Intervention Thresholds and the Diagnosis of Osteoporosis

John A Kanis,1 Eugene V McCloskey,1 Nicholas C Harvey,2 Helena Johansson,1 and William D Leslie3

1Centre for Metabolic Bone Diseases, University of Sheffield Medical School, Sheffield, United Kingdom
2MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom
3University of Manitoba, Winnipeg, Canada

ABSTRACT

A position paper of the National Bone Health Alliance recently recommended that diagnostic criteria for osteoporosis be redefined. We review the merits and demerits of this proposal and argue that the operational bone mineral density (BMD)-based definition be retained while clarity is brought to bear on the distinction between diagnostic and intervention thresholds. © 2015 American Society for Bone and Mineral Research.

KEY WORDS: OSTEOPOROSIS; DEFINITION; INTERVENTION THRESHOLDS; FRAX; FRACTURE RISK; DIAGNOSIS

Introduction

Conceptual and operational descriptions of osteoporosis have survived for 20 years. Recently, a position paper of the National Bone Health Alliance recommended that diagnostic criteria for osteoporosis be redefined1 to more closely reflect treatment thresholds espoused within guidance from the National Osteoporosis Foundation.2 This proposal has implications for the global management of osteoporosis. We briefly review the merits and demerits of redefining osteoporosis.

Conceptual Definition of Osteoporosis

A starting point for discussion is the current description of osteoporosis and how this has evolved. Medical dictionaries and reference books from 1972 to 1995 offered widely inconsistent definitions.3 Several attempts at international consensus failed4,5 but was eventually achieved at an international consensus conference held in Hong Kong in March 1993, sponsored by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, the European Foundation for Osteoporosis and Bone Disease (now the International Osteoporosis Foundation), and the American National Osteoporosis Foundation.6 This landmark Consensus Development Conference described osteoporosis as: “A systemic skeletal disease characterized by low bone mass and micro-architectural deterioration of bone tissue with a consequent increase in bone fragility and susceptibility to fracture.” Despite attempts to refine this landmark definition7 it has survived intact to the present day. This description captured the notion that low bone mass was an important component of the risk of fracture but that other abnormalities occur in the skeleton that contribute to skeletal fragility. Thus, ideally, clinical assessment of the skeleton should capture all these aspects of fracture risk. At that time, however, the assessment of bone mineral density (BMD) was the only aspect that could be readily measured in clinical practice, and it formed the cornerstone for the operational description of osteoporosis.

Operational Description of Osteoporosis

The World Health Organization (WHO) diagnostic criterion for osteoporosis is based on the measurement of BMD. In 1994, the operational definition of osteoporosis in postmenopausal women was given as a BMD that was greater than 2.5 standard deviations below the mean value of young healthy women, ie, a T-score < –2.5 SD8,9 using the T-score notation originally devised by Tom Kelly. The criterion was subsequently updated to remove the ambiguity of using multiple sites for BMD measurement, different reference values for calculating T-scores, and to provide definitions for men aged 50 years or older.10 Thus, osteoporosis is described as a value for BMD at the femoral neck of 2.5 SD or more below the young female adult mean (T-score less than or equal to –2.5 SD). The recommended reference range is the Third National Health and Nutrition Examination Survey (NHANES III) database for femoral neck measurements in white women aged 20 to 29 years11 as recommended by the International Osteoporosis Foundation, the National Osteoporosis Foundation, and the International Society of Clinical Densitometry.12–14 The referents in women apply equally to men aged 50 years or older because the gradient of risk and the age-adjusted risk of hip fracture for any given BMD at the femoral neck is similar in both sexes.15–18
**Intervention Thresholds**

The WHO diagnostic criteria for osteoporosis were intended primarily for descriptive epidemiology\(^8,9\) but were soon adopted as inclusion criteria for drug trials and subsequently proposed as intervention thresholds and a basis for health technology assessments. Given that a fragility fracture is the hallmark of the disease, intervention thresholds have also been based on a history of a prior fragility fracture (spine or hip fracture in North America) variously with or without a BMD threshold. More recently, the advent of risk assessment algorithms indicates that prevention of fractures is better targeted on the basis of fracture probability using multiple risk factors rather than BMD alone.\(^19,20\) FRAX is the most widely used tool to generate fracture probabilities\(^21\) and is incorporated in a large number of assessment guidelines\(^22\) and recommended by the Committee for Medicinal Products for Human Use (CHMP).\(^23\) Intuitively, it makes some sense that operational definitions of osteoporosis should match intervention thresholds. Intuition is not always based on logic (Sinak J, personal communication, 2015), and this proves to be the case in osteoporosis.

