Plain film radiography is the preferred method for evaluating disease progression in rheumatoid arthritis and for establishing the efficacy of new disease modifying antirheumatic agents. However, the relative efficacy of these agents cannot be determined by comparing radiographic data from different studies, and a standardised system is needed.

Several new disease modifying antirheumatic drugs (DMARDs) for the treatment of rheumatoid arthritis (RA) have recently been introduced, including etanercept, infliximab, leflunomide, and anakinra, and more will soon be available. These agents come at a time when early, aggressive treatment to modify the course of the disease has gained wide acceptance, and the question of whether to use a combination of agents to increase efficacy is increasingly being asked. Quantitative plain film radiography has been the most common imaging modality used to follow disease progression and to assess drug efficacy.

An ideal method for monitoring the progression of RA has several characteristics. Firstly, the method must have a high sensitivity for detecting early disease, when DMARDs have their greatest potential benefit. The ability to define single point damage, quantify severity, and monitor progression with accuracy, precision, and high reproducibility are also important characteristics of an ideal monitoring method. Other features of an ideal method include widespread availability, cost effectiveness, lack of patient harm, ease of use, and rapid generation of results. Creation of a permanent record that can be easily studied in a randomised and blinded fashion. Radiographs are reproducible, allow measurement of severity, and can identify single point damage and progression with fairly high precision and accuracy. In addition, the overwhelming majority of reported data related to radiological progression of RA comes from plain film radiography.

However, there are several disadvantages to plain film radiography. Firstly, technical limitations include the need for proper technique and positioning, which can falsely obscure or enhance various findings on an initial film or follow up radiographs. Secondly, there are “floor” and “ceiling” effects related to detection and scoring of RA induced disease seen on conventional radiographs. The floor effect stems from the fact that the hallmark radiographic findings of bony erosions and joint space narrowing may occur late in the pathophysiology of the disease. The ceiling effect refers to the fact that radiographic progression of disease can continue even after the highest damage score has been assigned. Another somewhat controversial issue with plain film radiography is the relationship of X-ray findings information is equivocal, contradictory, or non-diagnostic. This modality can also be used to document, quantify, and monitor the amount and location of joint disease.

The structural damage of RA disclosed in plain film radiographic findings usually first occurs in the hands, wrists, and feet. Soft tissue swelling, marginal erosions, juxta-articular osteopenia, and uniform joint space loss are often seen fairly early in the course of RA. In the hand, the changes of RA are primarily seen in the metacarpophalangeal and proximal interphalangeal joints, while the intercarpal, radial, and ulnar styloid joints are most commonly affected in the wrist. In the feet, changes are most commonly seen in the metatarsophalangeal and proximal interphalangeal joints.

“Plain film radiography has many advantages but also some disadvantages”

Using the reference characteristics of an ideal method for monitoring RA discussed above, many of the advantages of plain film radiography are clear. Plain radiographs are inexpensive, easy to generate, and widely available and accepted. They also give rapid results and provide a permanent record that can be easily studied in a randomised and blinded fashion. Radiographs are reproducible, allow measurement of severity, and can identify single point damage and progression with fairly high precision and accuracy. In addition, the overwhelming majority of reported data related to radiological progression of RA comes from plain film radiography.

Correspondence to:
P A Ory, Southwest Medical Imaging, Seattle, Washington, USA
Dr P A Ory, Southwest Medical Imaging, 16259 Sylvester Road SW, Seattle, WA 98166, USA; ory@attbi.com

Accepted 20 January 2003

Abbreviations: DMARDS, disease modifying antirheumatic drugs; FDA, Food and Drug Administration; MTX, methotrexate; NSAIDs, non-steroidal anti-inflammatory drugs; RA, rheumatoid arthritis

Plain film radiography is the preferred method for establishing the efficacy of new disease modifying antirheumatic agents. However, the relative efficacy of these agents cannot be determined by comparing radiographic data from different studies, and a standardised system is needed.

S

everal new disease modifying antirheumatic drugs (DMARDs) for the treatment of rheumatoid arthritis (RA) have recently been introduced, including etanercept, infliximab, leflunomide, and anakinra, and more will soon be available. These agents come at a time when early, aggressive treatment to modify the course of the disease has gained wide acceptance, and the question of whether to use a combination of agents to increase efficacy is increasingly being asked. Quantitative plain film radiography has been the most common imaging modality used to follow disease progression and to assess drug efficacy.

An ideal method for monitoring the progression of RA has several characteristics. Firstly, the method must have a high sensitivity for detecting early disease, when DMARDs have their greatest potential benefit. The ability to define single point damage, quantify severity, and monitor progression with accuracy, precision, and high reproducibility are also important characteristics of an ideal monitoring method. Other features of an ideal method include widespread availability, cost effectiveness, lack of patient harm, ease of use, and rapid generation of results. Creation of a permanent record that can be easily randomised and blinded is also desirable. A final important characteristic of an ideal method for monitoring the progression of RA is correlation with clinical disease course, which can often be fluctuating.

