World Health Organization and International League of Associations for Rheumatology Core Endpoints for Symptom Modifying Antirheumatic Drugs in Rheumatoid Arthritis Clinical Trials

MAARTEN BOERS, PETER TUGWELL, DAVID T. FELSON, PIET L.C.M. van RIEL, JOHN R. KIRWAN, JOHN P. EDMONDS, JOSEF S. SMOLEN, NIKOLAI KHALTAEV, and KENNETH D. MUIRDEN

ABSTRACT. The WHO/ILAR core set of endpoints for rheumatoid arthritis clinical trial signifies progress in a continuing worldwide effort. This core set includes the following measures: pain, patient global assessment, physical disability, swollen joints, tender joints, acute phase reactants, and physician global assessment; in studies of one or more years' duration, radiographs of joints should be performed. (*J Rheumatol 1994;***(suppl 41) 21:86-9)**

Key Indexing Terms:
ENDPOINTS RA CLINICAL TRIALS

Clinicians vary in the way they use clinical variables to make judgments about the efficacy of treatment[1]. In rheumatoid arthritis (RA) it has been common practice to use a selection of traditional measures to define the endpoints of most clinical trials. However, the measures chosen are not comprehensive, some are insensitive to change, and some overlap[2]. Despite conferences, reviews, and editorials in the last 10 ears, no consensus has yet emerged on the appropriate endpoints in RA clinical trials[3]. We summarize recent developments in the challenge to improve the quality of clinically relevant endpoints in RA clinical trials. In this context, an endpoint is any measure that is used in the evaluation of patients with RA in clinical trials. Some endpoints may be considered patient outcome variables, and some measures of the disease process. We focus on the worldwide effort to reach a consensus on a minimum set of endpoints to be used in all RA clinical trials. These efforts make it possible now to recommend what may be called the WHO/ILAR core set of endpoints for RA clinical trials.

From the Department of Internal Medicine/Rheumatology, University Hospital Maastricht, The Netherlands; the Department of Medicine, University of Ottawa, Ottawa, Canada; the Arthritis Center, Boston University School of Medicine, Boston, USA; the Department of Rheumatic Diseases, University of Nijmegen, The Netherlands; the Department of Medicine, Rheumatology Unit, Bristol University, Bristol, UK; the Department of Rheumatology, St. George Hospital, Sydney, Australia; the Department of Rheumatic Disease, 2nd Department of Medicine, Lainz Hospital, Vienna, Austria; the Department of Non-communicable Diseases, World Health Organization, Geneva, Switzerland; and the Department of Medicine, Royal Melbourne Hospital, Melbourne, Australia.

M. Boers, MD, PhD, MSc, Associate Professor of Rheumatology; P. Tugwell, MD, FRCP, Professor of Medicine, Head of Department; D.T. Felson, MD, MPH, Associate Professor of Medicine and Public Health; P.L.C.M. van Riel, MD, Associate Professor of Rheumatology; J.R. Kirwan, MD, FRCP, Senior Lecturer in Rheumatology; J.P. Edmonds, MB, BS, FRACP, Conjoint Professor of Rheumatology; J.S. Smolen, MD, Professor and Head of Department; N. Khaltaev, MD, Acting Chief; and K.D. Muirden, AO, MD, FRACP, Director of Rheumatology, Past-President ILAR.

Address reprint requests to Dr. M. Boers, Department of Internal Medicine/Rheumatology, University Hospital, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands.

A short history of recommendations. Before 1991, 5 meetings were held on endpoints in clinical trials: in Santa Barbara, USA in 1980[4]; in Hamilton, Canada in 1981[5]; in London, UK in 1983[6] and 1986[7], and in Droitwich, UK in 1987[8]. These meetings resulted in various recommendations. In Santa Barbara, it was concluded that a combination of articular index, pain, and global response were sufficient as endpoints[4]. In Hamilton, a methodological framework to select valid endpoints and indices was proposed[5,9,10]. It was suggested that joint count, pain, global assessment, morning stiffness, and grip strength were key measures. A separate measure of physical function was proposed. Finally, a pooled index aggregating several measures was proposed as a summary index for clinical trials. In Droitwich[8], the Ritchie articular index, pain, the Health Assessment Questionnaire[11], the erythrocyte sedimentation rate (ESR), C-reactive protein, and radiographs of hands and feet were selected as measures.

At the first meeting in London[6], the balance of opinion was in favor of pain, global assessment, and joint counts, together with ESR, and rheumatoid factor (RF); for longer studies radiographs and measures of disability were advised. At the second meeting in London[7], the development of a simple index was proposed to measure the response to antirheumatic drugs, based on the criteria for remission. It was also felt that measurement of serious morbidity (e.g., destruction of major joints, development of major extraarticular features, and major side effects of drug treatment)

must be standardized, and that the relation of these dimensions of health status with functional indices must be determined.

