Radiographic Hand Joint Space Width Assessed by Computer Is a Sensitive Measure of Change in Early Rheumatoid Arthritis

JANE ANGWIN, ANDREW LLOYD, GEOFF HEALD, GERALD NEPOM, MICHAEL BINKS, and MICHAEL F. JAMES

ABSTRACT. ***Objective.*** To compare changes in the computerized measurement of radiographic hand joint space width (JSW) to changes in modified Sharp scores in a retrospective 2-year study of early rheumatoid arthritis (RA).

Methods. First and last standard clinical hand radiographs of 245 patients with RA were analyzed blind using purpose-written computer software to measure changes in JSW for proximal interphalangeal (PIP) and metacarpophalangeal (MCP) joints in the 3 middle fingers of each hand. Before measurement, the radiographs were scored independently by 2 radiologists using a modification of Sharp scoring.

Results. The paired changes in JSW (–0.051 ± 0.005 mm) and Sharp score (+3.81 ± 0.50) were both significant over the study duration. In measured joints showing an increase in joint space narrowing (JSN) score, 92% had a corresponding reduction in JSW. In patients with an increase in total score, including JSN and erosion scores in fingers and wrists, 84% had a corresponding reduction in mean (PIP + MCP) JSW. Patients with no change in Sharp score (47%) still experienced a significant reduction in measured JSW (–0.027 ± 0.006 mm). HLA-DR genetic markers of severe disease progression were associated with significantly greater reductions in JSW but not increases in Sharp score. (Values: mean ± standard error of mean).

Conclusion. Measured JSW averaged over 6 PIP and 6 MCP joints was a valid and more sensitive measure of change than total Sharp score in this study of early RA. (J Rheumatol 2004;31:1050–61)

Key Indexing Terms:
RHEUMATOID ARTHRITIS DISEASE PROGRESSION JOINT SPACE WIDTH
HAND COMPUTER MEASUREMENT STANDARD RADIOGRAPHY

Disease progression in rheumatoid arthritis (RA) is routinely assessed using clinical measures, subjective assessments of the patient's condition, and radiographs of patient's hands and feet. In principle, radiographs are objective, cost-effective, and accessible. In clinical trials, radiographs are visually assessed by radiologists who assign scores to the severity and number of erosions and the severity of joint space narrowing (JSN) at a number of joint locations. Despite optimization[1-9], scoring remains subjective and suffers the inherent disadvantage that changes are only detected once they are sufficiently severe to be visible to the eye.

From GlaxoSmithKline Research and Development, Stevenage, Hertfordshire, UK.

J. Angwin, MA; M. Binks, MD, RA Biology, Respiratory and Inflammation Centre of Excellence in Drug Discovery; A.P. Lloyd, BSc, Biostatistics and Data Sciences; G. Heald, Information Technology Research and Development; M.F. James, PhD, MRI Group, Neurology & GI Centre of Excellence for Drug Discovery, GSK Research & Development, UK; G.T. Nepom, MD, PhD, Virginia Mason Research Center, Seattle, Washington, USA.

Address reprint requests to J. Angwin, GlaxoSmithKline R&D, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK. E-mail: jane.s.angwin@gsk.com

Submitted April 23, 2003; revision accepted November 3, 2003.

The development of new therapies is encouraged by a sensitive measure of their effect: "If you cannot measure it, you cannot improve it" (Lord Kelvin, 1883). An objective sensitive measure of radiographic joint space width (JSW), together with erosion and JSN scores, could provide such a stimulus to the development of RA therapies.

Methods to measure JSW either manually, semiautomatically, or automatically have been described[10-18]. Buckland-Wright, *et al* used magnified (×5) stereopair hand and wrist radiographs, a cursor to manually outline the joint margins, and a computer to obtain the margin separation, or JSW[10,11]. Higgs, *et al* used a magnifying glass (×7) with an internal scale and plastic overlay templates to measure change in JSW and erosion size, respectively. They considered paired serial radiographs from 10 patients taken over an average 34-month period[12]. They found that the measured change in JSW correlated well to both change in measured erosion size and change in total Sharp score, and appeared to offer greater discrimination.

Allander, *et al* initiated computerized measurement of JSW in hands[13]. Although tested on only 3 patients, he found that his computer method was more precise than manual measurements using a magnifying glass. Using a forerunner of our present computer program, James, *et al* investigated

the change in measured hand JSW with time for 34 patients with RA[14]. Different radiographic features identifying the distal joint margin were considered. JSW measured in proximal interphalangeal (PIP) and metacarpophalangeal (MCP) joints decreased significantly over a one-year period in female patients (n = 16), although not in males (n = 18). There was a corresponding significant increase in mean JSN score in the female, but not the male, patients (average of scores assigned by 6 readers).

Sharp, *et al* describe a computer-based method of measuring JSW and another for determining erosion volume in hand joints[15]. Using serial, Sharp-scored radiographs of 25 RA patients at 2 timepoints, they found good agreement between JSN score and JSW measurement changes and similar discrimination between treatment effects by both methods. The automatic location of joint boundaries was not always robust enough to cope with the diseased joints seen in the study, and boundaries were mostly located manually. They found the measurement method was more reproducible than scoring, although the former was more time-consuming. It was suggested that, although the measurements in this pilot study did not show greater discrimination than scores, the measurement method should be tested in future therapeutic trials.

In using JSW measurements, the radiographs are taken "at face value" and any changes in radiographic JSW due to equipment, techniques, or inadvertent changes in hand positioning become an important, and sometimes unavoidable, source of error. Our early work[14] was followed by improvements to the computer program and a study of the reproducibility of JSW measurements in repeat radiographs of healthy volunteers[16]. Under the relatively ideal conditions of the study, the smallest organic change in average JSW (95% probability) expected to be detectable in repeat radiographs for a single subject was 0.05 mm (averaged across 8 PIP or 8 MCP joints per patient), equivalent to ~3% of average JSW size. In an RA clinical trial, with a considerable time lapse between repeat radiographs, less controlled radiographic techniques, and disease affecting the positioning of the patient's hands, the smallest detectable organic change could well be larger.

Complementing the work on JSW measurement, Duryea, *et al* developed computer programs to automatically locate joints in digitized images of hand radiographs with the intention of developing a fully automated means of measuring JSW and erosions at these locations[17,18]. Such a program would be useful if the JSW and erosion measurements were found to be sensitive and discriminating.

The object of our study is to establish, in a much larger patient group with early RA than hitherto considered, whether measurements of radiographic hand JSW are more discriminating than scoring in measuring disease progression. For this purpose, we retrospectively analyzed serial radiographs from 245 patients taken during a multicenter US clinical trial (unpublished data) over the period 1987 to 1990.

If more discriminating information can be extracted from standard clinical radiographs, it may be possible to reduce the size or duration of clinical trials. It may also be easier to identify factors affecting the severity of joint disease, for example, genetic characteristics.

MATERIALS AND METHODS

Study design. The study evaluated hand and wrist radiographs of 247 patients with early RA, collected during a 2-year double-blind placebo-controlled clinical trial carried out under the auspices of GlaxoSmithKline (formerly Smith Kline & French) by 34 investigators at centers across the USA in the period February 1987 to December 1990. As part of the clinical trial, the first and last radiographs for each patient were scored manually for erosions and JSN using a modification of the Sharp method[1]. For this study, the same radiographs were digitized and, using image analysis software as described[16], JSW were measured in order to investigate the change with time. The effect of drug treatment was not considered.

