The Journal of Rheumatology Volume 28, no. 2

Discrimination of changes in osteoporosis outcomes.

A Cranney, V Welch, G Wells, J Adachi, B Shea, L Simon and P Tugwell

J Rheumatol 2001;28;413-421
http://www.jrheum.org/content/28/2/413

Discrimination of Changes in Osteoporosis Outcomes

ANN CRANNEY, VIVIAN WELCH, GEORGE WELLS, JONATHON ADACHI, BEVERLY SHEA, LEE SIMON, and PETER TUGWELL

ABSTRACT. The purpose of this paper was to identify existing work related to discrimination, responsiveness, and minimal clinically important differences (MCID) for 4 key clinical outcomes in osteoporosis, to serve as a background to discussions about how to define MCID for both individuals and groups. The outcomes assessed were bone density, fractures, quality of life, and function. We conducted a systematic literature search in MEDLINE, EMBASE, and Current Contents for articles that discussed responsiveness, detectable difference, improvement criteria, and clinical importance. We used the Beaton cube to classify the studies depending on whether they compared differences between or changes within individuals or groups. Although a number of studies were identified that presented data on detectable differences beyond error and observed differences, few studies presented data on how to define clinically important differences. A key priority for future research is to define minimally clinical important differences for clinically important osteoporosis outcomes using a consensus based approach that will be accepted by the osteoporosis community at large. Furthermore, these MCID will likely be different for individual patients seen in clinical practice than for individuals in a clinical trial. (J Rheumatol 2001;28:413–21)

Key Indexing Terms:
DISCRIMINATION VALIDITY
OSTEOPOROSIS MINIMAL CLINICALLY IMPORTANT DIFFERENCE

INTRODUCTION

During the OMERACT 3 and 4 conferences, the clinimetric properties of outcome measures in osteoporosis were used to identify a potential core set of outcome measures[1-3]. One of the properties assessed was responsiveness, measured by the effect size in randomized controlled trials (RCT). The set of outcomes, in ordering of increasing effect size, were fractures (both vertebral and nonvertebral), pain, quality of life, height, bone mineral density, and biochemical markers.

We aimed to identify existing work related to discrimination, responsiveness, and minimal clinically important differences (MCID) for 4 key clinical outcomes in osteoporosis to serve as a background to discussions about how to define MCID for both individuals and groups. The outcomes assessed are bone density, fractures, quality of life, and function.

From the Department of Medicine, Ottawa Hospital, and the Clinical Epidemiology Unit, Loeb Research Institute, Ottawa, Ontario, Canada.

A. Cranney, MD, MSc, FRCPC, Department of Medicine, Ottawa Hospital, Clinical Epidemiology Unit, Loeb Research Institute; V. Welch, MSc, Clinical Epidemiology Unit, Loeb Research Institute; J.D. Adachi, MD, FRCPC, St. Joseph's Hospital, McMaster University, Hamilton, Ontario; B. Shea, BSc, RN, Department of Medicine, Ottawa Hospital, Clinical Epidemiology Unit, Loeb Research Institute; L. Simon, MD, Beth Israel Deaconess Medical Center, Boston, MA, USA; P. Tugwell, MD, MSc, FRCPC, Department of Medicine, Ottawa Hospital.

Address reprint requests to Dr. A. Cranney, Department of Medicine, Ottawa Civic Hospital, 1053 Carling Ave., Ottawa, Ontario K1Y 4E9, Canada. E-mail: acranney@civich.ottawa.on.ca

METHODS

A systematic literature search was conducted in MEDLINE, EMBASE, and Current Contents up to December 1999, using text words for osteoporosis and minimal clinically important difference, minimum observable or detectable difference, responsiveness, and improvement criteria. In addition, recent conference proceedings and journals were searched for additional relevant studies. The literature search identified 238 articles. Two independent reviewers assessed the titles and abstracts to determine eligibility. Articles were included if they assessed one of bone density, fractures, quality of life, or function in terms of one of the concepts of the cube, i.e., defining and discriminating changes and differences between groups or individuals. Articles were excluded if they were correlation studies among different outcomes. A total of 78 articles were considered potentially relevant and were retrieved for closer examination.

The literature search identified some RCT that used the words "clinical importance" or "sample size calculation" in their abstract. Because the search was not designed to identify RCT in a comprehensive fashion, the inclusion of these RCT in this report would not be a systematic review of all RCT. Therefore, 4 recent RCT in osteoporosis were selected, based on their clinical relevance, for review of their sample size calculations.

Two independent reviewers assessed the articles and determined which concepts of the responsiveness cube were assessed. The methods and results of each article were extracted in a tabular format and assessed to determine how

their results addressed the concepts in the cube. Articles were classified according to the setting (individual or group) and type (change, difference, or both).

RESULTS

Bone mineral density (Table 1)

Bone mineral density (BMD) is a surrogate marker of therapeutic effectiveness and can be assessed by several different methods, including dual x-ray absorptiometry (DXA), single x-ray absorptiometry (SXA), quantitative computerized tomography (QCT), and ultrasound[4,5]. Ultrasound has been shown to be as good as DXA bone density at predicting hip fracture in large prospective, population based studies[6,7]. DXA bone density can discriminate between patients with fractures and nonfracture controls[8-10]. Changes in DXA bone density during a clinical trial have been shown to predict the clinical outcome of vertebral and hip fracture incidence[11].

