Measurement of Joint Space Width and Erosion Size

JOHN T. SHARP, DESIREE van der HEIJDE, JANE ANGWIN, JEFF DURYEA, HEINS J. BERNELOT MOENS, JOHANNES W.G. JACOBS, JEAN-FRANCIS MAILLEFERT, and C. VIBEKE STRAND

ABSTRACT. Measurement of radiographic abnormalities in metric units has been reported by several investigators during the last 15 years. Measurement of joint space in large joints has been employed in a few trials to evaluate therapy in osteoarthritis. Measurement of joint space width in small joints has been reported by several investigators but has not yet found a place in clinical trials in rheumatoid arthritis or osteoarthritis. We review methods for measuring joint space width in finger, toe, and wrist joints; special attention is given to how the joint edges are found, the method used to measure distance between joint margins, size of an area of the sampled joint, and reproducibility of measurements. Methods for measurement of erosion size, which have had less attention, are briefly discussed. (J Rheumatol 2005;32:2456–61)

Key Indexing Terms:
RHEUMATOID ARTHRITIS
OSTEOARTHRITIS
RADIOGRAPHY
EROSIONS
JOINT SPACE MEASUREMENT

Scoring erosions and joint space width (JSW) by trained physicians as an estimate of disease severity has proven to be a useful method of evaluating structural damage in rheumatoid arthritis (RA)1-3. Comparing estimates on serial radiographs of patients in controlled trials is considered necessary to determine whether a drug is accepted as having the capacity to slow progression of structural damage. A number of studies have demonstrated limited precision in scoring and have indicated that scoring is both highly evaluator-dependent and sensitive to variation in radiographic technique. There is considerable variation between readers even when those readers are close associates with the same training for scoring radiographic abnormalities. In spite of these limitations, scoring has proven to be highly successful in detecting effectiveness of drug therapy in RA when large numbers of patients are included in each treatment arm in a therapeutic trial.

Clearly, measurement in standard metric units would be preferable to scoring erosions on an ordinal scale if measurement methods could be shown to be highly reproducible and easily applied. Initial attempts have been focused on measuring JSW, with only a few attempts at measuring erosions. Early measurements were carried out on large joints, especially the hip and knee, using magnifying lenses with scales included in the lens system or with calipers4-6. More recently, computer based methods have been developed for measuring JSW, and some studies have reported measurement of erosion size.

Computer Based Methods

Dacre and Huskisson described a computer based method for measuring joint space in the knee about 15 years ago7. Buckland-Wright, et al reported methods for measuring joint space in the knee in osteoarthritis (OA) using a Sun Spark platform, which requires that the operator locate a few points on the joint margins8. Duryea, et al reported an automated algorithm to delineate the joint margins in digital radiographs of the knee9,10, and Conrozier, et al also reported a computer based method for measurement of hips in patients with OA11.

A method developed by James, et al measures joint spaces in proximal interphalangeal (PIP) and metacarpophalangeal (MCP) joints of the hand using a PC platform; it requires that the operator locate only one nearby point to identify the joint to be measured12. Duryea, et al reported a method using neural networks to measure PIP and MCP JSW automatically after the operator identifies nonanatomical structures in the image10. Sharp, et al reported a program to measure finger and wrist JSW that runs on US National Institutes of Health (NIH) Object Image using a Macintosh
computer system. This program was adapted to measure hip JSW and recently was evaluated by comparing with manual measurements. The program has been rewritten as a Java plug-in to use Image J, a freeware program that runs on a Macintosh, PC, or Linux platform. Angwin and colleagues have modified and updated the method of James, et al. Vischer, et al wrote a macro program to measure JSW based on NIH Object Image using a Macintosh computer; this is also available on the Internet as freeware.

Reliability

Numerous factors affect reproducibility of joint measurements, such as patient positioning, radiographic procedure, site of measurement, measuring methods, and the readers.

Attempts have been made to increase the reproducibility of joint space measurement of large joints by standardizing radiographic procedures. These attempts have been especially focused on knee measurements, since the variability of positioning is much greater at the knee than the hip. These protocols use fluoroscopy and/or positioning standards.

Compared with manual measurements, computer based methods should increase reliability through decreasing reliance on the operator and variation in the reading procedure. They also offer the possibility to evaluate the mean JSW across the entire joint, or the joint space area, rather than the JSW at the narrowest point. Intuitively, the mean measurement might be more relevant than JSW at the narrowest point, although this is not yet proven.

Reproducibility of measurements has been reported in several studies. Large joints can be measured within an error of 0.5 to 1 mm and small joints within an error of 0.1 to 0.5 mm. With this range of precision it is estimated that a difference in mean width of 0.04 to 0.07 mm between 2 treatment groups with 50 to 200 subjects per group can be detected with 95% confidence.

