Workshop Report

Guidelines of Osteoporosis Trials

Randomized clinical trials are the key source of information about the efficacy of drugs used in the prevention and treatment of osteoporosis. Currently, the measures of efficacy used in clinical trials in osteoporosis vary considerably from trial to trial, making it difficult to judge the relative efficacy of different therapies or pool results in metaanalyses. This is particularly important, since few direct "head to head" comparisons of active agents exist in the literature or are likely to be performed, and many studies are too small to answer some types of questions on their own, such as the effect of treatment on fracture rates. For these reasons it would be useful for trials to share a set of measures that would allow comparisons and pooling of results.

In April 1996, the third meeting of a series of conferences examining outcome measures in clinical trials in musculoskeletal diseases (OMERACT) was held in Cairns, Australia. OMERACT III focused on 2 diseases, osteoarthritis and osteoporosis. This article reports suggested outcome measures proposed for use in future randomized clinical trials of therapies in osteoporosis arising from that meeting. These recommendations are an attempt to standardize outcome measures used in future phase III plus and non-registration clinical trials to allow results to be pooled and compared, and should not be regarded in any sense as guidelines for drug development or regulatory approval.

Types of Trials

Outcome measures for osteoporosis trial were discussed according to 2 broad groupings of trials, namely (a) randomized trials where prevention of rapid bone loss was the primary aim and (b) randomized trials where prevention of fractures may be feasible outcome because patients were already at high risk for osteoporotic fractures either on the basis of low bone mass or previous osteoporotic fractures. Examples of studies in group (a) would include prevention of perimenopausal bone loss or corticosteroid associated bone loss in patients commencing corticosteroid therapy.

Outcome measures were considered in 3 categories: (a) core outcome measures of clinical benefit that would be recommended for inclusion in all randomized trials to allow comparisons and pooling, (b) core outcome measures of toxicity, and (c) non-core outcome measures of clinical benefit that should be considered for inclusion to further our understanding about the value of these outcomes. OMERACT participants recommended that a small number of measures be considered "core" measures. To be considered a core outcome measure of clinical benefit, the measure needed to be sufficiently important to influence clinical decisions about a new therapy in patients, sufficiently responsive to treatment so that the trial would be able to detect statistically significant and clinically important differences between groups with feasible sample sizes, and widely available in research settings. Non-core outcome measures of clinical benefit were judged to be measures that were considered of major clinical importance, but for which there were insufficient data on responsiveness to therapy in terms of statistical power or in which further research on the best actual endpoint was required.

Prevention of Bone Loss

Core measurements. For randomized trials where prevention of rapid bone loss was the primary aim, 2 core outcome measures of clinical benefit were considered appropriate: (a) bone mineral density (BMD) — measured at 2 sites, the lumbar spine and proximal femur; (b) biochemical markers — including at least one resorption marker (which should be based on urinary crosslink excretion) and at least one formation marker. The conferees recommended BMD of the lumbar spine because vertebral fractures are very common and an important outcome of osteoporosis, a common endpoint of trials, and measurement at this site has been very responsive to treatment, even in elderly women. Measurement of the hip was also recommended because hip fractures are the most disabling and costly consequence of osteoporosis; hip BMD appears to be a stronger predictor of hip fractures than measurements at other sites, and hip BMD has been responsive to several types of interventions. BMD would usually be measured by dual energy x-ray absorptiometry, but conferees concur that other measurements, such as QCT, were also appropriate and could be included in trials depending on local interest and availability. Conferees recommended biochemical markers for further evidence of efficacy and mechanism of action. Non-core outcome measures of clinical benefit were considered to be: (a) fractures, (b) quality of life (QOL), and (c) change in height (measured in a standardized fashion).

Non-core measurements. Although it was considered important that fracture be reported to allow pooling of data for potential future metaanalyses, it was similarly thought such studies need not be powered to show a significant difference in fractures in groups, since agents used in prevention studies would be assessed for antifracture efficacy in the 2nd type of randomized clinical trial, discussed below. With

respect to reporting fractures, conferees recommended that both nonvertebral and vertebral fractures be included and identified separately. For nonvertebral fractures, a record of whether the fracture was associated with low or high energy trauma should also be noted. Because self reports of fractures are inaccurate about 15—20% of the time, a radiograph or radiographic report would be required to document each fracture event. Conferees recommended that incident vertebral fractures should be recorded in one of 2 ways and preferably by both, namely, (a) morphometry (where an incident fracture was defined as a 20% change in vertebral height from a previous standardized radiograph with at least a 4 mm absolute decrease in height) or (b) using a semi-quantitative (SQ) index[1] (where an incident fracture was defined as a change in SQ grade > 1). It was recommended that QOL instruments should encompass both general and disease targeted aspects.

