Clin Rheumatol (2000) 19:174–183

Clinical Rheumatology

Review Article

The Fallacy of BMD: A Critical Review of the Diagnostic Use of Dual X-ray Absorptiometry

S. Pors Nielsen

Department of Clinical Physiology and Nuclear Medicine, Hillerød Hospital, Hillerød, Denmark

Abstract: The diagnostic use of BMD should be cautious as BMD is not an ideal measure of true bone density; it is not an ideal measure of bone strength; it does not predict fractures well; and it has inherent problems of accuracy and linearity. The limitations of BMD, based on the physical deficiencies of DXA, are further obscured by the introduction of T-scores.

It is suggested that BMD and BMC, when used diagnostically and for fracture risk classification, be used after correction for body size and/or bone size, age and sex, and that measured values be evaluated in the light of established mean fracture incidence data. BMD is not a parameter of sufficient validity to be the sole indicator of present and future fracture risk. A low BMD should be regarded one of several fracture risk factors.

It seems that there is a need to redefine the T-score based definition of osteoporosis.

Keywords: Bone mineral density; Dual X-ray absorptiometry; Fracture risk; Osteoporosis

Introduction

Bone densitometry with dual-X-ray absorptiometry (DXA) is eminently suited for the measurement of changes over time in an individual because of its low precision error, but there are problems with its diagnostic use. This review does not deal with spontaneous changes in bone mineral mass over time, or with the effects of treatment, but with problems related to the diagnosis of osteoporosis and fracture risk evaluation with the aid of DXA.

Three decades ago bone mineral content (BMC) in terms of grams was introduced in two-dimensional bone densitometry, but soon the term bone mineral density (BMD) followed, being BMC divided by the projected bone area, having the dimension of mass per area (g/cm^2). DXA has been used extensively during the last two decades, and has been called the best method for estimating fracture risk. Demonstration of low BMD by DXA might have major consequences for patients and society, i.e. medication over many years, unavoidable side effects, repeated physician visits and repeated measurements to study the effects of treatment. Therefore, the reliability of bone densitometry results is crucial. The estimation from one DXA measurement of future fracture risk and the conclusion regarding prevention of fractures in one patient is based on a knowledge of statistics. The introduction of T-scores (definition of osteoporosis is a T-score below –2.5) has confounded the situation. What the patient needs to know is what is his or her risk of future fracture and the development of osteoporosis?, and not how he or she compares with a reference population, of which we know little [1].

Consensus reports usually appear because there is disagreement. Many consensus statements on osteoporosis have been published during the last decade, one of them concluding the following: 1) bone-mass measurements predict a patient's future fracture risk; 2) osteoporosis can be diagnosed on the basis of bone-mass measurements even in the absence of prevalent fractures; 3) bone-mass measurements provide information that can affect patient management; 4) the choice of the appropriate measurement site(s) for the assessment

Correspondence and offprint requests to: Stig Pors Nielsen, MD, dr. med, Department of Clinical Physiology and Nuclear Medicine, Hillerød Hospital, 3400 Hillerød, Denmark.

of bone mass or fracture risk may vary depending on the specific circumstances of the patient; 5) the choice of the appropriate technique for bone-mass measurements in any given clinical circumstance should be based on an understanding of the strength and limitations of the different techniques; 6) bone-mass data should be accompanied by a clinical interpretation [2].

It is easy to agree on the last point, but there have been many comments on the others since they were published. It seems that there is a need for a critical re-examination of the general value of the diagnostic use of DXA, which in the minds of some opinion leaders is a gold standard. As the risk of fracture is multifactorial it is logical that a low BMD on DXA is one of several risk factors for later fractures. However, reimbursement policy in many countries focuses more on a low BMD than on other risk factors.

The problems associated with the diagnostic use of BMD and DXA are manifold. Some of them will be discussed here:

1. BMD might not be a good measure of three-dimensional bone density.
2. BMD might not be a good measure of bone strength.
3. The discriminatory ability of BMD for fractures is not good, i.e. low-energy fractures can occur at a normal BMD, and a patient with a subnomal BMD may never experience a fracture.
4. Falsely high spine BMD values are often encountered.
5. DXA devices have inherent problems of accuracy and linearity which are difficult to correct for, making inter-device comparisons and cross-calibration difficult or impossible.
6. The introduction of T-scores and Z-scores is threatening the credibility of bone densitometry, being the source of a diversity of potentially erroneous conclusions.

BMD Might not be a Good Measure of Bone Density

BMD and Bone Depth

Many physicians referring patients for DXA seem to believe that BMD is synonymous with true three-dimensional bone density, as obtained with CT (g/cm^3). It is not. BMD was introduced as a normalisation procedure for BMC which is proportional to bone size and body size. When BMC is converted to BMD by division with the projected bone area, a surrogate bone density is obtained. This parameter is also proportional with the bone size and body size, although less so than BMC [3]. In the following BMD will be the term for this two-dimensional variable only, whereas the three-dimensional bone density of CT will be termed 3D-BMD. The BMD of any bone is the mean bone mass of a number of cylinders vertical to the two-dimensional plane having an endplate of 1 cm^2. If the bone is large the depth (length of the radiation beam) is large and the BMD therefore higher than if the bone is small (the BMD of an elephant vertebra is not the same as the BMD of a mouse vertebra at the same 3D-BMD). This fact is well known [4,5], but largely neglected. This neglect has the obvious consequence that osteoporosis is overdiagnosed in persons of petite body stature, simply because the means of reference populations are calculated from the values of large and small people. It is usually difficult to correct for differences in bone depth, but a meaningful normalisation can be obtained by normalising with regard to body surface area (BSA), which can be calculated reasonably accurately from body height and weight using the DuBois equation, as bone size is proportional to body size:

