Osteoporosis Clinical Trials Endpoints: Candidate Variables and Clinimetric Properties

ANN CRANNEY, PETER TUGWELL, STEVEN CUMMINGS, PHILIP SAMBROOK, JONATHAN ADACHI, ALAN J. SILMAN, WILLIAM J. GILLESPIE, DAVID T. FELSON, BEVERLEY SHEA, and GEORGE WELLS

ABSTRACT. We reviewed evidence on endpoints used in osteoporosis clinical trials to assist in the development of a set of endpoints to be included in all trials. A MEDLINE search was conducted using the Cochrane Collaboration strategy for each endpoint. Additional published literature was obtained from content experts. A proposed list of endpoints was developed after consultation with experts in the field. Each endpoint was evaluated with respect to validity, reproducibility, redundancy, and feasibility. We classified the endpoints into 2 major categories: clinical health status outcomes and intermediate endpoints, and for each endpoint we present current evidence from the literature as pertains to defined methodologic criteria. Multiple endpoints have been used in osteoporosis clinical trials, and an agreement on a core set of measures needs to be evidence based with an emphasis on validity, reproducibility, and feasibility and to satisfy clinical credibility. (*J Rheumatol 1997;24:1222–9*)

Key Indexing Terms:
OSTEOPOROSIS OUTCOMES CLINICAL TRIALS

Clinical trials are a key source of knowledge about the efficacy of interventions to prevent and treat the clinical consequences of osteoporosis. In clinical trials of patients with osteoporosis, the endpoints measured vary from single measures (usually bone density) to assessing multiple types of endpoints.

Why is reliance on different sets of outcomes a problem? First, given the many different interventions available and under development, it is unlikely that the different therapies will all be directly compared "head-to-head" within the same study; the use of dissimilar outcome measures to assess different interventions in different trials makes it difficult, if not impossible, to judge therapies against a common standard. Second, assessing multiple outcomes increases the likelihood of a Type I error at conventional levels of statistical significance. Third, the absence of preset core endpoints allows the selection and reporting of only those endpoints that report impressive results. Fourth, endpoints with poor responsiveness to change (e.g., a poorly designed scale or set of questions regarding quality of life) may miss clinically important true benefits and risks. Fifth, the Cochrane Collaboration is committed to summarizing the evidence for interventions in major conditions, including osteoporosis, for use by clinicians and policy makers. Metaanalyses are not possible without similarities in the assessment and reporting of the clinically important endpoints.

We outline the data available on candidate variables for assessing the benefit of interventions in osteoporosis to serve as a background paper for the OMERACT III Conference, where participants were asked to: (1) propose a preliminary core set of endpoint measures for use in osteoporosis clinical trials; (2) propose specific methods for assessing each outcome measure that fulfill minimal psychometric criteria; (3) propose how the preliminary core set results should be reported for subsequent inclusion in meta-analysis and economic evaluation; and (4) define a research agenda to validate and establish consensus on the preliminary core set of osteoporosis outcomes.

The approval by regulatory authorities of new osteoporosis drugs requires evidence of benefit from phase 3 pivotal studies. The requirements for which endpoints are included in these vary from country to country. In the United States and Canada, a reduction in fracture rates is required for all classes of drugs except estrogen analogs, where bone density endpoints are sufficient. This raises the interesting concept that perhaps the first agent in a new class of drugs should be required to prove a reduction in fracture rate, but

From the Department of Medicine and Clinical Epidemiology Unit, University of Ottawa and Ottawa General Hospital, Ottawa, Canada; University of California, San Francisco, USA; Department of Rheumatology, Royal North Shore Hospital, Sydney, Australia; McMaster University, Hamilton, Canada; ARC Epidemiology Research, University of Manchester, Manchester; Department of Orthopedic Surgery, Edinburgh, UK; and Arthritis Center, Boston University, Boston, USA.

A. Cranney is the recipient of Arthritis Society Research Fellowship.

A. Cranney, MD, Rheumatologist; P. Tugwell, MD, Professor and Chairman, Department of Medicine, University of Ottawa and Ottawa General Hospital; S. Cummings, MD, Professor of Medicine and Epidemiology and Biostatistics, Chair of Internal Medicine, University of California; P. Sambrook, MD, Professor, Department of Rheumatology, Royal North Shore Hospital; J. Adachi, MD, Professor, Department of Medicine, Director of Rheumatology, McMaster University; A.J. Silman, MD, Professor, Director, ARC Epidemiology Research, University of Manchester; W. Gillespie, MD, Professor and Head of Orthopedic Surgery, University of Edinburgh; D.T. Felson, MD, Professor, Arthritis Center, Boston University; B. Shea, BScN, Research Associate, Department of Medicine; G. Wells, PhD, Associate Director and Senior Biostatistician, Clinical Epidemiology Unit, University of Ottawa and Ottawa Civic Hospital.

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

that subsequent drugs in the same class should only be required to show improvement in appropriate intermediate endpoints such as bone mineral density (BMD).

We review the published evidence on the endpoints most commonly used in trials, which was used as supporting documentation for the discussions in these proceedings.