Individual differences and changes. Bone density can be used for 2 purposes in an individual: (1) to diagnose osteoporosis and (2) to monitor changes over time. In 1994, the WHO proposed a definition of osteoporosis based on BMD. They reached consensus that a T score of –2.5 standard deviations (SD) below the young adult mean or presence of fractures be used as a diagnosis of osteoporosis, using DXA or QCT bone density. The prevalence of osteoporosis using

Table 1. Summary of bone density studies. Table continues opposite.

Individual

*	Ref	Method	Spine	Femoral neck
1	(4)	Clinician judgement of importance of standard deviation - proposed change by SD of group to be considered "responder"	5%	8%
2	(20-21)	Long-term precision of spine, femoral neck and total hip DXA calculated over 7 years with 8 repeat measurements in 40 untreated women as: CV= (Standardized error of the estimate divided by the mean)*100%	3CV: 3.36%	Fem neck: 3CV: 6.63% Total hip: 3CV: 3.96%
2	(18)	Clinician judgment of importance of precision error, they recommended a change needs to be ≥ 2.8 times the CV	4.2-5.6%	5.6-8.4%
2	(32)	Coefficient of short-term reliability= variance of true value/ variance observed BMD Long-term variability- measured by SD	64 mg/cm^2/year for 1 yr (≈7.7%/year)	70 mg/cm^2/year for 3 mth (≈9%/year)
2	(23)	Short-term variability calculated by CV and by SD Defined "limits of agreement" based on Bland and Altman as 1.96*sqrt(2)*CV and 1.96*sqrt(2)*SD	0.024 g/cm^2 if age ≤70 0.030 g/cm^2 if age > 70	0.030 g/cm^2 if age < 70 0.033 g/cm^2 if age > 70
3	(28)	Compared % change from baseline in year 1 and year 2 of therapy with alendronate or raloxifene	**Year 1 / N / Year 2** <-4% 317 4 (3.4 to 4.5) -4 to-2 412 1.8(1.4 to 2.2) 0 to 2 845 0.4(0.2 to 0.6) 6 to 8 242 -1.2(-1.7 to -.7) >8 239 -2.8(-3.4 to -2.2)	
4	(5)	Correlation of change in DPX spine, femur and calcaneous SOS and BUA	No significant correlation between ultrasound changes and DPX changes	
5	(27)	Clinician judgment of importance (based on error of measurement) for response in RCT	≥ 2%	-
5	(22)	Clinician judgement of important change in RCT	>2*CV	
6	(10)	Looked at classification of women and men as osteoporotic based on DXA or QCT	AUC of ROC for classification of case similar for QCTspine and DXAspine	% meeting -2.5SD criteria **Spine# / Hip #** DXAspine 71% 48.7% DXAfem neck 33% 52.8% QCTspine 94% 36.1%
7	(14)	Effect of using -2.5 SD criteria for DXA, QCT and ultrasound on prevalence of osteoporosis	Prevalence of OP ranged from 5.6 to 65.8%	
7	(15)	Area under the curve of ROC for ability of QUS to predict femoral neck BMD <2.5 SD below young, adult mean	**Sens / Spec** Calcaneal BUA 48% 88% Calcaneal SOS 44% 80% Tibia SOS 37% 72%	
7	(17)	AUC for discrimination of hip fracture from age-matched control for different combinations of quantitative ultrasound at different sites	Best combination was calcaneous with distal radius for AUC: sensitivity and specificity = 94%	
7	(16)	Vertebral fracture vs age-matched controls Compared differences in mean scores, AUC, sensitivity and specificity of BUA, SOS, Stiffness, and 5 BMD sites	Difference between groups using Z-scores (number of SDs difference) was similar for SOS, BUA, Stiffness, BMD. AUC of ROC for Stiffness and spine BMD were significantly better than BUA, but not different from SOS	

Group *	Ref	Method	Spine	Femoral neck
8	(32)	Calculated SD of short and long-term intra and intersubject variance for rates of change in g/cm^2/year as minimum detectable beyond error in group	Sample size decreased as length of followup and frequency of observations increased	
9	(2), (3)	Effect size= mean in treatment less mean in placebo divided by standard deviation of placebo	Better responsiveness for spine BMD than other sites	
9	(25)	Different cut-offs of % change in spine BMD used to define a responder in 3 year RCT (Liberman)	Choice of cut-off affects results of RCT	
9	(31)	Treatment response index= difference in %change between groups of RCT divided by standardized precision in 4 year RCT of HRT for early, nonosteoporotic postmenopausal women	Treatment response index DXAspine 10.4 (0.5) DXAtot hip 3.9 (0.4) BUA 3.1 (1.2) SOS 0.3 (13.7) Stiffness 4.2 (0.4)	
9	(30)	Compared dose-related % change in BMD in 2 RCTs, and differences between treated and placebo groups Correlated changes in new region of interest in forearm to changes in BMD at the spine and hip	New region of interest in forearm is as responsive to change over time during therapy in an RCT as spine or hip BMD	New region of interest in forearm is as capable of detecting differences between groups
10	(4)	Clinician judgment that group difference or change should be greater than SD to be important	5%	8%
11	(8)	Used 3 arbitrary cut-offs (0%, 0-3%, >3% change in Spine, hip and fem neck BMD)	>3% change predicted fewer vertebral fractures 0 0-3% >3% spine 6% 4% 3.7% f neck 5.5% 4.2% 3.1% tot hip 6.2% 4.6% 2.7%	
14	(36)	Working party of 14 European experts in osteoporosis	15% reduction in fracture freqency suggested as MCID, depending on unwanted effects	

* The first column indicates the number in the cube in Figure 1 where this work has been classified.
CV: coefficient of variation; SOS: Speed of sound; BUA: broadband ultrasound attenuation; ROC: receiver operating curve; AUC: area under the curve; DXA: dual x-ray absorptiometry; QCT: quantitative computerized tomography; Sens: sensitivity; Spec: specificity; BMD: bone mineral density; SD: standard deviation.