Patients. Patients met the subset of the 1958 revision of the American Rheumatism Association (ARA) 1956 criteria for definite or classical RA[19,20] not requiring invasive techniques and, in addition, had a history of symptoms compatible with RA for a period less than 12 months. Patients were recruited without regard to sex, race, or socioeconomic status. Patients gave informed written consent to participate and were free to withdraw at any time without stating a reason. Further, investigators were expressly allowed to withdraw patients whose symptoms either increased or showed no benefit from the coded treatment within the first 7 months of the study.

The ARA 1958 criteria for definite or classical RA were less specific than the current 1987 American College of Rheumatology (ACR, formerly ARA) criteria for RA[21] and, being less specific, used exclusion criteria to omit patients who showed arthritis-type symptoms but were suffering from another complaint. A current 1987 criterion specifically considers swelling in the MCP, PIP, and/or wrist joints as an indicator of RA. Although not part of the 1958 criteria, 241/247 patients in our study were reported to have swelling in these hand joints at study entry, and a further 4/6 patients had notable pain/tension. From an examination of the initial clinical data, all patients are expected to have met the current ACR 1987 diagnostic criteria for RA.

Patients were excluded from the clinical trial in case of: severe or total incapacitation due to the disease (Steinbrocker Function Class IV); the normal ARA 1958 exclusion criteria[20]; clinical or laboratory evidence of serious comorbid disease; previous toxicity to heavy metals; and past treatment with a disease modifying antirheumatic drug.

Wrongly labelled radiographs were sent in by investigators for 2 patients, and data for these patients were removed from the study (2/247). The remaining patient population (176 women, 69 men) had mean age 53 years at the start of the study (range 18–82 yrs). Patients were randomly assigned, roughly equally, to either oral gold plus nonsteroidal antiinflammatory drug (NSAID) or placebo plus NSAID for a duration of up to 2 years, median 17 months (range 3–28 mo). The age distribution and duration in the study showed no significant difference by sex.

Genetic typing. Peripheral blood lymphocytes (PBL), obtained from 50 cm^3 heparinized whole blood at the onset of the study, were frozen in liquid nitrogen for later use. DNA purified from the PBL was assayed using oligonucleotide hybridization to detect the presence of particular specific HLA alleles, as described[22]. Sequence-specific oligonucleotide probes identified the HLA-DR4 subtypes DRB1*0401 (Dw4) and DRB1*0404 (Dw14) and the HLA-DR1 specificity associated with RA, DRB1*0101 (Dw1).

Radiography. Radiographs used 8 × 10 inch (20 × 25 cm) Kodak Min-R radiographic film and a Kodak X-Omatic Cassette containing a single

Lanex Fine screen. A sample film showing the recommended film density was included with the instructions for each investigating center. The smallest focal spot available was used. To ensure accurate radiographic assessment, the protocol specified that each hand and wrist should be radiographed separately, in posteroanterior (PA) position, with the palmar surface pressed as flat as possible against the cassette; the fingers should be spread slightly apart and the beam centered for that hand; the second digit should form a straight line with the forearm, which should be laid flat against the radiographic table, using a light sandbag to help flatten the hand and wrist, if necessary.

Radiographs were taken at study commencement, at 6-month intervals, and on study withdrawal. On completion, the initial and final radiographs were extracted and blinded for patient's name, source, and radiographic sequence.

Radiological scoring. Initial and final radiographs were examined, in known sequence, by 2 experienced radiologists who independently graded them for severity of erosive disease and JSN using a modification of Sharp scoring[1]. In each hand radiograph, 5 PIP and 5 MCP joints were each scored for both erosions and JSN, together with a further 4 wrist joints scored for erosions and 3 wrist joints for JSN. Based on a scoring scale 0–4, a theoretical maximum score of 216 was possible for each patient per timepoint.

The 2 readers' scores were averaged for each joint in each radiograph. The difference in final and initial film scores indicated the patient's disease progression. No disease progression was indicated by a change in total erosion and JSN score < 1[23]. Similarly, no evidence of erosions, or JSN, at study start was indicated by an initial total erosion, or JSN, score < 1. Only averaged, not separate, readers' scores were available.

Image capture. Digitized images of the hand radiographs were recorded onto computer using an Hitachi KP-141 video camera[16]. Captured images had 8-bit pixel depth giving 256 grey levels. The highest level was displayed in red to assist adjustment of the camera aperture for optimal image brightness. The image scale was set to ensure optimal resolution of the relevant joint features[16].

Image calibration. Using standard millimeter graph paper, the camera height was adjusted until the field of view was precisely 32 mm vertically, giving a resolution of 17.8 pixels/mm, equivalent to 450 dots per inch. The pixel size was set at 0.05614 mm. The camera position was thereafter rigidly maintained and the resolution frequently checked.

In daily checks, a precision machined 1.10 mm-wide metal strip was viewed and the computer software used to determine its width [1.101 ± 0.003 mm, mean ± standard deviation (SD) used throughout this section]. Further checks of measurement reproducibility were based on a sample hand radiograph from another RA study. Before the start of the study, index, middle, and ring PIP and MCP joints were measured 10 times by each of the 2 technologists carrying out the measurements, and the results were recorded: for PIP 0.975 ± 0.019 mm; MCP 1.525 ± 0.014 mm (n = 120 measurements for each joint type). During the measurement process, at the beginning of each analysis session and after measurement of radiographs for 3 patients, 2 random joints (one MCP and one PIP joint) from the sample radiograph were remeasured. Recorded JSW were, for PIP 0.973 ± 0.014 mm (n = 114 measurements) and for MCP 1.525 ± 0.012 mm (n = 113). The SD are in accord with repeat measurements for healthy subjects: ± 0.022 mm and ± 0.014 mm for individual PIP and MCP joints, respectively[16]. There was no drift in calibration with time and no consistent variation in measurements between the 2 operators.

JSW measurement: procedure. Measurements were shared (not repeated) between 2 operators, who each received 2–3 hours training in use of the computer software.

Mean JSW was measured for index, middle, and ring finger PIP and MCP joints. The operators were blinded to radiographic sequence but sets of radiographs for each patient (left and right hand at 2 timepoints) were measured together in order to minimize the effect of any changes in image capture or operator technique. Measurements for 245 patients (2940 joints) took a total of 76 hours (average of 20 minutes per patient) including film handling and calibrations.

JSW measurement: method of computer analysis. The computer analysis method was as described[16], with results illustrated in Figure 1. Briefly, for both PIP and MCP joints, the proximal joint margin was identified as the line of maximum slope in radiographic density (red boundaries) while the distal margin was identified as the ridge of peak radiographic density (green boundaries), on the other side of the low density joint space. A software tracking procedure employing a Gaussian function contributed to reliably identifying valid anatomical margins irrespective of image digitization and noise[24,25].

The mean JSW was calculated as the linear separation of the margins averaged over the defined joint breadth. The horizontal straight red line, Figure 1B, corresponds to accepted valid measurements across the PIP joint, while MCP joints were measured radially within a 1-radian sector roughly centered on the metacarpal head (between the straight green lines, Figure 1D). Averaging radiographic and image noise, implicit in the calculation of mean JSW, gives a more reproducible measure than single width measurements, such as minimum or midline JSW. Values were stored automatically in a spreadsheet.