$$BSA = 71.84 \times W^{0.425} \times H^{0.725}$$

where BSA = body surface area in cm^2; W = body weight in kg; H = body height in cm. Normalisation was performed by us in a large population of normal early postmenopausal women, lumbar spine BMC and hip BMC being divided by normalised BSA, and BMD of the same sites being divided by the square root of normalised BSA, with the result that the Z-scores became independent of body size and the overdiagnosis of petite women abolished [3]. Among 1625 early postmenopausal women divided into BSA quartiles, the number having a lumbar spine BMD Z-score (Hologic normal reference population) less than –1.0 was 102 for the lowest BSA quartile, in contrast to 25 women in the highest BSA quartile. For femoral neck BMD the corresponding numbers were 112 and 31, respectively. When BMC/BSA was used, the number of women in the lowest and highest BSA quartiles were almost identical for both measuring sites (lumbar spine 62 and 62; femoral neck 57 and 59, respectively). Thus, it can be seen that the consequences of not using body size correction can be pronounced and deleterious.

In some countries a low Z-score ($\leqslant$–1.0) is the basis for reimbursement. Figure 1 exemplifies the significance of BSA correction of Z-scores for a large woman and a small one in case of lumbar spine BMD, using BSA quartiles and the data from [3]. It can be seen that if the BMD of the former (belonging to the highest quartile) were 0.900 g/cm^2, she would have a Z-score of –1.31 and be entitled to reimbursement. However, if, having the same BMD, she belonged to the lowest BSA quartile (petite body stature) she would have a Z-score of –0.64 and not be entitled to reimbursement. Similar examples can be made for T-score if a valid normal reference population for young people is available. It is obvious from the example above that the use of T-scores and Z-scores based on normal means which do not take body size into account can be misleading.

Mazess et al. [6] found no reasons to abolish the use of BMD, because BMD has good diagnostic sensitivity and low precision error. Others have argued that BSA should not be corrected as small women break their hips more

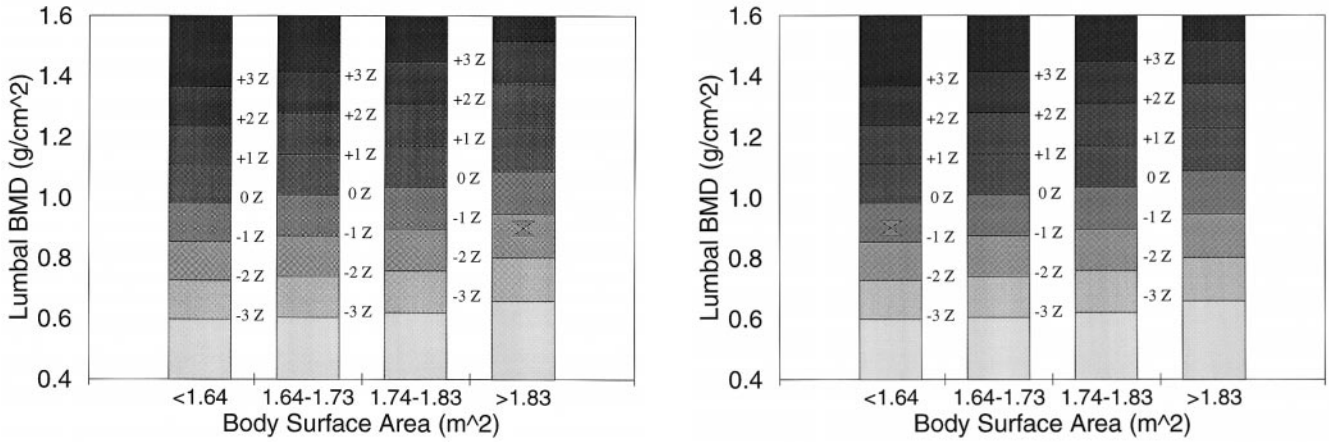

Fig. 1. The influence of body size on Z-score of measured lumbar spine BMD, using body size quartiles calculated according to the DuBois formula. A total of 1625 normal early postmenopausal women. Data from [3]. Left: A 52-year-old woman with a body weight of 90 kg, height 185 cm and a lumbar spine BMD of 0.900 g/cm^2. Right: A 52-year-old woman with a body weight of 50 kg, height 155 cm and lumbar spine BMD of 0.900 g/cm^2. *Note*: The Z-score of the former was –1.31 and of the latter –0.64. A BMD of 0.900 g/cm^2 is low for a big early postmenopausal woman, but nearly normal for a small woman of the same age and menopausal status.

frequently than tall women. This might well be true in some populations, but fractures may be caused by factors other than low BMD, such as propensity to falls and lack of subcutaneous fat. It seems that the literature does not confirm the impression that small people break their bones more often than do tall people. In some countries tall women seem to be more at risk for hip fractures than small ones [7,8]. Hip axis length and angulation might also play a part.

Rapid Bone Losers

Over the years it has been claimed that one could identify rapid bone losers by sequential DXA measurements, or by biochemical markers taken once. It has never been shown that there is a bimodal distribution of bone loss in normals, but naturally some lose bone faster than others at certain times. Pouillès et al. in 1996, studied the problem of axial bone loss in postmenopausal women longitudinally and found 1) great variability in vertebral and femoral bone loss within the first postmenopausal years; 2) that bone losses from the spine and hip are uncorrelated; and 3) that only a small minority sustained a fast rate of bone loss over those years [9]. We have also demonstrated a lack of correlation between bone loss at different measurement sites [10]. The rates of loss observed with DXA are a mixture of cortical and trabecular bone loss. True rates of bone loss differ for cortical and trabecular bone, as shown by Boonen et al. [11]. They found in elderly women that the cortical bone loss from the ultradistal radius averaged 0.41% per year, versus a trabecular loss of 0.65% per year [11]. Given the costs, variability and errors of bone densitometry, it seems wise to agree that individual loss rates should be evaluated with the utmost caution. Individual lifestyles, including exercise and eating habits, often change, and intercurrent diseases influencing bone loss tend to become more frequent with age. For these reasons alone it seems hazardous to predict long-term individual bone loss from two BMD measurements, or from a combination of one BMD measurement and a set of biochemical markers. To complete the picture of an uncertain prediction of future events from BMD it should be mentioned that changes in BMD can be underestimated owing to linearity errors in some DXA devices [12].