OBJECTIVES

The objective is to propose a set of endpoint measures that would be included in all trials. These need to meet 2 major criteria, namely, be clinically credible, and be evidence based, in that they should satisfy designated methodologic criteria.

Clinical credibility within a conceptual framework

The same model was used elsewhere in these proceedings for both osteoporosis and osteoarthritis; in the osteoarthritis background paper[1], Bellamy proposes that the selection of candidate variables be placed in the context of a conceptual framework that begins with the underlying cellular pathology and extends through to the clinical manifestations, and thence to the World Health Organization classification of outcomes (i.e., impairment, disability, handicap). The appropriate outcomes in clinical trials will depend to some extent at least upon the stage of the disease process defined by the inclusion criteria. The European Foundation for Osteoporosis and the Bone Guideline Group have argued for different endpoints for each of 3 groups: (1) Prevention: intervention in people with normal skeleton status, for example, regimens that modulate peak bone mass and its preservation; (2) treatment of osteopenia: intervention in patients with osteopenia but without fractures to reduce the probability of future fractures; and (3) treatment of established osteoporosis in patients with one or more fragility fractures to reduce the probability of future fractures. The concern with this classification is that the distinction between normal skeleton status and osteopenia is not clearcut, with bone density being a continuous variable, decreasing levels of which are associated with progressive increases in risk of clinical fractures; in addition different individuals are classified as osteopenic using different measurement techniques. A simpler classification would be (1) primary: treatment to prevent clinical fractures in individuals who have not had a previous clinical fracture (keeping in mind that primary prevention may also refer to calcium supplementation in early adulthood); and (2) secondary: treatment to prevent further clinical fractures in individuals who have sustained at least one fracture.

Other aspects that may influence the selection of a core set include whether the intervention works primarily upon trabecular or cortical bone mass; similarly, whether the intervention works on nonosseous factors such as neuromuscular coordination to reduce falls.

Methodologic criteria

A key feature of the OMERACT process is that the selection of clinical trial endpoints be "data driven," or evidence based. We propose using a modification of the same methodologic framework published by Tugwell and Bombardier[2], used to define the methodologic criteria for acceptability of endpoints for the preliminary core set in rheumatoid arthritis[3]. These are:

(1) Each of 4 types of Validity:

(a) Face: Are the outcomes clinically credible and the results easily understood by clinicians, policy makers, and patients/public?

(b) Content: Do the endpoints cover the multiple domains of improvement in osteoporosis: death, anatomic/morphometric evidence of osteoporosis (e.g., radiographic evidence of fracture, BMD, histology), physical disability, psychosocial function, symptoms (e.g., pain), and quality of life?

(c) Criterion: We chose 3 gold standards, death, physical disability, and radiological evidence of fracture; to have criterion validity, an outcome measure should correlate or predict one or more of these.

(d) Discriminant (Sensitivity-to-Change): Do the endpoints detect the smallest clinically important improvement?

The evaluation consisted of a structured review of the literature on the validity of the measures.

(2) Reproducibility: in ideal circumstances and in usual clinical practice situations.

(3) Redundancy: where one measure duplicated other(s).

(4) Feasibility: availability at sites where clinical trials are conducted.

To consider all studies addressing the validity of osteoporosis endpoint measures, we conducted a MEDLINE search using the Cochrane Collaboration strategy and bibliographic reviews, and sought additional sources from experts in the field. Additional analyses were conducted where there was insufficient published evidence.

REVIEW OF INDIVIDUAL ENDPOINTS

We have classified the endpoints into 2 major categories: namely, true clinical health status outcomes, where these directly assess the symptoms and the health related physical, emotional, and social quality of life of the individual; and intermediate endpoints, where these assess the disease process that is believed to result in the subsequent clinical health status outcomes. These may vary depending upon the perceived mechanism of the intervention, for example, whether it works primarily upon trabecular or cortical bone mass; also, different intermediate endpoints might well be chosen if the intervention works on nonosseous factors such as neuromuscular coordination to reduce falls.

The commonly included endpoints are reviewed below.

True clinical health status outcomes

Clinical fractures. Although all agree that clinical fractures and their sequelae are the outcomes that matter, there is concern that some vertebral deformities may not produce symptoms besides height loss and some mild radiographic "deformities" might not be true fractures. Thus, an argument can be made that the endpoint of clinical fracture should require that there be both radiological evidence meeting the requisite criteria as described below plus documented new onset of pain in the region affected.

A number of studies have included both vertebral and peripheral fractures as outcome measures[4–11]; however, none of these studies defined vertebral fracture as a clinical outcome, except for the recent Fracture Intervention Trial[12]. In this trial clinical fractures were categorized: all clinical fractures, non-spine clinical fractures, hip fractures, wrist fractures, and clinical vertebral fractures. Fractures due to excessive trauma (defined as trauma sufficient to cause a fracture in young individuals with normal bone mass) were excluded[12].

(a) Radiographic evidence of fracture: Although vertebral fractures are a common outcome, there has been a lack of agreement on how to define a vertebral fracture. Unfortunately, a gold standard for vertebral fracture does not exist. The radiographic detection of a vertebral fracture in clinical practice is subject to variability in interpretation and hence routine variability in the reported fracture rate.