JSW measurements: monitoring program performance. In the majority of cases, the distal margin was unambiguous[16] (Figure 2) and a precise JSW measurement was obtained. However, the finger joint anatomy varied (Figure 1) and subchondral regions of compacted bone could appear in the radiograph as thickened boundaries (Figure 1A) or as parallel ridges of high density lying adjacent to the distal margin (Figure 1C). The automatic detection procedure normally found appropriate joint margins, but when the parallel "ridge" had greater radiographic density than the joint margin itself, the line of peak densities found by the program sometimes moved between the 2 bands. Similarly, in cases where there was a thickened boundary, the program did not always find a consistent margin in serial radiographs. In both cases, the operator accepted the findings without attempting to constrain the program to find either the more appropriate margin or a consistent margin within pairs of radiographs. As a result, there could be apparent, but erroneous, "changes" in JSW, introducing a source of variability not inherent in the radiographs themselves. We intentionally required no informed input from the operators in order to simulate expected field trial conditions. However, this source of variability is avoidable with program improvements or quality control procedures.

Unmeasured joints. For some joints, the margins were insufficiently distinct to be reliably located by the program and JSW could not be measured. Joints unmeasured at either timepoint were eliminated from the investigation. Overall, 181/1470 PIP and 94/1470 MCP joints (9.3%) were eliminated, although the majority of these had been measurable at one timepoint (105/275, 80/275 were measurable at the first and second time-points, respectively).

Reasons for indistinct joint margins were: (1) disease (abutting margins, fusion, severely flexed joints); (2) radiographic quality (over or underexposed films); (3) incorrectly positioned hands (not PA view); or, usually, a combination of these conditions. Conditions (2) and (3) are avoidable, and software changes could improve measurements in poor contrast images. For joints unmeasurable due to abutting margins at the second timepoint (i.e., JSW = 0.00 mm), JSW loss could be defined from the first measurements. However, the more conservative approach of eliminating these joints from the study was adopted.

Eliminated joints affected a total 101/245 patients, with the majority (65/101) having only one or 2 joints affected. In total, 233/245 patients had at least 7/12 measured joints and a further 3/245 patients had at least 2 PIP *and* 2 MCP joints measured. Of the remaining 9/245 patients, 3/245 had no joints measured (poor quality radiographs), 5/245 had no PIP joints measured (3/5 due to radiography or hand positioning and 2/5 due to disease), and one patient had no MCP joints measured (disease). Scores were provided for all patients at both timepoints, and no patient was eliminated due to absence of scores.

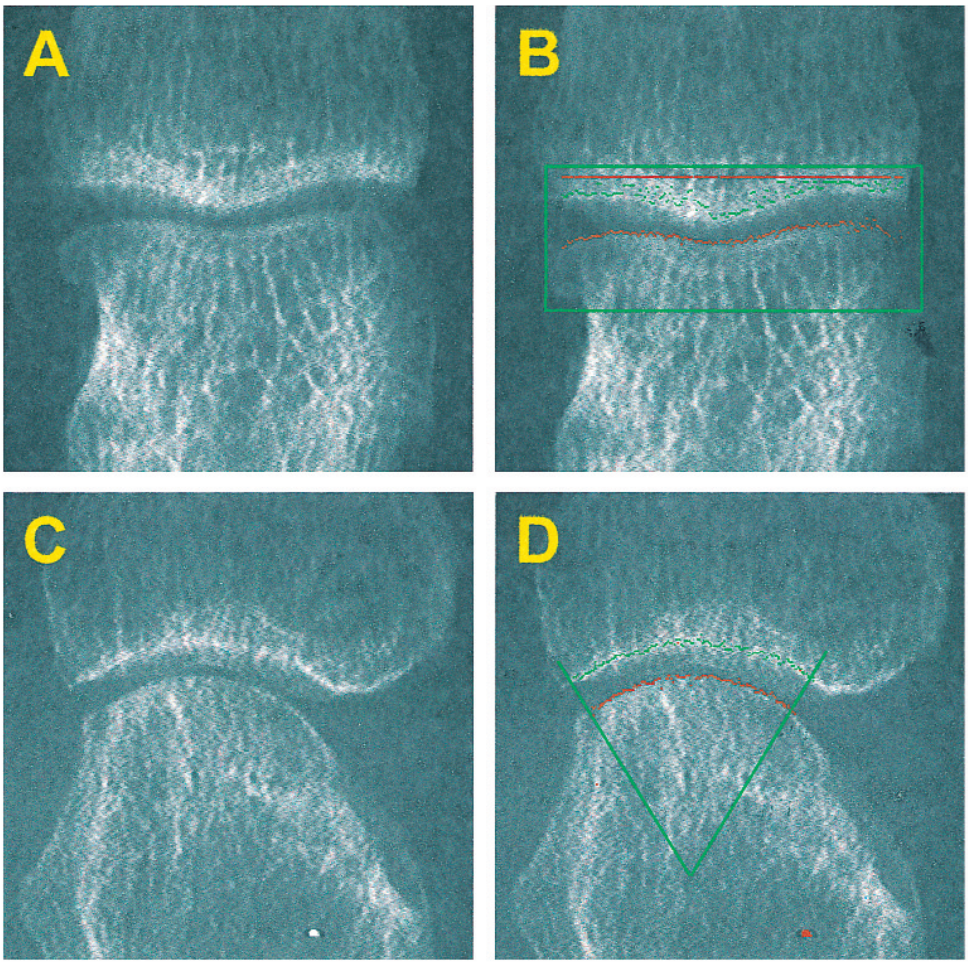

Figure 1. Anatomical landmarks used for computerized mean JSW measurement in PIP and MCP radiographs showing the proximal (red) and distal (green) joint margins detected. Rectangular box in (B) is inserted by the user to roughly locate the PIP joint. Horizontal red line corresponds to accepted valid measurements. Green lines in (D) identify the 1-radian sector of the MCP joint used to determine mean JSW. The effect of anatomical variation is illustrated: in A, B, the PIP distal margin is thickened, while in C, D, a parallel range of subchondral compacted bone lies adjacent to the MCP distal margin. The pixels distant from the joint margin colored red in (D) are at the brightest level of the image intensity scale. These points have not been identified as lying on the joint margin. Examples of joints with unambiguous margins are illustrated in Angwin, *et al,* Figure 2[16].

Joints showing JSN change over the study duration were eliminated considerably more often (31%, 41/134) than those with no JSN score change (8%, 234/2806). This loss of data affects the comparison of JSW and JSN change for individual joints but has little effect on the study conclusions, as 97% of patients (238/245) had at least 6/12 joints measured and 99% of patients (242/245) had at least 3 joints measured. Only one patient with no measured joints (1/3) was affected by Sharp score change. However, it is important to recognize that JSW measurement as a means of monitoring the progression of RA is most appropriate in early RA before severe joint damage has occurred.

Statistical analysis. Paired changes in scores and JSW measurements were derived from the initial and final radiographs for each patient. Due to the relative variability of the data, changes in hand averaged JSW were considered rather than changes in individual JSW. PIP-hand, MCP-hand, and PIP + MCP-hand are the measured changes between a patient's first and last radiographs averaged over 6 PIP joints, 6 MCP joints, and 6 PIP and 6 MCP joints, respectively, considering index, middle, and ring fingers, but only including joints measurable in both sets of radiographs. Variation in JSW paired measurements can arise from various sources: changes in radiographic technique, hand positioning, the measurement process itself, and real organic change. Variation in scores can derive from similar sources, except that some interpretive correction by the scoring radiologist is possible.