Interracial Differences

Much controversy dominates discussion of the influence of race, bone size and geometry on bone density measurement. Correction of BMD with regard to BSA would naturally change observed interracial differences in BMD, as body size and bone size differ considerably between races. In Table 1 some published values of BMD in Swedish and Japanese women of the same age, all investigated with the same DXA device [13–15], are shown before and after BSA correction by the author. It can be seen that uncorrected lumbar spine BMD and hip BMD are higher for Swedish than for Japanese women, but that after division of the uncorrected BMD by the square root of the BSA the situation is reversed: Swedish women have lower BMD corrected for BSA ($BMD_{corr.}$) than Japanese women, suggesting that the higher values of uncorrected BMD in Swedish women are due to their larger bone and body size, and that Swedish women might have lower true 3D-BMD than Japanese women. Swedish women have a much higher hip fracture incidence than Japanese women, a fact not explicable by BMD but in accordance with $BMD_{corr.}$ [16]. However, the racial differences in hip fracture risk are not all explicable by bone size.

Table 1. Interracial differences between Swedish and Japanese women. The effect of correcting lumbar spine BMD (LS-BMD) and femoral neck BMD (FN-BMD) for body surface area (BSA) differences from body height and weight according to the DuBois formula. Rel. BSA is the mean BSA in relation to that of the 'standard man' of 1.73 m^2. $LS\text{-}BMD_{corr.}$ and $FN\text{-}BMD_{corr.}$ are BMD divided by $\sqrt{BSA}$. *Note:* BMD values are higher for Swedish women, the $BMD_{corr.}$ lower (I, see [13]; II, see [14] ; III, see [15]. Measurements with Hologic QDR-1000 (pencil beam device)

	Swedish		Japanese			
	40–49 yrs *n* = 159 (I)	50–59 yrs *n* = 105 (I)	av. 45.7 yrs *n* = 36 (II)	av. 55.2 yrs *n* = 29 (II)	40–49 yrs *n* = 48 (III)	50–59 yrs *n* = 50 (III)
Height (cm)	166	163	153.9	150.9	154.7	154.9
Weight (kg)	64.3	65.7	52.2	53.0	54.5	59.4
BSA (m2)	1.716	1.709	1.486	1.475	1.520	1.578
Rel. BSA ($BSA/1.73m^2$)	1.004	1.006	1.079	1.083	1.067	1.047
LS-BMD (g/cm^2)	1.02	0.85	1.000	0.808	1.07	0.92
$LS\text{-}BMD_{corr.}$ (g/cm^2)	*1.024*	*0.855*	*1.079*	*0.875*	*1.142*	*0.963*
FN-BMD (g/cm^2)	0.80	0.69	–	–	0.79	0.70
$FN\text{-}BMD_{corr.}$ (g/cm^2)	*0.803*	*0.694*	–	–	*0.843*	*0.733*

Bone Size, Architecture, Growth and Geometry

BMC and BMD are crude expressions of bone mineral mass, not taking into account bone size or architecture. Eiffel, the constructor of the Eiffel tower in Paris (which incidentally has a similar shape and architecture to the distal human tibia), did not believe that the strength of his tower was critically dependent on the mass of iron in tons or the projected areal density ($tons/m^2$), nor that for his calculations of strength those parameters would suffice.

In a cadaver study of pigs it was shown that during growth two-dimensional BMD was heavily correlated with age, but volumetric density was not [17], the results showing that bone dimensions increase during growth, not true density. Growing children have virtually unchanged true volumetric bone density, but rising BMD owing to bone growth [18]. It has been demonstrated that men with vertebral osteoporotic fractures have reduced vertebral dimensions compared to age-matched controls, be it lack of periosteal increase or not, so that the load per unit area is relatively high [19]. As vertebral compression fractures are bound to occur when the load per horizontal unit area is high, it is likely that those who lack the adaptive horizontal vertebral growth with age are those who develop compression fractures, the body weight and load not tending to become smaller with age. Physical activity seems to augment BMC and projected area and width more than BMD [20]. Bone size augments during both childhood and adolescence. The transverse diameter of long bones and vertebrae also increases during adulthood [21,22]. Another important feature pertaining to long bones is that, after years of transverse growth with parallel augmentation of the inner and outer diameter, cortical thinning finally occurs, a phenomenon which can be seen with pQCT, but not with DXA.

Erroneous interpretations are possible if bone and body size are disregarded. This was emphasised by Prentice et al. [23] who stated: 'The size correction [calculating BMD from BMC by dividing by the projected area] assumes that BMC and projected bone area are directly proportional to one another, such that a 1% change in bone area is matched by a 1% change in BMC. This is rarely the case, and the exact relationship depends on the population group, skeletal site, body size, instrumentation, and scanning conditions'. They advocated that the use of BMD in epidemiological research be discontinued. In order to avoid the possibility of artefacts in the analysis of bone mineral data they recommended the use of BMC as the dependent variable and the inclusion of bone area, body weight and height as independent variables in all multiregression models. They rightly stressed that BMD does not adequately correct for bone and body size, and that many published correlations between BMD and factors such as obesity, calcium intake, biochemical indices, activity level etc. might be false, reflecting inadequate adjustment for bone area and body size. Whether they are in fact false might be difficult to detect, as BMD values are often published without BMC and bone size values.

Attention to the pathophysiology of growth might add to our understanding of bone fragility later in life. Perhaps health authorities, in their effort to improve our general health, should focus more on the events during growth, before peak bone mass is attained during adulthood [24].