More recently, there has been interest in the assessment of 2 techniques, semiquantitative and quantitative (or morphometry), for the determination of prevalent and incident fractures[13–44].

Semiquantitative assessments are performed by a trained radiologist, who grades the extent of each fracture (T4–L4) using mutually exclusive categories; for example, Grade 0 = normal vertebral shape; Grade 1 = 20–25% reduction of anterior height (H_a); Grade 2 = 25–40% reduction of H_a; Grade 3 = > 40% reduction of H_a. This technique has the advantage that anatomical variants and artifacts on radiographs can be detected by the reader. Intra and interobserver variability is a problem that can be minimized by training.

Quantitative morphometry looks at deformity (alteration in vertebral shape) and is performed using a lateral spine radiograph, with the placement of 6 points defining the anterior, middle, and posterior margins of the midplane of the vertebral bodies. It is recommended to measure the dimensions by recording points with a cursor on a computerized digitizing board. Vertebrae from T4 to L4 are measured. Different morphometric criteria have been outlined[15–18] and compared in a study by Black, *et al* and Jones, *et al*[14,19].

Fracture definitions are based on comparison with normal values of means and standard deviations (SD) for each vertebral level derived from the same measurement protocol. It has been suggested that prevalent vertebral fractures be defined on the basis of a reduction of 3 SD or more from normal mean ratios of dimensions for the corresponding vertebrae. Morphometry is reproducible (between 7 and 9% for vertebral heights); however, the accuracy of this method is limited by technique and geometric distortion. Some "fractures" will be artifacts. Also, if criteria to define a fracture are too stringent, this would reduce the false positive rate and compromise sensitivity.

Wu, *et al*[13] recently published a comparison of semiquantitative and quantitative techniques, and found that while quantitative morphometry was a reasonable approach for the diagnosis of prevalent fractures, the 2 methods do not always agree. This was largely attributed to the differences in baseline and followup films.

There has been little information on the best method for defining incident vertebral fractures. A change in vertebral height could represent a measurement error. A standardized protocol is increasingly being employed for both semiquantitative and quantitative (morphometric) assessment of vertebral fractures, with written protocols devised. Without this, it will be difficult to compare this endpoint in clinical trials.

(b) Pain: In up to 35% of patients, incident nontraumatic vertebral fractures will be asymptomatic[20]. Recording episodes of acute back pain without radiographic documentation of new spine fracture can both overestimate and underestimate the occurrence of vertebral fractures. Chronic pain is a feature of vertebral fractures and assessment of pain can be a useful endpoint, especially in rehabilitation programs or trials assessing medications, such as calcitonin, designed to reduce pain. Nonvertebral fractures, such as those of the wrist, generally cause self-limited pain but can also result in chronic pain.

The general literature on pain measurement is extensive. In osteoporosis, assessment of pain in clinical trials has been based on pain rating scales: 5 point Likert[21] or visual analog scales (VAS)[22,23]. These are popular due to their simplicity and ease of administration[24]. Because it is important that perceived pain be quantified, pain scales should be completed by the patient. In a cohort study by Ettinger[25,26], both VAS and a 24 item checklist[27] to elicit back related disabilities were utilized.

Complex pain questionnaires, such as the McGill Pain Questionnaire, or variations on basic scales, such as pain related behaviors, have not been used in clinical trials. Diaries for pain medication or pill counts can be kept to assess individuals' level of pain.

Ryan, *et al* have shown that higher pain and disability scores are found with more severe disease in osteoporotic patients, especially in association with vertebral deformations from T8 to T12[21]. Pain is an important outcome to adequately assess in clinical trials, particularly those that involve patients with pre-existing vertebral deformities. It may, however, be difficult to relate pain to fractures in some patients.

Health status instruments. A major result of many clinical fractures is a negative effect upon the health related quality of life — this is reflected in physical disability and psychosocial dysfunction. Patients, clinicians, and policymakers need information on both the risk and magnitude of short and longterm disability resulting from osteoporotic fractures, and the effect of therapy upon quality of life, if they are to make informed decisions. Most areas of medicine now require that quality of life instruments be included in clinical trials.

There are 4 major types of disability and quality of life assessments: (a) Performance measures: These assess the degree of disability when carrying out observed movements. (b) Self-report disease specific questionnaires: These focus upon activities known to be affected by osteoporotic related fractures. (c) Self-report generic questionnaires: These are general questionnaires covering a broad range of activities designed to reflect disability and psychosocial dysfunction across all types of disease. (d) Patient preference/utility scales: These involve asking the patient to take into account all the benefits, as well as the side effects, inconvenience, and costs and then estimate the amount they improved or deteriorated within the continuum between perfect health and death.

(a) Performance based measures: Although a variety of performance based measures have been included in many osteoporosis studies, they have been less widely used as endpoints than as predictors for future fractures[28,29]. Measures include: (i) Functional reach (difference between patient's arm length and maximal forward reach), ability to stand from a chair, mobility skills, timed tests of gait, and other tests of balance such as the "up and go test"[30]. (ii) Performance based measures of functioning have been shown to be valid and reliable measures of function in a nondisabled older population[31,32]. These measures include: balance, gait, lower extremity strength, upper extremity strength, hand performance, and lower extremity co-ordination[31,32].