The current “gold standard” for monitoring the progression of RA is plain film radiography, which has both advantages and disadvantages. The goals of this manuscript are to examine the applicability of quantitative radiographic evaluation to DMARD research and to explore the comparability of radiographic data across different trials, with a focus on data from the most recently approved DMARDs.

**IMAGING TECHNIQUES IN RA**

Plain film radiography is the most commonly used imaging modality for initial clinical evaluation and monitoring of RA. Plain radiographs can confirm a diagnosis of RA or even allow the diagnosis to be made when the clinical and laboratory
with clinical disease course. Multiple studies have been performed in an attempt to correlate commonly used clinical indices, such as the Health Assessment Questionnaire, grip strength, and the Ritchie index, with radiographic findings. The results have been variable. More recent data from clinical trials, however, indicate a clear relation between disease activity (particularly inflammation) and structural damage (reviewed by van der Heijde).

Other modalities, such as magnetic resonance imaging, ultrasound, and radionuclide imaging, have been used to assess joint changes in RA.14,15,16 Of these, magnetic resonance imaging holds the most promise for future use, but cost effectiveness standards17 and predictive value have yet to be established.

USING PLAIN FILM RADIOGRAPHY TO EVALUATE RA PROGRESSION

After reviewing 60 published reports and four data sets, Scott and colleagues recently published an analysis of the link between radiographic findings in RA and disability over the course of the disease.17 The authors concluded that a relatively strong, and most likely causal, relationship exists between joint damage and later disability. Both increase over time, as does the correlation between the two. x-ray damage and disability are not correlated in early RA; once radiological scores exceed 33% of maximum damage, the relationship between damage and disability becomes more linear. In late RA (>8 years), joint damage and disability are most strongly correlated (Pearson r = 0.30–0.70). Joint damage accounts for about 25% of disability in established RA.17 Radiographic progression is also associated with income loss and work disability in patients with RA.18 These findings support the notion that increased joint damage results in severe, long term consequences.

Because of the importance of radiographic progression in determining long term outcomes, a standardised, systematic method to evaluate and quantify the amount and progression of radiographic damage caused by RA is desirable. A quantitative approach to characterising joint damage and damage progression offers several advantages over a system that relies on a qualitative evaluation, such as “better,” “the same,” or “slightly worse.”

Data have shown that quantitative, systematic approaches to evaluating the status of RA induced joint damage can result in a high inter- and intraobserver correlation.19 Such methods also allow population means to be created, which enables more accurate comparisons between groups, possibly even across different studies.20

“Radiographic scoring systems differ in the joints examined and the method of scoring damage”

Using empirical reasoning that a systematic approach would be desirable, in 1949 Steinbrocker et al developed the first scoring system to quantify radiographic evaluation of RA.20 This system involved classifying RA induced damage into one of four gross stages. Since that time, many scoring systems have been developed (table 1).21 Most scoring systems look at certain joints within the hands and wrists or selected joints within the hands, wrists, and feet, primarily because these joints are easy to evaluate and a reasonable correlation between disease in these joints and total disease burden in other joints has been demonstrated.22 For clinical trials, investigators have increasingly advocated evaluation of the hands, wrists, and feet. Studies have shown that the joints of the feet usually become eroded earlier than the joints of the hands,23,24 including feet thus may help improve the sensitivity of joint damage assessment in early RA.

The scoring systems that have been designed to evaluate radiographic changes in RA can be divided into two main groups, global and detailed.25,26 Global scoring systems assign one score to the entire joint, taking into account all the abnormalities seen, whereas detailed systems assign scores on at least two separate variables for each joint evaluated. The most widely used detailed scoring system is the modified Sharp method and its variations, and the most widely used global scoring system is the Scott modification of the Larsen score.26