Analysis of variance was used to examine the effect of joint, sex, hand, and finger on the initial JSW and score values. Paired JSW and score changes were investigated using dependent t tests and the nonparametric Wilcoxon matched pairs test. The software package Statistica (v5.1, Statsoft Inc., Tulsa, OK, USA) was used. Non-zero changes in score were found to be independent of their initial value. Although paired changes in MCP-hand were significantly correlated to their initial value ($p = 0.01$), paired changes in PIP-hand were not ($p = 0.06$). Where there is such a correlation, it can be taken into account to reduce the variability. In this study, a conservative approach was adopted in the analysis of JSW data. Scores and JSW have both been considered without adjustment for initial values. Patient genotype subgroups were compared using the nonparametric Mann-Whitney U test for scores as well as independent t tests for

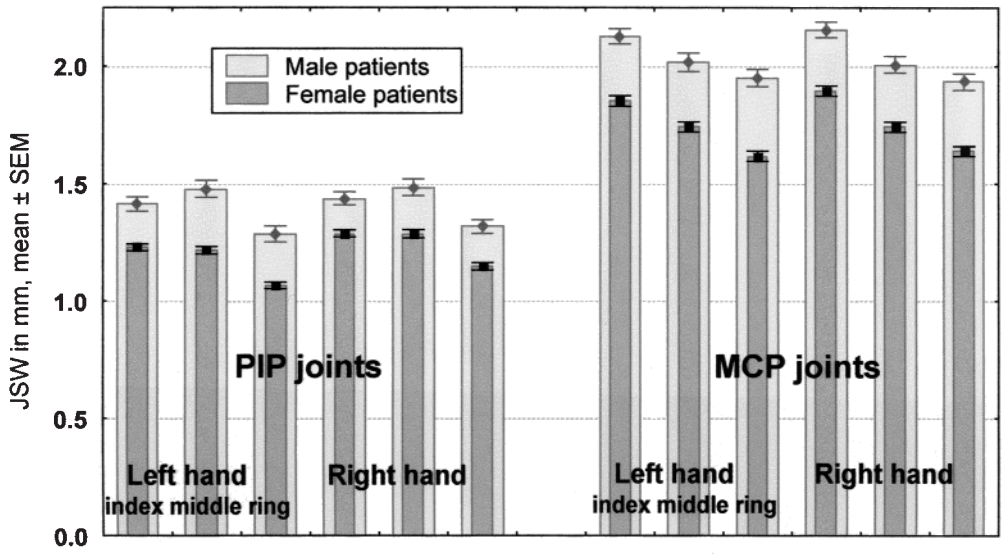

Figure 2. Variation in PIP and MCP joint space widths (JSW) (mm, mean ± SEM) showing the effect of sex, joint, and left/right hand. Values were measured from the initial set of radiographs for each patient. For male and female patients, PIP mean JSW were 1.41 ± 0.029 mm and 1.20 ± 0.013 mm, respectively. Corresponding MCP mean JSW were 2.03 ± 0.030 mm and 1.74 ± 0.019 mm, respectively. Values for the complete group are given in Table 1A.

JSW measurements and scores. In general, results were considered statistically significant if there was less than 5% ($p < 0.05$) probability of their occurrence due to chance.

Population mean changes, standard errors of the means (SEM), and standard deviations (SD) were determined both for the full data and for various subsets of the data. Signal to noise ratios (SNR), defined as ratios of the population mean change (taking the positive value) to the SEM change, enabled a comparison of the relative power of scores and measurements. The SNR is equivalent to the standardized change used in Jones, *et al*[26] and Goldsmith, *et al*[27]. When applied to normally distributed data, it is equivalent to the standardized "effect size"[28]. Although changes in score tend to follow a non-normal distribution, the SNR remains a useful means of comparing measures[26,27]. SEM was preferred as the nondimensionalizing factor, instead of SD, since the SNR then reflects the total effect of each measure including the reduced number of patients for whom JSW measurements were possible. An alternative ratio of median to interquartile range was not helpful, as median score changes were generally zero. SNR should be considered in conjunction with sensitivity since an insensitive measure, showing some change but little variation in that change, will have a misleadingly low SEM and high SNR[29].

As repeat radiographs at a single timepoint are not available, the smallest organic change that can be detected in repeat radiographs for an individual joint or an individual patient cannot be assessed (the smallest detectable difference and the specificity). Repeat measurements were carried out on a single set of radiographs under the field trial conditions used in this study. The measurements indicate a maximum error (at 95% probability level) of 0.038 mm for an individual joint and 0.009 mm for PIP + MCP-hand measurements including the effect of inter- and intra-operator error. However, repeat radiographs will introduce additional variation as a result of small changes in hand position and radiographic technique between radiographs in the absence of organic change. For JSW measurements, the effect of these changes is not moderated by the informed judgment of the radiologist and, as a result, changes in JSW may be less specific than changes in Sharp score, i.e., changes in JSW measurements may occur in the absence of real organic change. However, the radiographs were blinded to sequence and genotype before measurement, and thus the JSW changes are unbiased. JSW changes followed an approximately normal distribution.

Values quoted in the results are mean ± SEM unless stated otherwise.

RESULTS

Initial JSW and scores. The condition of patients' joints at the start of the study is summarized in Figures 2 and 3 and Table 1.

The variation in JSW between patients is considerable and is similar, in degree and range, to that found in healthy subjects[16]. Further, JSW varies significantly depending, among other things, on left or right hand, the finger joint considered, and patient's sex. This systematic variation is not surprising, but the ability of the program to detect it in retrospective radiographs is encouraging and gives confidence in the program's potential to measure small variations and changes. Although the smallest JSW tend to correspond to high JSN scores, initial JSW give little information on the initial disease state of patients.

Conversely, scores do indicate the disease state of patients at the start of the study. JSN was observed less frequently than erosions and 91% of patients with evidence of JSN also had evidence of erosions. JSN was significantly more evident in wrist joints than in finger joints, while erosions were most evident in PIP joints. Scores did not vary significantly between left and right hand nor, in general, were they dependent on patient's sex.

Change in JSW and scores over the study duration. The condition of patients' joints deteriorated significantly over the study duration as determined by both score and JSW changes (Figures 4 and 5, Table 1). Averaging change across joints reduced the variability.

JSW reduction was significantly greater in MCP than in PIP joints ($p < 0.001$), although only slightly greater as a percentage of initial value (Table 1A). In MCP, JSW reductions varied (not significantly) between fingers, with the

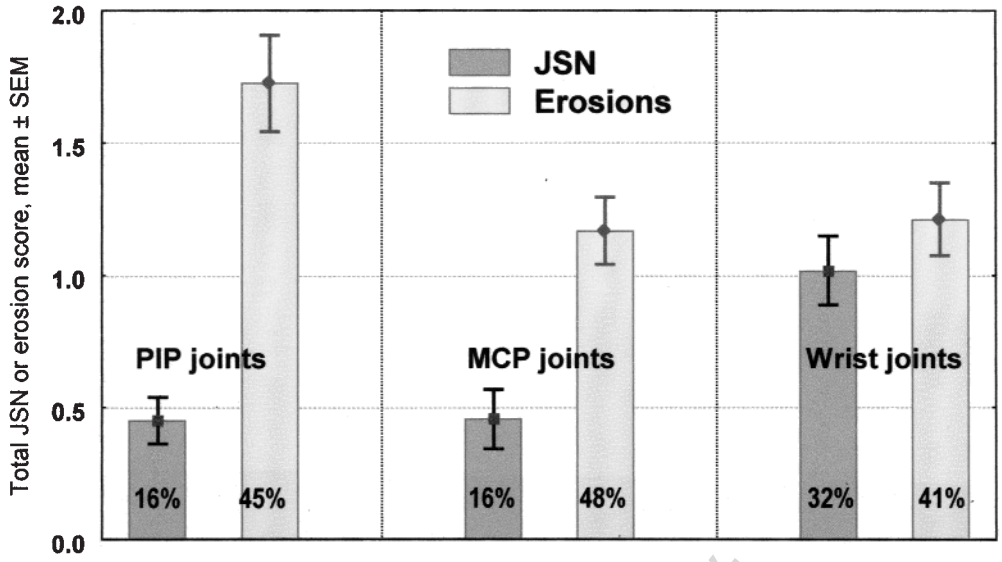

Figure 3. Variation in modified Sharp scores for joint space narrowing (JSN) and erosions for PIP, MCP, and wrist joints at the start of the study (mean ± SEM). MCP and PIP erosion and JSN scores each represent the aggregate of 10 joint scores for each patient. Wrist erosion and JSN scores represent the aggregate of 8 or 6 scores for each patient, respectively. Median and lower quartile values were all zero, while upper quartile values were in the range 0–2. Also shown is the percentage of patients with an initial score ≥ 1 for each component. Mean and SEM scores were determined from the full set of patient data including zero values. Combined scores are given in Table 1B.

middle finger tending to be worst affected (Figure 4). PIP JSW reductions were similar for all 3 fingers.