Problems with existing measures. The problems with existing measures are in their validity, their relation with individual patient outcomes, and in their multitude. Regarding validity, the traditionally used endpoints can be classified as process or outcome measures. Process measures represent the pathophysiological occurrences that follow from the cause of RA, such as inflammatory activity, whereas outcome measures represent the suffering or loss of health experienced by an individual as a result of the process of disease[12,13]. Outcome reflects the values of the patient and of society, and research over the last 20 years has shown that it may be best measured as a number of areas or dimensions that represent distinct but related disease effects. Summarized with a series of D's, these dimensions include distress (pain), disadvantages (drug side effects), disability/dysfunction, disharmony, dissatisfaction, dollar cost, and death[10,14]. These dimensions are intuitively obvious. For example, a patient wishes to be alive, functioning, and free of symptoms. While outcome is the most relevant measurement category, process measures such as the ESR are valuable insofar as they serve as proxies for outcome. For example, patients with a consistently normal ESR might have less destructive and progressive disease than patients with an elevated ESR. Some measures are hybrid: e.g., grip strength contain components of patient outcome (physical function), but is also an indicator of inflammatory activity.

The list of endpoints selected in the conferences mentioned above, and the endpoints recommended by the US Food and Drug Administration (FDA) and international bodies such as the International an European Leagues Against Rheumatism contain many process and hybrid measures. For example, the FDA recommends using the number of painful joints, the number of swollen joints, morning stiffness, grip strength, 50 foot walk time, ESR and physician and patient global assessments for all antirheumatic drug studies[15]. For chronic studies, Steinbrocker functional and anatomical classification, hand radiographs, and RF are recommended[15]. However, in such recommendations, most of the dimensions of health status are inadequately addressed, and many of the process measures suggested cannot serve as proxy for patient outcome such as death or disability.

Moreover, some recommended endpoints may be invalid because they are duplicative (e.g., tender/painful/swollen joints), unreliable (e.g, 50 foot walk time), or insensitive to change (e.g., RF). Many measures are not adequately standardized, so that each group uses its own variant under a common name (e.g., active joint count).

A separate issue is that the result of the trial is usually focussed on the mean value of its endpoints. Thus it is often hard to translate the result into an expected result for a prospective patient to be treated with the drug or regimen in question[2]. Finally, the multiplicity of outcome measures, assessments, and comparisons in most trials makes it extremely difficult to interpret the result of a particular trial.

POSSIBLE SOLUTIONS

A few high quality outcome measures. To improve the quality of endpoints in RA clinical trials, they should be carefully selected according to the purpose of the trial, using published validity criteria[10.] More dimensions in health status or outcome should be covered. Specifically, physical function and pain should be measured with one of the instruments currently available. There is increasing recognition of the importance of endpoints that reflect the perception of the patient, but the instruments to measure these endpoints are still under development. Second, the number of endpoints can be reduced by eliminating those of lesser quality. For example, Anderson, *et al* and Paulus, *et al* recently analyzed data from several studies, and concluded that a set of 4-6 measures of inflammatory activity was optimal to discriminate between patients treated with active drug and patients treated with placebo[16,17]. The recommended measures included joint tenderness count, ESR, grip strength, and physician global assessment. Such reports indicate that selection of instruments can be, at least partially, based on evidence from appropriately designed studies. Sufficient evidence has accumulated that it is now reasonable to require that only instruments meeting minimal levels of accuracy and responsiveness to change should be included as major endpoints in trials.

A single pooled outcome measure. Measures can be pooled into a single score or index based on retrospective or prospective criteria[18]; several indices of proven validity have become available recently[19]. One of these can be used as the endpoint. The advantage of a single measure lies in the increase in statistical power; possible disadvantages lie in the interpretation of an unfamiliar measure, and in pooling endpoints that truly measure disparate processes or outcomes.

RECENT DEVELOPMENTS

From 1991 to date, a series of meetings have hastened developments. In 1991 the Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials of the American College of Rheumatology (ACR) met in Boston to review the validity of RA endpoints[20,21]. They reviewed construct, face, content, criterion, and discriminant validity as suggested by Tugwell and Bombardier[10]. Data was presented from the literature, extended by analyses of several large datasets from trials of 2nd line antirheumatic drugs and nonsteroidal antiinflammatory drugs. Structured discussions took place on the basis of the nominal group technique[22]. The committee achieved consensus to include a minimum of 6 endpoints, or disease activity measures, in

every trial: pain, physical disability, tender and swollen joint counts, patient and physician global assessments. For trials lasting over 1 year in which drugs were tested as disease modifying antirheumatic drugs, imaging of joints was recommended. The committee also made suggestions on the preferred method of measurement.