It is not clear how fractures may affect performance of these various tests. If selected for inclusion as important endpoints, these measures would need to show responsiveness to reductions in fracture rates in large osteoporosis clinical trials.

(b) Self-report disease specific measures: Several instruments have been developed specifically to assess the disability in patients with established osteoporosis and symptomatic vertebral fractures. These instruments have been developed to assess domains such as fear of falling, independence, back pain, and self-image. Osteoporosis differs from other diseases such as rheumatoid arthritis and osteoarthritis in that the majority of patients are asymptomatic. These instruments are in various stages of validation and include (i) Osteoporosis Functional Disability Questionnaire, (ii) Osteoporosis Quality of Life Questionnaire, (iii) Osteoporosis Assessment Questionnaire, and (iv) Quality of Life Questionnaire of the European Foundation for Osteoporosis.

(i) The Osteoporosis Functional Disability Questionnaire[22] is a self-administered, disease specific health status measure with 5 separate domains. It deals with the current degree of back pain, contains a 20 item depression scale, CES-D, a 26 item list of activities of daily living, and assesses the current involvement in recreational and social activities. It has been validated through an exercise program, but not in a randomized clinical trial. Reproducibility, test–retest and internal consistency have been established. Criterion validity and sensitivity to change were proven. Correlation between vertebral fracture severity and quality of life was present. Further validation of this instrument in pharmacologic trials and longitudinal studies would be beneficial.

(ii) The Osteoporosis Quality of Life Questionnaire of Cook, *et al*[33] is a similar disease specific instrument that requires trained persons to administer. The authors were very thorough in their methodology in developing this questionnaire. It consists of 30 questions across domains of symptoms, activities of daily living, physical limitations, emotional function, and leisure activities. A 7 point Likert scale is used for each item. The authors were able to establish validity and sensitivity to change by following a group of eligible patients over a 6 month period[34]. This instrument needs further study in the setting of a clinical trial.

(iii) The Osteoporosis Assessment Questionnaire[35] of Silverman and Mason is a self-report questionnaire based on the Arthritis Impact Measurement Scale 2 core. It has been validated in a nonrandomized clinical trial, and is being studied in a number of multicenter trials.

(iv) The Quality of Life Questionnaire of the European Foundation for Osteoporosis is a self-report questionnaire in the process of being validated[36].

(c) Generic measures: A variety of instruments have been developed to assess various aspects of quality of life. These are being included in ongoing clinical trials but there is little evidence of their current use. Instruments that may be applicable to osteoporosis clinical trials are (i) the Short-Form (SF)-36[37] Health Status Questionnaire, (ii) the Nottingham Health Profile[38], and (iii) the European Quality of Life Questionnaire[37]. Two instruments that specifically assess physical function and disability are (iv) the Functional Status Index[40], and (v) the Days of Disability[41].

(i) Short-Form 36 Health Status Questionnaire is a self-administered questionnaire dealing with 3 major health attributes (functional status, well being, overall health), and 8 health concepts: limitations in physical activities because of health problems; limitations in social activities because of physical or emotional problems; limitations in usual role activities because of physical health problems; bodily pain; general mental health; limitations in usual role activities

because of emotional problems; vitality; and general health perceptions.

(ii) Nottingham Health Profile is designed to give simple indications of perceived physical, social, and emotional health problems. The design and content of this instrument were influenced by the Sickness Impact Profile. It contains 38 items grouped into 6 sections: physical mobility (8 items); pain (8 items); sleep (5 items); social isolation (5 items); emotional reactions (9 items); and energy (3 items). An overall score may be calculated or section scores may be presented as a profile.

(iii) European Quality of Life is a self-administered questionnaire comprising 6 factors: mobility, self-care, social relationships, pain, mood, and social relationships. Respondents rate their health in terms of each dimension; they then rate their health on a VAS[39]. This instrument may be used to derive utility measures for cost effectiveness studies.

(iv) Functional Status Index was developed as a clinical and evaluative tool to measure the degree of dependence, pain, and difficulty experienced in performing activities of daily living. It was developed by Jette and Deniston[40] to evaluate a pilot geriatric arthritis program designed to improve the quality of life in elderly arthritic patients. There is both a 45 and an 18-item version, both interviewer administered. The shortened version takes 20–30 min to administer. The validity, reliability, and sensitivity to change have been well studied; however, there are concerns about the validity of this instrument.

The Functional Status Index has also been used in a small study by Lyles, *et al*[29] to determine if vertebral compression fractures are associated with reduced levels of functional performance. Scores were significantly different in the vertebral fracture group compared to controls. The validity and responsiveness of the index in clinical trials of osteoporosis has not been confirmed.

(v) Days of Disability, designed by the UCSF Study of Osteoporotic Fractures Group to assess the effect of clinical fractures on disability, quantifies days of disability. The UCSF group have used this instrument to document the degree of disability associated with first and recurrent vertebral fracture[41]. It has the advantage of capturing self-limited episodes of pain and days of disability due to acute fractures.