Fan Beam or Pencil Beam?

DXA fan beam technology was introduced some years ago and became popular and of commercial interest because image quality was improved and examination time reduced to a few minutes. However, fan beam technology also introduced a magnification artefact differing in magnitude between brands. Critical studies failed to give encouraging results regarding the precision and accuracy of fan beam devices, and fan beam

technology is associated with higher radiation doses than is pencil beam technology. Pencil beam mode remains more precise and accurate [25–27].

Bone Mineral Apparent Density

Volumetric density can theoretically be estimated by combining anterior/posterior and lateral DXA measurements. However, pronounced accuracy errors remain, although smaller than for BMC. By comparing vertebral BMC to ashing data DXA systematically understimated ashing data by 15% for A/P vertebral BMC, by 33% for vertebral body BMC, by 23% for vertebral volume, and by 12% for the combined volumetric BMD [28].

An alternative to BMD is bone mineral apparent density (BMAD), which in theory is BMC per total bone volume; however, this cannot be measured correctly with DXA, but only estimated. According to Sievänen et al. [29] BMAD for an ellipsoid structure can be estimated as $6BMC/\pi\ BA$ (BA denoting projected bone area). This is clearly an approximation invented in the light of the deficiencies of DXA, recognising that the mechanical competence of bone is a combination of bone mass, macroscopic geometrical characteristics, cortical thickness, cross-sectional area, trabecular architecture, material properties and loading conditions for the site in question.

Another definition of BMAD is $BMD/\sqrt{BA}$. Using this definition Tsai et al. [30] found that BMAD was much lower in young Chinese men than in premenopausal women of similar age because of the higher BA and BMC of young men, and that men had higher BMD and BMAD than age-matched women only after the age of 50. It is a challenging thought that young women might have higher true bone density than young men, but it might be true (see below).

Sex Differences of BMD and 3D-BMD

It is often said that men have a higher peak bone mass than women, which explains the higher prevalence of fractures in women. Although true, this is a statement open to misunderstanding. Ebbesen et al. [31] in a carefully conducted study of cadaver vertebral bodies, clearly showed 1) that 3D-BMD (CT) in women was higher than in men in the younger decades, and 2) that men had higher 3D-BMD in the oldest decades. Regarding the higher 3D-BMD in young women, it is interesting to note that Schiessl et al. [32], by recalculating data published by others were able to show that in adolescence, females gained more bone per gram lean mass than did males. It could be hypothesised that this is a protection invented by nature to protect them from osteoporosis, counteracting the loss during later lactation periods from axial (trabecular) bone [33]. This relationship of lean body mass versus whole body BMC in adolescent females, so admirably highlighted by Schiessl et al. [32] is reversed after the menopause (H. Frost, personal communication).

BMD Might not be a Good Measure of Bone Strength

There seems to be universal agreement that at present BMD is the best predictor of hip fracture risk measurable in vivo, but this may change in the future with the advance of new techniques. The mechanical effectiveness of a solid body depends on its material properties, the amount of mass and the spatial distribution of that mass (architecture and geometry), and of cause the deforming force. The intrinsic stiffness of bone, described as Young's modulus of elasticity depends on the true volumetric density, i.e. 3D-BMD (and not the two-dimensional density BMD), the arrangement of the crystal and collagen fibres, the composition of collagen and ground substance etc., whereas high bone mass per se might not play a major part [34].

According to the 'Utah paradigm' neuromuscular function and anatomy dominate control of the biological mechanisms that control postnatal bone strength and 'mass' [35,36]. The consequence of the paradigm is that the present WHO definition of osteoporosis, based on T-score, is insufficient in that it ignores the biomechanical pathogenesis of osteoporotic fractures, and that it will have to be revised in the interest of patients [37].

It has recently been suggested that muscle cross-sectional area as an expression of force is a determinant of the mass of the corresponding bone, and that bone mass increases with age during childhood as a consequence of increased muscle mass and force [32]. This would be in accordance with the findings of Ito et al. [38], who demonstrated that cross-sectional muscle area was significantly related to the risk of fracture. Nordström et al. [39] found that in young ice hockey players the muscular strength of the thigh independently predicted BMD of the humerus and spine.

Turner [40], in a review article, drew our attention to the fact that that we often search meticulously for the correct answer in the wrong place: studies have focused on the effect of exercise on bone mass in adults, and even in elderly people. As the results of studies with DXA regarding BMD have been disappointing, some have concluded that exercise has only a moderate effect on osteoporosis. The correct question would have been: Does exercise prevent fractures? It certainly does, but perhaps (or probably) not via the effect on BMD, but via its effect on muscle strength, balance, postural stability and bone dimensions [40]. The solution therefore would be to focus on muscle training inclusive of the proprioceptors, as suggested by Schiessl (personal communication), rather than concentrating all efforts, in conjunction with the pharmaceutical industry, on finding and testing new drugs that act on bone only.

Singer et al. [41] in a study of the mechanical properties of thoracolumbar vertebrae, demonstrated that compressive vertebral bone strength was significantly correlated with BMD, but not with 3D-BMD. After multiplication of 3D-BMD with the midvertebral cross-sectional area the correlation was much improved, showing the importance of the dimensions of the vertebrae [41]. Non-invasive estimation of bone strength remains a problem for the spine and hip in that strength coming from the mechanical properties of the trabecular network is not being measured with DXA. The problem is less complicated for long bones, where pQCT is useful. Using this modality the cross-sectional bending moment of inertia can be measured. Feretti [34] introduced for long bones the bone strength index (BSI), which is the product of 3D-BMD and cross-sectional moment of inertia, easy to measure for each slice of a pQCT investigation of the radius or tibia. BSI has been validated in animal experiments and shown to be good indicator of breaking force, better than its two components alone. Interestingly, BSI of the distal forearm correlated very closely with the maximum muscle force of the forearm in both men and women [42].