JSN score changes were observed as frequently as erosion score changes and were of similar magnitude (Figure 5, Table 1B). The index and middle fingers tended to be worst affected by both JSN and erosions.

Score and JSW changes were found to be similar in men and women and in left and right hands. The only significant variation occurred in PIP erosion score increases, which were greater for this patient population in the right than in the left hand ($p < 0.001$).

Signal to noise ratios. The SNR indicates the strength of the signal compared to the noise. Increasing the SNR increases the power of a study when other factors (patient population and required significance level) are unchanged[26-28].

SNR for component JSW change increased when aver-

Table 1. Summary of initial values and changes (mean ± SEM) and SNR for JSW (A) and scores (B). * JSW change as percentage of initial value; † numbers (%) of patients with JSN, erosion, or total scores (or score changes) ≥ 1; IQR: interquartile range. Scores were determined from the full set of patient data including zero values. PIP + MCP-hand was only determined for patients if both PIP and MCP-hand were available.

Table 1A	PIP-hand, 237 Patients	MCP-hand, 241 Patients	PIP + MCP-hand, 236 Patients
JSW measurements			
Initial value, mm	1.26 ± 0.014	1.82 ± 0.018	1.55 ± 0.015
Change, mm	–0.039 ± 0.005	–0.065 ± 0.008	–0.051 ± 0.005
Change, %*	–3.1	–3.6	–3.3
SNR, mean/SEM	8.2	8.3	9.6

Table 1B	JSN	Erosion	Total
Scores at initial timepoint, 245 patients			
Mean ± SEM	1.92 ± 0.23	4.11 ± 0.36	6.03 ± 0.53
Score ≥ 1† (%)	n = 108 (44)	n = 159 (65)	n = 170 (69)
Median (IQR)	0 (0–2)	2 (0–6)	3 (0–8)
Score change, 245 patients			
Mean ± SEM	1.83 ± 0.25	1.98 ± 0.28	3.81 ± 0.50
Score ≥ 1† (%)	n = 92 (38)	n = 90 (37)	n = 109 (44)
Median (IQR)	0 (0–2)	0 (0–2)	0 (0–4)
SNR, mean/SEM	7.3	7.1	7.6

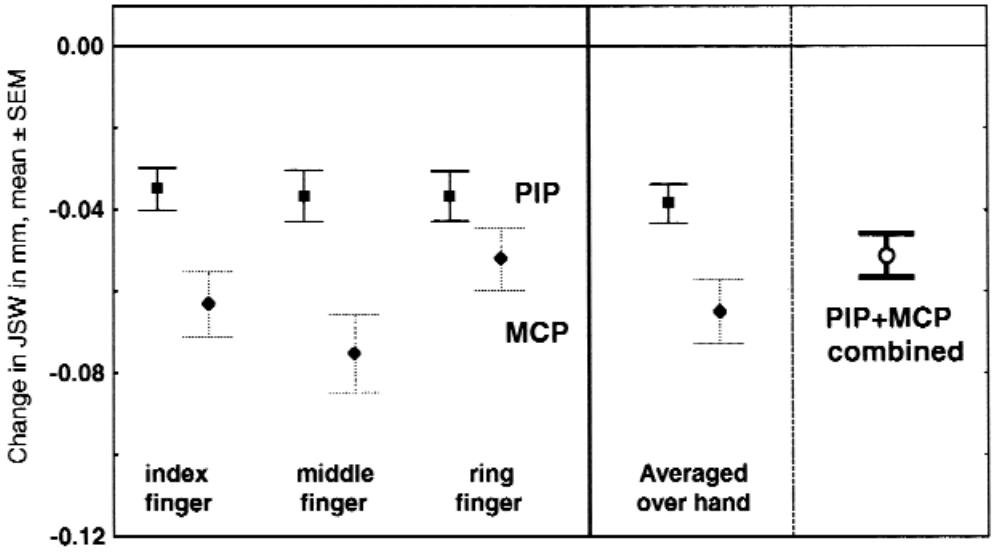

Figure 4. The paired changes in JSW (mm, mean ± SEM) determined from computer measurements of initial and final radiographs per patient. Negative values indicate a decrease in mean JSW with time. Left side shows changes averaged per patient over each finger (left and right hands). Right side shows changes averaged per patient over 6 PIP and/or 6 MCP joints (PIP-hand, MCP-hand, and PIP + MCP-hand).

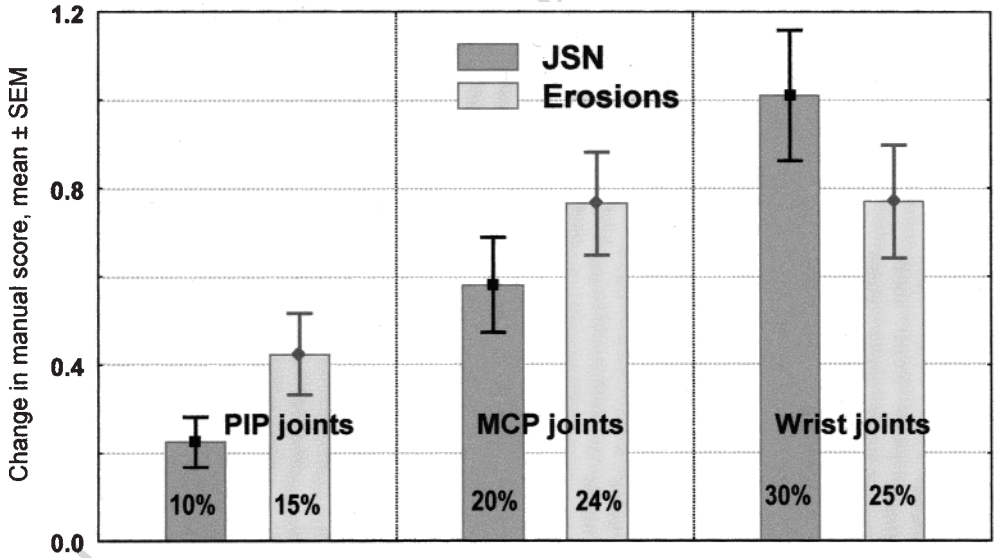

Figure 5. The paired change in modified Sharp joint space narrowing (JSN) and erosion scores (mean ± SEM) between the initial and final radiographs. Figure 3 legend gives details of the joints considered. A positive change reflects a deterioration in the patient's condition. Median and lower quartile values were all zero, while upper quartile values were in the range 0–1. Also shown is the percentage of patients with a change in score ≥ 1 for each component. Mean and SEM changes in score are determined from the full set of patient data including zero values. Combined score changes are given in Table 1B.

aged over all joints (compare PIP and MCP-hand to PIP + MCP-hand, Figure 6 and Table 1A). SNR for erosion and JSN score changes lie in the range 4.0–7.0 for PIP, MCP, and wrist joints (Figure 6), but increased when combined over all joints (7.1–7.6; Table 1B).

