Henry Ford Hospital Medical Journal

Volume 31 | Number 3 Article 8

9-1983

A Radiologic Method of Assessment of Bone and Joint Destruction in Rheumatoid Arthritis

Gilbert B. Bluhm

David W. Smith

Walter M. Mikulashek

Henry Ford Hosp Med J
Vol 31, No 3, 1983

Special Articles

A Radiologic Method of Assessment of Bone and Joint Destruction in Rheumatoid Arthritis†

Gilbert B. Bluhm, MD,* David W. Smith, PhD, and Walter M. Mikulashek, MD, PhD*****

We employed radiographic analysis and a numerical scoring system to evaluate bone and joint destruction in 39 patients who received aspirin and nonsteroidal, anti-inflammatory drugs or benoxaprofen therapy over a continuous study period of 36 months (mean). Our purpose was to compare the radiological changes in these patients over two observation intervals and to test the reproducibility of the scoring system. Our modified rating system has sufficient intra-rater reliability to provide a statistically adequate, reproducible method for evaluating bone and joint destruction in rheumatoid arthritis.

Rheumatoid arthritis is a systemic, progressive disease often associated with bone erosion and/or subchondral bone cysts and cartilage space or joint space narrowing. At present, its etiology is unknown, and the best treatment available is disease-modifying agents that can arrest the course of the disease process. However, these agents such as gold, penicillamine, antimalarials, and the immunosuppressants like cyclophosphamide have severe toxicity. While it is possible to alleviate the patient's pain and distress with aspirin, corticosteroids, or nonsteroidal, anti-inflammatory drugs, these do not arrest the development of bone and joint destruction. These destructive changes can be demonstrated radiographically, and it has become accepted clinical practice to diagnose and follow the course of rheumatoid arthritis with radiographic films (1-6). Serial radiographic assessment can also be used to study the efficacy of a drug in halting the progressive changes in the joints of carefully chosen patients who have established but not advanced disease (4,7-9).

In 1974, a two-year, double-blind clinical trial used both clinical and radiographic assessments to test the efficacy of injectable gold salts in halting the progression of bone and joint destruction in rheumatoid arthritis (10). Evidence of erosive changes, or osseous defects, and of cartilage space, or joint space narrowing, was determined by radiographic analysis of the patient's hand and wrist films. This study also employed a numerical method of scoring the radiographic films that had been originally described by Sharp, et al (11) in 1971.

We have developed a modification of that rating system for scoring disease progression. The present study was undertaken to test the reliability and reproducibility of our modified scoring system by using data on a selected group of patients who were receiving treatment for rheumatoid arthritis with benoxaprofen in a preliminary, preclinical drug study (12). Concurrently, we performed a series of experiments to test the reliability and reproducibility of the radiographic scoring system using radiographs from patients participating in this preclinical trial (13).

Materials and Methods

Patient selection

Four criteria were used to select patients. First, patients had to have evidence of definite or classical rheumatoid arthritis by the ARA criteria (9). Second, they had to have evidence of continuous disease for at least 12 months before the drug study treatment was started. Third, they could not have previously received drug therapy with gold salts, antimalarial compounds, or penicillamine.

*Department of Internal Medicine, Division of Rheumatology, Henry Ford Hospital

**Statistical Research Laboratory, The University of Michigan, Ann Arbor, Michigan

***Senior Clinical Investigator, The Eli Lilly Research Laboratories, Indianapolis, Indiana

Address reprint requests to Dr. Bluhm, Department of Internal Medicine, Henry Ford Hospital, 2799 W Grand Blvd, Detroit, MI 48202.

†Supported in part with funds from The Eli Lilly and Company, Indianapolis, Indiana.

Since any rheumatoid arthritis patient who has been under long-term treatment for continous disease would have received some form of medication, it was permissible for patients to have received aspirin, ibuprofen, or other nonsteroidal, anti-inflammatory drugs. Finally, each patient had to have available a single posterior anterior (P-A) radiograph of the hands and wrists that had been obtained more than six months before entry into the study.

Thirty-nine patients from three investigative centers were included in this study. There were 32 women and 7 men, with a mean age at entry of 50 years. Twenty-two patients (56%) had a rheumatoid factor titer greater than 160, and 11 (28%) showed rheumatoid nodules. The mean duration of rheumatoid arthritis for the entire group was 82 months from the time of onset until they entered the study. Each patient received 600-1000 mg of benoxaprofen a day in either a single or split dose during one of the sequential observation periods.

Radiographic technique

Single P-A films of the hands and wrists were taken by the standard method (2), and both hands were placed on one film (25 x 30 cm). It was important for the scoring system that the patient's hands and wrists be correctly positioned. An effort was made to have the palmar surface of the fingers pressed flatly upon the film surface, with the elbows level with the hands and wrists, to avoid hyperextension at the wrists. This reduces distortion, particularly of the relationship between the radial, ulnar, and carpal bones. Some patients were unable to perfectly flatten every finger on the film because of inflammatory swelling or deformity. Every reasonable effort, short of pressing or forcing an already deformed digit, was made to achieve this optimal position.

Radiographic scoring system

Each radiograph was scored for two features, once for osseous defects and once for joint space narrowing, as described below. Scores were determined for each joint by a single rater. The scores for each joint were added and standardized to obtain an osseous defect score and a joint space narrowing score for each radiograph.

Osseous defects

Erosive changes of bone, or osseous defects, occur at variable rates in the progression of rheumatoid arthritis. Since marginal erosions, subchondral bone cysts, and erosions of the articular surface represent the same process, they were scored together (1,11). The hands and wrists were chosen for this study rather than the feet or other joints because they contain certain sites most likely to show early erosive changes (3). These include the metacarpophalangeal joints (MCP), proximal interphalangeal joints (PIP), ulnar styloid process (medial aspect), lateral (radial) aspect of the navicular bone, and adjacent margins of the navicular and lunate bones. Normal notching of the navicular and capitate bones had to be carefully evaluated since it can be mistaken for erosion.