The Discriminatory Ability of BMD for Fracture is not Good

In all published graphs depicting BMD of non-fracture cases versus BMD of fracture cases there is an immense overlap of BMD of non-fracture cases and normals, stating that low-energy fractures can occur at high, normal or low BMD, and suggesting that factors other than BMD might be important (e.g. load, elasticity of bones, layer of subcutaneous fat, propensity to falls etc.).

Vertebral Compression Fractures and Hip Fractures

We have previously shown, using ROC analysis and logistic regression analysis, that when lumbar spine BMC was corrected for bone size and body size, the discriminatory ability of DXA for vertebral fracture was much improved, but that lumbar osteodensitometry could not be used to identify women with a history of peripheral low-energy fractures [43]. It was demonstrated for vertebral fractures that body and bone size correction improved the true positive fraction from 60% to 80% at a 5% false positive fraction. Looking at the 3D-BMD with pQCT, Nijs et al. [44] found that the discriminatory ability of ultradistal 3D-BMD of the radius for vertebral compression fractures was the same, namely 80% true positive fraction at a 5% false positive fraction. A more proximal site of the radius was far less good for discrimination of those fractures.

If, in spite of the drawbacks described, BMD is used without bone or body size correction for hip fracture prediction, one can choose the graphs presented by De Laet et al. [45]. The graphs concerning Caucasians are based on several thousand persons. They were the result of a prospective study rendering reliable mean fracture incidence data, distributing individuals according to gender and age, in risk groups from 0.05% to 10% risk or higher [45]. This approach to the problem of fracture prediction is undoubtedly better than the use of T-scores and Z-scores for the purpose (see below).

Colles' Fractures

It has been discussed over the years whether DXA of the forearm could be used to predict future fractures and diagnose osteoporosis. Distal forearm fractures begin to occur earlier in life than vertebral and hip fractures and, unlike axial fractures, the incidence of Colles' fractures does not rise exponentially. It has been postulated that women in their 50s tend to break their forearm when falling, because they can still react swiftly enough to stretch out their arms, whereas older women cannot, for which reason they fall directly on their hips. It is obvious that Colles' fractures can occur at normal or even high BMD of the forearm. Statistically there is a certain – although somewhat weak – connection between Colles' fracture and low BMD at axial measuring sites. It appears that DXA of the forearm alone does not have an impressive discriminatory ability for Colles' fractures. Nor is axial osteoporosis, as judged from DXA, found frequently in patients with Colles' fractures. The isolated use of a forearm DXA to identify individuals at risk of later fractures at different bone sites is gaining widespread use in some countries, as a result of low cost and clever marketing. It has been postulated that this is more to the benefit of the manufacturers of forearm scanners of the patients. It would seem more meaningful to examine the distal forearm with pQCT or computed radiogrammetry than with DXA, when evaluation of future Colles' fracture risk is the problem, as it has been shown that the loss of cortical bone is an important factor in postmenopausal women with this type of fracture [46]. Certainly 3D-BMD of cortical bone of the radius, as measured with pQCT, declines with age, like the transaxial speed of sound with quantitative ultrasonometry. Although Colles' fracture can occur at high and low 3D-BMD, cortical 3D-BMD and thickness are lower in the contralateral radius of Colles' fracture cases than in non-fracture cases (S. Pors Nielsen, in press). The minimum moment of inertia of the distal radius in vitro and the cross-sectional area have been shown to correlate strongly with the forearm fracture force, unlike BMD highlighting the importance of bone geometry rather than BMD for the estimation of fracture risk [47].

Falsely High Spine BMD Values are often Encountered

It is well known that aortic calcifications and osteoarthritis/spondylosis can give falsely high lumbar spine BMD values in AP projection, and that these accuracy errors can be difficult to identify without radiography. Accordingly, some centres do not use lumbar spine densitometry after the age of 60. It is also well known that vertebral compression fractures give falsely high BMD values without affecting BMC. Small vertebral compressions large enough to give falsely high BMD values might remain undetected by DXA.

DXA Devices have Inherent Problems of Accuracy and Linearity

It should be borne in mind that DXA devices do not measure the absolute bone mineral content, but rather a model-related equivalent of the calibration material. Intraunit variation between devices is a major problem, reaching 20%, mainly due to calibration procedures, edge detection algorithms and assumptions regarding fat distribution [48,49]. Accuracy errors of DXA exist and are rather large for some devices, but easier to correct for than linearity errors [50]. Not only is there a variability between different brands, but also an intermachine variability of a significant magnitude for the same brands [51]. These factors clearly point to a more sceptical attitude towards the diagnostic use of DXA results.

When methods of evaluation of future fracture risk are being discussed and accuracy errors mentioned, it is often forgotten that the reliability of the test for fracture risk stratification in a population depends on the accuracy of the test in relation to the biological variation of the variable measured. For fracture risk stratification a low ratio of accuracy error to biological variation is needed. For DXA the accuracy error is around 3% for the forearm, around 6% for A/P spine, around 7% for the total hip, and around 10% for the lateral spine. The biological variation varies from one study to another. In healthy premenopausal women it is around 9% for the forearm BMD and A/P spine, and about 13% for the lateral spine and total hip [52].

Unexpected accuracy errors can occur after machine repair which are difficult to disclose by recommended routine procedures [53]. The situation is even more complicated because the performance of DXA is hampered not only by accuracy errors, but also by linearity errors. It is a major problem for DXA that the projected bone area with some devices augments with increasing bone mass, and that osteoporotic patients do not have their entire bone measured, but only part of it, namely the regions with relatively high attenuation. This is a linearity error, so far seldom mentioned in the scientific literature, although some recent reports deal with this problem [12,50,54]. Peel and Eastell [12] concluded that the correlation between the apparent change in projected bone area and BMC meant that changes in BMD would be underestimated. This problem was bigger for Hologic QDR 1000/W than for Lunar DPX, making it unlikely that the rate of bone loss could be a useful parameter in multicentre trials using densitometers from different manufacturers.

