OMERACT Conference on Outcome Measures in Rheumatoid Arthritis Clinical Trials: Introduction

CLINICAL TRIALS ARE ONLY AS CREDIBLE AS THEIR ENDPOINTS

This conference addresses the ongoing challenge of improving the accuracy and responsiveness to change of clinically relevant (to patient and clinician) endpoints.

There are problems with the endpoints presently employed in clinical trials of patients with rheumatoid arthritis (RA). Such trials are best suited to detect short term efficacy of a treatment modality (usually a drug) by comparing it with placebo, or, more recently, with another modality. For this purpose, a set of traditional measures defines the endpoints. However, the measures chosen are not comprehensive, insensitive to change, and show overlap[1]. Despite several conferences, reviews, and editorials in the last 10 years, no consensus exists on the appropriate (set of) endpoints in RA clinical trials[2]. In the following paragraphs, an overview will be given of the existing problems and attempts to solve them. Our focus is on indices that pool information from traditional measures.

Five meetings have been held on endpoints in clinical trials: in Santa Barbara, USA in 1980[3]; in Hamilton, Canada in 1981[4]; in Droitwich, UK in 1987[5]; in London, UK in 1988[6]; and in Boston, USA in 1991[7] and in Arnhem, The Netherlands in 1992[8]. These meetings have resulted in various recommendations.

In Santa Barbara, it was concluded that a combination of articular index, pain and global response was sufficient as endpoints[3].

In Hamilton, a methodological framework to select valid endpoints and indices was proposed[4,9,10]. 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, described in more detail below, was proposed as a summary index for clinical trials.

In Droitwich[5], the Ritchie articular index, pain, the Health Assessment Questionnaire, the erythrocyte sedimentation rate (ESR), C-reactive protein, and radiographs of hands and feet were selected as measures.

In London[6], 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 two dimensions of health status with functional indices must be determined.

In Boston[7], a review of existing data led to the selection of a preliminary core set consisting of patient pain, patient and physician global assessments, self-assessed physical disability, and tender and swollen joint counts.

In Arnhem[8] the results of a prospective study of 282 patients from 12 European centers suggested that the most useful measures to assess disease activity were the number of swollen joints, number of tender joints, pain, patient's assessment of response and ESR.

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, it must be recognized that the measures available as endpoints can be classified as "process" or "outcome" measures. Process measures represent "what happens along the way," i.e., inflammatory activity, whereas outcome measures represent end results[11]. Outcome must reflect the values of the patient and of society, and are usually intuitively obvious. For example, a patient wishes to be alive, functioning, and free of symptoms. Conversely, a normal ESR has no intrinsic importance. Obviously, outcome is the more relevant measurement category. However, 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 disease than patients with an elevated ESR. Some measures are hybrid: painful joint count and grip strength both contain components of patient outcome (physical function), but are also indicators of inflammatory activity.

The list of endpoints selected in the conferences mentioned above, and the endpoints recommended by the Food and Drug Administration (FDA) and international bodies such as the International and European Leagues Against Rheumatism contain mostly 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[11]. For chronic studies, Steinbröcker functional and anatomical classification, hand radiographs, and rheumatoid factor (RF) are recommended[12].

In such recommendations, most of the dimensions of health status are inadequately addressed. Summarized with a series of D's, these dimensions include distress (pain),

disadvantages (drug side effects), disability/dysfunction, disharmony, dissatisfaction, dollar cost and death[10,13]. Moreover, some recommended measures are potentially duplicative (e.g., tender/painful/swollen joints), unreliable (e.g., measurement of RF), or insensitive to change (e.g., Steinbröcker functional class). 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[1]. Finally, the multiplicity of outcome measures, assessments, and comparisons in most trials, makes it extremely difficult to interpret the end result.

SOLUTIONS

A few high quality outcome measures. To improve the quality of outcome measurement in RA clinical trials, several suggestions can be made. First, endpoints should be carefully selected according to the purpose of the trial, using the validity criteria mentioned previously. More dimensions in health status should be covered. Specifically, physical function and pain should be measured with one of the instruments currently available. Second, the number of endpoints can be reduced by eliminating those which are of lesser quality. For example, Anderson, *et al* and Paulus recently conducted analyses on the 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[14,15]. The measures recommended were joint tenderness count, ESR, grip strength, and physician global assessment.