Our scoring system differed from that of Sharp by excluding the distal interphalangeal joints of the fingers. These are only infrequently involved in rheumatoid disease and are common sites for degenerative changes that may be difficult to distinguish from rheumatoid erosions (1). Osseous defects were rated on a scale from 0 to 5. When no erosions were found, the joint was scored 0 for normal. A single erosion was given a value of 1; two erosions a score of 2, and so forth, up to a maximum score of 5. A maximum score of 5 was given either for five or more separate erosions of the articular margin, or rated proportionally for loss of definition of a joint's articular surfaces (Fig. 1A). Because individual joints were occasionally unratable, each score was standardized by dividing the sum of the joint scores by the total number of joints evaluated and multiplied by the maximum joint score of five. Therefore, the maximum possible standardization score was 250 (50 joints x 5). A schematic drawing of a pair of hands and wrists was devised that outlined the bone and joint areas to be scored (Figs. 2A, 2B). The individual ratings for each radiograph were recorded on this schematic "map" for coding and analysis. Osseous defect and joint space narrowing ratings were recorded on separate maps.

Joint space narrowing

Cartilage of a joint is damaged by the inflammatory process of rheumatoid disease. Cartilage can be injured and repaired, but once destroyed, it cannot be significantly replenished. Rheumatoid disease may produce evidence of both joint space narrowing and/or osseous defects. However, the changes do not necessarily parallel each other in the same joint or the same patient (10,11). Even though joint space narrowing may at times seem difficult to judge, it is a useful measure of the progress of rheumatoid disease (1).

Joint space narrowing was scored on a scale from 0 to 4. A normal cartilage space width was given a 0 value. If the joint space narrowing affected only the radial or ulnar side, it was scored 1. Most often, the joint space narrowing is symmetrical, and when less than 50% narrowed, it was scored 2. When narrowing was greater than 50%, it was scored 3. A value of 4 was given whenever there was complete loss of joint space due either to articular de-

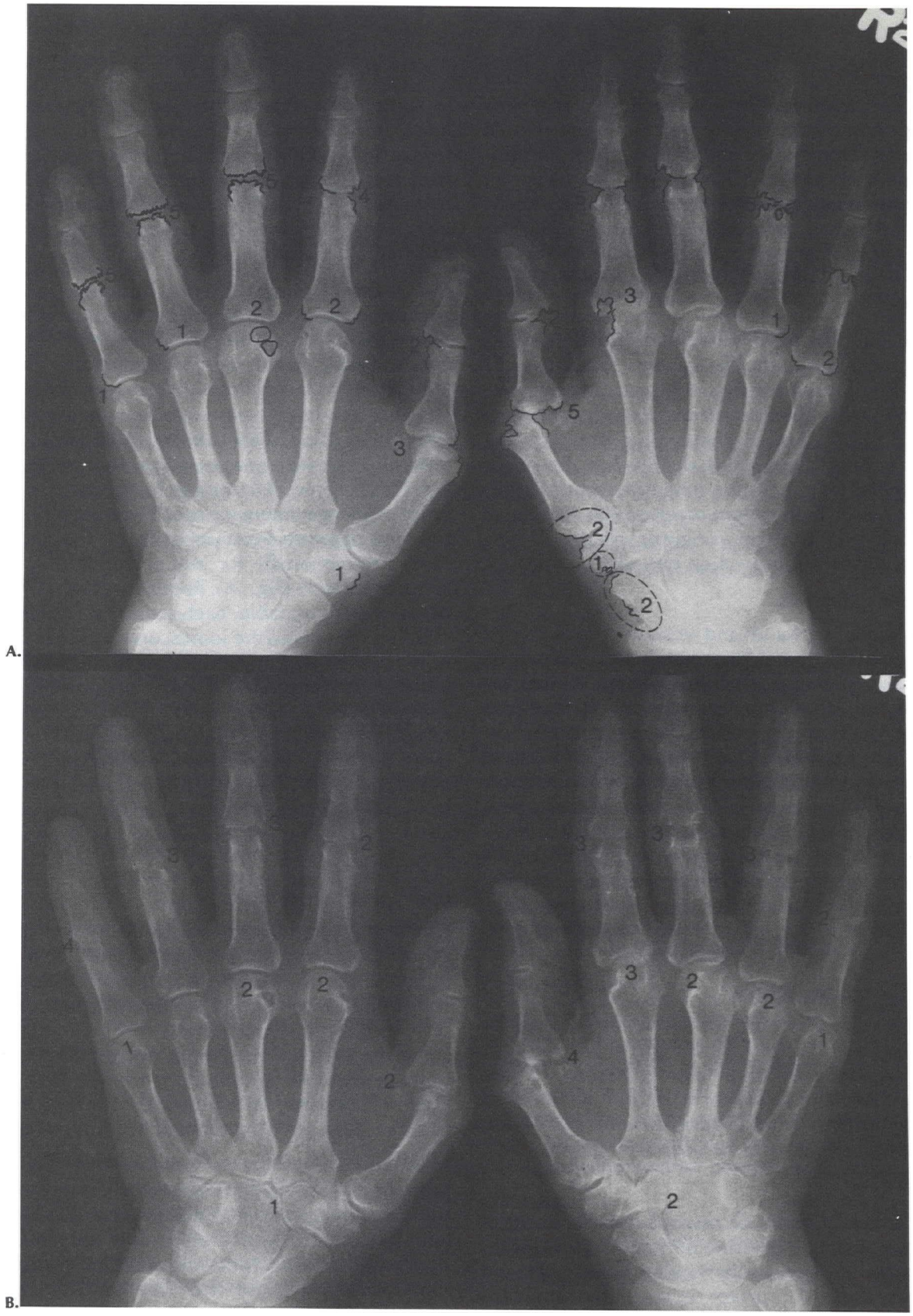

Fig. 1

Radiograph of the hands and wrists of a patient with established rheumatoid arthritis. A. the scoring numbers marked for osseous defects; B. scoring marked for joint space narrowing.