Table 2. Fractures, group changes, and differences.

*	Ref	Method	Outcomes
9	(2), (3)	Effect size= mean in treatment less mean in placebo divided by standard deviation of placebo	Better responsiveness for clinical vertebral fractures in high risk subgroup with femoral neck BMD >2.5 SD
12	(42)	Tested if different criteria for diagnosis of vertebral fracture would change results of RCT of Fluoride	Found no difference between difference between groups for any of the fracture definition criteria
14	(9)	Sample size calculation to detect difference in fractures associated with 1 SD decrease in bone mass	30% reduction in incidence of vertebral and nonvertebral fractures associated with low bone mass
14	(43)	Evaluated 7 methods of classifying/diagnosing vertebral fracture and calculated sample size for RCT	40% risk reduction in vertebral fracture
14	(47)	Sample size determination for FIT trial	40% risk reduction in vertebral fractures 90% power to detect 32% reduction
14	(44)	Sample size calculations for different definitions of vertebral fracture, from 5-30% reduction in height	50% risk reduction in vertebral fracture
14	(44)	Sample size calculation for 90% power, 2 tailed test, p<.05. MORE RCT of 7705 postmenopausal women with ≥ 1 vertebral fracture	40% risk reduction in vertebral fractures
14	(46)	Sample size calculation for Risedronate RCT of 2458 postmenopausal women with ≥ 1 vertebral fracture	40% risk reduction in vertebral fractures
14	(47)	Sample size calculation for RCT of HRT, vitamin D/ calcium in postmenopausal women	21% hip fracture reduction. 20% reduction in combined fractures (vertebra, proximal femur, distal forearm, proximal humerus, pelvis)
14	(36)	Working party of 14 European experts in osteoporosis	15% reduction in fracture frequency suggested as MCID, depending on unwanted effects
14	(49)	Cost per averted hip fracture used to determine important difference needed for policy to change	10% difference in hip fracture incidence with thiazides is cost-neutral

