Structural damage in rheumatoid arthritis as visualized through radiographs

Désirée van der Heijde

University Hospital Maastricht, Maastricht, The Netherlands; and Limburg University Center, Diepenbeek, Belgium

Correspondence: Désirée van der Heijde, MD, PhD, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands; Tel: +31 43 387 50 26; fax +31 43 387 50 06; e-mail: dhe@sint.azm.nl

Abstract

Several agents show an effect on reducing radiographic progression in rheumatoid arthritis. It is tempting to retrospectively compare the effects of these agents on radiographic progression across clinical trials. However, there are several limitations in interpreting and comparing radiographic results across clinical trials. These limitations, including study designs, patient characteristics, durations of follow-up, scoring methodologies, reader reliability, radiograph sequence, handling of missing data, and data presentation, will be discussed. The consequences are illustrated with several examples of recent clinical trials that show an effect on radiographic progression. A guide in the interpretation and clinical relevance of radiographic results is presented, with the Anti-TNF Trial in Rheumatoid Arthritis with Concomitant Therapy used as an example.

Keywords: radiography, infliximab, etanercept, leflunomide, methotrexate

Introduction

Radiographs are widely accepted as the 'gold standard' in assessing structural joint damage associated with rheumatoid arthritis (RA) and are therefore essential in evaluating the efficacy of experimental therapeutics. Both traditional drugs, such as sulfasalazine and methotrexate (MTX), and new drugs, such as leflunomide and biologic agents, can reduce the progression of radiologic damage [1–6]. With the development of agents that have a beneficial impact on structural joint damage, it has become tempting to retrospectively compare the efficacy results from various RA trials. However, it is difficult, if not impossible, to compare across clinical trials, and clinicians need to be aware of the limitations in comparing radiographic data. This paper discusses these limitations and offers guidance on how to interpret the results of clinical trials.

Limitations in comparing data across clinical trials

Knowledge of the limitations in comparing radiographic data across clinical trials is necessary for accurate interpretation of data in the absence of direct, head-to-head trials. These limitations include differences in study design, patient characteristics, severity of disease, duration of follow-up, scoring method used, reader reliability, order in which radiographs are read, handling of missing data, and, finally, data presentation. Each of these limitations is discussed below.

Study design

Some trials use a parallel design with patients who have not previously received treatment for RA. An example of this is the MTX arm in the etanercept trial, in which

---

patients with early RA who were naive to MTX treatment were randomized to an MTX treatment arm or an etanercept treatment arm [5]. Other trials included patients who were previously treated and who experienced a partial response to a disease-modifying antirheumatic drug. An example of this study design is the Anti-TNF [tumor necrosis factor] Trial in Rheumatoid Arthritis with Concomitant Therapy (ATTRACT), in which patients who had previously been treated with MTX were randomized to groups treated with infliximab plus MTX or with MTX alone [4]. Although both studies included treatment arms in which patients received MTX alone, the MTX arm in the etanercept trial is not comparable with the MTX arm in the ATTRACT trial, because of the different baseline characteristics and treatment histories of the patient populations. It can be expected that the radiographic response to MTX in patients naive to this drug will be more pronounced (on a population level) compared with patients who have previously shown a partial response to MTX and then continued treatment with this drug. Therefore, it is important to be aware of the study designs before comparing data between trials.

Patient characteristics

Several prognostic factors are known to predict an unfavorable outcome with respect to structural joint damage. The most important of these are the presence of rheumatoid factor, evidence of joint erosion early in the course of the disease, and rapid disease progression [7]. However, these predictors account for only a limited percentage of the variation between patients. Further, although these predictors are valid for groups of patients, they have little value when applied to individual patients. In addition, it is likely that other, currently unknown, factors will be associated with structural joint damage. By sampling from one patient population and randomizing the enrolled patients over two (or more) comparative trial arms, investigators can reasonably assume that both known and unknown factors are well balanced between the treatment arms. However, if trial arms from various studies are compared, this randomization has not taken place and many hidden differences between the patient populations may exist.