struction, extreme subluxation, and/or ankylosing (Fig. 1B). The number of joints rated for joint space narrowing from a single P-A film was 46. Again, because individual joints were occasionally unratable, the aggregate of the separate joint ratings was standardized by dividing it by the maximum possible score (4 times the number of joints rated). Each individual score was then recorded on a joint map (Fig. 2B).

Radiographic collection and scoring procedures

Radiographs were obtained from each patient a minimum of three times (Fig. 3): At least six months before drug therapy with benoxaprofen began (the "antetrial" radiograph), at the time the patient entered the study (the "pretreatment" radiograph), and at about six-month intervals thereafter, including a final radiograph taken at the end of the study (the "post-treatment" radiograph). The number of different radiographs from each patient varied from a minimum of three to as many as 12 in a few cases.

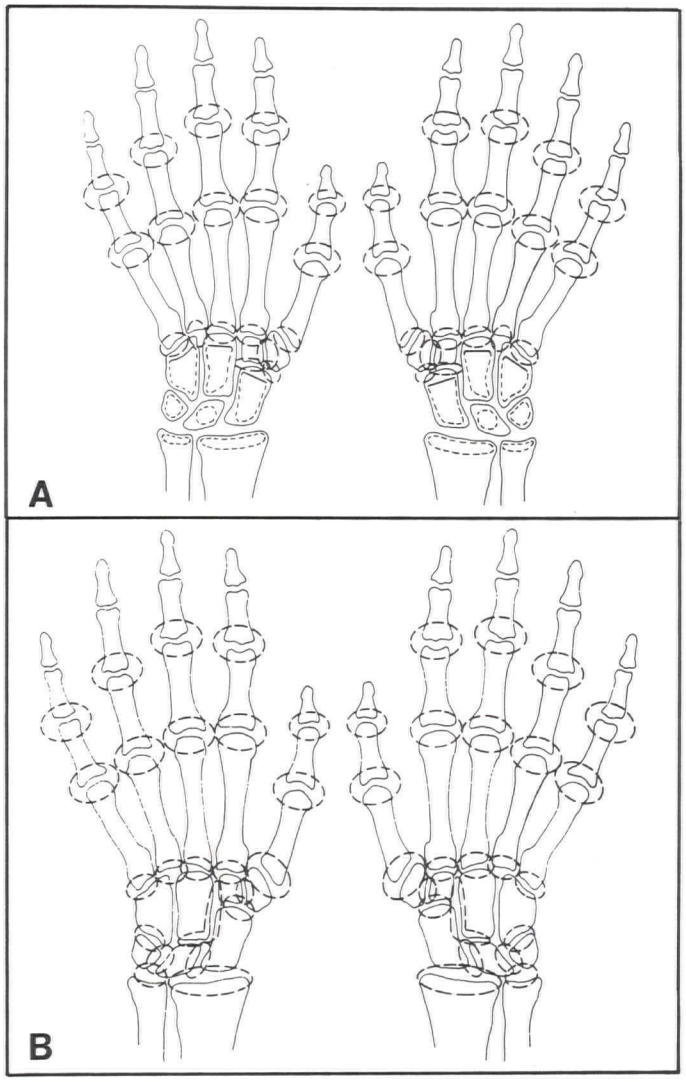

Fig. 2

Original scoring map devised for rating osseous defects (A) and joint space narrowing (B). For osseous defects, outlined joints were rated individually and given a numerical value from 0 to 5. For both hands and wrists, a total of 50 joints was rated. For joint space narrowing, outlined areas were rated individually and scored on a scale of 0 to 4. For both hands and wrists, a total of 46 joints was rated.

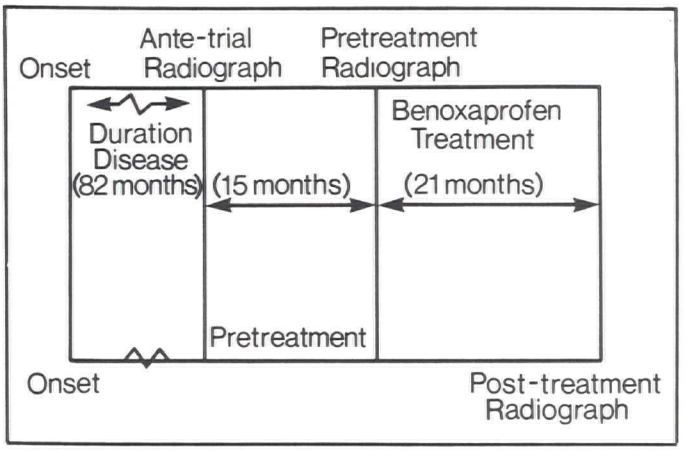

Fig. 3

Radiologic assessment of benoxaprofen treatment in rheumatoid arthritis. Drawing illustrates the three time periods for which radiographs were collected and compared, with the mean duration of each period.

Each radiograph was identified and coded to permit blind, randomized scoring. Before scoring, the coded radiographs were masked and randomly mixed. Films from both the pretreatment and drug treatment periods were intermixed and evaluated by single rater (GBB). To avoid fatigue and consequent carelessness on the part of the rater, no more than 10 to 15 radiographs were evaluated during the same day. The average time required to evaluate and score one film from a single patient was nine minutes. Each film was scored once for osseous defects, and once for joint space narrowing.