The use of BMD as an intervention threshold

In many countries, intervention thresholds have historically been based on the \(T\)-score for BMD and/or the presence of a prior fragility fracture. These strategies seem intuitively sound because they cover the operational definition of disease and/or its clinical expression. For example, the NOF in the United States recommends BMD assessment in women and treatment is advised in women with a \(T\)-score of \(< -2.5\) SD. Treatment is also recommended in women with a prior spine or hip fracture.\(^2\)

There are three considerations that make the use of a \(T\)-score problematic as an intervention threshold. The problems are illustrated using a \(T\)-score threshold of \(-2.5\) SD by way of example. First, although reduced bone mass is an important and easily quantifiable measurement, most fragility fractures occur in individuals with a BMD \(T\)-score above the operational threshold for osteoporosis.\(^8,24\) In the case of hip fractures, approximately 50% of cases will have osteoporosis so defined.\(^25,26\) The problem arises because BMD captures the likelihood of fracture incompletely. There is an appropriate analogy with several other multifactorial diseases, such as hypertension and stroke. Blood pressure is continuously distributed in the population (as is BMD), and hypertension is an important cause of stroke (high specificity). But a majority of individuals with stroke are normotensive (low sensitivity).\(^27\)

A second factor relates to the fact that any given \(T\)-score threshold has a different significance at different ages. At the age of 65 years, a \(T\)-score of \(-2.5\) SD confers a modest increase in the probability of fracture compared with women with no clinical risk factors and in whom BMD is not measured. With advancing age, the difference in the probability of fracture between the general population and those with a \(T\)-score of \(-2.5\) SD diminishes, and, indeed, from the age of 78 years onward in the United States, the fracture probability becomes progressively lower than that of the age- and sex-matched general population (Fig. 1). In other words, a \(T\)-score of \(-2.5\) SD becomes a protective factor from the age of 78 years and the \(T\)-score required to achieve the “fracture threshold” becomes progressively lower with age.\(^29\)

Third, it is well established that fracture rates vary widely from country to country—much more so than can be explained by variations in BMD.\(^26,30\) Thus, for any given fracture risk, the \(T\)-score will vary from country to country. For example, when an intervention threshold is set at a 10-year probability of a major fracture of 20% (as used in Canada and the United States), the femoral neck \(T\)-score ranges widely. Fig. 2 illustrates this for women aged 65 years, prior fracture, and a body mass index (BMI) of 24 kg/m\(^2\) using the FRAX tool. For this clinical scenario, a 10-year fracture probability of 20% is equivalent to a \(T\)-score of

---

**Fig. 1.** Ten-year probabilities (%) of a major osteoporotic fracture for white women from the United States according to a \(T\)-score of \(-2.5\) SD (open circle), prior fracture (solid triangle), or the combination (solid square) (BMI is set to 24 kg/m\(^2\)) [http://www.shef.ac.uk/FRAX]. Note the decreased probability after the age of 85 years attributable to the competing effect of mortality.
–4.6 SD in Venezuela, whereas the equivalent $T$-score in women from Iceland is –2.0 SD.

The use of prior fracture as an intervention threshold

In contrast to BMD, a prior fracture confers an increased risk over all ages (Fig. 1). However, the utility of a prior fracture in combination with BMD carries the same problems as the use of a BMD threshold alone (Fig. 1). When a prior fracture is used as a risk variable in the absence of BMD, the predictive value of a prior fracture decreases with age because of the higher mortality risk in men and women with a prior fracture. For example, the 10-year probability of a major fracture in women at the age of 50 years from the United Kingdom (BMI = 24 kg/m$^2$ and no clinical risk factors) is 3.4%. In the presence of a prior fracture, this rises to 7.3%, giving a probability ratio of 2.11 between the 10-year probability of major fracture with and without prior fracture. This probability ratio falls to 1.8 at the age of 70 years and to 1.5 at the age of 90 years. Notwithstanding, a prior fracture carries an increased risk of fracture at all ages. As would be expected, the fracture probability in the presence of a prior fracture varies on a worldwide basis. The variation is illustrated in Fig. 3 for the same countries shown in Fig. 2.
The use of a fixed FRAX probability as an intervention threshold