Duryea, et al evaluated their software for measuring knee joint space using images of normal and osteoarthritic patients. Reproducibility of the minimum JSW was about twice as good as manual measurement. Mazzuca, et al performed repeat measurements of minimal JSW on 174 osteoarthritic knees radiographed twice within 7 days in the same radiology unit. Standard error of the measurement (SEM) was 0.32 mm. The SEM for radiographs repeated at different radiology units was 0.45 mm. Gordon, et al reported the root mean square standard deviation (SD) for 19 duplicate measurements of the minimal JSW of hips to be 0.12 mm.

Maillefert, et al reported the smallest detectable difference (SDD, i.e., 1.96 × SD of the differences between repeated measurements) was 0.78 mm for minimal JSW using manual measurement on OA hips, 0.67 mm for computer based measurement of minimal JSW, and 0.47 mm for average JSW. Conrozier, et al reported SDD of 0.22 and 0.26 mm for minimal and mean hip JSW using a computer-based method. Vignon, et al found a SDD of 0.50 mm for the minimal knee JSW using a computer-based method and Lyon schuss radiographs.

Comparison between methods is very limited, and the considerable variation between methods is not readily explained. Nevertheless precision of measurement in these ranges is sufficiently good to encourage further exploration of these computer based methods.

Studies on small joints of the hands, feet, and wrists also have shown variability between methods. In a reliability study measuring JSW in fingers of healthy volunteers under relatively ideal conditions, Angwin, et al estimated the SDD for change in JSW to be 0.11 mm for an individual PIP or MCP joint. In a RA clinical trial, with less controlled radiographic techniques and disease affecting the positioning of the patient’s hands, the smallest detectable organic change could well be larger. However, precision in this range is encouraging and further exploration of these computer based methods is warranted.

There are minimal data on the expected progression rate of joint space narrowing in the normal population. Conrozier, et al studied 69 hips in 61 patients who had at least 2 radiographs 12 months apart between diagnosis of OA and total hip arthroplasty; mean yearly decrease in JSW was 0.43 mm. Auleley, et al reported a 3-year decrease in hip JSW ranging from 0.47 to 0.6 mm, using a manual measurement method. A 3-year mean change in minimal hip JSW of –0.53 and –0.6 mm (manual and computer based measurements, respectively) was observed by Maillefert, et al. Vignon, et al, using a computer based method, reported a 2-year decrease in knee minimal JSW of 0.17 mm using an extended anteroposterior view and 0.24 mm for the Lyon schuss view.

In the small joints of the hands in RA patients, James, et al in a retrospective study reported an average yearly change in JSW of –0.03 mm in PIP and –0.06 mm in MCP joints in 16 female patients with established disease. There was, however, no significant change in PIP or MCP JSW for 18 male patients in the study. Using a later version of James’s computer program in an 18 month retrospective study on 245 patients with early RA (176 female, 69 male) Angwin, et al found the average yearly JSW changes were –0.03 mm in PIP and –0.04 mm in MCP joints. In both studies, JSW was measured blind to time-order in 3 PIP and 3 MCP joints in each hand per patient per time point. Changes found were significant at p < 0.05. No attempt was made to divide clinically involved from noninvolved joints. Measurement changes agreed with scoring in that the reduction in JSW was significantly greater for the 93 PIP and MCP joints with an increase in joint space narrowing score than for the 2572 joints with no change in joint space narrowing score: –0.336 mm and –0.040 mm, respectively, over the 18 month mean study. Lacking robust data on the progression of narrowing...
Asymmetrical joint space is another frequently encountered problem. Many joints are slightly asymmetrical, which may be imperceptible to the naked eye and only detected by precise measurements across all regions of the joint. Metatarsophalangeal joints present another problem: often they are wider in the center than on either medial or lateral sides because of a flat metatarsal head seated in the base of an arc-shaped proximal phalanx. What is the proper measurement of these joints?

In cases with evident gross asymmetry in the PA projection there may be subluxation, but it is not possible to detect this in a 2-dimensional view unless there is bony overlap. In the Sharp and Sharp/van der Heijde scoring methods this asymmetry is scored as 1 on the joint space narrowing scale of 0–4. How should such grossly asymmetrical joints be measured? Neither an average across the breadth of the joint, medial to lateral, nor the shortest measurement across the joint at the narrowest point accurately represents the joint space.

Collecting and analyzing measurement data on cohorts that include such joints requires establishing a standard method of handling these abnormalities. Some of these questions can be answered with more experimental data. In other situations in which relevant data are indeterminate because of insufficient data or inconsistent results, standards will have to be based, at least temporarily, on a consensus among active investigators.