The improvements in the SNR on averaging reflect the degree to which the averaged elements complement, rather than strictly replicate, each other[27]. Although a proportion of noise is "unnecessary" (due to changes in hand position, radiography, scoring, or JSW measurement), a proportion is real inherent organic variation (due, for example, to disease affecting some patients and some joints more than others). The increases in SNR when erosion, JSN, and JSW changes are totalled, or averaged, across component joints reflect not

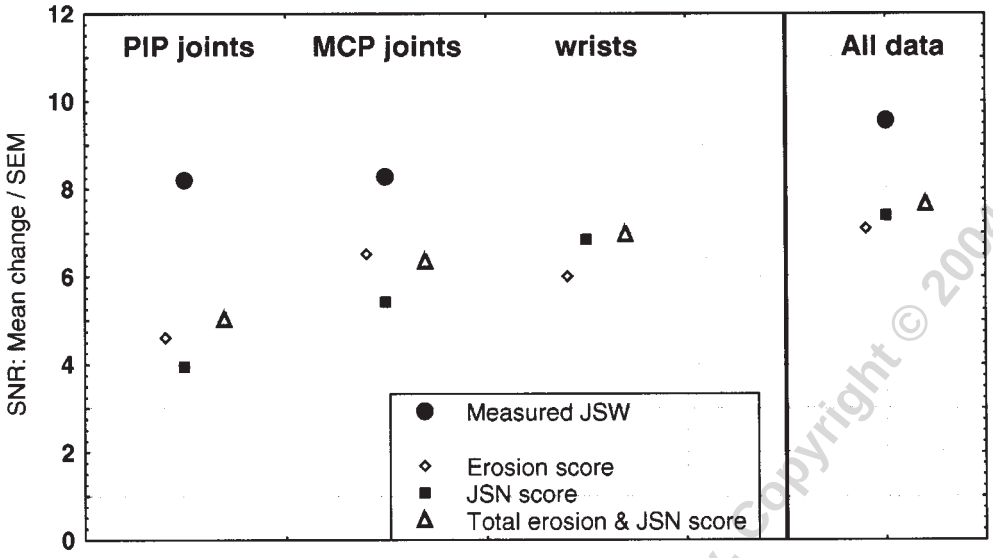

Figure 6. Comparison of signal to noise ratios (SNR) for measured joint space width (JSW) and for Sharp JSN and erosion scores. SNR is the ratio of the absolute mean change to its SEM. For all joints combined, the SNR for the change in JSW (decrease) is 25% greater than the SNR for the change in total Sharp score (increase), right panel. See legends for Figures 3 and 4 for details of the joints considered.

only a reduction in the unwanted noise but also the effect of averaging the differential responses of patients to RA, for example, the fact that MCP joints may be affected in one patient but PIP joints in another. If 2 measures precisely replicate each other there is no improvement in SNR on averaging[27].

The relatively small increase in SNR when JSN and erosion scores are combined (5% increase on average) indicates that the 2 scores give much the same information, i.e., the same patients are affected by erosions as are affected by JSN. The 25% improvement in SNR obtained by JSW measurement in contrast to scoring (Table 1) indicates the potential for increasing the study power.

Changes in JSW for patients with no change in score. Only a relatively small percentage of patients were affected by increases in erosion and/or JSN scores in any individual joint area (Figure 5). Overall, 44% of patients (109/245) experienced an increase in score over the study duration as agreed by both radiologists (Table 1B).

For those patients who had no change in score *in any joint* recorded by *either* reader, absolutely no differential information on the progress of joint disease could be derived from scoring alone. This occurred in 116/245 patients (47%). The patient group with a change in score (129/245) now comprised: 109/245 patients with score change ≥ 1; 8/245 patients with a recorded decrease in score; and 12/245 patients with a score change < 1, i.e., a unit change observed in a single joint by only one of the 2 readers. Decreases in score may have occurred due to an observed improvement in the joint condition or a recording error on the part of the readers.

The reductions in PIP and MCP-hand and combined PIP + MCP-hand were significantly greater for the patient group with an increase in score than for the group with no increase in score with PIP + MCP-hand changes –0.073 ± 0.008 mm and –0.027 ± 0.006 mm, respectively ($p < 0.002$; Figure 7). However, the JSW reductions were significant for both patient groups ($p < 0.001$). Thus, quantifiable information on the progress of disease in patients for whom there was no change in score could be obtained from JSW measurements.

Pairwise comparison of changes in JSW and scores. Only 134/2940 of the 6 PIP and 6 MCP joints measured per patient showed an increase in JSN score with time; 93/134 of these joints were measurable at both timepoints, and the majority of these (92%, 86/93) showed a corresponding reduction in measured JSW. For the remaining 7/93 joints, there was a recorded increase in JSW. However, this may have been due to small changes in hand position or individual measurement errors, as PIP + MCP-hand was reduced for 4/5 of the patients involved. The reduction in JSW was significantly greater for joints with JSN increase than for those with no change (–0.336 ± 0.033 mm, n = 93; –0.040 ± 0.003 mm, n = 2572, respectively).

Of the 92/245 patients who had an increase in JSN score ≥ 1 (totalled over 26 PIP, MCP, and wrist joints per patient), one patient had no measurable PIP or MCP joints, but 84% (76/91) of the remaining patients had a corresponding decrease in PIP + MCP-hand. Further, of the 109/245 patients who had an increase in total JSN *and* erosion score ≥ 1 (total of 54 joint scores per patient), one patient had no measurable PIP or MCP joints, but 84% (91/108) of the remaining patients had a corresponding decrease in PIP +

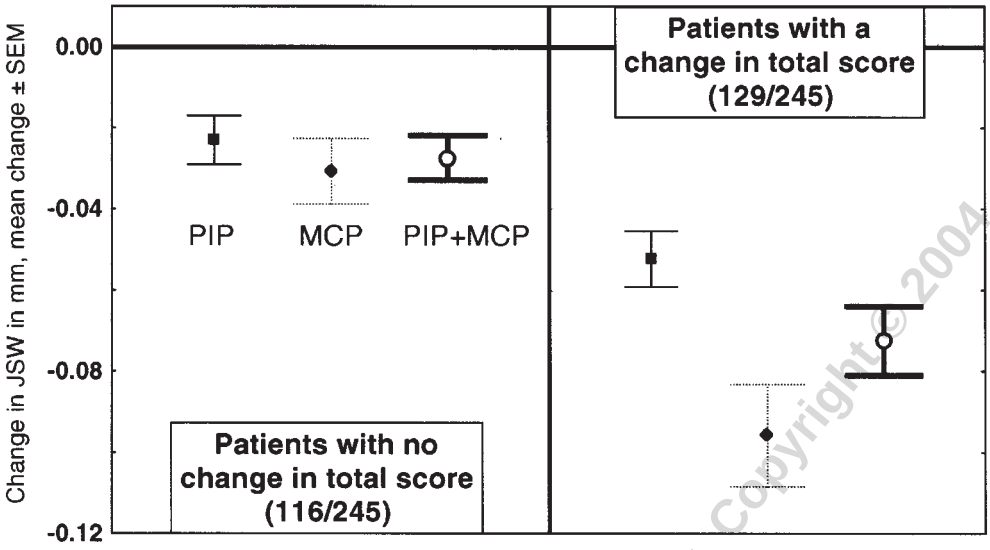

Figure 7. Comparison of changes in measured JSW (mm, mean ± SEM) for patients with and without a change in Sharp score (totalled for erosions and JSN over all joint areas). Decreases in JSW are significant for both patient groups for each joint area. However, changes are significantly greater for the patient group having a change in Sharp score (change ≠ 0) than for the group showing no change. See legends of Figures 3 and 4 for details of the joints considered.