**Table 1** A comparison of common radiographic scoring systems used in rheumatoid arthritis (RA)

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Description of scoring system</th>
<th>Advantages/disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949</td>
<td>Steinbrocker et al</td>
<td>Global, ordinal scale (range 1–4)</td>
<td>Simple, initial standard limited by a short scale and partiality for severely damaged joints</td>
</tr>
<tr>
<td>1963</td>
<td>Kellgren and Lawrence</td>
<td>Global, ordinal score for joint groups including hands, wrists, and feet (range 0–4), standard reference films</td>
<td>Simple, semiquantitative. Lacks sensitivity in detecting progressive damage</td>
</tr>
<tr>
<td>1971</td>
<td>Sharp et al</td>
<td>Detailed, assigns 2 individual ordinal scores, one for erosions (range 0–5) and one for joint space narrowing (range 0–4), for various joints of the hands and wrists</td>
<td>Sensitive, but requires training and is time consuming to apply. Later modified</td>
</tr>
<tr>
<td>1977</td>
<td>Larsen et al</td>
<td>Global, ordinal score (range 0–5) for limb joints, standard reference films</td>
<td>Semiquantitative, easier to learn, less sensitive to changes than the Sharp method</td>
</tr>
<tr>
<td>1983</td>
<td>Genant</td>
<td>Detailed, scores erosions and joint space narrowing in the hands, wrists, and feet, ordinal scale (range 0–4)</td>
<td>Semiquantitative, presents difficulties in assessing progression of structural damage</td>
</tr>
<tr>
<td>1987</td>
<td>Kaye et al</td>
<td>Erosions and joint space narrowing scored together for multiple joints of the hands and wrists, ordinal scale (range 0–2–5)</td>
<td>Simple, quantitative, requires training to apply efficiently</td>
</tr>
<tr>
<td>1989</td>
<td>van der Heijde et al</td>
<td>Modified Sharp scoring system for hands, wrists, and feet, uses a wider range (0–10) of ordinal scores for the feet</td>
<td>Semi-quantitative, easier to learn, less sensitive to changes than the Sharp method</td>
</tr>
<tr>
<td>1995</td>
<td>Scott et al</td>
<td>Modified Larsen scoring system, redefines grades of score, adds new reference films</td>
<td>Modifications improve ease of use and reliability, but non-structural (eg, swelling) and early erosive changes disproportionately affect the score</td>
</tr>
<tr>
<td>1995</td>
<td>Rau and Herborn</td>
<td>Modified Larsen scoring system which adds a semiquantitative description of the loss of joint surface area and provides standardised reference films for all stages at different anatomical sites</td>
<td>Good intra- and interobserver agreement. Omission of feet may compromise ability to detect signs of early RA</td>
</tr>
<tr>
<td>1998</td>
<td>Genant</td>
<td>Modified Sharp scoring system in which typically only hands are scored, with a range of 0 to 3 for the erosion score of each site and 0 to 4 for joint space narrowing</td>
<td>Table adapted from van der Heijde, Scott, and Lassen with permission.</td>
</tr>
</tbody>
</table>
system for scoring joint damage. Although this section will focus primarily on the differences between these approaches, it is important to note that radiographic scores obtained by the modified Sharp and Larsen methods have been shown to be significantly correlated. Nevertheless, some studies have found that compared with the Larsen method and its modifications, the Sharp method and its variations are more sensitive, particularly with respect to change over time.

The original Sharp method assessed 27 joints in each hand and wrist, with each joint being given a separate score for joint space narrowing and erosions. Sharpe and colleagues subsequently identified 17 areas for erosions and 18 areas for joint space narrowing that resulted in a high degree of intra- and interobserver accuracy. Van der Heijde later added feet to the radiographic analyses, a modification that has also been used by Sharp. Because of their similarities, these radiographic scoring systems will be referred to as “modified Sharp methods.” In modified Sharp scoring systems, each joint is given a separate score for joint space narrowing and erosions. Variations of these methods exist, each with subtle differences in the scales used and the criteria for scoring. In one commonly used scheme, 13 sites in each hand and wrist and six joints in each foot are examined for joint space narrowing on a scale of 0 to 4: 0 indicates no narrowing, 1 represents minimal narrowing, 2 indicates loss of 50% of the joint space, 3 indicates loss of 75% of the joint space, and 4 represents complete loss of the joint space. The erosions are counted individually, usually at 16 sites in each hand and wrist and six sites in each foot, with a maximum score of 5 given for a destroyed hand or foot joint. For joints in the feet, the van der Heijde version of the Sharp scoring system has a maximum score of 10 for a destroyed joint. A total Sharp score is generated based on the sum of the joint space narrowing and erosion scores.

The Genant modification of the Sharp method (Genant/Sharp method) focuses on 14 sites for erosions and 13 sites for joint space narrowing. Erosion scores range from 0 to 3.5 for each joint, and joint space narrowing is scored on a scale ranging from 0 to 4. Half grade scores (for example, 3+) are allowed to improve sensitivity. The total erosion score and the total joint score are each normalised based on a maximum score of 100, and these two normalised scores are added to give a total score in which erosions and joint space narrowing are evenly weighted.

The original Larsen method assessed all limb joints. Although comprehensive, the need for x-rays was not always practical. In the Scott modification of the Larsen method, multiple joints in the hands, wrists, and feet (in some studies) are evaluated. Each joint is given a grade based on the extent of bone loss, with 0 representing a normal joint. Grade 1 reflects slight, early, or non-specific findings of RA. Periarticular osteopenia/joint swelling must be major, and/or suggested erosions/cysts at two sites in the joint must be smaller than 1 mm. Grade 2 reflects a definite early abnormality; one or more erosions larger than 1 mm must be present, with a break in the cortical margin. Grade 3 reflects medium destructive abnormality; erosions at both sides of the joint must be of significant size with preservation of some joint surface. Grade 4 reflects severe abnormality; subluxation must be present. Grade 5 reflects mutilating abnormalities. The original articular surfaces must have disappeared and gross bone deformation must be present in the weightbearing joints. To assign a grade, the joints are compared with radiographic standards.