In a separate process, a group of rheumatologists joined in the Consensus Study Group of the European Workshop for Rheumatology Research. This group performed a prospective study across Europe in 12 centers, enrolling 282 patients in need of 2nd line antirheumatic therapy[23]. They noted high variability between centers in certain measures, and problems with assessing functional disability, partly because of unavailability of validated versions of the Health Assessment Questionnaire[11] in some European languages. On the basis of correlations noted between variables, a set of 5 endpoints to assess disease activity was selected: these were similar to the ACR core set, with the addition of ESR but excluding physician global assessment, physical disability, and imaging. The European League Against Rheumatism (EULAR) Standing Committee for International Clinical Studies later designated physical disability and imaging as outcome measures to be added to their recommendations[24]. This committee also suggested preferred methods of measurement for several of the endpoints.

The Outcome Measures in RA Clinical Trials (OMERACT) conference was held in Maastricht in 1992 under the auspices of the World Health Organization (WHO), the International and European Leagues Against Rheumatism (ILAR and EULAR), and several of the world's national colleges of rheumatology. This conference brought together 92 rheumatologists, methodologists, drug regulatory officials, and pharmaceutical physicians from all over the world, including most of those involved in formulating the recommendations outlined above[25]. One of the objectives of the conference was to develop consensus on the minimum number of outcome measures to be included in all RA clinical trials. Other objectives included the development of criteria for minimum clinically important improvement in RA patients, and minimum important difference between treatment groups in RA clinical trials; and study of the usefulness of aggregate outcome measures (indices) in the assessment of patients and trials.

In plenary sessions and in small groups, various techniques were used to elicit opinions and preferences: direct questioning, rating of sample profiles of patients and trials, and interactive voting before and after discussion.

The conference adopted a hybrid of the 2 recommendations, coming to a final set of 7 + 1 measures (Table 1). There was no full agreement on physician global assessment: a sizable group felt this measure was redundant in the presence of patient global assessment. However, in the end it was retained as a compromise until validation studies define its worth further. The conference started with 2 divergent recommendations from the ACR and the EULAR on the preferred method of measurement of each of the measures. For example, the ACR prefers full joint counts, but the EULAR reduced joint counts; also the ACR recommendations on other measures are less specific than EULAR's[21,24]. It was decided the postpone discussions on this issue pending further studies comparing validity of the different methods.

Table 1. *WHO/ILAR core set of endpoints for RA clinical trials*

1.	Pain
2.	Patient global assessment
3.	Physical disability
4.	Swollen joints
5.	Tender joints
6.	Acute phase reactants
7.	Physician global assessment
In studies of 1 or more years' duration	
8.	Radiographs of joints

A very important aspect of this conference was the explicit consideration of methodological issues that emerged, including measurement methodology; and the recognition that promising fields of measurement (e.g., psychosocial) are under development[13]. It was felt that future studies should also focus on the drawbacks of therapy (cost, toxicity), and its relation to efficacy. Improvement criteria and indices were explored, but decisions of these issues were postponed.

Finally, in Atlanta in 1992, the Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials of the ACR voted to accept the recommendations made during the Maastricht conference[21].

FINAL RECOMMENDATIONS

Core set of endpoints. We suggest that the core set of endpoints agreed on at the OMERACT conference deserves World Health Organization and International League Against Rheumatism endorsement as a preliminary core set for use in RA clinical trials. This set is not permanent: a proactive program is planned to test the validity of these endpoints and the methods for their measurement. The core set should be actively reviewed on a regular basis by an international group of rheumatologists to accommodate emerging data on the validity of existing and new measures, regardless of their presence in the core set. Those publishing in this area should be encouraged to refine this core set through prospective studies (by confirmation, extension, reduction, or substitution). Meanwhile, adoption of this core set defines a minimum standard of measures for all clinical trials in RA. It is not meant to be exhaustive, and should not prevent inclusion in a trial of other measures that may be necessary to answer specific research questions.

Future research. In relation to the refinement of the core set, the conferences noted several areas where study is necessary: (1) further validation of existing measures, especially in the field of measurement methodology; (2) validation of measures in other domains: e.g., patient utilities, psychosocial aspects, toxicity, costs; (3) definition of improvement criteria; and (4) use of aggregate measures, or indices instead of single measures.

We hope that the remarkable level of international cooperation seen in the last few years may be sustained and strengthened, and lead to answers in these fields.

ACKNOWLEDGMENT

We thank Professor Sj van der Linden for his thoughtful comments.