Although prognostic factors listed previously relate to the progression of structural joint damage, they are not necessarily transferable to predicting response to therapy. Anderson *et al.* [8] investigated the prognostic factors for response to treatment in 14 randomized clinical trials. The investigators concluded that patients whose RA is of longer duration do not respond as well to treatment as do patients with early disease. Moreover, female gender, previous treatment with a disease-modifying antirheumatic drug, poorer functional class, and higher disease activity affect the likelihood of patient response to treatment. According to Anderson *et al.* [8], these factors should also be considered when interpreting data from clinical trials.

Baseline radiographic damage

Patients enrolled in clinical trials have different levels of structural joint damage. Depending on the eligibility criteria for a study, patients in one study may have substantially more baseline radiographic damage than patients in another study. Therefore, baseline radiographic damage represents another important obstacle to comparisons across clinical trials. Expressing this baseline damage in terms of disease duration results in a radiographic progression rate. Recently, in two different studies, the radiographic progression rate before entering the study was shown to be an important predictor of treatment outcome [9,10]. The ATTRACT trial was conducted in patients with established disease who had achieved a partial response to MTX. Patients were randomized to four active treatment arms with infliximab and a control arm with placebo, with all patients continuing on MTX [4]. The Combinatietherapie Bij Reumatoide Artritis [Combination Therapy in Rheumatoid Arthritis] (COBRA) trial was performed in patients with early disease who were randomized to either a combination of high-dose prednisone (quickly tapered) combined with MTX and sulfasalazine, or to sulfasalazine alone [3]. Although these trials were based on patients from different populations (patients with established disease versus those with early disease), both trials showed retardation of radiographic progression (infliximab compared with placebo on an MTX background and combination therapy compared with sulfasalazine alone, respectively). Further, it was evident that within each trial, patients with the highest radiographic progression rate at the onset of the trial benefited most from treatment [9,10].

Duration of follow-up

At the group level, radiographic progression in RA is a linear process [11,12]. However, progression rates and patterns differ markedly from patient to patient [13,14]. Because radiographs show cumulative damage, differences in duration of follow-up are expected to have a large impact on the results. Moreover, because of differences in the patterns of progression, patient-to-patient variability cannot be easily corrected for by dividing progression scores by follow-up duration – for example, to calculate a monthly rate of progression. Therefore, it is important that the duration of follow-up be similar when radiographic progression is compared across trials.

Scoring methodology

Another important consideration when comparing radiographic scores across clinical studies is the scoring system used to assess structural joint damage. Several scoring methods exist to assess joint radiographs. These scoring methods evaluate different bony features, assess different joints, and have different scoring ranges. The most widely used methods are the Larsen and Sharp methods, along with their modifications [15–18]. The Larsen method uses a global grading system that mainly

assesses erosive damage. The scoring range is from 0 to 200. The Sharp method assigns separate scores for erosions and joint space narrowing, which are combined to obtain a total score. The scoring range for the Sharp method is from 0 to 314 or 448, depending on which modified version of the method is used. Because of the differences in scoring ranges and in the abnormalities included in the assessment, a score of, for example, 5 in the Larsen method cannot be directly compared with a score obtained using the Sharp method. In some trials, scores obtained from hand radiographs are included, whereas in other studies, radiographs of both the hands and feet are used. Therefore, it is important to compare scores obtained on the same films: joints of either the hands or the feet, or a combination of both.

Reader reliability

Clinical trials are typically designed to have one or two observers read and score each radiograph. The use of two observers reduces the variability in scoring and decreases the error of measurement. Interobserver reliability is high for the progression of scores. However, the absolute scores from reader to reader may be significantly different. In other words, each observer has his or her own reading level (and is consistent with his or her own readings), and this reading level may be clearly different from that of another observer, but the progression seen is fairly consistent between the observers. Trials are analyzed making use of these progression scores, scored by one (pair of) observer(s). However, when absolute scores are compared across trials, another variable besides treatment, design, and patient characteristics is introduced: a different observer. This further complicates the comparison of scores across trials that used different readers.

Radiograph scoring sequence

Radiographs are read either in a known sequence or in random order. The order in which radiographs are scored has a significant effect on the measurement error of scores and on the ability of scores to capture disease progression accurately [19,20]. Consequently, the order in which radiographs are read is an important factor when comparing results across trials. However, earlier published trials often did not present this information [19]. Therefore, comparing results from new trials with those already published is often problematic.