Differences Between Methods
Some substantial differences between methods exist, but their effect has not been tested. In principle, results between methods should be interchangeable, but several factors must be well controlled for this to be true. Radiographic technique should be standardized with regard to distance from x-ray source to object and object to film. Positioning of the joints to be measured must be well controlled to limit rotation in any plane. This is particularly important in measuring large joints because of the considerable distance between the joint and the image cassette, and of minor importance in small joints that are imaged very close to the film or phosphor plate. Selection of the area of the joint to be measured should be uniform between measurements. Methods of identifying the joint margin vary between measurement methods. As long as the same method is used for the baseline and followup radiographs, this difference should not matter in following progression of disease. In collecting data on change in measurement, both measurements should be done by the same method. Comparing such change data obtained by different methods should be valid if the difference between methods is known to be an offset that is constant for all types of joints. For example, the James method identifies the joint margin as the pixels with the greatest density in the dense line at the base of the proximal or middle phalanges in the finger joints, whereas the Sharp...
method locates the joint margin nearer the shoulder of the joint edges. The distance between the shoulder and the peak density of the joint margin might be relatively constant for specific joints, but can vary between different types of joints, for example, toe, finger, and wrist joints.

Calculation of the JSW has utilized several different
methods. In the Lynch method the largest circle that is completely contained between joint margins is fitted at regularly spaced intervals, and the diameter of these circles is averaged to determine the distance between the joint margins. Sharp, et al obtain the shortest distance across the joint for every pixel on the x-axis of the distal margin of the joint by iterative measurement, moving a line anchored on the outer margin one pixel at a time on the inner margin. Vischer creates a series of trapezoids contained within the joint space to measure the shortest distance between the 2 joint margins (unpublished observations).

All the above methods that make multiple, evenly spaced measurements generate data that would allow calculation of a correlation coefficient between the location of each measurement and the JSW for that location. Sharp, et al routinely report this calculation along with the slope, which together make it possible to evaluate the extent of the asymmetry.

Before employing JSW measurement in clinical trials extensive basic data on normal joints and progression of JSW loss and variance of measurements are needed in cohorts stratified for age, sex, physical activity, and disease. Determining which joints to measure in which disease and disease stage in terms of sensitivity to change would be of great value. For example, the joints to measure in OA and RA would likely be different, as well as for early versus long-standing RA.

In summary, limited relevant JSW measurement data are currently available and few comparisons have been made between methods. Nevertheless it is already clear that computer based measurement of JSW is highly reproducible within limits of less than 1 mm for large joints and less than 0.4 mm for small joints. Reproducibility and sensitivity to change can probably be improved by further refinement of computer programs. The data already available suggest that measurements are sufficiently sensitive to be useful in following disease progression in the hands of patients with RA, but whether measurements would be more sensitive to change than scoring has not yet been determined.

Erosions

Three methods to measure erosion size have been reported. Buckland-Wright, et al, using digitized macroradiographs and a projection unit, traced the erosion margins on a digitizing tablet to measure erosion area in the 2-dimensional representation of the erosion. Higgs, et al created templates of different shapes and known area to match observed erosions for measuring erosion size. Sharp, et al scanned plain radiographs and measured optical density (OD) in an area outlined to include the erosion. OD was expressed in units standardized for the third finger proximal phalanx and compared to a similar but uninvolved anatomical area. The latter method gives an estimate of the volume of the eroded area expressed in units calibrated for the density in the same patient’s third proximal phalanx, provided certain conditions are met. The method is based on linear absorption of x-rays over a range of about 5 log, adequate dynamic range of the x-ray film used to capture the image, and a scanner with sufficient range to encompass the range of bone density in the calibration area and region of interest. These conditions are not always met unless careful attention is paid to radiographic technique and appropriate quality of film or light-sensitive imaging plate.

Development of useful methods of measuring erosions is not as far along as measurement of JSW and is clearly more challenging. Finding the distance between 2, usually nearly parallel, lines is not a difficult problem, given the tools available to skilled computer programmers. Finding the edge of an erosion is a more nebulous proposition. Active erosions are characterized by no perceptible edge. Instead, normal bone density gradually fades into the background soft tissue density. Because this challenge is daunting, investigators tend to look for more tractable problems for which, with reasonable effort, a definitive answer is expected.

Measurement in Clinical Trials

As of now, measurements have not made a significant impact on clinical trials. Only a few have been reported. Within the last few years several investigators have focused attention on measuring changes in OA in the hopes that new agents can be developed to slow the loss of cartilage. If measurement proves to be more sensitive than scoring, measurement will play a major role in the evaluation of putative OA disease modifying agents. In RA, our greatest attention has been given to progression, stabilization, and repair of erosions. Since there is some evidence that destruction of cartilage and bone are not tightly linked in RA, there is a potential role for studies that evaluate cartilage loss in a more precise fashion than is possible with scoring. If a more sensitive and reproducible method than scoring can be developed, there will be a role for such methods in reducing the number of patients required in studies evaluating disease modifying agents in RA. Further effort to improve measurement methods is warranted.

REFERENCES