REFERENCES

1. Kirwan JR, Chaput de Saintonge DM, Joyce CRB: Clinical judgment analysis. *QJ Med 1990;*76:935-49.
2. Felson DT, Anderson JJ, Meenan RF: Time for changes in the design, analysis, and reprint of rheumatoid arthritis clinical trials. *Arthritis Rheum 1990;*33:140-9.
3. Bellamy N: Methods of clinical assessment of antirheumatic drugs. *Clin Rheumatol 1988;*2:339-62.
4. Wright V: Do we need so many assays to detect suppression of inflammation? In: Paulus HE, Ehrlich GE, Lindenlaub E, eds. *Controversies in the Clinical Evaluation of Analgesic-Antiinflammatory-Antirheumatic Drugs*. New York: FK Schattauer Verlag, 1981;208-10.
5. Bombardier C, Tugwell P, Sinclair A, *et al:* Preferences for endpoint measures in clinical trials: Result of structured workshops. *J Rheumatol 1982;*9:798-801.
6. Huskisson EC, Sturrock RD, Tugwell P: Measurement of patient outcome. In: Kirwan JR, Chaput de Saintonge DM, Joyce CRB, Currey HLF, eds. Advances in Assessing Rheumatoid Arthritis. *Br J Rheumatol 1983;*(suppl)22:86-9.
7. Scott DL, Spector TD, Pullar T, McConkey B: What should we hope to achieve when treating rheumatoid arthritis? *Ann Rheum Dis 1989;*48:256-61.
8. Symmons DPM, Dawes PT: Summary and consensus view. *Br J Rheumatol 1988;*27(S1):76-7.
9. Bombardier C, Tugwell P: A methodological framework to develop and select indices for clinical trials. *J Rheumatol 1982;*9:753-7.
10. Tugwell P, Bombardier C: Methodological framework for developing and selecting endpoints in clinical trials. *J Rheumatol 1982;*9:758-62.
11. Fries JR, Spitz P, Young D: The dimensions of health outcomes: The Health Assessment Questionnaire disability and pain scales. *J Rheumatol 1982;*9:789-93.
12. Fries JF: Towards an understanding of patient outcome measurement. *Arthritis Rheum 1983;*26:697-704.
13. Kirwan JR: A theoretical framework for process, outcome and prognosis in rheumatoid arthritis. *J Rheumatol 1992;*19:333-6.
14. White KL: Improved medical care: Statistics and the health services system. *Public Health Rep 1967;*82:847-54.
15. Anonymous. Guidelines for the clinical evaluation of antiinflammatory and antirheumatic drugs (adults and children). US Department of Health and Human Services, Public Health Service, Food and Drug Administration, 1988;1–35.
16. Anderson JJ, Felson DT, Meenan RF, Williams J: Which traditional measures should be used in rheumatoid arthritis clinical trials? *Arthritis Rheum 1989;*32:1093-9.
17. Paulus HE, Egger MJ, Ward JR, Williams HJ: The Cooperating Systematic Studies of the Rheumatic Diseases Group: Analysis of improvement in individual rheumatoid arthritis patients treated with disease modifying antirheumatic drugs, based on the findings in patients treated with placebo. *Arthritis Rheum 1990;*33:477-84.
18. Smythe HA, Helewa A, Goldsmith CH: Selection and combination of outcome measures. *J Rheumatol 1982;*9:770-4.
19. Boers M, Tugwell P: The validity of pooled outcome measures (indices) in rheumatoid arthritis clinical trials. *J Rheumatol 1993;*20:568-74.
20. Felson DT: Choosing a core set of disease activity measures for rheumatoid arthritis clinical trials. *J Rheumatol 1993;*20:531-4.
21. Felson DT, Anderson JJ, Boers M, *et al:* The American College of Rheumatology Preliminary Core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum 1993;*36:729-40.
22. Delbecq AL, Van de Ven AH, Gustafson DH: Group techniques for program planning — A guide to nominal group and delphi processes. Glenview: Scott, Foresman, 1975.
23. Scott DL, Panayi GS, van Riel PLCM, Smolen J, van de Putte LBA, and the Consensus Study Group of the European Workshop for Rheumatology Research: Disease activity in rheumatoid arthritis: Preliminary report. *Clin Exp Rheumatol 1992;*10:521-5.
24. van Riel PLCM: Provisional guidelines for measuring disease activity in clinical trials on rheumatoid arthritis. *Br J Rheumatol 1992;*31:793-4.
25. OMERACT Committee: Conference on outcome measures in rheumatoid arthritis clinical trials, Maastricht, The Netherlands, April 29–May 2, 1992. *J Rheumatol 1993;*20:525-91.