In addition to the type of scoring system used, the way in which the films are viewed may have an important impact on the sensitivity and reliability of radiographic scoring. Three basic methods are available: single readings, paired readings that are blinded to sequence, and chronologically ordered paired readings. Studies of these three approaches suggest that single readings may provide less reliable data than paired readings.

“Scoring methods must be sufficiently sensitive to detect changes in early RA.”

Several issues in developing and selecting an ideal scoring system should be considered. Reader disagreement and inter/intraobserver variation are important issues, though they can often be minimised by a training period to ensure familiarity with the scoring method. Problems are compounded by the need to assign a discrete number or score to a continuum of damage. Questions about the sensitivity of scoring methods in detecting change over time have also been raised.

A recent report by an international panel of experts found that with the modified Sharp method, the smallest detectable difference (5.0 units) corresponded closely with the minimal clinically important difference (defined as radiographic progression that makes a rheumatologist change treatment). In contrast, the smallest detectable difference by the Larsen/Scott method was too insensitive to use as the threshold for clinically relevant change. As a result, changes in patients with early RA and patients with late RA and high disease activity in some cases went undetected. The ability of scoring systems to assess radiological healing may also be an issue. Healing phenomena can be seen in about 6% of joints; this figure may be increased by the new DMARDs, particularly the biological agents. Accordingly, radiographic scoring systems used in clinical trials should be able to take potential healing into account.

The table below compares the modified Sharp, Genant/Sharp, and Larsen radiographic scoring systems.

<table>
<thead>
<tr>
<th>Type of scoring method</th>
<th>Modified Sharp</th>
<th>Genant/Sharp</th>
<th>Modified Larsen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joints evaluated</td>
<td>Detailed</td>
<td>Detailed</td>
<td>Global</td>
</tr>
<tr>
<td>Scoring</td>
<td>Hands, wrists, feet</td>
<td>Hands, wrists, sometimes feet</td>
<td>Hands, wrists, sometimes feet</td>
</tr>
<tr>
<td>Strengths/weaknesses</td>
<td>More sensitive, more tedious to use</td>
<td>Easier to learn and use, less sensitive</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2** A Comparison of the modified Sharp, Genant/Sharp, and Larsen radiographic scoring systems used in rheumatoid arthritis (RA)

---

PROGRESSION IN RA

STUDIES OF JOINT DAMAGE/DISEASE PROGRESSION IN RA

Multiple studies using variations of the Sharp or Larsen scoring systems have shown that joint damage occurs early in the
course of RA, often within the first two years after presentation. Clinical predictors of radiographic damage and/or disease progression include positive rheumatoid factor titre status, raised erythrocyte sedimentation rate or C reactive protein level, joint swelling, and duration of disease. The reported rate of RA induced radiographic damage over time varies in different studies and in different populations. Some studies have found that the rate of disease progression is most rapid early in the disease, and then gradually decreases to a steady rate. In one recent prospective study that followed 183 patients with early RA (mean duration of symptoms, 11 months), the rate of disease progression was found to be threefold higher during the first two years after study entry compared with subsequent years. The rate of progression becomes a more important variable in light of several studies that have shown that aggressive treatment can delay joint damage. These data have suggested a possible “therapeutic window” early in the course of disease during which medical intervention with DMARDs may have a more significant impact than DMARD treatment given later.

To understand the significance of a clinical trial that tests the effects of an agent on RA induced radiographic joint damage, one must be aware of the method used to quantify radiographic damage and the context within which it is applied. Comparing the effectiveness of two agents based on anything besides a head to head trial is currently difficult. Comparisons are less likely to be meaningful if trials used different scoring systems, examined films in a chronological versus random order, or had experienced versus naive readers. Also, subtle differences may exist between studies that use the same general method of radiographic scoring, such as the number of joints examined or the total scoring range. Additional factors, particularly differences in patient populations, such as the amount of baseline radiographic damage at trial entry and the amount of disease activity and/or progression at baseline, also prevent accurate comparisons between trials. The different ways of reporting data can add to the confusion. Adopting specific standards for reporting radiographic data may help correct this. A round table discussion on this topic involving a panel of experts from academia, the pharmaceutical industry, and the U.S. Food and Drug Administration (FDA) was held in December 2000, and the guidelines for reporting radiographic data developed by this group were recently published. Some of the most important of these guidelines include the reporting of absolute numbers for radiographic change along with the maximum score possible, mean change values together with the standard error or standard deviation for group comparisons, and the percentage of patients with disease progression. Radiographs of both hands and feet are preferred, and the smallest detectable difference for the study should be reported in order to provide quality control and to allow the reader to gauge the clinical significance of the data presented.