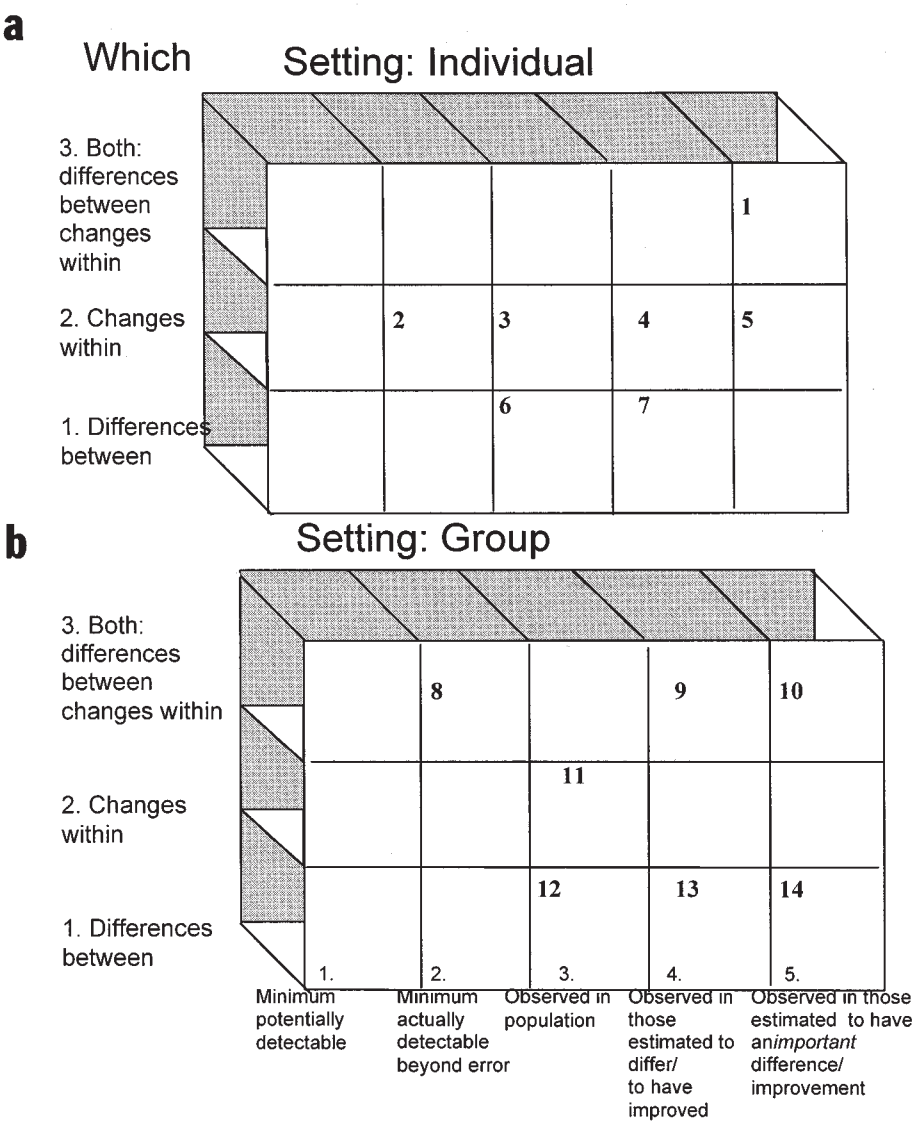

Figure 1. Classification of discrimination and changes in studies of osteoporosis. (a) *Individual*. 1. Clinician judgment that individual change should be > group SD to be considered important = 5% spine[4]. 2. A. Individual change beyond error for DXA BMD = 3 * CV = 4.5% spine[20,21]. 3. Regression to the mean is a threat to using BMD to monitor individual changes over short-term[28]. 4. No correlation in change in DXA BMD and ultrasound in RCT with demonstrated improvement[5]. 5. Clinician judgment = 2% change in spine BMD for responder to therapy[27]. 6. The prevalence of osteoporosis in the population differs according to method of BMD assessment[10,13,14]. 7. Discrimination between women with spine or hip fractures and nonfracture controls was similar for DXA-spine and QCT-spine using area under the curve (AUC) of ROC[10].
(b) *Group*. 8. The longterm intra and intersubject variability in DXA spine and femur were used to calculate sample size required to detect a difference in the rate of change per year (mg/cm^2/yr)[32]. 9. The responsiveness, using effect size of percentage change, of the core set outcomes was calculated from RCT, where differences were shown between groups[2,3]. 10. A minimum important difference of 0.5 units was proposed for the osteoporosis quality of life questionnaire[54]. 11. Hochberg, *et al* found a decreasing rate of vertebral fractures in groups of women with a change in BMD of 0%, 0–3%, or > 3%, respectively[8]. 12. Different criteria for the diagnosis of vertebral fracture did not change the observed lack of difference between 2 groups in a fluoride trial[42]. 13. The mean scores and SD of groups of osteoporotic patients and controls for QOL questionnaires have shown differences in pain, energy, and function[49–52,54,55]. 14. Vertebral fracture reductions of 30–50% in women with prevalent vertebral fractures have been used for sample size calculations[9,44-48].

these criteria varies widely depending on the method of assessment (both the site and whether BMD or bone mineral content is used)[12-14].

The discrimination of hip or vertebral fracture patients from controls has been examined for QCT, DXA, and ultrasound. Spine DXA classified only 48.7% and 71% of women with spine and hip fractures, respectively, as osteoporotic using the –2.5 SD criteria. In contrast, 94% of women with spine fractures met the –2.5 SD criteria using QCT[10]. The discrimination of osteoporotic patients from controls is poor with calcaneal broadband ultrasound attenuation, speed of sound (SOS) and tibia SOS with poor sensitivity as low as 37%[15]. However, another ultrasound variable, stiffness, was found to be as good as spine DXA[16]. Also, a combination of calcaneal and distal radius ultrasound found a sensitivity and specificity of 94% for the discrimination of hip fracture patients[17].

For minimal detectable differences beyond error, a consensus panel on bone density endorsed the use of 2.8 times the precision error (measured by coefficient of variation) as a cutoff for individual changes that are larger than measurement error[19]. More recent studies of longterm error (7 years) also suggest cutoffs of 2–3 times the CV[20-22]. Because of better precision, the minimum percentage change beyond error detectable in an individual is lower for DXA spine (ranging from 1.8 to 5.6%) than DXA femoral neck (ranging from 3.2 to 8.4%).

The minimal detectable difference for individual change has been looked at using the Bland and Altman limits of agreement to calculate the smallest detectable difference for 2 methods: coefficient of variation (as percentage change from baseline) and SD in absolute units (g/cm^2)[23]. This approach has also been used in rheumatoid arthritis to examine the smallest detectable difference in studies of imaging[24]. For bone density, Ravaud, *et al* found that the smallest detectable difference in absolute units (g/cm^2) is independent of baseline BMD, in contrast to percentage change, which depends on baseline bone density and age.

In the context of clinical trials, there is a trend to defining "responders"[25]. This is a similar approach taken by the American College of Rheumatology for definition of improvement in rheumatoid arthritis[26]. The cutoff for a "responder" can be considered an MCID for an individual. Clinician judgment of individual MCID varies. Based on the SD of group measurements, changes of 5% at the spine and 8% at the femoral neck have been suggested as MCID[4]. A lower cutoff of 2% at the spine was proposed as criteria for "responders" to alendronate by Hosking, *et al*[27].

A problem of using BMD to monitor individual responses to therapy over time is that the measurements tend to regress toward the mean. The women with the largest decreases in bone density in the first year in RCT of alendronate or raloxifene demonstrated the largest increases in the second year[28].

Group differences and changes. DXA BMD has good responsiveness in RCT, requiring sample sizes of less than 150 patients to detect a significant difference between groups, depending on the intervention[2,3].

Of the various sites, lumbar spine BMD is the most sensitive to longitudinal change[29]. A new region of interest in the forearm has also been shown to be sensitive to change in an RCT of alendronate[30]. Calcaneal broadband ultrasound attenuation and stiffness have group response indices (difference in percentage change between groups at end of study divided by the standardized precision) about 40% smaller than with DXA at the spine, suggesting that the time period for followup would need to be 2–3 times longer for QUS than DXA spine because of poorer precision[31]. There is no significant correlation between longitudinal changes in ultrasound variables and changes in femoral neck or spine DXA, suggesting that ultrasound cannot be used to predict changes in spine or femoral neck bone density[5].

The sample size required to detect rates of change in DXA BMD was calculated, using intra and intersubject SD measured from a population of postmenopausal women. The sample size decreased as the length of followup increased (from 1 to 10 yrs) and the frequency of measurements increased from one to 10[32].