Discrepancies between the performance of different bone densitometers have been described and largely neglected. Any standardisation procedure is bound to meet with difficulties, as cross-calibration of DXA devices is difficult or impossible because of the use of different edge detection algorithms implemented by the manufacturers [50]. Tothill et al. [49] demonstrated pronounced accuracy errors with Lunar, Hologic and Norland devices, using a moderately anthropomorphic model. They concluded that there were differences in calibration of up to 8%, that routine changes of software by the manufacturers introduced accuracy errors, and that different assumptions had been made by the manufacturers concerning fat distribution, all precluding interchangeability of results from different instruments. Cross-calibration with the European spine phantom (ESP) was attempted for spine DXA at 18 centres: after cross-calibration the centre with the highest age-adjusted normal density value averaged 23% more than the centre with the lowest [55].

Peripheral versus Axial Measuring Sites

Osteoporosis is a general bone disease but bone loss takes place at different rates in different bones. Although there is a reasonably good correlation between distal forearm BMC and lumbar spine BMC in normal subjects, it is now widely accepted that the ability to diagnose axial osteoporosis is less good for forearm DXA (low axial BMD at normal forearm BMD is a common finding). This is to be expected, as bone loss with age seems to start and be more pronounced in axial trabecular bone. Manufacturers often try to convince potential customers that their forearm DXA device should be chosen, because its correlation with axial DXA is good. This is of course an invalid argument. What is important is the prediction of axial BMD from the peripheral device, as judged from the standard error of estimate (SEE), not the correlation coefficient. Even more important is the ability of the device to identify fracture cases from the peripheral measurement. Without this it is hard to believe that the device can be used to predict events of the future (fractures). Although DXA at a peripheral site might be of limited value, a low bone mass of the distal forearm may be used by a physician as one among several arguments to convince a patient that hormone replacement therapy might be a good idea. A normal DXA of the forearm cannot be used to exclude axial osteoporosis, and should typically lead to further testing, e.g. DXA of the hip or lumbar spine. The problem of discordance between measuring sites seems to be larger in early postmenopausal women than in the elderly [56].

Introduction of T-scores and Z-scores might be Threatening the Credibility of Bone Densitometry

DXA T-scores and Z-scores are increasingly used by physicians for fracture risk evaluation and by health authorities as a basis for reimbursement, after the publication of the position paper by Kanis et al. [57]. The measured absolute values of BMC and BMD are compared to mean values of young people, expressing the measured deviation from the mean as standard deviations of the young reference population (T-score) and/or of age-matched controls (Z-score). A certain low T- or Z-score corresponds to a certain average fracture risk. A T-score below –2.5 (measured BMD below –2.5 SD of the mean of the normal reference population of young people of the same sex) is synonymous with osteoporosis, according to the new definition recommended by Kanis et al. and apparently endorsed by the WHO [57].

It can be seen that the value of the T-score presupposes a correct normal mean value at the time of peak bone mass. This might not be so easy to attain. The normal reference material given to the customers by the manufacturers might not always be ideal: biases and mistakes may occur, and inclusion and exclusion criteria are often hidden in the haze of the past. One of the more serious problems with a normal reference population was the discrepancy of hip T-scores between Hologic and other DXA devices, owing to a reference population problem resulting in lower T-scores for Hologic devices. The solution to that problem was found only after some years by the introduction of the NHANES III reference material, ending up with T-scores close to those obtained with Lunar devices.

When a new definition of osteoporosis is introduced there should be unanimous acceptance of it by the medical community, which naturally can only be obtained if the method used, namely DXA, exhibits few false positives and false negatives and gives a safe prediction of fracture risk. However, bone densitometry (DXA) is not associated with an insignificant number of false positives or false negatives, nor does it give solid information on the fracture risk. In fact, published results of prospective studies over many years on the relation between BMD and fracture risk are few. Establishing the relationship between BMD (or size corrected BMD) and fracture incidence is what is needed. Such a relationship has been established for calcaneus BMD, spine BMD and distal radius BMC versus vertebral fracture incidence [1]. De Laet et al. [45] published the relationship between hip fracture risk and hip BMD, data that can be used for individual estimation of hip fracture risk, demonstrating that bone density values can be translated into fracture incidence rates. This is done without using T-scores, which can be regarded as an indirect statistical way of complicating the situation. The idea that by using T-scores and Z-scores different DXA devices could be compared is not valid, because 1) the reference populations are bound to be different; 2) the variability (inaccuracy, non-linearity and imprecision) of commercial DXA devices is far from being identical, with the result that a patient measured with an inferior-quality device must have a lower value of BMD to obtain a T-score below –2.5; 3) T-scores and Z-scores depend on body size (see above; Table 1 and Fig. 1); 4) another problem of cause is the very different T- and Z-scores obtained for different measuring sites. Which one should be chosen, the lowest or the highest or a mean, and then the mean of how many? This problem was highlighted by Abrahamsen et al. [58], who studied 2000 normal perimenopausal women with Hologic QDR 1000 and 2000, and found, using the WHO –2.5 T-score definition, that 4.3% of the population had osteoporosis as judged from the lumbar spine BMD alone, but that 12.5% fulfilled the –2.5 SD criterion in at least one of the regions scanned: lumbar spine, femur, total upper femoral end, femoral neck, trochanter, proximal distal forearm, ultradistal forearm or the whole body [58].

The so-called WHO definition of osteoporosis implies that either BMC or BMD can be used to define osteoporosis as a T-score below –2.5 [59]. Shipman et al., however [59], have shown that this is not so. In a normal female reference population of 8789 individuals aged 33–73 years and using a standard pencil beam DXA device for measurement of the lumbar spine, they found that using BMD around 20% of the 70-year-old women would be defined as osteoporotic, but using BMC only 10% would [59].