Intra-rater reliability

Using the scoring system as described, we conducted three experiments to test intra-rater reliability and compare variations in rating procedures (13). In the first experiment, a single rater (GBB) interpreted and scored each of 30 radiographs separately and in a completely randomized, blind sequence. He evaluated the same radiographs using the same method about three months later in order to estimate intra-rater reliability, i.e., the ability to reproduce the original scores (Fig. 4) (4). A second experiment evaluated the effect of using a magnifying lens and high intensity light where necessary to improve the accuracy of the results (Fig. 5). In the third

experiment (Fig. 6), the rater rescored 111 radiographs from 37 patients in sets. That is, he evaluated and rated three radiographs from the same patient simultaneously. The sequence of radiographs was blinded and randomized both within and among each set of three. The left and right hamate-capitate joints were not scored for joint space narrowing in this experiment. In the previous experiments, joints were scored if any rating at all could possibly be assigned, but in this experiment joints were given a score only if a rating could be assigned with complete confidence. This experiment permitted us to evaluate both an alternative method of evaluating radiographs and to determine which joints were the most difficult to score. The results of all three experiments

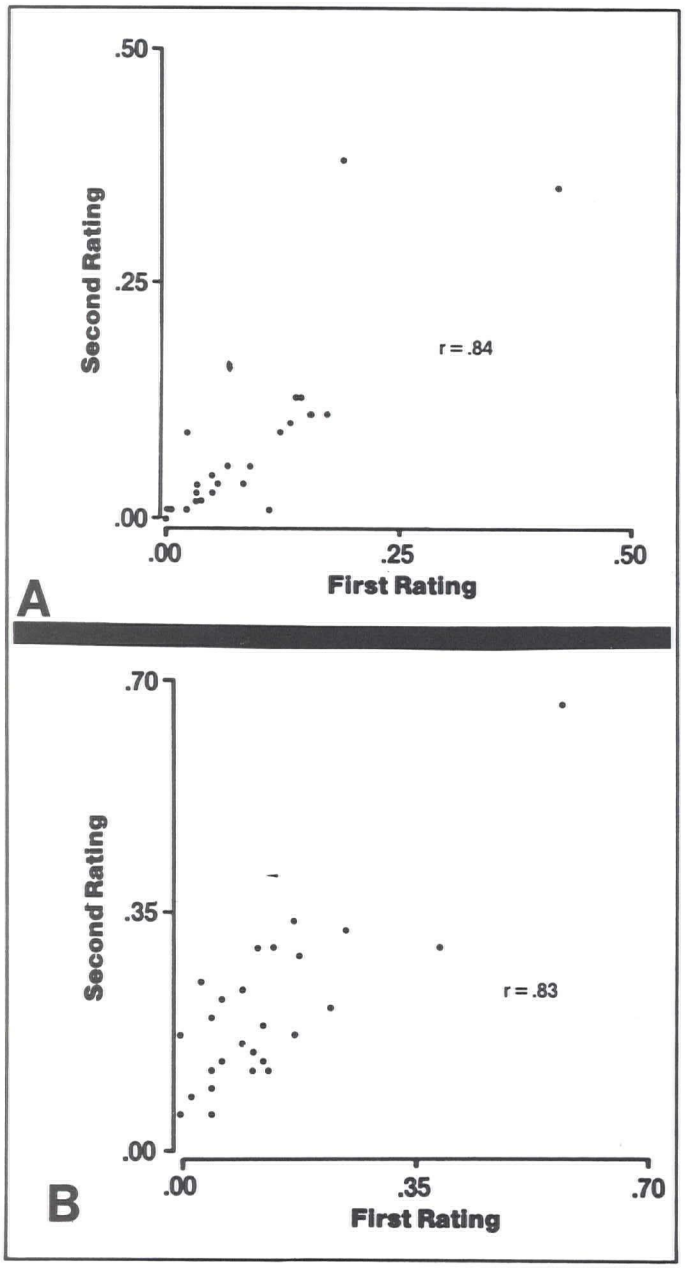

Fig. 4

Comparison of two ratings of individual radiographs performed three months apart by one judge using the same technique for osseous defects (A) and joint space narrowing (B).