Responder analysis uses cutoffs for response to therapy to compare groups in an RCT[33,34]. However, the choice of cutoff has been shown to affect the results[25]. For example, low cutoffs underestimate the difference between groups in an RCT of alendronate in postmenopausal women[25,35].

Fractures

Fractures can be detected by clinical presentation of pain and disability or radiographically. The Study of Osteoporotic Fractures population has shown that vertebral fractures (both radiographic-asymptomatic and clinically detected) are significantly associated with mortality, back pain, and disability[37-39]. Data from the alendronate and raloxifene trials support this association[40,41].

Individual differences and changes. There have been no articles on how or whether to determine individual MCID in terms of fracture incidence. However, discussion at OMERACT 5 clearly showed that clinicians consider one fracture as an important event.

Group differences and changes (Table 2). Clinical vertebral and nonvertebral fractures should be reported as the number of people with new (or worsening) fractures. Hence, to compare 2 groups, the relative risk of a new fracture is calculated in the treatment group compared to the control group. The relative risk of any type of fracture occurrence is the least responsive outcome to change, requiring sample sizes of 150 to greater than 900 patients for most interventions in the treatment of postmenopausal osteoporosis[2,3]. Although there has been some debate about how the defini-

tion of a radiographic vertebral fracture (for example by 15% or 20% reduction in height) affects the rate of false positives, a reanalysis of a fluoride RCT demonstrated no difference in the outcome of the trial with 9 different methods of classifying a vertebral fracture[42].

The sample size and power calculations of RCT are not explicitly MCID. However, they reflect the clinician judgment of an important change that is worth detecting with statistical confidence. Sample size calculations have used a 30% reduction in fractures associated with low bone mass[9] and a 40% to 50% risk reduction in vertebral fractures in women[43,44]. The sample size calculations of the 4 RCT reviewed used a 40% reduction in vertebral fracture incidence to assess the efficacy of fracture reduction of raloxifene[45], risedronate[46], alendronate[47], and hormone replacement therapy[48] for postmenopausal osteoporosis.

A working party of European experts proposed a 15% reduction in all types of fracture frequency as an MCID, depending on the unwanted effects[36].

A different approach to MCID is to calculate what difference would be needed for the intervention to be cost-effective, using the incremental cost per averted hip fracture, compared to existing therapies (relevant to the country and setting). The calculation for thiazides for the prevention of hip fracture showed that a difference of 10% in hip fracture incidence would be cost-neutral in the United Kingdom[49].

Quality of life

Osteoporosis has psychosocial consequences that include reduced self-esteem, anxiety, and chronic pain. Recent research has focused on the development and validation of disease-specific quality of life (QoL) instruments for osteoporosis. Disease-specific instruments include the OPAQ, Osteoporosis-Targeted Quality of Life Survey Instrument (OPTQoL), Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO), and OQLQ[50–55]. Furthermore, there is interest in using patient derived preferences as outcomes of clinical trials and for economic assessments[56].

Individual differences and changes. Very little research has been done on MCID of quality of life in individuals.

Group differences and changes. In most studies, the difference between fracture and nonfracture or non-osteoporotic patients has been estimated on the various domains of both disease-specific and generic QoL instruments. These are outlined in Table 3. A recent cross sectional study by Cortet, *et al* demonstrated lower quality of life in domains of physical function and energy in women with recent vertebral fractures[57].

Table 3. Quality of life group differences and changes.

*	Ref	Method	Outcomes
10	(54)	Determined by other quality of life research as 0.5 units for mini-OQLQ	0.5 units
11	(58)	Followed patients with Colle's fracture over time VAS-current health- Mean score initial and last visit	**Average VAS Current Health** Initial: 71.6 Final: 90.8 Only lost 2%
13	(52)	Mean scores of OPTQOL and SF36	3 domains of OPTQOL- physical, adaptations, fears
13	(50)	Capacity to discriminate between fractures cases and nonfracture controls using conditional logistic regression Mean scores and variance Receiver operating curve for all possible cut-off values of the questionnaire scores	Odds ratios for each domain were significant QUALEFFO score: Control: 20.3 (12.3 to 29.5); Fracture: 35.3 (24.9 to 48.4) Pain, physical function and general health perception were most discriminative
13	(53)	Adjusted mean scores on 3 different OPTQOL domains	**OP Non-OP** Physical 49 (27) 74 (24) Adaptations 44 (17) 49 (14) Fears 49 (17) 52 (10)
13	(56)	Mean scores (SD), utilities (TTO, HUI, QWB)	**TTO results Fracture Non-#** Established OP 0.84(.29) 0.43(.40) Disabling hip# 0.65(.45) 0.28(.37) Mult vertebral # 0.68(.40) 0.31(.38) HUI, hip #: 0.70 (0.41) HUI, vert #: 0.81 (0.32)
13	(51)	Mean scores of QUALEFFO for subjects with 0, 1, 2, 3, 4, or 5 vertebral fractures	**QUALEFFO mean No. fracture median, 25-75 %ile** 0 22, 15-33 2 31, 22-40 5 42, 31-58
13	(57)	Mean scores of Nottingham Health Profile in women with vertebral fracture vs non-fractured osteoporotic women and controls	**Vert # <3mth Vert # > 3 mth Control** Pain 48(24) 34(27) 17 Physical mobility 51(21) 35(22) 15 Energy 73 (36) 39(42) 28

*First column indicates the number in Figure 1 which indicates where this work has been classified in the cube.
TTO: time tradeoff; #: fracture; %ile: percentile; Vert: vertebral; HUI; health utilities index; QWB: quality of well-being; SD: standard deviation

Assessment of QoL in RCT of therapeutic interventions has been included in ongoing RCT. Very little published work has been done on MCID for QoL in osteoporosis patients. Cook and Guyatt, *et al* estimated that a 0.5 unit change in their mini OQLQ was a reasonable MCID based on earlier work that they did with disease-specific quality of life instruments[54].