(d) Patient preference/utilities. These include standard gamble, time trade-off, rating scales, and use of the Health Utilities Index (HUI)[41] to represent the net effect on quantity and quality of life. They reflect patient preferences for treatment processes and outcome, and can be incorporated into cost-utility analyses. Some trials are including these measures[43].

Height. Height is an endpoint that can be used to monitor progression of disease, as loss of height is a feature of vertebral osteoporosis[44,45]. Height can be most accurately measured with a stadiometer (mm) with excellent precision. Kleerekoper, *et al*[44,45] found that in osteoporotic women (average age of 67) in whom no new vertebral fractures were observed, the rate of height loss was 1.8 mm/year compared to women with one or more new vertebral fractures, whose rate of height loss was 4.6 mm/year ($p < 0.05$).

Height has been assessed as an endpoint in clinical trials[4,7,9]. In these studies, there was less loss of height in the treatment group, and this finding was significant, except in the Storm paper[7]. Height as an endpoint is most useful in older women (> 65) or in patients with previous vertebral fractures. Height loss can be a useful endpoint to assess the effects of vertebral fractures but it may be the result of other processes.

Kyphosis. Thoracic kyphosis is an outcome that also reflects progression of spinal osteoporosis. Mechanical stresses produced by kyphosis might contribute to chronic back pain. Lyles, *et al*[29] showed that subjects with fractures had significantly more thoracic kyphosis and less lumbar spine lordosis. Various methods have been employed to quantify the degree of kyphosis. In a study by Ettinger, *et al*[46], thoracic curvature was measured using an architect's semiflexible rule (flexi-curve). A kyphosis index was then calculated as 100 times the maximum horizontal distance divided by the vertical length of the upper back curve. This method[47] has been shown to be both valid and reproducible. The kyphosis index was shown to be inversely related to bone density and kyphosis was associated with definite height loss. In contrast, Leidig, *et al*[20] reported only slightly more back pain and disability in women with kyphosis. The study dealt with women hospitalized for vertebral fracture, while Ettinger's study[46] involved a community based population.

This measurement is not widely used in clinical trials and there is little information regarding its sensitivity to change. Another problem inherent with kyphosis is that a number of women with kyphosis do not have vertebral deformities. Kyphosis as an endpoint needs further study, and it may be difficult to incorporate in large multicenter trials.

Intermediate outcomes

Bone densitometry. Prospective studies have shown that women with low bone density are at increased risk of clinical fractures[19,48–53]. Dual x-ray absorptiometry (DXA) rapidly and accurately measures bone density and has become the measure of choice for most clinical trials. For the assessment of appendicular bone, single photon absorptiometry (SPA) and more recently single x-ray absorptiometry have been used. Peripheral quantitative computerized tomography (CT) of the wrist is being evaluated[54]. BMD is strongly correlated statistically with subsequent fractures; for example, Cummings found that each SD decrease in femoral neck bone density increased the age adjusted risk of hip fracture 2.6 times[48]. However, from a clinical perspective, this intermediate outcome is an imperfect proxy for true clinical endpoint. Furthermore, the benefit of treatment

with different agents correlates to different degrees of differences in bone density.

CT scans have been shown to be accurate in assessing bone density but are not recommended for inclusion in a core set of endpoint assessments due to their expense and limited availability. Ultrasound is an attractive new technology that has the potential to be easier, more portable, and less expensive; a number of different machines are being evaluated, most using a variant of broad band ultrasonic attenuation of the os calcis with the foot immersed in a water bath[54–57]. Initial reports suggest the relative risks of associated fractures of the spine and hip are very similar to those of SPA and DXA; these must be confirmed before this technology can be recommended. It is unclear if ultrasound can replace existing bone density measurements given there is inadequate data to recommend its use in monitoring treatment and variability in ultrasound measurement is greater than DXA.

Biochemical markers. In contrast to medical imaging, assessment of biochemical markers allows a more frequent determination of bone metabolism. The biochemical markers are either enzymes involved in bone remodeling or they may be bone matrix components released into the circulation during formation or resorption[58]. High bone turnover is reflected by high concentrations of markers. Bone markers are the most sensitive method for monitoring acute changes in bone metabolism. Riis, *et al*[59] reported a correlation between changes in biochemical markers and bone loss in women undergoing hormonal replacement therapy (HRT). They concluded biochemical markers of bone turnover may be of value in monitoring response of bone to HRT. The rate of bone turnover in postmenopausal osteoporotic women was compared to healthy premenopausal women. When compared to healthy premenopausal women, all formation markers except carboxyterminal propeptide of type I collagen and all resorption markers except carboxyterminal telopeptide of type I collagen exceeded normal concentrations. Also, Garnero, *et al*[60] examined biochemical markers and the changes that occurred with alendronate therapy. With alendronate therapy, resorption markers decreased earlier than formation markers, consistent with antiresorptive therapy. They also found a significant correlation between percentage change in biochemical markers at 3 months and spinal BMD at 24 months.