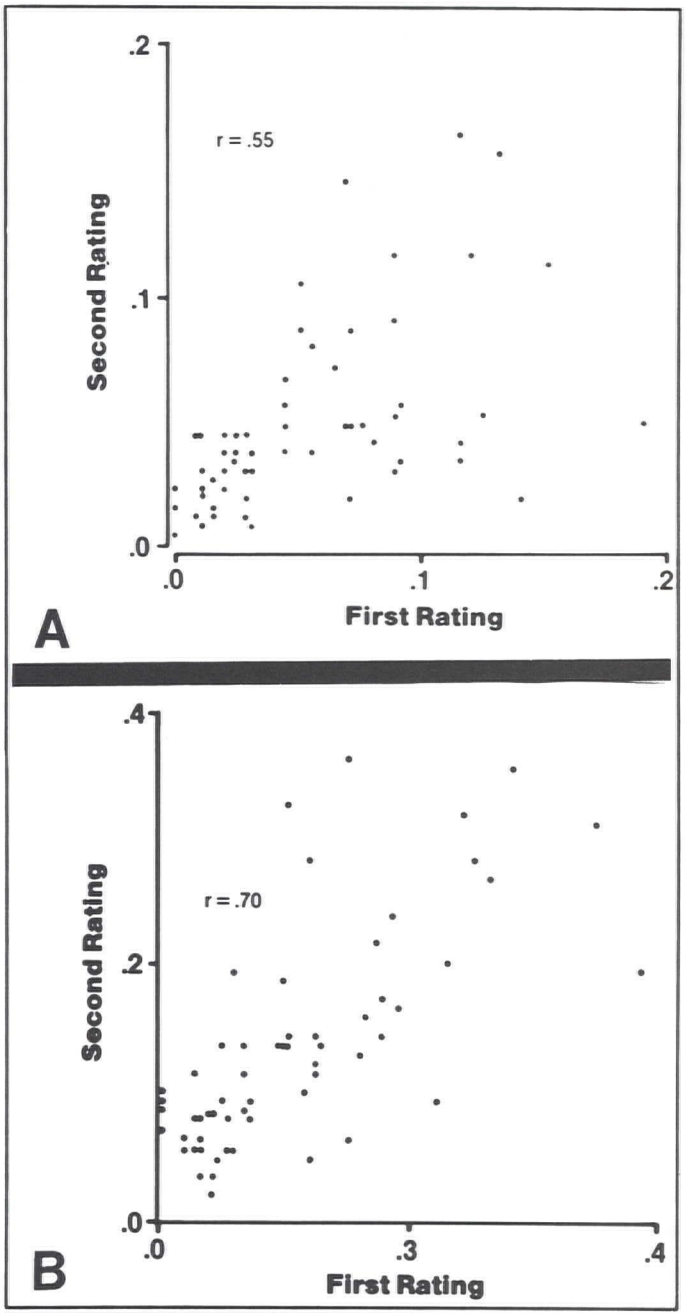

Fig. 5

Comparison of two ratings of individual radiographs by one judge for scoring osseous defects (A) and joint space narrowing (B). The second score was obtained by the additional use of a magnifying lens and high intensity light, which facilitated scoring radiographs of lesser quality.

were analyzed with scatter plots and Spearman's rank correlations.

Results

Radiographic technique

Our reliability experiments demonstrated that if the quality of the radiographs was improved, the accuracy of the rating was increased. Therefore, we used a standard method of producing radiographs to achieve the best quality of radiologic assessment consistently. Three phase x-ray equipment and Kodak XTL packet film should be used with the x-ray tube held 101.6 cm (40 inches) from the film. An exposure technique of 300 Ma 50 KV and an exposure time of one-half second are recommended. Exposure factors should be adjusted to reproduce an equivalent result of the density and quality required whenever other equipment is used.

Radiographic scoring procedures

When the study started in 1977, we used the scoring system developed by Sharp, et al (11), with the modifications described above (Fig. 2). Sharp included all the finger joints for osseous defects (for a total of 29 for each hand and wrist) and 27 articular relationships in each hand for joint space narrowing. Initially, we modified our system to eliminate the distal joints of the fingers; our scoring method included 50 joints for osseous defects (25 per hand and wrist) and 46 for joint space narrowing (23 per hand and wrist).

To determine the level of disease activity for each joint and carpal area evaluated, we computed and compared the average scores of each rated joint and area separately in 111 radiographs (Table I). Although the scores were slightly lower for the joints and areas in the wrist than in the PIP and MCP finger joints, the differences were not consistent across all areas.

To determine whether the reliability of the scoring system would be compromised if we eliminated certain joints and carpal areas that were difficult to evaluate, including some of those with low average scores (Table I), we computed correlations from the radiographs evaluated in the third experiment (Fig. 6). The first pair of correlations was determined by using the original number of joints and areas that were rated (50 for osseous defects, 46 for joint space narrowing), and the

TABLE I

Radiologic Assessment

Mean Rates from All First Ratings

Osseous Defects			Joint Space Narrowing		
Proximal Interphalangeal	LEFT	RIGHT	Proximal Interphalangeal	LEFT	RIGHT
1	.33	.26	1	.53	.36
2	.60	.41	2	.50	.41
3	.72	.53	3	.66	.69
4	.70	.64	4	.64	.73
5	.30	.48	5	.45	.40
Metacarpophalangeal			Metacarpophalangeal		
1	.48	.50	1	1.36	1.11
2	.70	.65	2	.93	.65
3	.69	.55	3	1.12	1.07
4	.50	.36	4	.76	.56
5	.50	.50	5	.50	.51
Metacarpal-Carpal			Metacarpal-Carpal		
1	.51	.49	1	.41	.41
2	.47	.54	2	.73	.77
3	.25	.19	3	.38	.34
4	.22	.14	4	.21	.13
5	.10	.13	5	.17	.18
Trapezium	.41	.45	Navicular Multangulars	.61	.52
Capitate	.55	.55	Capitate-Navicular & Lunate	.58	.68
Hamate	.17	.32	Capitate-Hamate	.23	.22
Navicular	.41	.45	Navicular-Lunate	.05	.10
Lunate	.24	.31	Lunate-Pisiform/Triquetrum	.25	.21
Pisiform/Triquetrum	.11	.25	Hamate-Pisiform/Triquetrum	.25	.25
Radius	.46	.48	Radial-Navicular & Lunate	.76	.86
Ulnar	.58	.68	Ulnar-Lunate & Triquetrum	.06	.07

second pair of correlations was determined by eliminating certain joints and carpal areas. The pisiform and triquetrum bones were eliminated, because their superimposition on a single P-A film makes evaluation difficult. For joint space narrowing, the metacarpocarpal relationships of the first and fourth joints were the most difficult to judge reliably. Similarly, carpal joint relationships that yielded more variation in rating were the ulnar-carpal, navicular-lunate, capitate-hamate, and greater and lesser multangular cartilage spaces.