Preferences or utilities have been evaluated in fracture patients using the time-tradeoff technique, health utilities index, and Quality of Well-Being scale (Table 3). Dolan, *et al* assessed QoL in 50 Colles fracture patients with the EQ-5D. They then calculated the number of quality adjusted life years (QALY) lost as a result of a Colles fracture. The QALY loss was about 2% in comparison to the 4% estimated by the National Osteoporosis Foundation guidelines[58].

The majority of the work in QoL has focused on women with established osteoporosis. There is very little research on the psychosocial effects of the awareness of osteoporosis/osteopenia[59]. Recent work has shown that these patients have lower scores on self-esteem and life satisfaction than healthy elderly women[60].

Function/performance based measures

Both hip and vertebral fractures are associated with significant functional impairment[59,61,62]. Some work has been done on the reliability, validity, and sensitivity to change for physical performance tests (e.g., strength, range of motion, and timed functional tests) as well as questionnaires (e.g., the Osteoporosis Functional Disability Questionnaire, Geriatric Depression Scale, Barthel Index)[63,64].

Psychological well being, self-esteem, and functional performance tests have been used as outcome measures in RCT of osteoporosis interventions[62,65-67]. However, little work has been done on what would be considered an MCID for such outcome measures.

DISCUSSION

Fractures

There is no specific work on deriving MCID for fractures, in particular for hip fractures. Most of the clinical trials have based sample size calculations on incident vertebral fractures. Furthermore, the difference between groups used for sample size calculations of RCT tends to be inflated in order to keep the size of clinical trials reasonable. Further research could include convening a consensus panel of experts and public members to determine an acceptable MCID for vertebral and hip fractures. This decision would include consideration of cost and adverse effects of the medications in addition to consideration of clinical sequel of fractures.

Bone density

There has been no definite consensus reached on an individual clinically important difference on bone density or ultrasound. Moreover, there has been no definite recommendation on how frequently bone density should be evaluated in an individual after treatment is commenced. Since fracture reduction is not solely dependent on bone density, the use of MCID for bone density to predict fracture reduction is of questionable value.

Quality of life

There are a number of quality of life measures (disease-specific and generic measures) that had been validated, but a paucity of information on MCID for quality of life, and in particular for pain. There is a need for research linking change in individual quality of life to both clinical and patient improvement to assist in the determination of MCID.

Functional measures

Bone independent factors play an important role in the risk of hip fractures. Therefore, future research needs to determine what is an important MCID for various performance based measures such as muscle strength and balance. Another priority for future research is to use existing functional scales to determine a clinically relevant MCID for hip fracture patients.

It was evident from discussion that occurred during the breakout session that the assessment of MCID for the various osteoporosis outcomes was not as well developed as MCID for rheumatoid arthritis or osteoarthritis outcomes. We agreed the MCID for surrogate markers such as bone density are useful if they predict a relationship to magnitude of fracture reduction.

It was also suggested that MCID derived from an RCT might not necessarily apply to the individual patient in clinical practice. For this reason, it was recommended that it was essential to develop tools for discriminating means versus tools for assessing success. We also agreed that MCID should be defined in terms of both absolute and percentage change since the use of both provides complementary information.

At OMERACT 5, we reviewed the current status of MCID for osteoporosis outcomes. We felt that it was important that more research be directed at developing MCID for outcomes of fracture, BMD, and quality of life. We recognized that consensus on MCID needs to be agreed upon by the osteoporosis research community at large.

REFERENCES

1. Cranney A, Tugwell P, Shea B, Wells G. Implications of OMERACT outcomes in arthritis and osteoporosis for Cochrane metaanalysis. J Rheumatol 1997;24:1206-7.
2. Wells G, Cranney A, Shea B, Tugwell P. Responsiveness of endpoints in osteoporosis clinical trials. J Rheumatol 1997; 24:1230-3.
3. Cranney A, Welch V, Tugwell P, et al. Responsiveness of endpoints in osteoporosis clinical trials — An update. J Rheumatol 1999;26:222-8.
4. Eastell R. Treatment of postmenopausal osteoporosis. N Engl J Med 1998;338:736-46.