Safety and toxicity. With regard to core toxicity outcomes, conferees recommended that adverse events be recorded even in randomized clinical trials of non-drug based therapies and standardized according to OMERACT II guidelines[2]. It was considered important that inclusion/exclusion criteria and patient recruitment procedures be defined in enough detail to allow comparisons between studies.

Most conferees believed it would be useful for trials to record health service utilization costs related to adverse events due to trial medication. The duration of such studies should be 2-3 years, but core outcome measures should be recorded at yearly intervals. In trials of some drugs, a 3rd BMD measurement site (a cortical site) and histomorphometry would also be appropriate if the preclinical and earlier phase II/III studies raised any issues with respect to adverse effects on the material and structural properties of bone.

Randomized Trials of Fracture Prevention in High Risk Populations

For randomized clinical trials of therapies for treating high risk patients for osteoporotic fractures, e.g., in patients with a previous osteoporotic fracture or low bone mass, the following core outcome measures of benefit were considered appropriate: (a) fractures, (b) hip and spine BMD, (c) biochemical markers, and (d) change in height. The principal difference between core outcomes for fracture prevention studies and bone loss prevention studies is the inclusion of fractures as an endpoint in this type of trial. The method of reporting fractures would be identical to that discussed above. Other core measurements, such as BMD, would be identical to those described in the previous section. Biochemical markers should again include at least one resorption marker and one formation marker. Conferees recommended that non-core outcome measures of benefit to be considered include (a) a QOL instrument, (b) a back pain measure, (c) an economic evaluation including health service utilization, e.g., hospitalization, co-therapy costs, nursing home days related to trial medication, and (d) a measure of incident falls. As for prevention studies, core toxicity outcome measures must be reported in a standardized fashion according to OMERACT II guidelines. These studies should be 3-5 years in duration. As noted above, a 3rd cortical measurement site and histomorphometry may be appropriate for some drugs.

Core measures would be assessed yearly (although not necessarily radiographs for vertebral fractures), but must include 3 year data to allow comparison with other agents at this time point. Consideration for poolability would need to take account of the inclusion/exclusion criteria and patient recruitment procedures. Both intention to treat and dropouts should be reported.

Research Priorities

The following items of high priority for future research were also identified: (a) the role of ultrasound for monitoring changes, (b) the need for comparative studies of the responsiveness of various bone markers to therapy —conferees encouraged studies to store urine and sera for future comparative studies of the responsiveness of markers of bone remodeling, (c) comparative studies of the responsiveness of different QOL instruments to therapy, and (d) comparative studies of the responsiveness of the various measures of incident vertebral deformity to therapy.

If randomized clinical trials are to provide relevant information about the efficacy of drugs used in the prevention and treatment of osteoporosis, the measures of efficacy must be standardized. The suggested outcome measures proposed at OMERACT III are not intended to be in any way prescriptive and clearly other outcome measures will develop over time. Similarly, the recommendations are not intended for use in the context of drug development or regulatory approval, in which other bodies such as the World Health organization, EFFO, and GREES are currently developing guidelines. The research priorities outlined above should allow non-core measures to be included as core measures in the future. Agreement about these core measures for use in future randomized clinical trials would allow comparisons of efficacy at least in regard to these few common endpoints in all future studies against a common standard.

These suggested core measures are being submitted to the key international groups involved in osteoporosis research for comment to encourage the widest possible acceptance and implementation of these tools in future randomized clinical trials.

REFERENCES

1. Watts NB, Harris ST, Genant HK, *et al:* Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. *N Engl J Med 1990;*323:73-9.
2. Brooks PM, Day RO: Toxicity of antirheumatic drugs. *J Rheumatol 1995;*22:998-9.

Philip N. Sambrook, MD, FRACP, Professor of Rheumatology, Royal North Shore Hospital and University of Sydney, Sydney, Australia; Steven R. Cummings, MD, Professor of Medicine, University of California, San Francisco, USA; John A. Eisman, Professor of Endocrinology, Garvan Institute of Medical Research, Sydney, Australia; Harry K. Genant, MD, Professor of Radiology, University of California, San Francisco, USA; Richard Eastell, MD, MRCP, Professor of Medicine, University of Sheffield, Sheffield, UK; Professor Jean Y. Reginster, MD, PhD, Bone/Cartilage Metabolism Unit, CHU Brull Liege, Belgium; Professor Jan Dequeker, MD, PhD, Afdeling Rheumatologie, Academisch Ziekenhuis, Pellenberg, Belgium; Professor Stuart Silverman, MD, University of California, San Francisco, USA; Professor Jonathan D. Adachi, MD, FRCPC, St. Joseph's Hospital, MacMaster University, Hamilton; Peter Tugwell, MD, FRCPC, Professor of Medicine, University of Ottawa, Ottawa, Canada.

These guidelines have been endorsed by the US National Osteoporosis Foundation.

Address reprint request to Dr. P. Sambrook, Department of Rheumatology, Royal North Shore Hospital, St. Leonards, Sydney, Australia 2065.