KEY FINDINGS FROM RADIOGRAPHIC STUDIES
Effects of early, aggressive treatment
Clinical trial experience with modified Sharp scoring systems has been extensive. An example is a comparison study between “aggressive” early treatment with methotrexate (MTX) or sulfasalazine and conventional stepwise treatment (that is, starting treatment with non-steroidal anti-inflammatory drugs (NSAIDs) and progressing to DMARD treatment only if response to NSAIDs is insufficient). This investigation used the modified Sharp scale to monitor radiographic progression. A single observer unaware of the patients’ clinical and laboratory data analysed radiographs taken every six months in sequential order. Progression of radiographic damage was expressed as the difference between the absolute Sharp scores at 6, 12, 18, and 24 months and at study entry. The study found a significant reduction in radiographic progression after six months in the high risk group who received early MTX or sulfasalazine compared with the high risk group who received traditional stepwise treatment. This trial helped to establish the value of starting aggressive DMARD treatment early in the course of RA.

Comparison of monotherapy and combination therapy
A second important trial that used the modified Sharp scoring system was the 80 week European study by Boers et al, which compared sulfasalazine monotherapy with combination therapy with sulfasalazine, MTX, and prednisolone. In the combination therapy group the MTX and prednisolone doses were gradually tapered. The mean Sharp score, determined by two trained and blinded observers who viewed the films in chronological order, was used to analyse radiographs of the hands and feet at baseline, 28, 56, and 80 weeks. The study found a statistically significant improvement in radiographic progression for the group treated with combination therapy compared with those who received sulfasalazine monotherapy. The results of this study are often used to support the use of combination therapy in RA.

RADIOGRAPHIC STUDIES OF NEW DMARDs
Etanercept, infliximab (in combination with MTX), leflunomide, and anakinra are recently introduced agents that have shown efficacy in the treatment of RA and have been approved by the FDA for the treatment of the signs and symptoms of this disease. Additionally, etanercept, infliximab (in combination with MTX), and leflunomide are approved for inhibiting or retarding structural damage associated with RA. Anakinra has not yet received this additional indication. Attempting to compare these four agents, based on the trials that examined their ability to inhibit radiographic progression, highlights the problems that arise in trying to interpret radiographic data in RA. Here we review data from trials in which these agents were compared with MTX, the conventional gold standard DMARD. For anakinra, the only radiographic data currently available are from a trial comparing anakinra with placebo. Table 3 shows the relevant trial characteristics.

Leflunomide
Two randomised, controlled trials examined radiographic progression in patients treated with leflunomide versus MTX. A 12 month trial reported by Strand et al used a modified Sharp scoring method to determine the effects of these DMARD treatments on radiographic disease progression. Radiographs of the hands and feet were obtained at baseline and 12 months (or early exit) for 352 of the 482 patients enrolled in the study. Forty six joints were evaluated for erosions (five point scale) and 48 for joint space narrowing (four point scale). Possible total scores ranged from 0 (no damage) to a maximum score of 422 (severe damage). Patients enrolled in the trial had RA for at least six months. Patients could not have previously received MTX treatment; treatment with other DMARDs was discontinued at least 30 days before study entry. Compared with placebo, monotherapy with MTX or leflunomide significantly slowed radiographic progression as assessed by total Sharp score, erosion and joint space narrowing scores. In patients receiving leflunomide, total Sharp scores increased by a mean of 0.53 over 12 months, compared with 2.16 (p=0.007) in the placebo group and 0.89 (p=0.05) in the MTX group.

In the second trial, reported by Emery et al, patients received leflunomide or MTX for one year; at this point, patients could choose to continue to receive double blind treatment for a second year. All patients had active RA (at least six swollen and six tender joints) of between four months’ and 10 years’ duration. DMARDs were discontinued at least 28 days before trial.
enrolment; approximately 66% of patients had failed to respond to at least one DMARD. Radiographs were obtained at baseline and at one and two years, and were assessed by the modified Larsen technique. Forty joints in the hands and feet were scored on a scale of 0 (no damage) to 5 (severe damage); the total score was divided by 40 to give a mean Larsen score for each joint. The number of eroded joints was also counted. Although the publication by Emery et al does not state the proportion of patients for whom radiographs were available, a subsequent reanalysis using the Sharp method reported that baseline and one year films were available for 64% of the 999 patients in the study (the two year data were not analysed, and therefore this figure was not reported). The leflunomide and MTX groups had comparable overall mean Larsen scores at baseline, and both groups showed a small (0.03) increase in Larsen score during the first year of the trial. During the second year, there was no further increase in joint damage in the leflunomide group and a slight improvement in the MTX group, resulting in a small but statistically significant treatment difference in change in radiographic scores in favour of MTX at two years. An analysis of one year radiographic data by the modified Sharp method (same scoring as in the Strand leflunomide study discussed above) also concluded that both leflunomide and MTX slowed disease progression and were comparable during the first year of treatment (6.74 increase in total Sharp score in the leflunomide group vs 6.47 increase in the MTX group).

