Recent Rheumatoid Arthritis Clinical Trials Using Radiographic Endpoints — Updated Research Agenda

VIBEKE STRAND, MARISSA LASSERE, DÉSIRÉE van der HEIJDE, KENT JOHNSON, and MAARTEN BOERS

ABSTRACT. Recent randomized controlled trials of traditional and newly developed therapies provide evidence that we have interventions that potentially slow or prevent structural damage in active rheumatoid arthritis, as measured using radiography. These trials also provide a unique opportunity for exploratory data analysis to generate hypotheses apropos the pathogenesis and determinants of radiographic progression and functional disability; they also facilitate further study of the methodological issues regarding imaging measurement. (J Rheumatol 2001;28:887–9)

Key Indexing Terms:
RHEUMATOID ARTHRITIS RADIOGRAPHS RADIOGRAPHIC SCORING METHODS

INTRODUCTION

With the increasing ability to demonstrate "disease control" or "prevention of structural damage" in rheumatoid arthritis (RA), it is important that we develop and attain consensus on appropriate methodology for generating, analyzing, and reporting these data[2,3]. This is relevant both for therapeutic decisions by clinicians and patients and for regulatory assessment.

RANDOMIZED CONTROLLED TRIALS IN RA — AN OPPORTUNITY

Recent randomized controlled trials (RCT) provide evidence for the slowing of structural damage as measured using radiography[4-8]. These trials also provide an opportunity to generate hypotheses about the pathogenesis and determinants of radiographic progression, and to explore the relation of radiographic change to functional disability. Moreover, they facilitate further study of the methodological issues regarding imaging measurement identified at OMERACT 4 such as how (and to what degree) the number, specific joints assessed, scoring methods, and chronologic sequence of interpretation influence agreement, accuracy, sensitivity, and specificity of radiographic analyses[9,10].

From the Division of Immunology, Stanford University, San Francisco, CA, USA; Department of Rheumatology, St. George Hospital, Sydney, Australia; Department of Rheumatology, University Hospital Maastricht, Maastricht, The Netherlands; Limburg University Center, Diepenbeek, Belgium; CDER, Food and Drug Administration, Rockville, MD, USA; and Department of Clinical Epidemiology, VU University Hospital, Amsterdam, The Netherlands.

V. Strand, MD, Clinical Professor, Division of Immunology, Stanford University; M. Lassere, MB, BS, FRACP, PhD, FAFPHM, Research Staff Specialist, St. George Hospital; D. van der Heijde, MD, PhD, Professor of Rheumatology, Department of Rheumatology, University Hospital of Maastricht; K. Johnson, CDER, Food and Drug Administration (Note: views are not necessarily those of the FDA); M. Boers, MSc, MD, PhD, Rheumatologist, Professor of Clinical Epidemiology, Department of Clinical Epidemiology, VU University Hospital.

Address reprint requests to Dr. C.V. Strand, 306 Ramona Road, Portola Valley, CA 94028, USA. E-mail: vstrand@aol.com

THE RESEARCH AGENDA

The clinical goal of structural improvement, of course, is to induce a major alteration in the progression of RA to structural deformity and disability. How to structure the research agenda to this end; what questions are best asked in what venues; and what conceptual ambiguities might merit attention independent of, and even preceding, empirical work are all fundamental questions. Given the currently available RCT databases, the following merit attention in deliberations about future research.

1. Validating a radiographic measure in RA — what gold standard?

What is, for a radiographic trial, the gold standard? The most straightforward answer — mortality — is not entirely applicable to RA. In a chronic disabling disease such as RA it is morbidity that affects the patient most and that should be taken into account. Mortality is only assessable in RA in longterm open designs. It is not feasible with the usual duration of controlled studies. Data on morbidity, such as functional disability, are important in our understanding of longterm factors important to patients. However, the measurement of physical function can also be problematic, thereby moderating its application as a gold standard against which we can validate radiographic measures.