Several guidelines that use FRAX have recommended that a fixed probability threshold be used as an intervention threshold. Examples include a 20% 10-year probability of a major fracture in Canada and the United States, and a 15% probability in Japan and Sweden. It should be noted that these fixed thresholds are not necessarily used in isolation. For example, in the United States, FRAX is reserved for patients with low bone mass (osteopenia) and treatment is recommended, based on cost-effectiveness, when the 10-year probability of a major fracture is 20% or above or where the probability of a hip fracture exceeds 3%. These thresholds are, however, relevant only to the United States and are inappropriate for use elsewhere because of differences in the importance of osteoporosis, the heath care budget allocated, current practice guidelines, reimbursement, and health economic considerations. Although this states the obvious, several guidelines (eg, Austria, Greece, Hungary, Hong Kong, Malaysia, and Taiwan) have chosen a 20% fracture probability without any justification other than its use in North America. Within Europe, the proportion of the population aged 50 years or older with a FRAX probability of a major fracture >20% varies from 2% (Romania) to 19% (Denmark). Indeed, the proportion of the population aged 50 years or older with a FRAX probability of a major fracture >10% varies from 10% (Romania) to 48% (Denmark). This variation in risk needs to be balanced against the health care spend (4.5% of gross domestic product in Romania and 10.8% in Denmark, equivalent to €309 and €4759 per capita per year, respectively). Given also that the cost of intervention varies little between countries, no one fixed FRAX threshold is applicable to all countries, and if fixed thresholds are to be used, they need to be country specific.

A further problem with the use of fixed thresholds alone arises in the proportion of the population eligible for treatment. The impact of using different intervention threshold is shown in Fig. 4 for postmenopausal women in Japan. At high thresholds, eg, >20% fracture probability, 20.5% of postmenopausal women would be eligible for treatment. A problem that arises is that very few women under the age of 60 years would ever attain this threshold (less than 1%). On the other hand, if a less stringent threshold were chosen, say 10%, then approximately 5% of women at the age of 50 years would exceed this threshold, and a majority of women over the age of 65 years would be eligible and the treatment threshold would be exceeded in 50% of all postmenopausal women. Both scenarios are counterintuitive to clinical practice.

The use of age-dependent FRAX probability as an intervention threshold

Before the advent of FRAX, many guidelines in Europe, North America, and elsewhere recommended treatment in the absence of information on BMD in women with a previous fragility fracture (a prior vertebral or hip fracture in North America). For this reason, the intervention threshold in women without a prior fracture can be set at the age-specific fracture probability equivalent to women with a prior fragility fracture and, therefore, rises with age; for example, the threshold rises from a 10-year probability of 8% at 50 years to 33% at 80 years in the United Kingdom. In other words, the intervention threshold is set at the “fracture threshold.” This approach to intervention thresholds, first adopted by the National Osteoporosis Guideline Group (NOGG) for the United Kingdom, is now used in France, Switzerland, Romania, and Finland. Incidentally, the same intervention threshold is applied to men because the effectiveness and cost-effectiveness of intervention in men are broadly similar to that in women for equivalent risk. The approach used has been well validated and the intervention strategy shown to be cost-effective. Using the same criteria, the intervention threshold will vary from country to country because the population risks (of fracture and

---

**Fig. 4.** The impact of a fixed treatment threshold in postmenopausal women in Japan according to threshold values for the probability of a major osteoporotic fracture. The left panel shows the proportion of the postmenopausal population exceeding the threshold shown at each age. The right panel shows the proportion of the total postmenopausal population that exceeds a given threshold (with kind permission from Springer Science and Business Media).
death) vary. The fracture probability in women with a prior fracture in the five major EU are highest in the United Kingdom and lowest in Spain. The difference between countries is most evident at younger ages and becomes progressively less with advancing age. In Europe, the proportion of men and women above this threshold varies little from 11% to 13%.