References

1. Wasnich RD. Consensus and the T-score fallacy. Clin Rheum 1997;16:337–9.
2. 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.
3. Pors Nielsen S, Kolthoff N, Bärenholdt O, et al. Diagnosis of osteoporosis by planar bone densitometry: can body size be disregarded? Br J Radiol 1998;71:934–43.
4. Morita R, Orimo H, Yamamoto M, et al. Some problems of dual-energy X-ray absorptiometry in the clinical use. Osteoporosis Int 1993;Suppl 1.S87–90.
5. Harris SS, Dawson-Hughes B. Weight, body composition, and bone density in postmenopausal women. Calcif Tissue Int 1996;59:428–32.
6. Mazess RB, Barden H, Mautalen C, Vega E. Normalization of spine densitometry. J Bone Miner Res 1994;9:541–8.
7. Gunnes M, Lehmann EH, Mellström D, Johnell O. The relationship between anthropometric measurements and fractures in women. Bone 1996;19:407–13.
8. Meyer HE, Falch JA, O'Neill T, Tverdal A, Varlov J. The European Vertebral Osteoporosis Study Group. Height and body mass index in Oslo, Norway, compared to other regions of Europe: do they explain differences in the incidence of hip fracture? Bone 1995;17:347–50.
9. Pouillès JM, Trémollières F, Ribot C. Variability of vertebral and femoral postmenopausal bone loss: a longitudinal study. Osteoporosis Int 1996;6:320–4.
10. Pors Nielsen S, Bärenholdt O, Hermansen F, Munk-Jensen N. Magnitude and pattern of skeletal response to long-term continuous and cyclic sequential oestrogen/progestin treatment. Br J Obstet Gynaecol 1994;101:319–24.