5. Rosenthall L, Caminis A, Tenehouse A. Calcaneal ultrasonometry: response to treatment in comparison with dual X-ray absorptiometry measurements of the lumbar spine and femur. Calcif Tissue Int 1999;64:200-4.
6. Bauer DC, Gluer CC, Cauley JA, et al. Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women. A prospective study. Study of Osteoporotic Fractures Research Group. Arch Intern Med 1997;157:629-34.
7. Pluijm SM, Graafmans WC, Bouter LM, Lips P. Ultrasound measurements for the prediction of osteoporotic fractures in elderly people. Osteoporos Int 1999;9:550-6.
8. Hochberg MC. Larger increases in bone mineral density during alendronate therapy are associated with a lower risk of new vertebral fractures in women with postmenopausal osteoporosis. Fracture Intervention Trial Research Group. Arthritis Rheum 1999;42:1246-54.
9. Seeley DG, Browner WS, Nevitt MC, et al. Which fractures are associated with low appendicular bone mass in elderly women? Ann Intern Med 1991;115:837-42.
10. Kroger H, Lunt M, Reeve J, et al. Bone density reduction in various measurement sites in men and women with osteoporotic fractures of spine and hip: the European Quantitation of Osteoporosis Study. Calcif Tissue Int 1999;64:191-9.
11. Wasnich RD, Miller PD. Antifracture efficacy of antiresorptive agents are related to changes in bone density. J Clin Endocrinol Metab 2000;85:231-6.
12. Faulkner KG, von Stetten E, Miller PD. Discordance in patient classification using T-scores. J Clin Densitom 1999;2:343-50.
13. Shipman AJ, Guy GWG, Smith I, et al. Diagnosis of osteoporosis and osteopenia by bone mineral density and content. [abstract]. J Bone Miner Res 1998;13:534.
14. Black D. Use of T-scores to establish comparable diagnostic categories for bone densitometers. NIH Consensus Development on Osteoporosis, Prevention, Diagnosis, and Therapy, Bethesda, MD, USA, March 27–29, 2000. Bethesda: National Institutes of Health; 2000.
15. Tromp AM, Smit JH, Deeg DJ, Lips P. Quantitative ultrasound measurements of the tibia and calcaneus in comparison with DXA measurements at various skeletal sites. Osteoporos Int 1999; 9:230-5.
16. Mikhail MB, Flaster E, Aloia JF. Stiffness in discrimination of patients with vertebral fractures. Osteoporos Int 1999;9:24-8.
17. Hans D, Srivastav SK, Singal C, et al. Does combining the results from multiple bone sites measured by a new quantitative ultrasound device improve discrimination of hip fracture? J Bone Miner Res 1999;14:644-51.
18. Miller PD, Bonnick SL, Rosen CJ. Consensus of an international panel on 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.
19. Ryan PJ. Overview of role of BMD measurements in managing osteoporosis. Semin Nucl Med 1997;27:197-209.
20. Blake GM, Fogelman I. Interpretation of bone densitometry studies. Semin Nucl Med 1997;27:248-60.
21. Patel R, Blake GM, Rymer J, Fogelman I. Long-term precision of DXA scanning assessed over seven years in forty postmenopausal women. Osteoporos Int 2000;11:68-75.
22. Hillard TC, Witcroft SJ, Marsh MS, et al. Long-term effects of transdermal and oral hormone replacement therapy on postmenopausal bone loss. Osteoporos Int 1994;4:341-8.
23. Ravaud P, Reny JL, Giraudeau B, et al. Individual smallest detectable difference in bone mineral density measurements. J Bone Miner Res 1999;14:1449-56.
24. Lassere M, Boers M, van der Heijde D, et al. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.
25. Oppenheimer L, Kher U. The impact of measurement error on the comparison of two treatments using a responder analysis. Stat Med 1999;18:2177-88.
26. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
27. Hosking DJ, Chilvers CED, Christiansen C, et al. Prevention of bone loss with alendronate in postmenopausal women under 60 years of age. N Engl J Med 1998;338:485-92.
28. Cummings SR, Palermo L, Browner WS, et al. Monitoring osteoporosis therapy with bone densitometry: misleading changes and regression to the mean. JAMA 2000;283:1318-21.
29. Faulkner KG, McClung MR, Ravn P, et al. Monitoring skeletal response to therapy in early postmenopausal women: which bone to measure? J Bone Miner Res 1996;11 Suppl 1:S96.
30. Ravn P. Monitoring of alendronate treatment and prediction of effect on bone mass by biochemical markers in the early postmenopausal intervention cohort study. J Clin Endocrinol Metab 1999;84:2363-8.
31. Sahota O, San P, Cawte SA, Pearson D, Hosking DJ. A comparison of longitudinal changes in quantitative ultrasound with dual-energy X-ray absorptiometry: the four year effects of hormone replacement therapy. Osteoporos Int 2000;11:52-8.
32. Nguyen TV, Sambrook PN, Eisman JA. Sources of variability in bone mineral density measurements: implications for study design and analysis of bone loss. J Bone Miner Res 1997;12:124-35.
33. Overgaard K, Lindsay R, Christiansen C. Patient responsiveness to calcitonin salmon nasal spray: a subanalysis of a 2-year study. Clin Ther 1996;17:680-5.
34. Watts NB, Becker P. Alendronate increases spine and hip bone mineral density in women with postmenopausal osteoporosis who failed to respond to intermittent cyclical etidronate. Bone 1999;24:65-8.
35. Liberman UA, Weiss JB, Broll J, et al. Effect of oral alendronate on bone mineral density and the incidence of fractures in post-menopausal osteoporosis. The Alendronate Phase III Osteoporosis Treatment Study Group. N Engl J Med 1998;333:1437-43.
36. Kanis JA, Geusens P, Christiansen C, on behalf of the Working Party of the Foundation. Guidelines for clinical trials in osteoporosis: A position paper of the European Foundation for Osteoporosis and Bone Disease. Osteoporos Int 1991;1:182-8.
37. Kado DM, Browner WS, Palermo L, et al. Vertebral fractures and mortality in older women. Arch Intern Med 1999;159:1215-20.
38. Nevitt MC, Ettinger B, Black DM, et al. The association of radiographically detected vertebral fractures with back pain and function: a prospective study. Ann Intern Med 1998;128:793-800.
39. Ensrud KE, Thompson DE, Cauley JA, et al. Prevalent vertebral deformities predict mortality and hospitalization in older women with low bone mass. Fracture Intervention Trial Research Group. J Am Geriatr Soc 2000;48:241-9.
40. Nevitt MC, Thompson DE, Black DM, et al. Effect of alendronate on limited-activity days and bed-disability days caused by back pain in postmenopausal women with existing vertebral fractures. Fracture Intervention Trial Research Group. Arch Intern Med 2000;160:77-85.
41. Silverman SL, Minshall M, Shen W, Harper KD, Xie S. The impact of incident vertebral fracture on health related quality of life in established postmenopausal osteoporosis: results from the Multiple Outcomes of Raloxifene Evaluation Study [abstract]. J Bone Miner Res 1999;14 Suppl 1:S159.
42. Melton LJ, Egan KS, O'Fallon WM, Riggs BL. Influence of fracture criteria on the outcome of a randomized trial of therapy. Osteoporos Int 1998;8:184-91.
43. Black DM, Palermo L, Nevitt MC, et al. Defining incident vertebral