Studies of biochemical markers have shown good correlation[61–63] between predicted and measured bone mass measurement for groups of patients. However, the variability is large and the predictive ability of markers is not as certain for individual patients. Newer markers such as urinary N-telopeptide or C-telopeptide (crosslaps) of Type I collagen may help to assess a patient's response to therapy. Garnero, *et al* reported that urinary excretion of C-telopeptide and free deoxypyridinoline predicted hip fracture in elderly women, independent of femoral neck BMD[64].

Falls. Most osteoporotic related fractures in older people are associated with a fall, although it is important to note that only around 25% of falls result in injury, of which 4–6% are fractures[65–67]. As it is plausible that interventions targeting fracture prevention might be effective, at least in part, through reducing the number of falls, falling may be an important secondary outcome in trials of fracture prevention.

Risk factors for falling and falls as an outcome have been reported widely, including a number of trials with a range of interventions, such as exercises, reduction of environmental hazards, and targeted medical interventions[68–72].

A consensus definition of falls for research purposes has been published[73]. However, subsequent reports have used other definitions[66,74]. Falling status at enrolment (e.g., no falls, one fall, multiple falls) has been reported and is a variable worth recording in trials of agents that might affect neuromuscular performance. Useful presentations of data have included the number of individuals sustaining any category of fall, the mean number (with SD) of falls sustained by each participant, and the time to first fall.

We have presented an evidence based review of the endpoints, both clinical health status outcomes and intermediate outcomes, to facilitate the selection of a core set of endpoints for osteoporosis clinical trials.

REFERENCES

1. Bellamy N: Osteoarthritis clinical trials: Candidate variables and clinimetric properties. *J Rheumatol 1997;*24:768–78.
2. Tugwell P, Bombardier C: A methodologic framework for developing and selecting endpoints in clinical trials. *J Rheumatol 1982;*9:758–62.
3. Felson DT: Choosing a core set of disease activity measures for rheumatoid arthritis clinical trials. *J Rheumatol 1993;*20:531–4.
4. Liberman UA, Weiss SR, Broll J, *et al:* Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. *N Engl J Med 1995;*333:1437–43.
5. Lufkin EG, Wahner HW, O'Fallon WM, *et al:* Treatment of postmenopausal osteoporosis with transdermal estrogen. *Ann Intern Med 1992;*117:1–9.
6. Overgaard K, Hansen MA, Jensen SB, Christiansen C: Effect of calcitonin given intranasally on bone mass and fracture rates in established osteoporosis: A dose-response study. *BMJ 1992;*305:556–61.
7. Storm T, Thamsborg G, Steinche T, Genant HK, Sorenson OH: Effect of intermittent cyclical etidronate therapy on bone mass and fracture rate in women with postmenopausal osteoporosis. *N Engl J Med 1990;*322:1265–71.
8. Watts NB, Harris ST, Genant HK, *et al:* Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. *N Engl J Med 1990;*323:73–9.
9. Pak CYC, Sakhaee K, Adams-Huet B, Poiziak V, Peterson RD, Poindexter JR: Treatment of postmenopausal osteoporosis with slow-release sodium fluoride. *Ann Intern Med 1995;*123:401–8.
10. Tilyard MW, Spears GFS, Thomson J, Dovey S: Treatment of postmenopausal osteoporosis with calcitriol or calcium. *N Engl J Med 1992;*326:357–62.
11. Chapuy MC, Arlot ME, Duboeuf F, *et al:* Vitamin D_3 and calcium to prevent hip fractures in elderly women. *N Engl J Med 1992;*327:1637–42.