When 50 joints were rated for osseous defects, the correlation was .88. When the correlation was computed on the basis of 46 joints, it remained the same at .88. With the 46 areas for joint space narrowing of the original scoring system, the correlation was .85. When the number was reduced to 32 areas by eliminating those that were difficult to evaluate, the correlation was only slightly higher at .86. There is no loss of reliability if these troublesome joints and areas are dropped from the scoring system. Consequently, we now recommend a scoring system in which 46 joints are rated for osseous defects (23 per hand

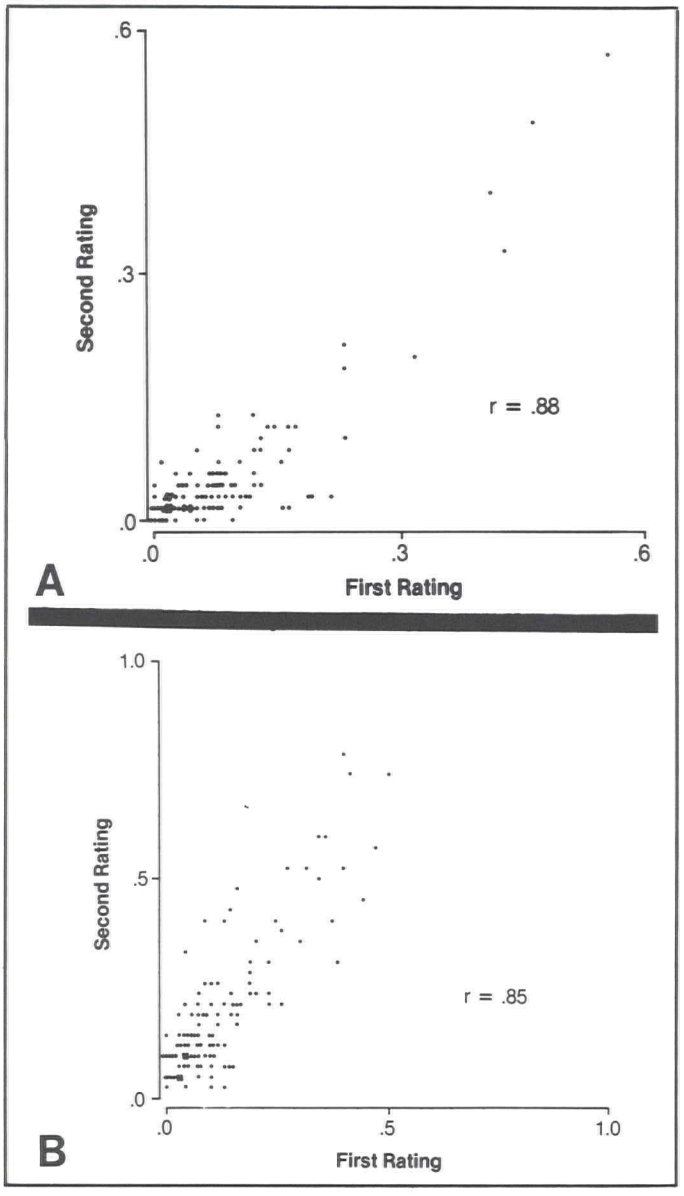

Fig. 6

Comparison of two ratings by one judge for osseous defects (A) and joint space narrowing (B). The first rating was obtained by reading a randomized mixture of all radiographs and the second rating by reading serial radiographs from the same patient with their sequence masked.

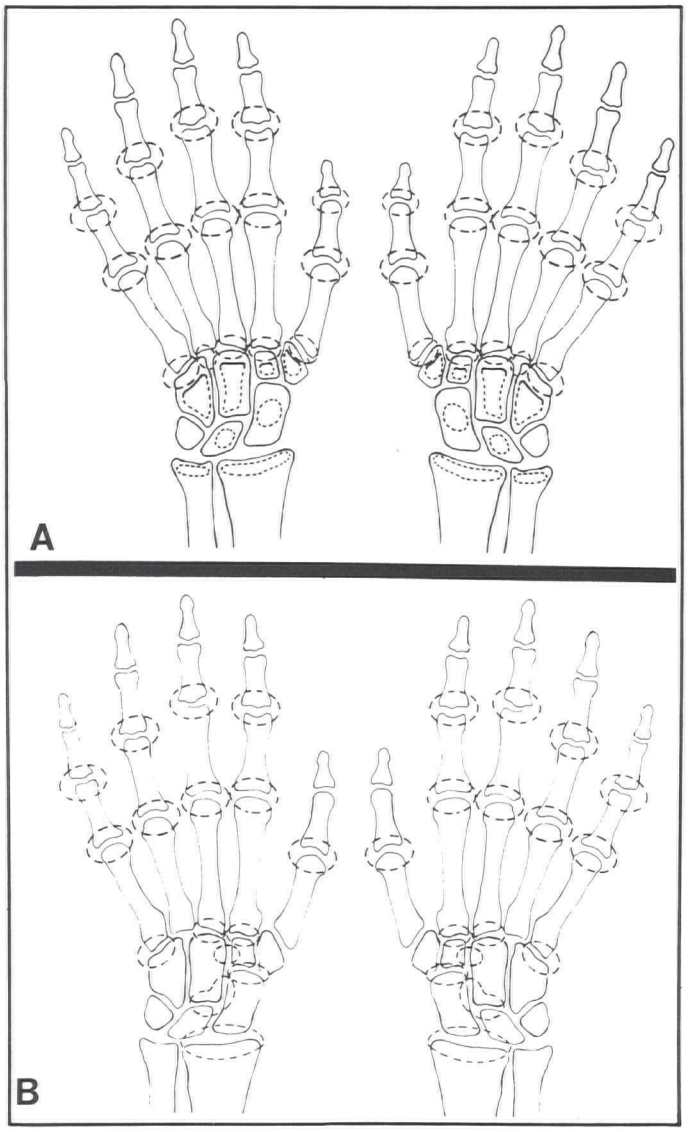

Fig. 7

Revised scoring map for rating osseous defects (A) and joint space narrowing (B). For osseous defects, a total of 46 joints is rated. For joint space narrowing, a total of 32 areas is scored.

and wrist), for a maximum possible score of 230 (Fig. 7A). For joint space narrowing, 32 are rated (16 per hand and wrist), for a maximum possible score of 128 (Fig. 7B).