**Table 3 Differences in radiographic studies of new rheumatoid arthritides agents**

<table>
<thead>
<tr>
<th></th>
<th>Leflunomide</th>
<th>Etanercept</th>
<th>Infliximab</th>
<th>Anakinra</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key entry criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis of RA for 6 months or longer, no prior MTX treatment</td>
<td>Active RA for at least 4 months but no longer than 10 years</td>
<td>Early RA (&lt;3 yr), no prior MTX treatment</td>
<td>Active RA despite treatment with at least 12.5 mg MTX a week</td>
<td>Active RA</td>
</tr>
<tr>
<td><strong>Mean disease duration at baseline</strong></td>
<td>6.5–7.0 years</td>
<td>3.5–3.8 years</td>
<td>11–12 months</td>
<td>9–12 years (MTX)</td>
</tr>
<tr>
<td>Receiving other DMARDs during study</td>
<td>No</td>
<td>Yes (MTX)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Comparator arm</strong></td>
<td>Placebo, MTX</td>
<td>MTX</td>
<td>MTX (in MTX failures)</td>
<td>Placebo</td>
</tr>
<tr>
<td><strong>Scoring method used</strong></td>
<td>Modified Sharp</td>
<td>Modified Sharp</td>
<td>Modified Sharp</td>
<td>Larsen</td>
</tr>
<tr>
<td><strong>Number of patients in radiographic substudy/total number of enrolled patients</strong></td>
<td>352/482</td>
<td>639/999 at 1 year</td>
<td>583/632 at 1 year</td>
<td>349/428</td>
</tr>
<tr>
<td><strong>Mean baseline total radiographic score (range of treatment groups)</strong></td>
<td>22.76–25.37</td>
<td>50.0–51.6</td>
<td>11.2–12.4</td>
<td>67–82</td>
</tr>
<tr>
<td><strong>Maximum possible total score</strong></td>
<td>422</td>
<td>200</td>
<td>398</td>
<td>440</td>
</tr>
</tbody>
</table>

DMARDs, disease modifying anti-rheumatic drugs; RA, rheumatoid arthritis; MTX, methotrexate; NR, not reported.

Data were also evaluated by the modified Sharp method.4

**Etanercept**

Ballon et al compared etanercept treatment and MTX treatment over a 12 month period. The modified Sharp scoring system was used to evaluate radiographic progression of RA every six months in the hands, wrists, and feet. A total of 46 joints were evaluated for erosions on a six point scale (0–5), and 42 joints were examined for joint space narrowing on a 5 point scale (0–4), with a total possible score ranging from 0 (no damage) to 398 (severe joint destruction). Each image was scored by two blinded radiologists or rheumatologists, who were first trained in the scoring method. Interobserver correlation was good (Pearson r = 0.85). Patients in the study had been diagnosed with RA for less than three years and had never been treated with MTX. If patients had been receiving other DMARDs, these were discontinued four weeks before the study began.

Radiographs were available for 583 of the 632 randomised patients. Compared with MTX, etanercept treatment resulted in less radiographic progression over the course of one year, as reflected by both the erosion score and the total Sharp score. A significant difference was found between etanercept 25 mg and MTX in the mean change in erosion scores at six months (0.30 ± 0.68; p = 0.001) and one year (0.47 ± 1.03; p = 0.002), while the total Sharp score showed a significant difference at six months (0.57 ± 1.06; p = 0.001), but not at one year (1.00 ± 1.59; p = 0.11). The most profound differences between patients receiving these two agents were seen during the first six months, during which the mean erosion score increased by more than twice as much in the MTX group as it did in those treated with etanercept 25 mg.4 No significant difference in joint space narrowing scores was found between these two treatment groups, with both groups showing only minor increases (<0.5 units) in joint space narrowing over the course of one year. This finding may relate to the early RA population enrolled in this study. A prospective study of radiological progression in patients with RA of less than three years’ duration found that erosions progressed more rapidly than joint space narrowing early in the disease.4

Two year data from this study, in which patients were allowed to continue to receive treatment in an open label manner, have recently been reported. The differences between the etanercept and MTX groups seen at earlier time points became more evident with a longer study duration. At two years, patients taking etanercept 25 mg showed significantly less radiographic progression than patients in the MTX group, as indicated by the mean change from baseline in the total Sharp score (1.3 ± 3.2; p = 0.001), and significantly less erosion, as indicated by the mean change from baseline in the erosion score (0.7 ± 1.9; p = 0.001). Joint space narrowing scores did not show a significant difference between treatment groups.