Handling of missing data

Because radiographs show cumulative structural joint damage, missing radiographs become an important issue in the analysis of a progressive disease such as RA. Missing data cannot be replaced by a simple 'last-observation-carried-forward' procedure, as is often applied to other data, especially in the case of long term trials. Sensitivity analyses to investigate the effect of the missing radiographs are warranted [6]. The aim of each trial should

be to have films of randomized patients at baseline and at follow-up, regardless of patient status. However, this often is not feasible, because patients, especially those who have withdrawn, may refuse to submit to follow-up films. Therefore, missing radiographs will continue to pose an obstacle and data need to be analyzed in various ways to rule out an effect of selectively missing films.

Data presentation

Clinical trials present radiographic data in a variety of ways, which makes the comparison of data across trials difficult. To minimize this obstacle, a roundtable conference was held to establish a minimum set of radiographic results that should be presented in each trial [21]. Most data are presented on a group level, either by mean ± standard deviation (SD) or by median and interquartile range (IQR). Because of the skewed nature of radiographic data, the two ways of presenting the data provide important, but completely different, information. If a large proportion of patients in a group shows no or minimal progression and a few patients show a significantly higher rate of disease progression, the latter set of patients gives much weight to the mean ± SD of the overall group. The presentation of these data as a median with IQR provides information on the proportion of patients showing a specific progression. Both the mean ± SD and the median with IQR give important and additive information on a group level.

Other important information is the analysis at the level of individual patients. By dichotomizing the data, statistical power is lost. Therefore, such analysis is advised as a secondary analysis [21]. It is useful to know the percentage of patients who show progression above a certain clinically important level. However, the decision about what level to use as a cutoff is often arbitrary and can result in incomparability between trials. Although some trials simply define no progression as a score of zero, this finding does not take into account measurement error, which is always present. Others use arbitrary cutoff values or base the cutoff level on the SDD (smallest detectable difference apart from measurement error), which is a trial-specific number [22,23]. In the leflunomide trials published by Sharp *et al.* [6], the cutoff value regarded as indicating progression in erosion was a score of 3, which resulted in progression of erosions being reported in 3% to 11% of treated patients (receiving leflunomide, sulfasalazine, or MTX), versus 12% to 17% of patients receiving placebo. In the ATTRACT trial, an SDD of the total score (in this case, 8.6) was selected as the cutoff value in reporting progression. In this trial, 6% of treated patients were reported to have shown progression of erosion, versus 31% of patients receiving placebo [4]. In the etanercept trial, 0 was selected as the cutoff value for the erosion score. Applying this cutoff value, 28% of etanercept-treated patients were judged to have erosion that pro-

gressed, versus 40% of MTX-treated patients. Within each trial, these figures are meaningful and show that the active treatment was effective. However, there is little value in comparing progression between these trials, because all have assigned different cutoff values.

Interpretation of radiographic results

Clinicians often question whether the measured differences in radiographic progression between treatment arms are clinically relevant. To answer this question, long-term follow-up of other outcomes such as functional disability and loss of work is required. However, collection of these long-term data takes several years; therefore, it is useful to look for circumstantial evidence. Structural joint damage in clinical trials is assessed in small joints. However, there is a good correlation between the damage in small joints with the damage in large joints [24]. Therefore, an observed reduction in disease progression in small joints is likely a reflection of the disease course in large joints. Moreover, there is an association between structural joint damage and physical function that is stronger with increasing disease duration [25]. Lastly, it is important to consider that RA is a chronic disease, and it can be expected that without treatment, patients will continue to show progression of structural damage.

As an example, the interpretation of the radiographic results of the ATTRACT trial are presented here. Are the findings clinically relevant? All films were scored by the Sharp/van der Heijde method (range 0 to 440), by two independent observers, and without knowledge of the radiograph sequence. The average score of two observers was used. The median increase in the modified Sharp score in all patients treated with infliximab plus MTX was 0.5 (IQR −2.0, 2.5), versus 4.3 (IQR 0.5, 10) in patients treated with MTX alone [26]. These data imply that at least 50% of patients treated with infliximab achieved a progression score of 0.5 or less and that 75% of patients progressed to a maximum value of 2.5. In patients treated with MTX alone, 50% of patients showed an increase of 4.3 and 75% progressed to a maximum value of 10.