The merits of this approach are that it embraces secondary prevention of fracture and that it can be readily applied to all countries regardless of the availability of BMD. In countries with a more conservative approach, the threshold can be uplifted, say by 10% to 20%. Conversely, an intervention threshold can be downward adjusted where a more liberal approach is desired. A difficulty that arises with age-specific thresholds is that the probability at which treatment is recommended is country-specific, though varies little in the western world.

Where does this leave us?

The considerations above indicate that neither a fixed BMD nor a fixed FRAX threshold is universally appropriate as an intervention threshold. The use of an age-specific threshold allied to a “fracture threshold” appears to be the least problematic approach. The setting of these thresholds does not depend on the ubiquitous availability of BMD and can be readily applied to all countries where a FRAX model is available.

It is axiomatic that different intervention thresholds will identify different patients at different risk. Some examples are given in Table 1 based on the National Health and Nutrition Examination Survey (NHANES) 2005–2008. It is of interest that the guidelines of the US National Osteoporosis Foundation select the greatest number of patients eligible for treatment and that the application of an age-specific threshold with or without prior fracture identifies a smaller proportion of the population but at higher risk.

**Redefining Osteoporosis**

The National Bone Health Alliance appointed a working group charged to expand the criteria by which osteoporosis can be diagnosed. The group recommended that postmenopausal women and men aged 50 years should be diagnosed with osteoporosis if they have sustained a hip fracture (irrespective of BMD); or a BMD T-score of ≤−2.5 SD at the spine or hip; or osteopenia and a prior clinical vertebral, proximal humeral, pelvic, or, in some cases, distal forearm or morphometric vertebral fracture; or a 10-year FRAX probability of hip fracture ≥3% or the 10-year probability of major osteoporotic fracture ≥20% in individuals with osteopenia. The criteria follow the NOF guidelines with some minor modifications.

The aims of the proposed revision are laudable, which is to target treatments to those at high risk of fracture and prevent more fractures. But the devil is in the detail. First, the inclusion of fracture outcomes in diagnostic criteria for multifactorial diseases is anachronistic in much the same way as would be the inclusion of stroke in the diagnosis of hypertension. Second, the use of BMD as the principal gateway to assessment raises the problems articulated above that a T-score of −2.5 SD is a protective factor rather than a risk factor in the elderly. Third, the inclusion of prior fracture together with BMD does not necessarily improve the sensitivity for fracture (Table 1). Fourth, the use of fixed FRAX thresholds based on health economics are only relevant for one intervention and only at a cost prevailing at the time of health economic assessment. Fifth, the diagnostic criteria are at best cumbersome and likely to deter the management of osteoporosis in primary care. The most important difficulty is that acceptance of these criteria on a worldwide basis is not possible because of the heterogeneous significance of T-score thresholds and widely differing fracture probabilities.

These considerations argue that diagnostic criteria should still be based on BMD, given the conceptual description of osteoporosis. The strength of the diagnostic category as a reference standard has been the fashioning of a common approach to description of the disease. The imperfect performance of BMD in risk assessment has been known for years in much the same way as the imperfect capture of cardiovascular events by hypercholesterolemia. The solution to the problem is to make a clear distinction between diagnostic thresholds and intervention thresholds as has been successfully managed in cardiovascular disease. There is a particular problem in the

<table>
<thead>
<tr>
<th>Table 1. Number Selected as Being Above the Intervention Threshold and the Proportion Who Will Fracture Over 10 Years (Mean 10-Year Fracture Probability of Major Osteoporotic Fracture [MOP] and Hip Fracture) in Men and Women Aged 50 Years or Older From the NHANES Cohort According to Different Intervention Thresholds (Reanalysis of Data From Dawson-Hughes et al., 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selection</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>FRAX fixed thresholds</td>
</tr>
<tr>
<td>FRAX at fracture threshold</td>
</tr>
<tr>
<td>FRAX fixed thresholds + prior fracture</td>
</tr>
<tr>
<td>FRAX at fracture threshold + prior fracture</td>
</tr>
<tr>
<td>NOF</td>
</tr>
<tr>
<td>Prior fracture</td>
</tr>
<tr>
<td>T-score ≤ −2.5°</td>
</tr>
<tr>
<td>Prior fracture and T-score ≤ −2.5°</td>
</tr>
</tbody>
</table>

*aFRAX with 20% and 3% probability thresholds for major fracture and hip fracture, respectively.