**Infliximab**

A trial comparing MTX alone with MTX plus infliximab also used the modified Sharp scoring system to evaluate radiographic changes. In this study, radiographs of the hands and feet were obtained at baseline, 30, and 54 weeks and were scored by two blinded readers. For erosions, 32 joints in the hand (5 point scale) and 12 joints in the feet (10 point scale) were examined. A total of 40 joints were examined for joint space narrowing (4 point scale). Possible total scores ranged from 0 (no damage) to 440 (severe damage). The mean of the two scores was used in the analyses. Patients in this study had active, longstanding disease despite MTX treatment of at least 12.5 mg per week.

Patients were randomly allocated to receive the same dose of MTX as they had been receiving before the study, plus either...
placebo or infliximab (3 or 10 mg/kg body weight every four or eight weeks).\textsuperscript{57} Radiographs were examined for 349 of 428 patients. Compared with placebo plus MTX, infliximab plus MTX resulted in significantly less progression of joint damage as judged by the total radiographic score, erosion score, or joint space narrowing score at 54 weeks. Total Sharp scores increased by a mean of 7.0 in the placebo plus MTX group, compared with 1.3 for the currently recommended infliximab plus MTX regimen (3 mg/kg every eight weeks; p<0.001).

**Anakinra**

Both the modified Larsen method and the Genant/Sharp method were used to assess data from a trial of anakinra. Patients were randomly allocated to receive placebo or a single daily dose of subcutaneous anakinra at 30, 75, or 150 mg for six months. Radiographs of the hands and wrists were obtained at weeks 0 and 24 and scored by two radiologists according to the Larsen method, with the lower of the two scores chosen to provide a conservative assessment. Fifteen areas in each hand and wrist were scored on a global scale of 0 to 5 by comparison with standardised images, with total scores ranging from 0 (no damage) to 150 (severe damage in all joints).\textsuperscript{65} Feet joints were not assessed.\textsuperscript{66} Patients with active, severe RA and a disease duration of less than eight years were enrolled in this study. Treatment with DMARDs was discontinued at least six weeks before enrolment.\textsuperscript{67}

Radiographs for 347 of 472 patients were assessed. Compared with placebo, the total Larsen score of the pooled anakinra treatment groups was significantly lower (6.4 \( v \) 3.8; p=0.03). Although other clinical parameters, such as swollen and tender joints, showed a clear dose response, the lowest dosage group (30 mg/day) had the best radiographic results and was the only treatment group that by itself showed a significant difference from placebo.\textsuperscript{68}

Radiographs from this study were subsequently analysed by the Genant modification of the Sharp method for 333 of the 472 patients.\textsuperscript{69} A single observer scored 14 sites for each hand for joint erosions on a scale of 0 to 3.5, with half grades allowed, and 13 sites for each hand for joint space narrowing on a scale of 0 to 4, with half grades allowed.\textsuperscript{70} Using this scoring system, anakinra was found to reduce progression in the total Genant/Sharp score by about 50% compared with placebo (3.52 \( v \) 1.85; p=0.0004) during the six month trial. All active treatment groups showed significantly reduced scores compared with placebo; in this analysis the highest dosage group (150 mg) showed the best results.\textsuperscript{71} Both joint space narrowing scores and erosion scores were significantly higher in the placebo group than in the active treatment group.

**Comparing new agents: which is most effective in stopping radiographic progression?**

All four of the new agents discussed here have demonstrated DMARD activity and have shown the ability to inhibit the progression of structural damage (although it should be noted that the FDA has requested more data for anakinra before approving this indication). Having new DMARDs for the treatment of RA will certainly be a boon to both clinicians and patients. But which agent is the best? All clinicians want to help their patients choose the “best” treatment, and therefore, it is tempting to try to use these trial data to determine the most efficacious agent for retarding radiographic progression. However, these data were designed only to answer questions of efficacy within a particular trial. The differences among the trials are too great to extrapolate their data to answer other questions. For instance, although three of the trials used the modified Sharp method, different possible scoring ranges were reported and, in some cases, different scales were used for scoring erosions. These trial-specific modifications produce non-comparable test results.

Perhaps the most important differences among the trials were the baseline patient characteristics. In the etanercept trial, average disease duration was roughly one year;\textsuperscript{72} in the anakinra trial, it was 3.7–4.3 years;\textsuperscript{73} in the leflunomide trials, it was about 3.5\textsuperscript{74} and 7.0 years;\textsuperscript{75} and in the infliximab trial, 9 to 12 years.\textsuperscript{76} This may be particularly important in studies of radiographic progression, as some studies have suggested a more rapid rate of progression in patients with early RA.\textsuperscript{77, 78} Radiographic damage at baseline was also markedly different, consistent with the different duration of disease in enrolled patients. Finally, patients in the infliximab trial showed an inadequate response to MTX at study entry. Accordingly, the MTX arm in this study was more similar to a placebo arm than to an active treatment group. The impressive impact on radiographic progression of infliximab plus MTX relative to MTX must thus be interpreted with caution. In contrast, both etanercept and leflunomide (in one study) were compared with MTX in an MTX-naïve population. Data comparing the effect of anakinra on radiographic progression relative to MTX have not yet been reported.