MCP-hand (Table 2). This agreement of decrease in JSW with increase in both JSN and erosion scores suggests that sensitively measuring JSW may act as a surrogate for progressive erosive disease as well as loss of articular cartilage and joint space closure.

Changes in JSW and scores dependent on patient's expected genetic susceptibility. Genetic data relating to patients' expected susceptibility to more severe disease progression were available for 83% of the patients (203/245); 109/203 patients (group DR4+) were positive for the HLA-DR4 subtypes DRB1*0401 (Dw4) and DRB1*0404 (Dw14); 29/203 patients (group DR1+) were positive for the HLA-DR1 subtype DRB1*0101 (Dw1) but not heterozygous for the HLA-DR4 subtypes. By contrast, 65/203 patients (group DR4– + DR1–) were negative for both HLA-DR4 and DR1 subtypes. 8/203 of the genetically classified patients were not included because all PIP and/or all MCP JSW were eliminated: 4 patients in group DR4+ (2/4 with score change); one patient in DR1+ and 3 patients in group DR4– & DR1– (2/3 with score change). Inclusion of these patients in the scores study did not affect the results.

Table 3 shows the reduction in JSW and the increase in total Sharp score were significant for all patient groups. However, the reductions in measured JSW were significantly greater for patient groups susceptible to severe disease progression (DR4+, DR1+, DR4+ or DR1+) than for the less susceptible DR4– & DR1– group. Using either parametric or nonparametric statistical tests, the changes in scores for the patients more and less susceptible to severe progression were not significantly different.

Table 2. Comparison of changes in JSW (PIP + MCP-hand) and total Sharp score (JSN and erosions) over the study duration. JSW change could not be measured in any joint for 3 patients due to poor quality radiographs. * Here, PIP + MCP-hand includes changes for 5 patients with no PIP-hand measures and one patient with no MCP-hand measures, exclusion of these 6/242 patients did not significantly affect the results (Table 1). † JSW reduction in each of these 2 groups is significantly larger than in the group to its left ($p < 0.004$). The correlation between total score change and JSW change was significant ($p < 0.005$, correlation coefficient –0.63). There was also significant correlation between JSW change and component JSN and erosion score changes: correlation coefficients –0.58, –0.61 for JSN and erosions, respectively.

Score Change	No Change	1–10	11–20	21–30	> 30	Total
No. patients	136	78	18	9	4	245
Unmeasured patients	2	0	0	0	1	3
JSW*						
Change, mm	–0.026	–0.056†	–0.122†	–0.177	–0.331	–0.052
SEM, mm	± 0.006	± 0.008	± 0.023	± 0.024	± 0.150	± 0.006

Table 3. The effect of HLA-DR genotype on changes in measured JSW (PIP + MCP-hand) and total Sharp score (JSN and erosions) over the study duration. Patients in groups: DR4+, DR1+ and combined (DR4+ or DR1+) are expected to be susceptible to severe RA disease progression. Patients in group DR4– & DR1– are expected to be less susceptible. For each group, there is a significant reduction in JSW and a significant increase in Sharp score ($p < 0.001$). JSW is significantly decreased in the RA susceptible groups compared to the less susceptible group (t test: * $p < 0.01$, † $p < 0.05$). However, scores do not vary significantly between the susceptible and the less susceptible groups using either parametric (t test) or nonparametric (Mann-Whitney U test) statistics ($p > 0.14$). See legends, Figures 3 and 4, for details of the joints considered and Results for details of the genetic groups. IQR: interquartile range.

	Groups			
	DR4– & DR1–	DR4+	DR1+	DR4+ or DR1+
No. patients in group	62	105	28	133
PIP + MCP-hand, mm				
Mean	–0.032	–0.066*	–0.065†	–0.066*
SEM	0.008	0.009	0.016	0.008
Total score				
Mean	2.79	4.44	4.18	4.39
SEM	0.65	0.84	1.77	0.76
Median (IQR)	0 (0–3)	0.5 (0–6)	0 (0–4)	0.5 (0–6)

DISCUSSION

We set out to determine whether a computerized measurement system can quantify organic changes in radiographic JSW due to RA. Retrospective radiographs were obtained from a clinical trial (unpublished data) conducted at centers across the USA in the period 1987 to 1990. The radiographs provide a good database, as they record the first and last visit for a group of 245 patients with early RA (disease duration < 1 year at recruitment) collected over a period up to 2 years following carefully specified hand positioning procedures. Clinical and HLA genetic data were available from the majority of patients. The radiographs had been blinded for patient name, center, treatment, and time sequence and had been scored independently by 2 radiologists, who had, however, been aware of the time sequence. At a later date, the mean JSW was measured for 3 MCP joints and 3 PIP joints on each hand for each patient using the same pairs of first and last radiographs but, in this case, blinded to time sequence.

The deterioration in patients' joints was evident both in the significant increase in erosion and JSN score and in the significant decrease in measured JSW over the study duration (Figures 4 and 5). In 92% of measured joints where there was an increase in JSN score, there was a corresponding reduction in JSW measurement. More remarkably, in 84% of patients where there was an increase in total score, including erosion and wrist joint scores as well as JSN scores, there was a corresponding reduction in PIP + MCP-hand. Thus, the reduction in JSW in the 6 PIP and 6 MCP joints measured per patient was indicative of the general degree of joint damage in that patient as represented by the change in score.

To investigate whether JSW measurement is a more sensitive index of joint progression than JSN scoring, the SNR of the change in PIP + MCP-hand was compared to the SNR of the change in total erosion and JSN score. We found the JSW SNR (9.6) was 25% greater than the total score SNR (7.6), reflecting the increase in study power obtainable when radiographic JSW is measured (Figure 6). The number of patients required in a study in order to obtain a significant result is roughly proportional to $(1/SNR)^2$, and hence increasing the SNR by 25% enables a theoretical reduction in patient numbers of 36%, i.e., from 245 to 157 patients in the present case. Rather than reduce patient numbers, it would perhaps be preferable to use the increased power to enable more information to be obtained from the study or to reduce study duration. For example, it is shown that JSW measurement enables differential information to be obtained on the 47% of patients with early RA for whom there was no change in score and to differentiate subgroups of patients with genetic susceptibility to severe disease progression, neither of which could be differentiated by scoring.

The problems that could arise when the JSW measuring procedure is applied to clinical studies where the joints are affected by disease are considered in the light of experience gained from this study. The SD of change in hand-averaged JSW for the RA patients (Table 1A) are 4–5 times greater than the SD previously found for 8 healthy volunteers[16], when radiographs were taken on 5 separate occasions over a period of 3 weeks with no organic change in JSW expected. The increased variability in the RA study is not surprising, considering the longer separation between radiographs (11–113 wks, median 69 wks) with consequent greater expected variability in hand positioning and radiographic techniques; additional difficulties in positioning hands due to disease; the less stringent standard of quality control required from the technologists measuring JSW under "field trial" conditions; and significant variations between patients in the rate of JSW loss due to disease.

In general, there was no difficulty in measuring either the smaller JSW encountered in women or joints affected by osteoporosis or erosions. However, fused joints and joints with abutting margins or excessive finger flexion could not be measured. In most cases, only individual joints were affected and the patient had other measurable joints (242/245 patients). Radiographic technique and hand positioning were clearly specified in the clinical trial protocol. However, errors did occur and more stringent standards of quality control could be implemented.