deformity: a prospective comparison of several approaches. The Study of Osteoporotic Fractures Research Group. J Bone Miner Res 1999;14:90-101.
44. Kleerekoper M, Nelson DA, Peterson EL, Tilley BC. Outcome variables in osteoporosis trials [review]. Bone 1992;13 Suppl 1:S29-S34.
45. Cummings SR, Eckert S, Krueger KA, et al. The effect of raloxifene on risk of breast cancer in postmenopausal women — Results from the MORE randomized trial. JAMA 1999; 281:2189-97.
46. Harris ST, Watts NB, Genant HK, et al. Effects of risedronate treatment on vertebral and nonvertebral fractures in women with postmenopausal osteoporosis: A randomized controlled trial. JAMA 1999;282:1344-52.
47. Black DM, Reiss TF, Nevitt MC, et al. Design of the Fracture Intervention Trial. Osteoporos Int 1993;3 Suppl 3:S29-S39.
48. Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials 1998;19:61-109.
49. Torgerson DJ, Ryan M, Ratcliffe J. Economics in sample size determination for clinical trials. Q J Med 1995;88:517-21.
50. Lips P, Cooper C, Agnusdei D, et al. Quality of life in patients with vertebral fractures: validation of Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO). Working Party for Quality of Life of the European Foundation of Osteoporosis. Osteoporos Int 1999;10:150-60.
51. The impact on health related quality of life (HRQOL) in postmenopausal women with low BMD and prevalent fractures. American Society for Bone and Mineral Research. Bone 1993;23:(55) S398.
52. Chandler JM, Martin AR, Girman C, et al. Reliability of an Osteoporosis-Targeted Quality of Life Survey Instrument for use in the community: OPTQoL. Osteoporos Int 1998;8:127-35.
53. Lydick E, Zimmerman SI, Yawn B, et al. Development and validation of a discriminative quality of life questionnaire for osteoporosis (The OPRQoL). J Bone Miner Res 1997;12:456-63.
54. Cook DJ, Guyatt GH, Adachi JD, et al. Development and validation of the mini-osteoporosis quality of life questionnaire (OQLQ) in osteoporotic women with back pain due to vertebral fractures. Osteoporosis quality of life study group. Osteoporos Int 1999;10:207-13.
55. Cantarelli FB, Szejnfeld VL, Oliveira LM, Ciconelli RM, Ferraz MB. Quality of life in patients with osteoporosis fractures: cultural adaptation, reliability and validity of the Osteoporosis Assessment Questionnaire. Clin Exp Rheumatol 1999;17:547-51.
56. Gabriel SE, Kneeland TS, Melton LJ, et al. Health-related quality of life in economic evaluations for osteoporosis: whose values should we use? Med Decis Making 1999;19:141-8.
57. Cortet B, Houvenagel E, Puisieux F, et al. Spinal curvatures and quality of life in women with vertebral fractures secondary to osteoporosis. Spine 1999;24:1921-5.
58. Dolan P, Torgerson D, Kakarlapudi TK. Health-related quality of life of Colles' fracture patients. Osteoporos Int 1999;9:196-9.
59. Gold DT. The clinical impact of vertebral fractures: quality of life in women with osteoporosis. Bone 1996;18 Suppl:185S-9S.
60. Purser JL, Pieper CF, Branch LG, et al. Spinal deformity and mobility self-confidence among women with osteoporosis and vertebral fractures. Aging-Clin 1999;11:235-45.
61. Hall SE, Criddle RA, Comito TL, Prince RL. A case-control study of quality of life and functional impairment in women with long-standing vertebral osteoporotic fracture. Osteoporos Int 1999;9:508-15.
62. Gold DT, Stegmaier K, Bales CW, et al. Psychosocial functioning and osteoporosis in late life: Results of a multidisciplinary intervention. J Women's Health 1993;2:149-55.
63. Purser JL, Pieper CF, Duncan PW, et al. Reliability of physical performance tests in four different randomized clinical trials. Arch Phys Med Rehabil 1999;80:557-61.
64. Helmes E, Hodsman A, Lazowski D, et al. A questionnaire to evaluate disability in osteoporotic patients with vertebral compression fractures. Series A. Biological Sciences & Medical Sciences. J Gerontol 1995;50:M91-M98.
65. Bravo G, Gauthier P, Roy PM, et al. Impact of a 12-month exercise program on the physical and psychological health of osteopenic women. J Am Geriatr Soc 1996;44:756-62.
66. Bravo G, Gauthier P, Roy PM, Payette H, Gaulin P. A weight-bearing, water-based exercise program for osteopenic women: its impact on bone, functional fitness, and well-being. Arch Phys Med Rehabil 1997;78:1375-80.
67. Kerschan K, Alacamlioglu Y, Kollmitzer J, et al. Functional impact of unvarying exercise program in women after menopause. Am J