At first glance, when considering the median increase in joint damage observed in patients treated with MTX alone in the context of the total range of the scoring system (0 to 440), a median increase of 4 appears clinically insignificant. In practice, however, it is extremely rare for patients to have complete destruction of all joints in both hands and feet and thereby receive a maximum score. Scores around 100 already represent major destruction. Usually, the progression score of 4 represents an increase in erosion and joint space narrowing in several joints. However, it is difficult to envision how this will affect the patient. As the maximum erosion score per hand joint is 5, one could imagine that an increase of 4 would represent

an almost completely eroded hand joint. Thus, a median increase of 4 is actually a substantial finding. Furthermore, this especially makes sense if the long duration of the disease, resulting in an increase of 40 over 10 years, is taken into account. Assuming a continuation of what was observed in the trial, 50% of patients receiving MTX alone will develop eight completely eroded hand joints in the following 10 years and 25% of these patients will reach a score exceeding 100 (if they started with normal films), which represents marked joint destruction. In contrast, 50% of patients treated with infliximab will have no progression of joint destruction in the following 10 years, and 25% of patients will reach a score of 25 points, which represents five completely eroded hand joints. Furthermore, recent research has shown that clinical experts consider an increase of 5 Sharp/van der Heijde points a clinically meaningful change [23]. Therefore, on the basis of this expert opinion, 50% of patients treated with MTX alone had clinically meaningful disease progression, whereas 75% of patients treated with infliximab did not [23].

The ATTRACT trial also analyzed radiographic progression in individual patients by using the SDD as a cutoff level. This value (8.6) represented the progression of disease that was distinguishable from measurement error. Measurements >8.6 represented significant radiographic progression. From these results, the number of patients needed to be treated (NNT) to prevent major progression can be calculated, where NNT = 1/(% of MTX-only-treated patients with progression above the SDD [31%] − % of infliximab-treated patients with progression above the SDD [6%]) × 100, which yields an NNT of 4. Therefore, four patients need to be treated with infliximab to prevent major radiographic progression in one patient. The NNT value associated with infliximab treatment compares favorably with that of many treatments used to prevent fractures due to osteoporosis, which have an NNT value of 100 to 200.

Conclusion

In summary, although a therapeutic effect on structural joint damage within a clinical trial setting can be evaluated, interpreting and comparing radiographic results across clinical trials can be very hazardous.

References
1. van der Heijde DM, van Riel PL, Nuver-Zwart IH, Gribnau FW, vad de Putte LB: **Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis**. *Lancet* 1989, 1:1036-1038.
2. Hannonen P, Mottonen T, Hakola M, Oka M: **Sulfasalazine in early rheumatoid arthritis. A 48-week double-blind, prospective, placebo-controlled study**. *Arthritis Rheum* 1993, **36**:1501-1509.
3. Boers M, Verhoeven AC, Markusse HM, van de Laar MA, Westhovens R, Van Denderen JC, van Zeben D, Dijkmans BA, Peeters AJ, Jacobs P, van den Brink HR, Schouten HJ, van der Heijde DM, Boonen A, van der Linden S: **Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis**. *Lancet* 1997, **350**:309-318.

4. Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, Smolen JS, Weisman M, Emery P, Feldmann M, Harriman GR, Maini RN: **Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group**. *N Engl J Med* 2000, **343**:1594-1602.

5. Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Schiff MH, Keystone EC, Genovese MC, Wasko MC, Moreland LW, Weaver AL, Markenson J, Finck BK: **A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis**. *N Engl J Med* 2000, **343**:1586-1593.

6. Sharp JT, Strand V, Leung H, Hurley F, Loew-Friedrich I: **Treatment with leflunomide slows radiographic progression of rheumatoid arthritis: results from three randomized controlled trials of leflunomide in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Investigators Group**. *Arthritis Rheum* 2000, **43**:495-505.

7. Young A, van der Heijde DM: **Can we predict aggressive disease?** *Baillière's Clin Rheumatol* 1997, **11**:27-48.

8. Anderson JJ, Wells G, Verhoeven AC, Felson DT: **Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration**. *Arthritis Rheum* 2000, **43**:22-29.

9. van der Heijde DM, Landewé RB, Lipsky PE, Maini RN, ATTRACT Investigators: **Radiologic progression rate at baseline predicts treatment differences: results from the ATTRACT-trial [abstract]**. *Arthritis Rheum* 2001, **44**:S80.