12. Black DM, Cummings SM, Karpf DB, *et al:* Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. *Lancet 1996;*348:1535–41.
13. Wu CY, Li J, Jergas M, Genant HK: Comparison of semiquantitative and quantitative technique for the assessment of prevalent and incident vertebral fractures. *Osteoporosis Int 1995;*5:354–70.
14. Black DM, Palermo L, Nevitt M, *et al:* Comparison of methods for defining prevalent vertebral deformities: The study of osteoporotic fractures. *J Bone Miner Res 1995;*10:890–902.
15. Genant HK, Jergas M, Palermo L, *et al:* Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis. *J Bone Miner Res 1996;*11:984–96.
16. Minne WH, Leidig G, Wuster C: A newly developed spine deformity index to quantitate vertebral crush fractures in patients with osteoporosis. *Bone Mineral 1988;*3:335–49.
17. McCloskey EV, Spector TD, Eyres KS, *et al:* The assessment of vertebral deformity: A method for use in population studies and clinical trials. *J Bone Miner Res 1993;*3:138–47.
18. National Osteoporosis Foundation Working Group on Vertebral Fractures: Report: Assessing vertebral fractures. *J Bone Miner Res 1995;*10:518–23.
19. Jones G, White C, Nguyen PN, *et al:* Prevalent vertebral deformities: Relationship to bone mineral density and spinal osteophytosis in elderly men and women. *Osteoporosis Int 1996;*6:223–39.
20. Leidig G, Minne WH, Sauer P, Wuster C, *et al:* A study of complaints and their relation to vertebral destruction in patients with osteoporosis. *Bone Mineral 1990;*8:217–29.
21. Ryan PJ, Blake G, Herd R, Fogelman I: A clinical profile of back pain and disability in patients with spinal osteoporosis. *Bone 1994;*15:27–30.
22. Helmes E, Hodsman A, Lazowski D, *et al:* A questionnaire to evaluate disability in osteoporotic patients with vertebral compression fractures. *J Gerontol 1995;*50(a):M91–8.
23. Lyritis GP, Tsakalakos N, Magiasis B, Karachalios T, Yiatzides A, Tsekoura M: Analgesic effect of salmon calcitonin in osteoporotic vertebral fractures: A double-blind placebo controlled clinical study. *Calcif Tissue Int 1991;*49:369–72.
24. Bellamy N: *Musculoskeletal Clinical Metrology.* Lancaster, UK: Kluwer Academic, 1993.
25. Ettinger B, Black DM, Nevitt M, *et al:* Contribution of vertebral deformities to chronic back pain and disability. *J Bone Miner Res 1992;*7:449–56.
26. Ettinger B, Block JE, Smith R, Cummings SR, Harris ST, Genant HK: An examination of the association between vertebral deformities, physical disabilities and psychosocial problems. *Maturitas 1988;*10:283–96.
27. Roland M, Morris R: A study of the natural history of back pain. I. Development of a reliable and sensitive measure of disability in low back pain. *Spine 1983;*8:141–4.
28. Cummings SR, Nevitt M, Browner W, *et al:* Risk factors for hip fracture in white women. *N Engl J Med 1995;*332:767–73.
29. Lyles KW, Gold DT, Shipp KM, Pieper CF, Martinez S, Mulhausen PL: Association of osteoporotic vertebral compression fractures with impaired functional status. *Am J Med 1993;*94:595–601.
30. Podsiadlo D, Richardson S: The timed "up and go": A test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc 1991;*39:142–8.
31. Guralnik JM, Seeman TE, Tinetti ME, Nevitt M, Berkman LF: Validation and use of performance measures of functioning in a non-disabled older population: MacArthur studies of successful aging. *Aging Clin Exp Res 1994;*6:410–9.
32. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB: Lower extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med 1995;*332:556–61.
33. Cook DJ, Guyatt GH, Adachi JD, *et al:* Quality of life issues in women with vertebral fractures due to osteoporosis. *Arthritis Rheum 1993;*36:750–6.
34. McClung MR, Love B, Rosen CJ, *et al:* Evaluation of a new osteoporosis quality of life questionnaire (OQLQ) for women with osteoporosis and back pain (abstr). *J Bone Miner Res 1995;*S419.
35. Silverman SL, Mason J, Greenwald M: The osteoporosis assessment questionnaire (OPAQ): A reliable and valid self assessment measure of quality of life in osteoporosis (abstr). *J Bone Miner Res 1993;*(suppl 1)8:S343.
36. Lips P, Agnusdei D, Caulin F, *et al:* The validation of the EFFO questionnaire for quality of life in patients with vertebral osteoporosis (QUALEFFO). *Osteoporosis Int 1996;*(suppl)6:575.
37. Ware JE Jr, Sherbourne CD: The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual frame-work and item selection. *Med Care 1992;*30:473–81.
38. Hunt SM, McEwen J, McKenna SP: Measuring health status; a new tool for clinicians and epidemiologists. *J R Coll Gen Pract 1985;*35:185–88.
39. Economic and Health Outcomes Research Group, Hurst NP, Jobanputra P, Hunter M, *et al:* Validity of EuroQol — a generic health status instrument — in patients with rheumatoid arthritis. *Br J Rheumatol 1994;*33:655–62.
40. Jette AM, Deniston OL: Inter-observer reliability of a functional status assessment instrument. *J Chron Dis 1978;*31:573–80.
41. Nevitt MC, Ettinger B, Black D, Stone K, Genant H, Cummings S: Functional impact of first and recurrent vertebral fracture: A prospective study (abstr). *J Bone Miner Res 1996;*11:M690.
42. Feeny D, Furlong W, Barr RD, *et al:* A comprehensive multiattribute system classifying the health status of survivors of childhood cancer. *J Clin Oncol 1992;*10:923–8.
43. Tosteson AN, Kneeland TS, Nease RF, Sumner W, Minshall M, Draper D: Feasibility of automated utility assessment in phase III clinical trials: U-titer and a raloxifene osteoporosis prevention trial (abstr). *Med Decis Making 1995;*15:413.
44. Kleerekoper M, Nelson DA, Peterson EL, Tilley BC: Outcome variables in osteoporosis trials. *Bone 1992;*13:S29–34.
45. Kleerekoper M, Nelson DA: Vertebral fracture or vertebral deformity. *Calcif Tissue 1992;*50:5–6.
46. Ettinger B, Black DM, Palermo L, Nevitt M: Kyphosis in older women and its relation to back pain, disability and osteopenia: The study of osteoporotic fractures. *Osteoporosis Int 1994;*4:55–60.
47. Chow R, Harrison J, Dornan J: Prevention and rehabilitation of osteoporosis program: Exercise and osteoporosis. *Int J Rehab Res 1989;*12:49–56.
48. Cummings SR, Black DM, Nevitt M, *et al:* Bone density at various sites for prediction of hip fractures. *Lancet 1993;*341:72–5.
49. Nguyen T, Sambrook P, Kelly P, *et al:* Prediction of osteoporotic fractures by postural instability and bone density. *BMJ 1993;*307:1111–5.
50. Ross PD, Genant HK, Davis JW, Miller PD, Wasnich RD: Predicting vertebral fracture incidence from prevalent fractures and bone density among non-black, osteoporotic women. *Osteoporosis Int 1993;*3:120–6.
51. Hui SL, Slemenda CW, Johnston CC Jr: Age and bone mass as predictors of fracture in a prospective study. *J Clin Invest 1988;*81:1804–9.
52. Wasnich RD, Ross PD, Davis JW, Vogel JM: A comparison of single and multi-site BMC measurements for assessment of spine fracture probability. *J Nucl Med 1989;*30:1166–71.
53. Miller PD, Bonnick SL, Rosen F: Consensus of an international panel of the clinical utility of bone mass measurements in the

