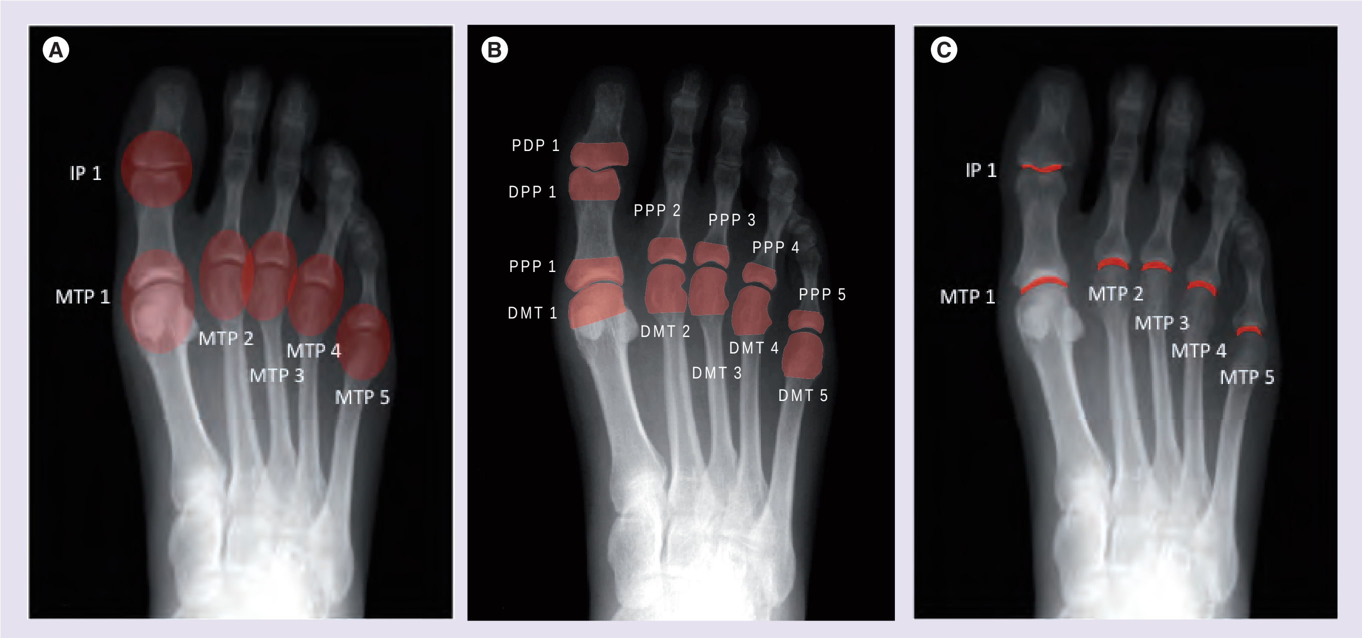

Figure 2. Locations in the feet for erosion scoring with (A) Genant–Sharp and (B) van der Heijde–Sharp and (C) for joint space narrowing scoring with both methods. Note that Genant–Sharp and van der Heijde–Sharp methods score the same bones and joint spaces, but the van der Heijde–Sharp method scores erosions in the proximal and distal articular components of the MTP and IP joints separately, whereas the Genant–Sharp method scores these joints as single units.

DMT: Distal metatarsal; DPP: Distal proximal phalanx; IP: Interphalangeal joint; MTP: Metatarsophalangeal joint; PDP: Proximal distal phalanx; PPP: Proximal proximal phalanx.

JSN scores were summed to create a combined, or total, score. The maximum total score achievable per patient was 292, or 584 increments (Table 1). Change scores (ΔE, ΔJSN, ΔTotal) were calculated by subtracting the baseline scores (E_0, JSN_0, $total_0$) from the corresponding 1-year follow-up scores (E_{1Y}, JSN_{1Y}, $total_{1Y}$).