Intra-rater reliability

The scores from the repeated evaluations by the same rater correlated with those of the first rating (Figs. 4-6), although not perfectly. While there were changes in the scores between the first and subsequent evaluations, the correlations were high enough to indicate that the system can be used reliably by a single judge to score radiographic changes in rheumatoid arthritis patients over time.

The first experiment evaluated the reproducibility of the method. The rater re-evaluated 30 of the radiographs a second time, under similar circumstances and using the same method (Fig. 4). Because four radiographs were rejected as being too dark, the results of this second, duplicate rating suggested that technical improvements in radiography were necessary if the rating process was to be considered reliable. The two scores of the remaining 25 films were highly correlated for both osseous defects and joint space narrowing (Fig. 4), although they were not identical; some shift was noted from the first evaluation to the second. Osseous defects decreased about 11% (sign test, $p = .011$); joint space narrowing increased about 22% (sign test $p<.001$). While there is a difference in means between ratings, the correlations are more than adequate.

The second experiment evaluated the effect of using a magnifying lens and high-intensity light where necessary to improve the accuracy of the ratings. The new scores correlated with the earlier ones, although not perfectly, indicating that changes occurred (Fig. 5). Sign tests indicated that the second scores for joint space narrowing, although consistently higher at 1.5% ($p = .016$), were minimal, and osseous defects were unchanged on average ($p = .29$).

The third experiment showed that there were no substantial differences between interpreting single radiographs and a set of radiographs from a single patient (Fig. 6). The correlations of the second scores with the original scores were high for both osseous defects and joint space narrowing. The second osseous defect scores were about 26% lower (sign test $p<.001$), and the second joint space narrowing scores were about 44% higher (sign test $p<.001$) than the original scores. Despite the change in mean values at the second evaluation, the correlation is still high.

Discussion

Following the course of rheumatoid arthritis in a patient over several years by evaluating and scoring disease progression rates on radiographs presented us with two major problems.

One problem concerns the nature of rheumatoid arthritis as a chronic disease process. Clinically, rheumatoid arthritis may follow any one of several patterns, as schematically represented in Fig. 8 (9). The episodic pattern is characterized by acute attacks that initially may last a few days or a few weeks, followed by a return to normal. In the progressive patterns, a patient may have a flare-up of the disease that occurs semi-annually or more frequently, followed by a relatively quiescent stage. Another pattern of rheumatoid arthritis, represented at the bottom of Fig. 8, is characterized by a slow, insidious onset and a slow, but continuous progression of the disease process.

When we designed our study, it was necessary to select patients carefully. A patient with advanced bone and joint destruction (Stage IV disease) (9) would be excluded because radiographic changes in such cases have already reached a point beyond which further destruction is minimal. Hence, the retarding effect of any intervening drug would also be minimal at best. However, a patient who exhibits the episodic pattern of rheumatoid arthritis is also a poor candidate for any kind of controlled study that measures disease progression radiographically. There would be periods with little or no changes, and it would be impossible to predict when these flares would occur.

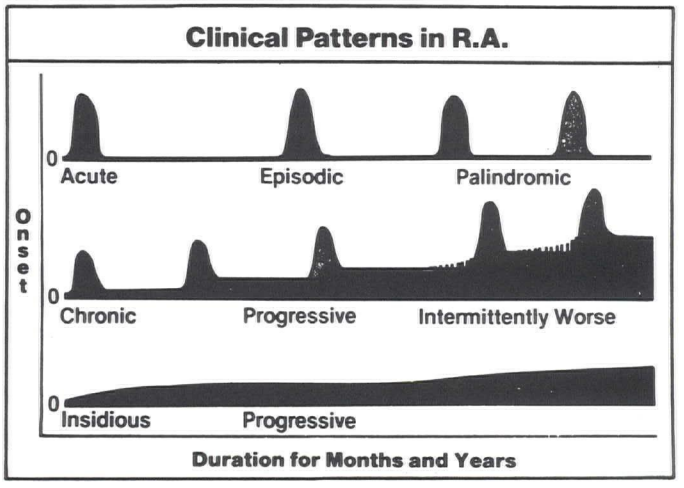

Fig. 8

Schematic illustration of the clinical patterns of disease progression in rheumatoid arthritis.

The ideal patient is one who exhibits the continuous pattern of rheumatoid arthritis; it is then possible to make radiographic measurements that reliably reflect the progress of bone and joint destruction. However, this is a very difficult group of patients to isolate unless they have had their disease for a protracted period of time, namely, several years. Consequently, one essential criterion of our study was that patients had to have demonstrated continuous disease and active synovitis for at least 12 months, and preferably longer (12). In our final study group of 39, the mean duration was 82 months with a range between 2 and 10 years.

Previous experience from the gold salt study (6,10,14) also demonstrated that radiographic measurements in a slowly developing disease like rheumatoid arthritis require that the patient be followed for a long period of time, two years at least. Since erosive changes in rheumatoid arthritis progress slowly over years as a rule rather than over a few months, radiographs that record such changes should be compiled over long intervals (5,9). In our study, we required a minimum of six months between the antetrial and the pretrial radiographs. Thereafter, during the drug treatment period, radiographs were taken at six-month intervals. However, our experience indicates that this is too short a time for radiographic changes to be clearly manifest. In a study like this, a time interval of a minimum of 18 months between ante- and pretrial radiographs is reasonable. In retrospect, therefore, fewer radiographs taken at longer intervals would have served the purposes of our study equally well. It is not as important to have a continual monthly or even semi-annual radiographic record of the destructive changes as it is to have a set of radiographs taken far enough apart so that the scores can be used to compute statistically reliable rates of change.