10. Landewé RB, van der Heijde DM, Verhoeven A, Boonen A, Boers M, van der Linden S: **Radiological progression rate at baseline predicts treatment differences: results from the COBRA trial [abstract]**. *Arthritis Rheum* 2001, **44**:S371.

11. Wolfe F, Sharp JT: **Radiographic outcome of recent-onset rheumatoid arthritis: a 19-year study of radiographic progression**. *Arthritis Rheum* 1998, **41**:1571-1582.

12. Hulsmans HM, Jacobs JW, van der Heijde DM, van Albada-Kuipers GA, Schenk Y, Bijlsma JW: **The course of radiologic damage during the first six years of rheumatoid arthritis**. *Arthritis Rheum* 2000, **43**:1927-1940.

13. Fex E, Jonsson K, Johnson U, Eberhardt K: **Development of radiographic damage during the first 5-6 yr of rheumatoid arthritis. A prospective follow-up study of a Swedish cohort**. *Br J Rheumatol* 1996, **35**:1106-1115.

14. Plant MJ, Jones PW, Saklatvala J, Ollier WE, Dawes PT: **Patterns of radiological progression in early rheumatoid arthritis: results of an 8 year prospective study**. *J Rheumatol* 1998, **25**:417-426.

15. Larsen A, Dale K, Eek M: **Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films**. *Acta Radiol Diagn (Stockholm)* 1977, **18**:481-491.

16. Rau R, Herborn G: **A modified version of Larsen's scoring method to assess radiologic changes in rheumatoid arthritis**. *J Rheumatol* 1995, **22**:1976-1982.

17. Sharp JT, Young DY, Bluhm GB, Brook A, Brower AC, Corbett M, Decker JL, Genant HK, Gofton JP, Goodman N, Larsen A, Lidsky MD, Pussila P, Weinstein AS, Weissman BN: **How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis?** *Arthritis Rheum* 1985, **28**: 1326-1335.

18. van der Heijde DM, van Leeuwen MA, van Riel PL, van de Putte LB: **Radiographic progression on radiographs of hands and feet during the first 3 years of rheumatoid arthritis measured according to Sharp's method (van der Heijde modification)**. *J Rheumatol* 1995, **22**:1792-1796.

19. van der Heijde D, Boonen A, Boers M, Kostense P, van der Linden S: **Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials**. *Rheumatology (Oxford)* 1999, **38**:1213-1220.

20. Salaffi F, Carotti M: **Interobserver variation in quantitative analysis of hand radiographs in rheumatoid arthritis: comparison of 3 different reading procedures**. *J Rheumatol* 1997, **24**: 2055-2056.

21. van der Heijde D, Simon L, Smolen J, Strand V, Sharp J, Boers M, Breedveld F, Weissmann M, Weinblatt M, Rau R, Lipsky P: **How to report radiographic data in randomized clinical trials in rheumatoid arthritis? Guidelines from a roundtable discussion**. *Arthritis Rheum* 2002, in press.

22. Lassere M, Boers M, van der Heijde D, Boonen A, Edmonds J, Saudan A, Verhoeven AC: **Smallest detectable difference in radiological progression**. *J Rheumatol* 1999, **26**:731-739.

23. Bruynesteyn K, van der Heijde D, Boers M, Lassere M, Boonen A, Edmonds J, Houben H, Paulus H, Peloso P, Saudan A, van der Linden S: **Minimal clinically important difference in radiographical progression of joint damage over 1 year in rheumatoid arthritis: preliminary results of a validation study with clinical experts**. *J Rheumatol* 2001, **28**:904-910.

24. Drossaers-Bakker KW, Kroon HM, Zwinderman AH, Breedveld FC, Hazes JM: **Radiographic damage of large joints in long-term rheumatoid arthritis and its relation to function**. *Rheumatology (Oxford)* 2000, **39**:998-1003.

25. Drossaers-Bakker KW, de Buck M, van Zeben D, Zwinderman AH, Breedveld FC, Hazes JM: **Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time**. *Arthritis Rheum* 1999, **42**:1854-1860.

26. US Food and Drug Administration. [http://www.FDA.gov/ohrms/dockets/AC/00/transcripts/3623t1.rtf].