Van der Heijde-modified Sharp scoring method

Erosion score

A total of 16 locations in each hand and wrist were scored (Figure 1B). These included the same 14 locations as in the Genant–Sharp scoring method, except that each of the two components of the CMC joint of the thumb was scored separately and the lunate bone was added. In total, 12 locations in each foot were scored (Figure 2B). These were the same six joints included in the Genant–Sharp method, except that each of the two components of these joints was scored separately rather than together as a unit.

Each location in the hand/wrist and foot was scored individually using a 6-point scale from 0 to 5 based on the number and size of discrete erosions in each location. Discrete erosions were graded 1 if small and 2 or 3 if larger. A score of 3 was given if an erosion was large and extended across the imaginary middle of the bone. Discrete erosions of each grade were summed to a maximum of 5 to give a total score for each location. When erosions in the carpal bones of the wrist were confluent rather than discrete and focal, the surface eroded was scored from 0 to 5 in approximately 20% intervals.

Table 1. Comparison of maximum increments and scores attainable per patient.

Score type	Genant–Sharp	van der Heijde–Sharp
E (hand)	196 (98)	160 (160)
E (foot)	84 (42)	120 (120)
E (total)	280 (140)	280 (280)
JSN (hand)	208 (104)	120 (120)
JSN (foot)	96 (48)	48 (48)
JSN (total)	304 (152)	168 (168)
Total (hand)	404 (202)	280 (280)
Total (foot)	180 (90)	168 (168)
Total (combined)	584 (292)	448 (448)

Values in parentheses are scores.
E: Erosion; JSN: Joint space narrowing.

JSN score

A total of 15 locations in each hand and wrist were scored. As with Genant–Sharp scoring, these included the proximal interphalangeal joints of digits 2–5, metacarpophalangeal joints 1–5, CMC joints 3–5, a pericapitate space and a radiocarpal space. By contrast to the Genant–Sharp method, the interphalangeal joint of the thumb was not included and the scaphoid–trapezium joint was added. Other changes from the Genant–Sharp method included scoring the CMC joints of digits 3–5 separately instead of as a single unit, including only the capitate–scaphoid joint rather than both the capitate–scaphoid and the capitate–lunate joints in the pericapitate space, and including only the scaphoid–radius joint rather than both the scaphoid–radius and the lunate-radius joints in the radiocarpal space. For the foot, the same six locations that were included in the Genant–Sharp method were scored using the van der Heijde–Sharp method. By contrast to the 9-point scale used in the Genant–Sharp method, however, JSN was scored using a 5-point scale from 0 to 4 in the van der Heijde–Sharp method:

- 0: normal
- 1.0: focal or minimal and generalized narrowing
- 2.0: generalized narrowing of less than 50%
- 3.0: generalized narrowing greater than 50%, or subluxation
- 4.0: ankylosis or complete dislocation

As with the Genant–Sharp method, individual location scores were summed to create a total E score and a total JSN score for each patient, and total scores were calculated by summing the E and JSN scores. The maximum total score achievable per patient was 448, or 448 increments (TABLE 1).

Analysis

Intrareader agreement for each scoring method was expressed as the intraclass correlation coefficient (ICC) for the duplicate readings. Smallest detectable change (SDC) was calculated according to the method of Bruynesteyn *et al.* [13]:

$$SDC = \frac{1.96 \times SD_{changescore}}{\sqrt{2} \times \sqrt{k}}$$

where $SD_{change\ score}$ was the standard deviation of the differences between independent readers' determinations of change scores, and k was the number of independent readers (in this case two). Since the scales used in the two scoring methods are different, direct comparison of the values or SDCs are not valid unless the scores are first normalized. Normalized SDC (nSDC) was calculated using scores normalized for the reader and cohort according to the method recommended by Sharp *et al.* [14]:

$$normalized\ score = \frac{raw\ score - minimum\ score}{minimum\ score - maximum\ score} \times 100$$

where the minimum and maximum scores are those observed for the particular reading.

Owing to the paired-sample design of the two reading methods, the explicit distributions of the two nSDCs were difficult to derive. Therefore, we used the bootstrap method [15] to calculate significance for the difference in nSDCs. Data were resampled with replacement 2000 times. The distribution of differences in nSDCs was determined and a two-sided bootstrap p-value was derived. The McNemar test was used to compare the proportion of patients with changes greater than 0 or greater than SDC. Owing to small sample size, we used the exact binomial distribution to calculate the p-value of McNemar's test. Statistical significance was defined as 5%.