A single, pooled outcome measure. Measures can be pooled into a single score or "index" based on retrospective or prospective criteria[16]. This index can be used as the endpoint. The advantage of a single measure is obvious; the disadvantages lie in the interpretation, i.e., translation back to more familiar measures. It must be kept in mind that both clinicians ("consumers" of research) constantly pool information for their own use; the process is implicit, intuitive, and based on experience. It is retrospective, taking place with the results of data collection in hand. The investigator may also pool information retrospectively, from multiple studies using metaanalytic techniques or doing a multivariate analysis on the collected data to find the best combination of measures to include in an index.

The field has progressed considerably over the past decade. First, there has been increasing acceptance by clinicians of the inclusion of endpoints that reflect the perception of the patient; and secondly, selection of instruments can be at least partially evidence-based — i.e., in appropriate studies using appropriate research design, 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.

The OMERACT conference was planned with 3 goals: (1) To attempt to obtain agreement on the minimum number of outcome measures to be included in all RA clinical trials. This was implemented by a preconference questionnaire, presentation of the evidence on their validity, both small group and plenary discussions on their performance in trials and in individual patients, and then by voting using an electronic voting procedure. (2) To review the range of magnitude of differences judged to be clinically important by experienced clinicians and clinical investigators. This was implemented by a baseline questionnaire and rank ordering of a series of clinical trials and individual patient scenarios, using a nominal group technique. (3) To review the extent to which experienced clinicians and clinical investigators feel that aggregate measures (indices) are useful in the assessment of trials and individual patients. This was implemented by presentation of the concepts behind a variety of examples of indices, by questionnaire and a scenario ranking exercise incorporating the results of 3 indices.

The following papers describe the details of each of these components.

PETER TUGWELL, MD,
Departments of Medicine and Epidemiology and Community Medicine,
University of Ottawa,
Department of Medicine,
Ottawa General Hospital,
Ottawa, ON, Canada;
MAARTEN BOERS, MD,
Departments of Internal Medicine and Rheumatology,
University Hospital,
Maastricht, The Netherlands,
for the OMERACT Committee.

Address reprint requests to Dr. P. Tugwell, Department of Medicine, Room LM12, Ottawa General Hospital, 501 Smyth Road, Ottawa, Ontario K1H 8L6, Canada.

REFERENCES

1. Felson DT, Anderson JJ, Meenan RF: Time for changes in the design, analysis, and reporting of rheumatoid arthritis clinical trials. *Arthritis Rheum 1990;*33:140-9.
2. Bellamy N: Methods of clinical assessment of antirheumatic drugs. *Clin Rheumatol 1988;*2:339-62.
3. Wright V: Do we need so many assays to detect suppression of inflammation? In: Paulus HE, Erhlich GE, Lindenlaub E, eds. *Controversies in the Clinical Evaluation of Analgesic-Antiinflammatory-Antirheumatic Drugs*. New York: FK Schattauer Verlag 1981;208-10.
4. Bombardier C, Tugwell P, Sinclair A, *et al:* Preference for endpoint measures in clinical trials: result of structured workshops. *J Rheumatol 1982;*9:798-801.
5. Symmons DPM, Dawes PT: Summary and consensus view. *Br J Rheumatol 1988;*27(S1):76-7.
6. 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.
7. Felson DT: A core set of disease activity measures for rheumatoid arthritis (RA) clinical trials (abstr). *Arthritis Rheum 1992;*35:S203.
8. Scott DL, Panayi GS, van Riel PLCM, Smolen J, van de Putte LBA and the Consensus Study Group of the European Workshop for Rheumatology: Disease activity in rheumatoid arthritis: preliminary report of the Consensus Study Group of the European Workshop for Rheumatology Research. *Clin Exp Rheumatol 1992;*10:521-5.
9. Bombardier C, Tugwell P: A methodological framework to develop and select indices for clinical trials. Statistical and judgmental approaches. *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 JF: Towards an understanding of patient outcome measurement. *Arthritis Rheum 1983;*26:697-704.
12. 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.
13. White KL: Improved medical care: statistics and the health services system. *Public Health Rep 1967;*82:847-54.
14. 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.
15. Paulus HE, Egger MJ, Ward JR, Williams HJ, the Cooperating Systematic Studies of Rheumatic Disease 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.
16. Smythe HA, Helewa A, Goldsmith CH: Selection and combination of outcome measures. *J Rheumatol 1982;*9:770-4.