The second major problem our methodology had to address was the seemingly subjective nature of our scoring system for measuring radiographic development (4,15). Scoring by judges, however expert, must show statistically reliable results, or they are useless as a tool of measurement. To deal with this problem, we designed several intra-rater reliability experiments that would evaluate the reliability of the system when used repeatedly. As a result, we found that standardized radiographs are crucial to the accuracy of the rating procedure (16). The variable quality of the radiographs in our study increased the difficulty of producing consistent rating scores. Hence, we strongly recommend that our method of producing high quality radiographs (see Results section) be carefully followed. We also found that it is essential to train raters to use the scoring system consistently (16). For this purpose, we developed the schematic scoring maps for osseous defects and joint space narrowing. These can be used as learning tools to train and test raters, with diverse backgrounds and different levels of professional skill and experience, to use the scoring procedures. This can be a continuing process, with information being reviewed with individual raters for their own improvement. Such feedback helps maintain quality control. A set of guidelines with illustrated radiographs to permit review of the rating of osseous defects and joint space narrowing is of benefit for a single rater to maintain scoring consistency. It can also be helpful for multiple raters scoring a large clinical trial (17).

The most significant outcome of our experiments was the gradual development of a major modification in the scoring system of Sharp, et al (11). We found that the difficulty in evaluating certain joints and carpal areas consistently added considerable variability to the scoring system, compromised its reliability, and increased the time required to read radiographs. Our revised and simplified scoring system is easier to use, requires less time to evaluate individual radiographs, and reduces the difficulties encountered with poorer quality radiographs. These advantages are all achieved with no loss of intra-rater reliability.

We previously tested the individual variability between raters using the same scoring system. The same trends occurred in the rating results reported by both raters, but wide variability of ratings occurred on the same radiographs, particularly when more extensive disease was present (13). Hence, whenever raters use the same scoring system, we believe prior agreement must be reached on how to score and rate defects. Furthermore, the degree of variability between the raters should be periodically tested to determine that statistically accepted limits of variability remain between the raters.

Acknowledgments

This project was supported in part by funds from The Eli Lilly and Company of Indianapolis, Indiana. The authors would also like to thank Ms. Sue Kardasz and Mr. Wayne Pitchford for research assistance.

References

1. Berens DL, Lockie LM, Lin RK, Norcross BM. Roentgen changes in early rheumatoid arthritis. Radiology 1964;82:645-54.
2. Bunim JJ. The early radiological signs of rheumatoid arthritis. Bull Rheum Dis 1960;11:231-4.
3. Genant HK. Radiology of rheumatic diseases. In: McCarty DJ, ed. Arthritis and allied conditions. 9th ed. Philadelphia: Lea and Febiger, 1979:70-130.
4. Kellgren JH, Bier F. Radiological signs of rheumatoid arthritis. A study of observer differences in the reading of hand films. Ann Rheum Dis 1951;15:55-60.
5. Martel W, Hayes JT, Duff IF. The pattern of bone erosion in the hand and wrist in rheumatoid arthritis. Radiology 1965;84:204-14.
6. Bluhm G. The treatment of rheumatoid arthritis with gold. Semin Arthritis Rheum 1975;5:147-66.
7. Cooperating Clinics Committee of the American Rheumatism Association. A controlled trial of cyclophosphamide in rheumatoid arthritis. N Eng J Med 1970;283:883-9.
8. Cooperating Clinics Committee of the American Rheumatism Association. A controlled trial of gold salt therapy in rheumatoid arthritis. Arthritis Rheum 1973;16:353-8.
9. Steinbrocker O, Traeger CH, Batterman RD. Therapeutic criteria in rheumatoid arthritis. JAMA 1949;140:659-62.
10. Sigler JW, Bluhm GB, Duncan H, Sharp JT, Ensign DC, McCrum WR. Gold salts in the treatment of rheumatoid arthritis: A double-blind study. Ann Intern Med 1974;80:21-6.
11. Sharp JT, Lidsky MD, Collins LC, Moreland J. Methods of scoring the progression of radiologic changes in rheumatoid arthritis: Correlation of radiologic, clinical and laboratory abnormalities. Arthritis Rheum 1971;14:706-20.
12. Bluhm GB, Smith DW, Mikulaschek WM. Radiologic assessment of benoxaprofen therapy in rheumatoid arthritis. J Europ Rheum Inflam 1982;5:186-97.
13. Smith DW, Bluhm GB. Rater reliability in reading P-A films of hands for bone and cartilage changes in rheumatoid arthritis. J Europ Rheum Inflam 1982;5:198-205.
14. McCrum WR, Sharp JT, Bluhm GB. Use of the Weibull distribution for analysis of a clinical therapeutic study in rheumatoid arthritis. Henry Ford Hosp Med J 1976;24:173-82.
15. Larsen A, Edgren J, Harju E, Laasonen L, Reitamo T. Interobserver variation in the evaluation of radiologic changes of rheumatoid arthritis. Scand J Rheumatol 1979;8:109-12.
16. Gore SM. Assessing clinical trials: Between-observer variation. Br Med J 1981;283:40-3.
17. Bluhm GB. A Method of Radiologic Assessment for Scoring Single P-A Films of the Hands and Wrists in Rheumatoid Arthritis. Detroit, 1983.