Results

Readings demonstrated high intra- and inter-reader reproducibility for both methods (TABLE 2). As shown in TABLE 3 & FIGURE 3, because of the different arbitrary units used by each method, raw van der Heijde–Sharp scores were numerically larger than raw Genant–Sharp scores. However, normalized scores correlated highly both cross-sectionally and for change (FIGURE 4), with correlation coefficients ranging from 0.95 to 0.99.

Genant–Sharp scoring demonstrated somewhat smaller SDC and nSDC than did van der Heijde–Sharp scoring (TABLE 4). Except for a significantly ($p = 0.001$) lower nSDC for Genant–Sharp JSN score by the single reader and a lower nSDC for Genant–Sharp E score, which approached significance ($p = 0.08$) for the averaged reading, there were no significant differences in nSDC between van der Heijde–Sharp

Table 2. Comparison of intra- and inter-reader agreement.

Score type	Genant–Sharp		van der Heijde–Sharp	
	$ICC_{intrareader}$	$ICC_{inter-reader}$	$ICC_{intrareader}$	$ICC_{inter-reader}$
E_0	0.94	0.92	0.94	0.88
JSN_0	0.97	0.87	0.96	0.93
$Total_0$	0.98	0.91	0.96	0.92
E_{1Y}	0.93	0.94	0.95	0.87
JSN_{1Y}	0.97	0.84	0.96	0.91
$Total_{1Y}$	0.97	0.93	0.96	0.90
ΔE	0.94	0.92	0.92	0.85
ΔJSN	0.97	0.90	0.91	0.91
ΔTotal	0.98	0.95	0.93	0.92

ΔE: $E_{1Y} – E_0$; ΔJSN: $JSN_{1Y} – JSN_0$; ΔTotal: $Total_{1Y} – Total_0$; E_0: Erosion score at baseline; E_{1Y}: Erosion score at 1 year; ICC: Intraclass correlation coefficient; JSN_0: Joint space narrowing score at baseline; JSN_{1Y}: Joint space narrowing score at 1 year; $Total_0$: Total score at baseline; $Total_{1Y}$: Total at 1 year.

Table 3. Comparison of raw scores.

Score type	Genant–Sharp				van der Heijde–Sharp			
	Mean	SD	Median	Range	Mean	SD	Median	Range
E_0	21.8	11.8	18.9	3.0–54.3	41.1	25.4	35.3	3.5–108.0
JSN_0	10.6	8.6	10.4	0.5–34.5	20.3	16.6	19.0	0.5–69.0
$Total_0$	32.4	18.2	29.5	3.8–83.3	61.4	37.5	59.0	4.0–160.0
E_{1Y}	24.8	12.9	21.5	6.3–57.0	47.9	27.4	41.0	8.5–111.5
JSN_{1Y}	12.8	9.2	12.8	0.5–36.5	23.8	17.4	22.3	1.0–69.0
$Total_{1Y}$	37.6	20.2	33.9	8.3–89.5	71.6	40.4	67.5	12.0–166.5
ΔE	3.0	4.1	2.5	-2.3–17.3	6.8	8.0	5.3	-1.0–34.5
ΔJSN	2.2	3.9	0.5	0–16.0	3.4	7.2	0.5	0.0–29.0
ΔTotal	5.2	7.7	3.5	-2.3–33.3	10.3	14.6	7.3	-1.0–63.5

Values are average scores from the two independent readers.
ΔE: $E_{1Y} – E_0$; ΔJSN: $JSN_{1Y} – JSN_0$; ΔTotal: $Total_{1Y} – Total_0$; E_0: Erosion score at baseline; E_{1Y}: Erosion score at 1 year; JSN_0: Joint space narrowing score at baseline; JSN_{1Y}: Joint space narrowing score at 1 year; SD: Standard deviation; $Total_0$: Total score at baseline; $Total_{1Y}$: Total at 1 year.

and Genant–Sharp scoring. The proportions of patients showing change greater than 0 or above SDC were not statistically different between the two methods (Table 5).