The differential changes in JSW that correlated with HLA genotype are consistent with the role these genes may play in disease pathogenesis. Particular HLA-DRB1 alleles are associated with susceptibility for erosive disease in both retrospective and prospective studies[30-33]. Patients enrolled in our study were within one year of onset of initial symptoms, representing an early disease cohort. In this group, JSW appears to be a sensitive indicator of early disease progression in genetically at-risk subjects, consistent with a progressive course of disease initiating at a very early stage.

A new RA measurement technique should be assessed against the OMERACT filter of truth, discrimination, and feasibility[29]. The "truth" of JSW measurements is established through the correlation of JSW changes with changes in Sharp score. The ability of JSW measurements to discriminate between different patient groups has been demonstrated. Within the framework of clinical trials, JSW measurements are feasible and can be carried out by relatively untrained personnel, although improvements to the technique are possible. More detailed attention to the hand positioning in repeat radiographs and to the standards of measurement quality control would be required if the progress of individual patients were to be monitored. Repeat radiographs at a single timepoint are not available for this cohort of RA patients and thus the smallest change due to disease that can be detected (smallest detectable difference, SDD) could not be determined[23]. Knowledge of the SDD is necessary if the progress of individual patients is to be tracked or if the number of patients with "disease progression" (according to this criterion) is to be assessed. The SDD, or individual cutoff, was considered in an earlier study of healthy subjects[16] and reasons for the increase in variability for the present cohort of patients have been suggested. Discrimination combines qualities of sensitivity and repeatability, i.e., strong signal, low noise, as described in this report, and it is concluded that radiographic JSW measurements satisfy the OMERACT filter for monitoring disease progression in RA clinical trials.

This study shows that, in hand radiographs of patients with RA, the measurement of changes in mean JSW for 6 PIP and 6 MCP joints per patient gives important quantitative information on the progress of disease that is not available from consideration of changes in Sharp erosion (28 joints) and JSN (26 joints) scores alone. The measured JSW changes reflect the general degree of joint damage to the patient, as represented by changes in Sharp score, but provide increased differential information, particularly for the 47% of patients (116/245) for whom there was no change in score. The measurements significantly increase the power of the study, enabling a greater yield of information. The application of such technology to clinical trials of RA therapy could reduce sample size, detection limits, or the required duration for detection of joint progression.

ACKNOWLEDGMENT

The authors thank Bruce Wallin for coordinating the original clinical trial, Darren Larsen and Stephanie Parvin for ably performing the radiological JSW measurements, Martin Rimmer for statistics advice, and Peter Hunter for computing support.

REFERENCES

1. 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.
2. Scott DL, Houssien DA, Laasonen L. Proposed modification to Larsen's scoring methods for hand and wrist radiographs. Br J Rheumatol 1995;34:56.
3. Larsen A. How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in long-term studies. J Rheumatol 1995;22:1974-5.
4. Rau R, Herborn G. A modified version of Larsen's scoring method to assess radiologic changes in rheumatoid arthritis. J Rheumatol 1995;22:1976-82.
5. van der Heijde D. Plain X-rays in rheumatoid arthritis: overview of scoring methods, their reliability and applicability. Baillieres Clin Rheumatol 1996:10:435-53.
6. 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.
7. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.
8. Sharp JT. An overview of radiographic analysis of joint damage in rheumatoid arthritis and its use in metaanalysis. J Rheumatol 2000;27:254-60.
9. Rau R, Wassenberg S. Scoring methods. J Rheumatol 2002;29:653-5.
10. Buckland-Wright JC, Carmichael I, Walker SR. Quantitative microfocal radiography accurately detects joint changes in rheumatoid arthritis. Ann Rheum Dis 1986;45:379-83.
11. Buckland-Wright JC, MacFarlane DG, Lynch JA, Clark B. Quantitative microfocal radiographic assessment of progression in osteoarthritis of the hand. Arthritis Rheum 1990;33:57-65.
12. Higgs JB, Smith D, des Rosier KF, Charlesworth RW Jr. Quantitative measurement of erosion growth and joint space loss in rheumatoid arthritis hand radiographs. J Rheumatol 1996;23:265-72.
13. Allander E, Forsgren PO, Pettersson H, Seideman P. Computerized assessment of radiological changes of the hand in rheumatic diseases. Scand J Rheumatol 1989;18:291-6.
14. James MF, Heald G, Shorter JH, Turner RA. Joint space measurement in hand radiographs using computerized image analysis. Arthritis Rheum 1995;38:891-901.
15. 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.
16. Angwin J, Heald G, Lloyd A, Howland K, Davy M, James MF. Reliability and sensitivity of joint space measurements in hand

radiographs using computerized image analysis. J Rheumatol 2001;28:1825-36.
17. Duryea J, Jiang Y, Countryman P, Genant HK. Automated algorithm for the identification of joint space and phalanx margin locations on digitized hand radiographs. Med Phys 1999;26:453-61.
18. Duryea J, Zaim S, Wolfe F. Neural network based automated algorithm to identify joint locations on hand/wrist radiographs for arthritis assessment. Med Phys 2002;29:403-11.
19. Ropes M, Bennett GA, Cobb S, Jacox R, Jessar R. Proposed diagnostic criteria for rheumatoid arthritis. Bull Rheum Dis 1956;7:121-4.
20. Ropes M, Bennett GA, Cobb S, Jacox R, Jessar R. 1958 revision of diagnostic criteria for rheumatoid arthritis. Bull Rheum Dis 1958;9:175-6.
21. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
22. Nepom GT, Byers P, Seyfried C, et al. HLA genes associated with rheumatoid arthritis. Identification of susceptibility alleles using specific oligonucleotide probes. Arthritis Rheum 1989;32:15-21.
23. van der Heijde D, Simon L, Smolen J, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion. Arthritis Rheum 2002;47:215-8.
24. Canny JF. A computational approach to edge detection. IEEE Trans on Pattern Analysis and Machine Intelligence 1986; PAMI-8:679-98.
25. Torre V, Poggio T. On edge detection. MIT Artificial Intelligence Laboratory, Cambridge, MA, 1984; Report AIM-768.
26. Jones G, Halbert J, Crotty M, Shanahan EM, Batterham M, Ahern M. The effect of treatment on radiological progression in rheumatoid arthritis: a systematic review of randomized placebo-controlled trials. Rheumatology Oxford 2003;42:6-13.
27. Goldsmith CH, Smythe HA, Helewa A. Interpretation and power of a pooled index. J Rheumatol 1993;20:575-8.
28. Cohen J. Statistical power analysis for the behavioral sciences. Revised ed. Orlando: Academic Press Inc.; 1977.
29. Boers M, van der Heijde DM. Prevention or retardation of joint damage in rheumatoid arthritis: issues of definition, evaluation and interpretation of plain radiographs. Drugs 2002;62:1717-24.
30. Nepom GT, Nepom BS. Prediction of susceptibility to rheumatoid arthritis by human leukocyte antigen genotyping. Rheum Dis Clin North Am 1992;18:785-92.
31. Gough A, Faint J, Salmon M, et al. Genetic typing of patients with inflammatory arthritis at presentation can be used to predict outcome. Arthritis Rheum 1994;37:1166-70.
32. Valenzuela A, Gonzalez-Escribano MF, Rodriguez R, Moreno I, Garcia A, Nunez-Roldan A. Association of HLA shared epitope with joint damage progression in rheumatoid arthritis. Hum Immunol 1999;60:250-4.
33. Wagner U, Kaltenhauser S, Sauer H, et al. HLA markers and prediction of clinical course and outcome in rheumatoid arthritis. Arthritis Rheum 1997;40:341-51.