Many questions about these agents are thus left unanswered. How do infliximab and anakinra compare with MTX in MTX-naïve patients? Will all of these agents be equally effective both in early RA and in later stages of the disease? And how do these agents compare with each other in halting radiographic progression? Until head to head radiographic studies among these agents are performed, the relative merit of these different DMARDs must remain unknown.

**CONCLUSION**

Although it has limitations, the current “gold standard” for radiological evaluation of disease progression in RA is plain film radiography. This standard has been used in multiple studies, and the World Health Organisation, American College of Rheumatology, and others recommend using joint imaging as one of the standards for classifying and investigating DMARDs.\textsuperscript{79, 80} Radiographic scoring systems provide valuable information on the ability of different treatment regimens to impact joint destruction. However, in many cases, data from one study cannot be directly compared with data from other studies, particularly if different scoring methodologies or different patient characteristics are involved. Creating a standardised system for radiographic scoring in clinical trials will help to improve the comparability of data in some cases, but even such a system may not allow comparison of data from diverse patient populations.

**ACKNOWLEDGEMENTS**

This research was supported by an unrestricted educational grant from Wyeth.

**REFERENCES**

Interpreting radiographic data in RA 603


van der Heijde D, Boonen A, Boers M, Kostense P, van der Linden S. Reading radiographs in chronic oligoarthritis, in pairs or as single films has important implications for the discriminative power of radiographic arthritis clinical trials. Rheumatology (Oxford) 2001;40:56–63.


van der Heijde D, Boonen A, Boers M, Kostenste P, van der Linden S. Reading radiographs in chronic oligoarthritis, in pairs or as single films has important implications for the discriminative power of radiographic arthritis clinical trials. Rheumatology (Oxford) 2001;40:56–63.


www.annrheumdis.com

Ann Rheum Dis was first published as 10.1136/ard.62.7.597 on 1 July 2003. Downloaded from http://ard.bmj.com/ on April 23, 2022 by guest. Protected by copyright.
Clinical Evidence—Call for contributors

Clinical Evidence is a regularly updated evidence based journal available worldwide both as a paper version and on the internet. Clinical Evidence needs to recruit a number of new contributors. Contributors are health care professionals or epidemiologists with experience in evidence based medicine and the ability to write in a concise and structured way.

Currently, we are interested in finding contributors with an interest in the following clinical areas:

- Altitude sickness
- Autism
- Basal cell carcinoma
- Breast feeding
- Carbon monoxide poisoning
- Cervical cancer
- Cystic fibrosis
- Ectopic pregnancy
- Grief/bereavement
- Halitosis
- Hodgkin's disease
- Infectious mononucleosis (glandular fever)
- Kidney stones
- Malignant melanoma (metastatic)
- Mesothelioma
- Myeloma
- Ovarian cyst
- Pancreatitis (acute)
- Pancreatitis (chronic)
- Polymyalgia rheumatica
- Post-partum haemorrhage
- Pulmonary embolism
- Recurrent miscarriage
- Repetitive strain injury
- Scoliosis
- Seasonal affective disorder
- Squint
- Systemic lupus erythematosus
- Testicular cancer
- Varicocoele
- Viral meningitis
- Vitiligo

However, we are always looking for others, so do not let this list discourage you.

Being a contributor involves:

- Appraising the results of literature searches (performed by our Information Specialists) to identify high quality evidence for inclusion in the journal.
- Writing to a highly structured template (about 2000–3000 words), using evidence from selected studies, within 6–8 weeks of receiving the literature search results.
- Working with Clinical Evidence Editors to ensure that the text meets rigorous epidemiological and style standards.
- Updating the text every eight months to incorporate new evidence.
- Expanding the topic to include new questions once every 12–18 months.

If you would like to become a contributor for Clinical Evidence or require more information about what this involves please send your contact details and a copy of your CV, clearly stating the clinical area you are interested in, to Claire Folkes (cfolkes@bmjgroup.com).

Call for peer reviewers

Clinical Evidence also needs to recruit a number of new peer reviewers specifically with an interest in the clinical areas stated above, and also others related to general practice. Peer reviewers are health care professionals or epidemiologists with experience in evidence based medicine. As a peer reviewer you would be asked for your views on the clinical relevance, validity, and accessibility of specific topics within the journal, and their usefulness to the intended audience (international generalists and health care professionals, possibly with limited statistical knowledge). Topics are usually 2000–3000 words in length and we would ask you to review between 2–5 topics per year. The peer review process takes place throughout the year, and our turnaround time for each review is ideally 10–14 days.

If you are interested in becoming a peer reviewer for Clinical Evidence, please complete the peer review questionnaire at www.clinicalevidence.com or contact Claire Folkes (cfolkes@bmjgroup.com).