Discussion

Of all the factors affecting the performance of radiographic scoring of RA in clinical trials, by far the most important is reader competency [16]. Determining whether a lucency within a bone on a radiographic projection represents a true erosion or simply a confluence of shadows mimicking an erosion, and whether an apparent change in the size of an erosion on serial radiographs is real or simply a result of variation in patient positioning, image projection or film exposure requires considerable perceptual skill and experience on the part of the reader. However, the focus of the present study was the potential impact of the scoring method used.

This direct comparison of the Genant-modified Sharp and van der Heijde-modified Sharp scoring methods using the same set of hand, wrist and foot images from a single cohort of patients with RA read by the same competent readers, demonstrated a high degree of agreement between the two methods (Figure 4), with high intra- and inter-reader reproducibility, expressed as ICC, observed for both methods (Table 2).

The differences between Genant's and van der Heijde's modifications of the original composite Sharp scoring method are relatively subtle. Both methods score the same locations for erosion in the hands, wrists and feet, except that the van der Heijde–Sharp method adds the lunate bone in the wrist and separates the CMC-1 joint at the base of the thumb into its two articular components. Adding the lunate bone potentially increases sensitivity for detecting bone erosion. However, prominent vascular channels and subentheseal changes related to scapholunate ligament damage can

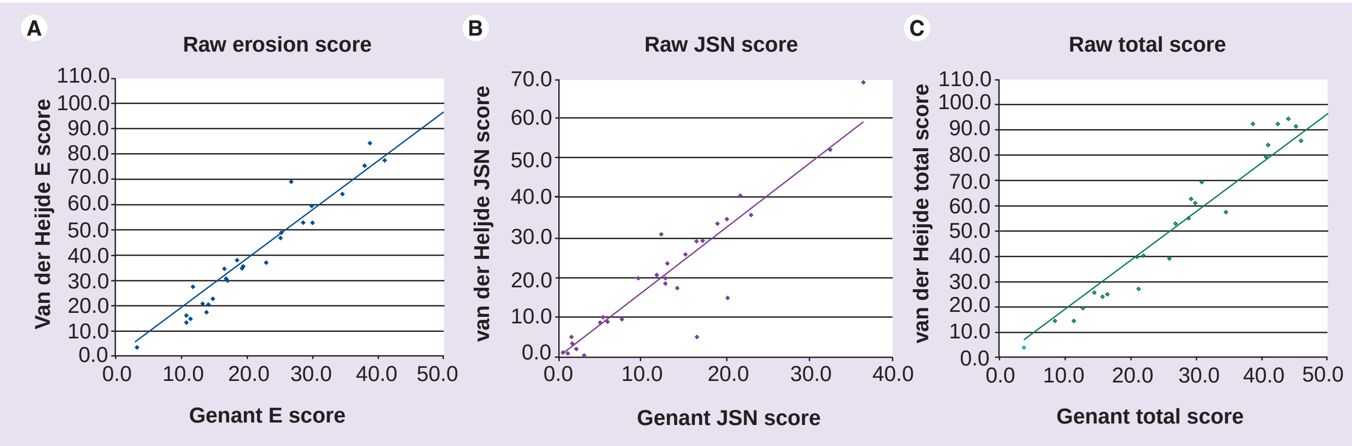

Figure 3. Raw baseline van der Heijde–Sharp scores plotted against raw baseline Genant–Sharp scores. (A) Erosion, **(B)** JSN and **(C)** total based on averaged values from the two readings. Slopes are 1.9, 1.6 and 1.9, respectively.
E: Erosion; JSN: Joint space narrowing.