11. Boonen S, Cheng XG, Nijs J, et al. Factors associated with cortical and trabecular bone loss as quantified by peripheral computed tomography (pQCT) at the ultradistal radius in aging women. Calcif Tissue Int 1997;60:164–70.
12. Peel NFA, Eastell R. Comparison of rates of bone loss from the spine measured using two manufacturers' densitometers: J Bone Miner Res 1995;10:1796–801.
13. Löfman O, Larsson L, Ross I, Toss G, Berglund K. Bone mineral density in normal Swedish women. Bone 1997;20:167–74.
14. Okano H, Mizunuma H, Soda M, et al. The long-term effect of menopause on postmenopausal bone loss in Japanese women: results from a prospective study. J Bone Miner Res 1998;13:303–9.
15. Yoshimura N, Hashimoto TY, Morioka S, Kasamatsu T, Cooper C. Determinants of bone loss in a rural Japanese community: the Taiji study. Osteoporosis Int 1998;8:604–10.
16. Villa ML, Nelson L. Race, ethnicity and osteoporosis. In: Marcus R, Feldman D, Kelsey J, eds. Osteoporosis. Academic Press, 1996:436–43.
17. Ott SM, O'Hanlan M, Lipkin EW, Newell-Morris L. Evaluation of vertebral volumetric vs. areal bone mineral density during growth. Bone 1997;20:533–56.
18. Compston JE. Bone density: BMC, BMD, or corrected BMD? Bone 1995;16:5–7.
19. Vega E, Ghiringhelli G, Mautalen C, Rey Valzacchi G, Scaglia H. Zylberstein C. Bone mineral density and bone size in men with primary osteoporosis and vertebral fractures. Calcif Tissue Int 1998;62:465–89.
20. Brahm H, Mallmin H, Michaëlsson, Ström H, Ljunghall S. Relations between bone mass measurements and lifetime physical activity in a Swedish population. Calcif Tissue Int 1998;62:400–12.
21. Smith RW, Walker RR. Femoral expansion in aging women: implications for osteoporosis and fractures. Science 1964;145:156–7.
22. Mosekilde Li, Mosekilde L. Sex differences in age-related changes in vertebral body size, density and biomechanical competence in normal individuals. Bone 1990;11;67–73.
23. Prentice A, Parsons TJ, Cole T. Uncritical use of bone mineral density in absorptiometry may lead to size-related artifacts in the identification of bone mineral determinants. J Clin Nutr 1994;60:837–42.
24. Seeman E. Reduced bone density in women with fractures: contribution of low peak bone density and rapid bone loss. Osteoporosis Int 1994(Suppl I):S15–25.
25. Eiken P, Bärenholdt O, Jensen L Bjørn, Gram J, Nielsen SP. Switching from DXA pencil-beam to fan-beam: I: studies in vitro at four centers. Bone 1994;15:667–70.
26. Eiken P, Bärenholdt O, Jensen L Bjørn, Gram J, Nielsen SP. Switching from DXA pencil-beam to fan-beam: II: studies in vivo. Bone 1994;15 671–6.
27. Pocock NA, Noakes KA, Majerovic Y, Griffiths MR. Magnification error of femoral geometry using fan beam densitometers. Calcif Tissue Int 1997;60:8–10.
28. Sabin MA, Blake GM, MacLaughlin-Black SM, Fogelman I. The accuracy of volumetric bone density measurements in dual X-ray absorptiometry. Calcif Tissue Int 1995;56:210–14.
29. Sievänen, H, Kannus P, Nieminen V, Heinonen A, Oja P, Vuori I. Estimation of various mechanical characteristics of human bones using dual energy X-ray absorptiometry: methodology and precision. Bone 1996;18:17S–27S.
30. Tsai KS, Cheng WC, Chen CK, et al. Effect of bone area on spine density in Chinese men and women in Taiwan. Bone 1997;21:547–51.
31. Ebbesen EN, Thomsen JS, Beck-Nielsen H, Nepper-Rasmussen H J, Mosekilde L. Vertebral bone density evaluated by dual-energy X-ray absorptiometry in vitro. Bone 1998;23:283–90.
32. Schiessl H, Frost H, Jee SS. Estrogen and bone-muscle strength and mass relationships. Bone 1998;22:1–6.
33. Kolthoff N, Eiken P, Pors Nielsen S. Bone mineral changes during pregnancy and lactation: a longitudinal cohort study. Clin Sci 1998;94:405–12.
34. Feretti JL. Biomechanical properties of bone. In: Genant HR et al., eds. Bone densitometry and osteoporosis. Berlin: Springer, 1998:143–60.
35. Frost HM, Feretti JL, Jee WSS. Perspectives: Some roles of mechanical usage, muscle strength, and the mechanostat in skeletal physiology, disease, and research. Calcif Tissue Int 1998;62:1–7.
36. Frost HM. The Utah paradigm, bone, and osteoporosis. A 1999 overview. In: Osteoporosis. Update 1999. Proceedings. Third International Congress on Osteoporosis. Xián, China, 1999:71–8.
37. Frost H M. Osteoporosis: a rationale for further definitions? Calif Tissue Int 1998;62:89–95.
38. Ito M, Ohki M, Hayashi K, Yamada M, Uetani M, Nakamura T. Relationship of spinal fracture to bone density, textural, and anthropometric parameters. Calcif Tissue Int 1997;60:240–3.
39. Nordström P, Thorsen K, Bergström E, Lorentzon R. High bone mass and altered relationship between bone mass, muscle strength, and body constitution in adolescent boys on a high level of physical activity. Bone 1996;19:189–95.
40. Turner CH. Exercise as a therapy for osteoporosis: the drunk and the street lamp, revisited. Bone 1998;23:83–5.
41. Singer K, Edmondston S, Day R, Breidahl P, Price R. Prediction of thoracic and lumbar vertebral body compressive strength: Correlation with bone mineral density and vertebral region. Bone 1995;17:167–4.
42. Schiessl H, Feretti JL, Tysarczyk-Niemeyer G, Willnecker J. Noninvasive bone strength index as analyzed by peripheral quantitative computed tomography. In: Schönau E, ed. Paediatric osteology: new developments in diagnostics and therapy. Amsterdam: Elsevier, 1996:141–6.
43. Pors Nielsen S, Hermansen F, Bärenholdt O. Interpretation of lumbar spine densitometry in women with fractures. Osteoporosis Int 1993;3:276–82.
44. Nijs J, Westhovens JJ, Cheng XG, Borghs H, Dequeker J. Diagnostic sensitivity of peripheral computed tomography measurements at ultradistal and proximal radius in postmenopausal women. Bone 1998;22:659–64.
45. De Laet CEDH, Van Hout BA, Burger H, Weel AEAM, Hofman A, Pols HAP. Hip fracture prediction in elderly men and women: validation in the Rotterdam study. J Bone Miner Res 1998;13;1587–93.
46. Crespo R, Revilla M, Usabiago J, Crespo E, García-Arino, Villa LF, Rico H. Metacarpal radiogrammetry by computed radiography in postmenopausal women with Colles' fracture and vertebral crush fracture syndrome. Calcif Tissue Int 1998;62:470–3.
47. Myers ER, Hecker AT, Rooks DS, Hipp JA, Hayes WC. Geometric variables from DXA of the radius predict forearm fracture load in vitro. Calcif Tissue Int 1993;52:199–204.
48. Fischer M, Kempers B. Phantom studies in osteoporosis. Eur J Nucl Med 1993;20:434–9.
49. Tothill P, Avenell A, Reid DM. Precision and accuracy of measurements of whole-body bone mineral: comparisons between Hologic, Lunar and Norland dual-energy X-ray absorptiometers. Br J Radiol 1994;67:1210–17.
50. Pors Nielsen S, Bärenholdt O, Diessel E, Armbrust S, Felsenberg D. Linearity and accuracy errors in bone densitometry. Br J Radiol 1998;71:1062–8.
51. Formica CA. Standardization of BMD measurements. [Editorial]. Osteoporosis Int 1998;8:1–3.
52. Hassager C, Christiansen C. Measurement of bone mineral density. Calcif Tissue Int 1995;57:1–5.
53. Blake GM, Preston NG, Patel R, Herd RJM, Fogelman I. An unexpected change in DXA calibration not detected by routine quality control checks. Osteoporosis Int 1999;9:115–20.
54. Tothill P, Avenell A. Anomalies in the measurement of changes in bone mineral density of the spine by dual-energy X-ray absorptiometry. Calcif Tissue Int 1998;63:126–33.
55. Dequeker J, Pearson J, Reeve J, et al. Dual-X-ray absorptiometry – cross-calibration and normative reference ranges for the spine: results of a European Community concerted action. Bone 1995;17:247–54.
56. Miller PD. Controversies in diagnostic classifications. In:

Osteoporosis. Update 1999. Proceedings. Third International Congress on Osteoporosis. Xián, China, 1999:134–6.
57. Kanis JA, Devogelaer J-P, Gennari C. Practical guide for the use of bone mineral measurements in the assessment of osteoporosis: A position paper of the European Foundation for Osteoporosis and Bone Disease. Osteoporosis Int 1996;6:256–61.
58. Abrahamsen B, Hansen TB, Jensen LB, Hermann AP, Eiken P. Site of osteodensitometry in perimenopausal women: correlation and limits of agreement between anatomical regions. J Bone Miner Res 1997;12:1471–9.
59. Shipman AJ, Guy GWG, Smith I, Ostlere S, Greer W, Smith R. Vertebral bone mineral density, content and area in 8789 normal women aged 33-73 years who never had hormone replacement therapy. Osteoporosis Int 1999;9:420–6.

Received for publication 18 June 1999
Accepted in revised form 30 September 1999