*bFRAX with age-specific thresholds plus prior fracture.

*cPrior hip or spine fracture.

*dNational Osteoporosis Foundation Guidelines.

*eT-score at proximal femur or lumbar spine.
United States in that reimbursement is conditional on fulfilling a diagnostic code, whereas in most countries reimbursement criteria very commonly differ from diagnostic criteria. The solution is not to disadvantage the rest of the world but to educate health care payers in the United States.

The measurement of risk most suited for assessment is the absolute risk, expressed as the probability of fracture within a given time frame, eg, the 10-year fracture probability. Thus, intervention thresholds will be based on fracture risk and differ, therefore, from diagnostic thresholds. Intervention thresholds have been established based on BMD, prior fracture, other clinical risk factors, fracture risk, and fracture probability. None is ideal from all perspectives and there is no “gold standard” established by international consensus. Of those developed thus far, the targeting of patients with a prior fracture combined with age-dependent FRAX thresholds has the greatest balance between sensitivity and clinical justification, a conclusion echoed recently for cardiovascular disease.\(^{54}\)

It is relevant to question the need for diagnostic criteria when the field is moving toward risk-based assessment and intervention. These developments will certainly decrease the clinical utility of the T-score, but they will, however, take time to implement into routine clinical practice. Notwithstanding, diagnostic criteria remain of value in quantifying the burden of disease and the development of strategies to combat osteoporosis in the foreseeable future.

Disclosures

As director of the World Health Organization (WHO) Collaborating Centre for Metabolic Bone Diseases, JAK led the team that developed FRAX; he has no financial interest in FRAX. Dr Kanis has served as a consultant to Amgen, USA, Switzerland, and Belgium; Celtrix, USA; D3A, France; the European Federation of Pharmaceutical Industries and Associations; GSK, UK and USA; Hologic, Belgium and USA; Lilly, USA, Canada, Japan, Australia, and UK; Merck Research Labs, USA; Merlin Ventures, UK; MRL, China; Novartis, Switzerland and USA; ProStrakan, UK; Roche, Germany, Australia, Switzerland, and USA; Rotta Research, Italy. He has provided expert testimony for the High Court (UK). His institution has received grants from the Medical Research Council (UK), the Arthritis and Rheumatism Council (UK), the European Union, and the European Federation of Pharmaceutical Industries and Associations. He has received speaker’s fees from Amgen, Novartis, Takeda, and Servier. He has also worked with the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis, the Government of Manitoba, the Group for the Respect of Ethics and Excellence in Science, INSERM (France), the Medical Research Council (UK), the Ministry of Public Health of China, the Ministry of Health of Australia, the National Institute for Health and Clinical Excellence (UK), the National Osteoporosis Guideline Group (UK), the National Osteoporosis Society (UK), the International Osteoporosis Foundation, the Japanese Osteoporosis Society, Osteoporosis 2000 (UK), Osteoporosis Australia, the Swiss Osteoporosis Society, and the WHO. EM has received consultancy and/or speaker fees and/or unrestricted research grants from Active-Signal, Amgen, Arthritis Research UK, AstraZeneca, Bayer, Consilient, GE Lunar, GSK, Hologic, IDS, Innovus i3, Internis, IOF, Lilly, Merck, MRC, Novartis, Pfizer, Roche, Servier, Synexus, Tethys, UCB, Uniliever, and Warner Chilcott. NH has received consultancy, lecture fees, unrestricted research grants, and/or honoraria from Alliance for Better Bone Health, Amgen, MSD, Eli Lilly, Servier, Shire, Consilient Healthcare, and Internis Pharma. HJ has no disclosures. WL has served on the speaker bureaus of Amgen, Eli Lilly, and Novartis, and has received research grants from Amgen and Genzyme.

Acknowledgments

Authors’ roles: Manuscript concept: JK, EM, NH, and HJ. Revising manuscript content: JK, EM, NH, HJ, and WL. Statistical analysis: HJ. Approved final version of manuscript: JK, EM, NH, HJ, and WL.

References