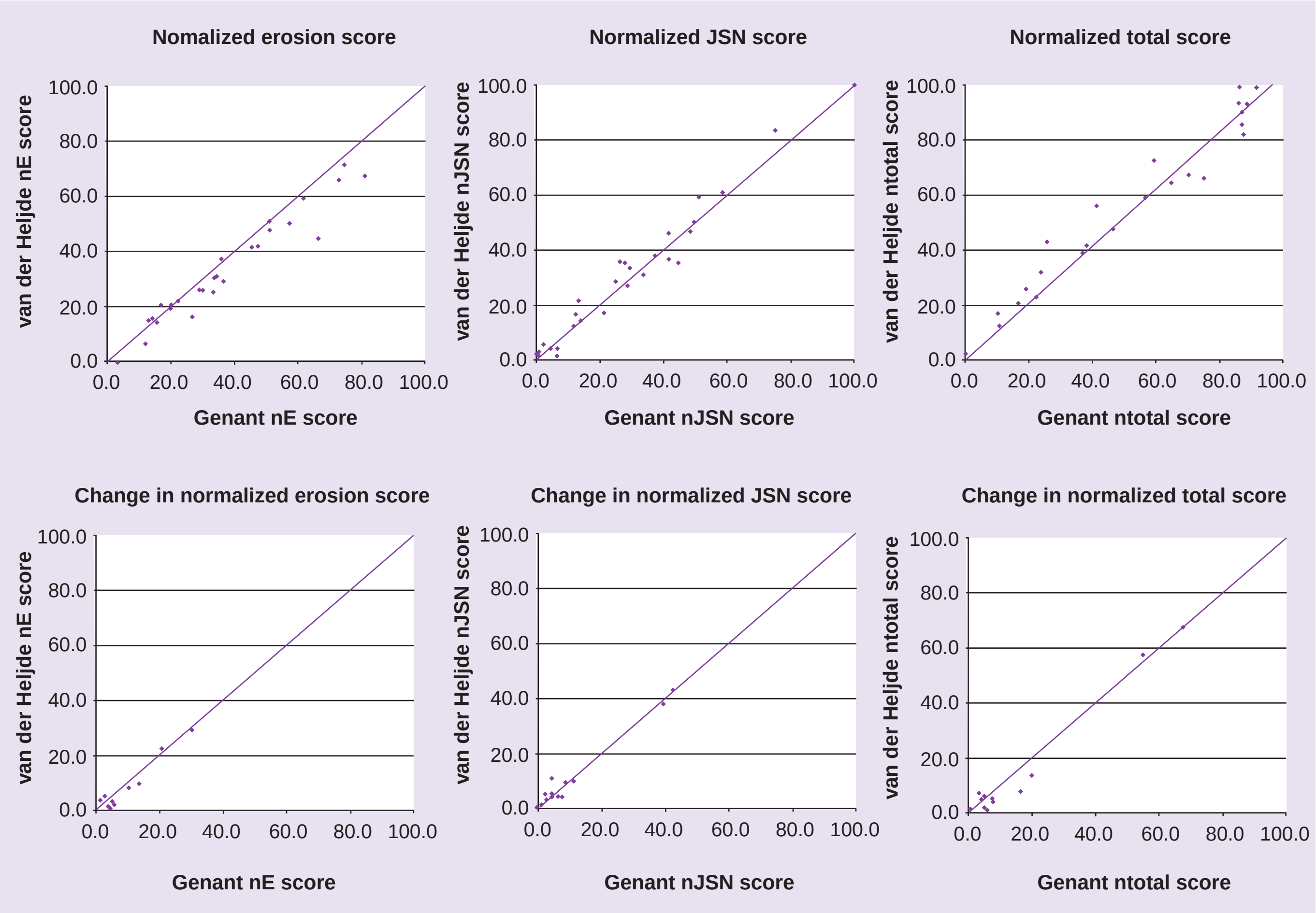

Figure 4. Normalized baseline van der Heijde–Sharp scores plotted against normalized baseline Genant–Sharp scores. The line depicts theoretically perfect correlation based on averaged values from the two readings. Correlation coefficients are 0.98, 0.98, 0.99, 0.95, 0.99 and 0.99, respectively.
nE: Normalized erosion scores; nJSN: Normalized joint space narrowing scores; ntotal: Normalized total scores.