2. Is it advisable to separate study designs that test measurement hypotheses from those that test interventional hypotheses? Clearly, this would be optimal. Otherwise trial failure cannot be unequivocally attributed either to failure of the measure, or, alternatively, to failure of the intervention. Given that optimal demonstration of treatment and/or disease effects itself is, inherently, a "moving target," clinical researchers as well as drug developers should be encouraged to systematically evaluate new measures as carefully as they do new therapeutic interventions.

3. How can important determinants of radiographic prognosis be efficiently discovered and described? Incomplete understanding of the pathogenesis of radiographic damage

and its measurable determinants (risk factors) hinders efficient trial design. Correlations between disease activity measures and radiographic progression in RCT in RA may be only low or moderate. Yet there is evidence from cohort studies that joint swelling predicts radiographic progression. However, there may be many explanations for the limited concordance seen in trials between activity, important clinical outcomes, and radiographic outcomes. The noncontinuous feature of the American College of Rheumatology (ACR) responder index may contribute to this poor correlation, but it is currently unclear whether continuous measures such as the Disease Activity Scale perform any better in this regard. Radiographic outcomes show a restricted sensitivity because of measurement inadequacies. Also, the degree of correspondence of different classes of outcomes may, in reality, differ across agents targeted to different RA pathogenetic pathways, and may differ in trials of short versus long duration. Finally, another outcome class, patient self-reports, are also predictive (of future morbidity and mortality), but they engage again different "domains" of outcome and are subject to variability by factors external to the disease. Thus, there are numerous methodological hurdles to an easy resolution of this aspect of the research agenda. One practical suggestion has been to encourage rapid release into the public domain of radiographic (and clinical) data from recent RCT.

4. What are the issues intrinsic to the characterization, validation, and performance of RA radiographic measures? Methodological issues intrinsic to RA radiographic measures include the number, degree, and distribution of joints assessed, the effects of scaling, aggregation, weighting, scoring method utilized, and sequencing of the films. Both Larsen and Sharp/van der Heijde methods have undergone considerable testing, and been shown to be reliable and sensitive measures. Modified Sharp/van der Heijde analyses have been utilized in the majority of recently completed RCT; only one series of trials employed both Larsen and Sharp methods. However, few patients actually attained radiographic scores beyond the second quartile of the potential distribution, and many individual patient radiographs in active *and* control groups demonstrated no disease progression, variously defined. (The definitions of "no disease progression" and furthermore "improvement" and/or "radiographic healing" all require further discussion.) However, despite low mean or median values reported for change from baseline radiographic scores in these trials, standard deviations and interquartile ranges remained large — due to significant heterogeneity within treatment groups, or problems with the measurement process (e.g., scale distribution, weighting, floor effects, nonlinearity), or both.

Recent trials have employed multiple readers of radiographs blinded as to treatment and sequence to reduce measurement error and bias[11,12]. Radiographs read in chronological sequence are reported to increase sensitivity to change[13]. Reading radiographs blinded to sequence and intervention reduces the bias produced by the expectation that baseline films have lower scores than subsequent films, but this reduction of bias comes at a cost. Small treatment effects are missed, and the ability to adjust readings for technical problems such as film exposure and positioning is lost. Another factor influencing the responsiveness of reading radiographs is the source and magnitude of technical and reader variability.

5. What role should a radiographic minimally clinically important difference (MCID) play in trial design, analysis, and interpretation? Recently the smallest detectable difference (SDD) was suggested as a useful starting point in determining what is a minimum clinically important difference in radiographic progression[10]. The SDD is that difference beyond random error of the measurement. In a clinical trial it is applied and interpreted in the context of individual responder status, and is therefore similar in concept to that of the ACR response criteria[14]. Further evaluation of the validity and robustness of the SDD, and whether it should, in principle, serve as a minimum clinically important difference, are topics for the OMERACT 5 Imaging Module agenda.

6. Can we improve our capacity to deal with certain recurrent, and seemingly intractable, design and analytic problems? Randomized controlled trials in active RA of 12 months or more duration were recommended to include radiographic outcomes by OMERACT at a consensus meeting in 1992[15]. Inevitably there will be missing radiographs in studies of this duration, the analytic consequences of which are much greater than with the more frequently assessed clinical endpoints. Although sensitivity analyses have been performed to account for missing data, allowing statistical assessment of "robustness" of results in the active treatment groups, there is no consensus regarding these methods[6,8,16]. Therefore how to convincingly account for the effects of missing information remains an important problem.