detection of low bone mass in the adult population. *Calcif Tissue Int 1996;*58:207–14.
54. Porter RW, Miller CG, Grainger D, Palmer SB: Prediction of hip fracture in elderly women: A prospective study. *BMJ 1990;*301:638–41.
55. Ross PD, Huang C, Davis JW, *et al:* Predicting vertebral deformity using bone densitometry at various skeletal sites and calcaneous ultrasound. *Bone 1995;*16:325–32.
56. Bauer DC, Gluer CC, Genant HK, Stone K, for the Fracture Intervention Group: Quantitative ultrasound and vertebral fracture in postmenopausal women. *J Bone Miner Res 1995;*10:353–8.
57. Hans D, Dargent-Molina P, Schott AM, *et al:* Ultrasonographic heel measurements to predict hip fracture in elderly women: The EPIDOS prospective study. *Lancet 1996;*348:511–4.
58. Riis BJ: Biochemical markers of bone turnover. II: Diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med 1993;* (suppl 5A)95:17S–21S.
59. Riis BJ, Overgaard K, Christiansen C: Biochemical markers of bone turnover to monitor the bone response to postmenopausal hormone replacement therapy. *Osteoporosis Int 1995;*5:276–80.
60. Garnero P, Weichung JS, Gineyts E, Karpf DB, Delmas PD: Comparison of new biochemical markers of bone turnover in late postmenopausal osteoporotic women in response to alendronate treatment. *J Clin Endocrinol Metabol 1994;*79:1693–700.
61. Hansen MA, Weis MAE, Bollen A-M, Maslan SL, Singer FR, Eyre DR: A specific immunoassay for monitoring human bone resorption: Quantitation of type I collagen cross-linked N-telopeptides in urine. *J Bone Miner Res 1992;*7:1251–7.
62. Hansen MA, Overgaard K, Riis BJ, Christiansen C: Role of peak bone mass and bone loss in postmenopausal osteoporosis: 12 year study. *BMJ 1991;*303:964–7.
63. Adachi J: The correlation of bone mineral density and biochemical markers to fracture risk. *Calcif Tissue Int 1996;*(suppl)59:S16–S19.
64. Garnero P, Haysherr E, Chapuy C, *et al:* Markers of bone resorption predict hip fractures in elderly women — the EPIDOS prospective study. *J Bone Miner Res 1996;*11:1531–8.
65. Nevitt MC, Cummings SR, Hudes ES: Risk factors for injurious falls: A prospective study. *J Gerontol 1991;*45:M164–70.
66. Tinetti ME, Speechley M, Ginter SF: Risk factors for falls among elderly persons living in the community. *N Engl J Med 1988;*319:1701–7.
67. Greenspan SL, Myers ER, Maitland LA, Resnick NM, Hayes WC: Fall severity and bone mineral density as risk factors for hip fracture in ambulatory elderly. *JAMA 1994;*271:128–33.
68. Lord SR, Ward JA, Williams P, Strudwick M: The effect of a 12 month exercise trial on balance, strength, and falls in older women: A randomized controlled trial. *J Am Geriatr Soc 1995;*43:1198–206.
69. Tinetti ME, Baker DI, McAvay G, *et al:* A multifactorial intervention to reduce the risk of falling among elderly people living in the community. *N Engl J Med 1994;*331:821–7.
70. Province MA, Hadley EC, Hornbrook MC: The effects of exercise of falls in elderly patients: A preplanned meta-analysis of the FICSIT trials. *JAMA 1995;*73:1341–7.
71. Mulrow CD, Gerety MB, Kanten R, *et al:* A randomized trial of physical rehabilitation of very frail nursing home residents. *JAMA 1994;*271:519–24.
72. Mahoney J, Sager M, Dunham NC, Johnson J: Risk of falls after hospital discharge. *J Am Geriatric Soc 1994;*42:269–74.
73. Gibson MJ, Andres RO, Isaacs B, *et al:* The prevention of falls in later life. A report of the Kellog International Work on the prevention of falls by the elderly. *Dan Med Bull 1987;*(suppl 4)34:1–24.
74. Nevitt MC, Cummings SR, Kidd S, Black D: Risk factors for recurrent nonsyncopal falls. A prospective study. *JAMA 1989;*261:2663–8.