mimic bone erosion and thereby decrease specificity in this location. Scoring the two components of the CMC-1 joint separately doubles the maximum erosion score attainable for that joint by the van der Heijde–Sharp method from 5 to 10, thereby weighting the joint more heavily than with the Genant–Sharp method. However, this does not add additional opportunity for detecting erosion or change in erosion, since the Genant–Sharp method evaluates both of these bones as well. Both methods also score erosion in the same joints of the foot; however, the van der Heijde–Sharp method scores each of the two components of these joints separately. As for CMC-1, this increases the maximum erosion score attainable for each joint in the foot from 5 to 10, and thus raises the ceiling of the van der Heijde scale closer to that of the Genant–Sharp scale, but it does not add greater opportunity for detecting erosion or change in erosion, as all of these bones are also evaluated in the Genant–Sharp method. The clinical implications of erosion on one side of

Table 4. Comparison of sensitivity to change.

Score type	Genant–Sharp		van der Heijde–Sharp		p-value
	SDC	nSDC	SDC	nSDC	nSDC
ΔE	2.4 (2.0)	4.7 (4.3)	7.3 (4.6)	6.4 (5.1)	0.08 (0.55)
ΔJSN	2.4 (1.6)	5.6 (2.6)	4.1 (4.2)	7.3 (5.9)	0.61 (0.01)
ΔTotal	3.8 (2.5)	7.9 (4.7)	8.1 (7.5)	10.5 (9.7)	0.27 (0.27)

Values are based on average scores from two independent readers. Values in parentheses are for a single reader.
ΔE: Erosion score at 1 year – erosion score at baseline; ΔJSN: Joint space narrowing score at 1 year – joint space narrowing score at baseline; ΔTotal: Total score at 1 year – Total score at baseline; nSDC: Normalized smallest detectable change; SDC: Smallest detectable change.

Table 5. Comparison of percentages of patients showing change.

Score type	Genant–Sharp		van der Heijde–Sharp		p-value	
	Change >0 (%)	Change >SDC (%)	Change >0 (%)	Change >SDC (%)	Change >0	Change >SDC
ΔE	89	50	93	39	1.00	0.45
ΔJSN	64	29	68	21	1.00	0.50
ΔTotal	89	39	93	43	1.00	1.00

Values based on average scores from two independent readers.
ΔE: Erosion score at 1 year – erosion score at baseline; ΔJSN: Joint space narrowing score at 1 year – joint space narrowing score at baseline; ΔTotal: Total score at 1 year – Total score at baseline; SDC: Smallest detectable change.

a joint but not the other is unknown; however, as all regional scores are combined in both methods to determine total erosion scores, this does not affect discrimination of subjects or treatments on the basis of total erosion score.

The joints scored for JSN in the hands, wrists and feet are also quite similar for each method. Differences include elimination in the van der Heijde–Sharp method of the interphalageal-1 joint of the thumb, the capitate–lunate joint of the wrist (scored as part of the pericapitate space in the Genant–Sharp method) and the lunate–radius joint (scored as part of the radiocarpal joint in the Genant–Sharp method). In addition, the van der Heijde–Sharp method adds the scaphoid–trapezium joint and separates the CMC 3–5 compartment into three individual joints. Subdividing CMC 3–5 raises the relative score ceiling for the van der Heijde method closer to that in the Genant–Sharp scale, but again, does not alter the opportunity to detect change, as these joints are also included in the Genant–Sharp method. The joints scored for JSN in the feet are the same for the two methods.

The erosion scales used differ somewhat between the methods. The van der Heijde–Sharp method scores each location on a 6-point scale with integer values from 0 to 5, whereas the Genant–Sharp method uses an 8-point scale from 0 to 3.5 in increments of 0.5. Accordingly, the van der Heijde–Sharp method typically gives a larger numerical score for a given degree of erosive damage (maximum erosion score: 280) than the Genant–Sharp method does (maximum erosion score: 140), although both scales include the same number of increments (280). The JSN scale also differs for each method, but primarily with respect to the number of increments. Both range from 0 to 4, the van der Heijde–Sharp scale however does so in five integer increments, whereas the Genant–Sharp scale uses nine intervals of 0.5. As for erosion score, the van der Heijde–Sharp scale yields numerically larger values (maximum JSN score: 168) than the Genant–Sharp method does (maximum JSN score: 152), but the Genant–Sharp scale includes more increments (304 vs 168). Finally, the Genant–Sharp total score, which combines erosion and JSN scores, has more increments (584) than the van der Heijde total score does (448).