What is the optimal analysis to apply to radiographic change scores? Should means or medians be used? Good trial design, as pertains to protecting type 1 error, would dictate that whatever analytic strategy is elected, that it be prespecified in the protocol (as otherwise the multiplicity issue will arise). Nonparametric analyses utilizing median values may be more suitable for datasets not normally distributed. However, parametric analyses offer more flexibility, especially in subgroup analyses. Skewed data may be "normalized" by data transformation. Finally, the central limit theorem allows parametric analysis when the sample size is sufficiently large. Although progression scores are usually more normally distributed than status scores, they still remain very skewed.

7. What about across trial comparisons? In the absence of a much clearer understanding of the risk factors for radi-

ographic progression, these comparisons will always be flawed, and even if such an understanding existed, trial particulars (specifics of design, differences in conduct and protocol populations) make such comparisons risky.

SUMMARY

Determining what is a clinically meaningful "no disease progression" or "prevention of structural damage" remains challenging. However, the availability of recent high quality RCT with radiographic endpoints, which show an effect on structural damage, provides a major opportunity to study several methodological issues. These data should allow researchers to better understand the determinants of damage in RA, further refine measurement methodology, validate better imaging techniques, and develop recommendations for more uniform reporting of data.

REFERENCES

1. Paulus HE. Heterogeneity in rheumatoid arthritis radiographic trials. Issues to consider in a metaanalysis. J Rheumatol 2000;27:249-50.
2. Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices and biological products for the treatment of rheumatoid arthritis. Washington, DC: DHHS; February 1999.
3. Committee for Proprietary Medicinal Products. Points to consider on clinical investigation of slow-acting anti-rheumatic medicinal products in rheumatoid arthritis, European Agency for the Evaluation of Medicinal Products; January 1998.
4. Boers M, Verhoeven AC, Markusse HM, et al. Randomized comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 1997;350:309-18.
5. Bresnihan B, Alvara-Gracia JM, Cobby M, et al. Treatment of rheumatoid arthritis with recombinant human IL-1 receptor antagonist. Arthritis Rheum 1998;41:2196-204.
6. Sharp JT, Strand V, Leung H, Hurley F, Loew F, on behalf of the Leflunomide Rheumatoid Arthritis Investigators Group. Treatment with leflunomide slows radiographic progression of RA — results from three randomized controlled trials of leflunomide in patients with active rheumatoid arthritis. Arthritis Rheum 2000;43:495-505.
7. Finck B, Martin R, Fleischmann R, Moreland L, Schiff M, Bathon J. A phase III trial of etanercept vs methotrexate in early rheumatoid arthritis [abstract]. Arthritis Rheum 1999;42 Suppl:S117.
8. Lipsky P, St Clair W, Furst D, et al. 54 week clinical and radiographic results from the ATTRACT trial: A phase III study of infliximab in patients with RA despite methotrexate [abstract]. Arthritis Rheum 1999;42:S401.
9. van der Heijde D, Boers M, Lassere M. Methodological issues in radiographic scoring methods in rheumatoid arthritis. J Rheumatol 1999;26:726-30.
10. Molenaar ETH, Boers M, van der Heijde DMFM, et al. Imaging in rheumatoid arthritis: Results of group discussions. J Rheumatol 1999;26:749-51.
11. Sharp JT, Bluhm GB, Brook A, et al. Reproducibility of scoring radiologic abnormalities in the hands and wrists of patients with rheumatoid arthritis by multiple observers. Arthritis Rheum 1985;28:16-24.
12. 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.
13. 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 1999;38:1213-20.
14. Lassere M, Boers M, van der Heijde D, et al. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.
15. Boers M, Tugwell P, Felson DT, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol 1994; 21 Suppl 41:86-9.
16. Leung H, Hurley F, Strand V. Issues involved in a metaanalysis of RA radiographic progression. J Rheumatol 2000;27:544-8.