Owing to these scale differences, direct comparisons of scores generated by the two methods are not meaningful. Accordingly, in this study comparisons were made using scores normalized to the ranges of scores observed with each method, as recommended by Sharp *et al.* [14]. In the end, the differences between the two methods were found to be relatively minor and the normalized scores from one method correlated closely with those from the other (FIGURE 3). The main limitation of this study was the relatively small numbers of patients included in the analysis. Corroboration of the results in a larger sample would, therefore, be useful. Also, the radiologists reading the images in this study had more experience with the Genant–Sharp method than with the van der Heijde–Sharp method in clinical trials. However, based on the intra- and inter-reader reproducibilities observed, both cross-sectionally and longitudinally (TABLE 2), reader performance was similar and very high for both methods. As noted previously, the greatest challenge in scoring by either method is accurately determining whether a lucency observed on a radiograph is truly a bone erosion or simply a normal anatomical feature, such as a bone tubercle viewed en face, or a perceptual phenomenon, such as negative mach effect [16], and whether an apparent change in the size of an erosion or the width of a joint space on serial radiographs is pathologically real or a result of variation in x-ray beam centering or angulation, or some other technical factor. Accordingly, we do not believe that the readings in this study were biased towards one or the other method.

In conclusion, the findings of this study indicate that the Genant–Sharp and van der Heijde–Sharp methods show relatively similar performance for scoring erosion and JSN in the hands, wrists and feet of patients with RA. Additional direct comparisons of these two scoring methods in different and larger cohorts of patients with RA would be useful to further elucidate the relative strengths and weakness of each method.

Future perspective

Aside from minor study-specific modifications in some trials, the Genant–Sharp and van der Heijde–Sharp scoring methods have not changed substantially over the more than 2 decades that the two methods have been in use. While some attempts have been made to reduce the complexity and effort needed to perform discriminative readings with such methods, most of the imaging attention in RA research over the past several years has shifted to MRI and ultrasound. These newer technologies offer broader pathophysiological information and greater sensitivity to change, and thus allow questions about drug efficacy to be answered with fewer patients, less time and therefore lower cost, than with radiography. Particularly in light of the recent shift in clinical trials to active-comparitor study designs, which show slower progression and smaller differences between treatment arms, it is likely that the trend towards MRI and ultrasound will increase over the next 5–10 years, and that the role of radiography in clinical trials will diminish considerably.

Acknowledgements

Images used in this study were provided by F Hoffmann, LaRoche, Ltd.

Financial & competing interests disclosure

Charles Peterfy is an employee and shareholder of Spire Sciences, LLC and a shareholder of Synarc Inc, which provide image analysis services for clinical trials to multiple pharmaceutical, biotechnology and medical device companies. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Executive summary

- Genant–Sharp and van der Heijde–Sharp have been the two most common radiographic scoring methods used in clinical trials for the past 2 decades, yet their relative performances had not been formally compared.
- In this investigation, Genant–Sharp and van der Heijde–Sharp methods demonstrated similar performance for scoring erosion and joint-narrowing space in the hands, wrists and feet of patients with rheumatoid arthritis using a common set of radiographs and common pair of readers.
- Readers demonstrated similar proficiency with each method based on inter-reader precision (intraclass correlation coefficient) across a broad range of severities.
- Since the arbitrary scales used in the two scoring methods are different, direct comparison of the values is not valid unless the scores are first normalized.
- The van der Heijde–Sharp method yields higher raw values than Genant–Sharp scoring does, but the Genant–Sharp scale has more increments and a higher ceiling for erosions.

Bibliography

Papers of special note have been highlighted as:
▪ of interest
▪▪ of considerable interest

1 Sharp JT, Lidsky MD, Collins LC, Moreland J: Methods of scoring the progression of radiologic changes in rheumatoid arthritis: correlation of radiologic, clinical and laboratory abnormalities. *Arthritis Rheum.* 14, 706–720 (1971).

2 Genant HK: Methods of assessing radiographic change in rheumatoid arthritis. *Am. J. Med.* 30, 35–47 (1983).

3 Genant H, Jiang Y, Peterfy C, Lu Y, Reidei Y, Countryman P: Assessment of rheumatoid arthritis using a modified scoring method on digitized and original radiographs. *Arthritis Rheum.* 40, 1583–1590 (1998).

4 Genant HK, Peterfy CG, Westhovens R *et al.*: Abatacept inhibits progression of structural damage in rheumatoid arthritis: results from the long-term extension of the AIM trial. *Ann. Rheum. Dis.* 67(8), 1084–1089 (2008).

5 van der Heijde D: How to read radiographs according to the Sharp/van der Heijde method. *J. Rheumatol.* 27(1), 261–263 (2000).

▪ **Very clear description of the van der Heijde scoring method.**

6 van der Heijde D, van Riel P, Nuver-Zwart I, Grinbnau F, van de Putte L: Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. *Lancet* 1, 1036–1038 (1989).

7 Breedveld FC, Weisman MH, Kavanaugh AF *et al.*: The PREMIER study: a multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, agressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis Rheum.* 54(1), 26–37 (2006).

8 van der Heijde D, Klareskog L, Rodriguez Valverde V *et al.*: Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis: two-year clinical and radiographic results from the TEMPO study, a double-blind, randomized trial. *Arthritis Rheum.* 54(4), 1063–1074 (2006).

9 Smolen JS, Han C, Bala C *et al.*: Evidence of radiographic benefit of treatment with infliximab plus methotrexate in rheumatoid arthritis patients who had no clinical improvement: a detailed subanalysis of data from the anti-tumor necrosis factor trial in

rheumatoid arthritis with concomitant therapy study. *Arthritis Rheum.* 52(4), 1020–1030 (2005).

10 Kremer J, Genant H, Moreland L *et al.*: Effects of abatacept in patients with methotrexate-resistant active rheumatoid arthritis: a randomized trial. *Ann. Intern. Med.* 144(12), 933–935 (2006).

11 Keystone E, Emery P, Peterfy CG *et al.*: Prevention of joint structural damage at 1 year with rituximab in rheumatoid arthritis patient with an inadequate response to one or more TNF inhibitors (REFLEX study). *Ann. Rheum. Dis.* 65(Suppl. 2), 58 (2006).

12 Strand V, Sharp J: Radiographic data from recent randomized controlled trials in rheumatoid arthritis: what have we learned? *Arthritis Rheum.* 48(1), 21–34 (2003).

13 Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D: Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. *Ann. Rheum. Dis.* 64, 179–182 (2005).

▪▪ **Explains the methodology and rationale for measuring smallest detectable change in clinical trials, and why smallest detectable change is more appropriate than smallest detectable difference under the current trial paradigm.**

14 Sharp J, Wolfe F, Lassere M *et al.*: Variability of precision in scoring radiographic abnormalities in rheumatoid arthritis by experienced readers. *J. Rheumatol.* 31(6), 1062–1072 (2004).

▪ **Asserts among other things that readers who are more experienced with one radiographic scoring method perform better with a new scoring method than do readers who are more familiar with the second method but less experienced overall.**

15 Efron B, Tibshirani RJ: *An Introduction to the Bootstrap.* Chapman & Hall, NY, USA (1993).

16 Daffner R: Visual illusions in the interpretation of the radiographic image. *Curr. Probl. Diagn. Radiol.* 18(2), 62–87 (1989).

Website

101 Title 21 Code of Federal Regulations (21 CFR Part 11): electronic records; electronic signatures 2000 www.fda.gov/ora/compliance_ref/part11/ (Accessed 15 January 2008)