Osteoporos Int (2000) 11:321–330


**Osteoporosis International**

*Original Article*

# Can the WHO Criteria for Diagnosing Osteoporosis be Applied to Calcaneal Quantitative Ultrasound?

M. L. Frost, G. M. Blake and I. Fogelman

Osteoporosis Unit, Guy's Hospital, London, UK

**Abstract.** With the increasing number of quantitative ultrasound (QUS) devices in use worldwide it is important to develop strategies for the clinical use of QUS. The aims of this study were to examine the age-dependence of *T*-scores and the prevalence of osteoporosis using the World Health Organization Study Group criteria for diagnosing osteoporosis and to examine the *T*-score threshold that would be appropriate to identify women at risk of osteoporosis using QUS. Two groups of women were studied: (i) 420 healthy women aged 20–79 years with no known risk factors associated with osteoporosis; (ii) 97 postmenopausal women with vertebral fractures. All subjects had dual-energy X-ray absorptiometry (DXA) measurements of the spine and hip and QUS measurements on three calcaneal ultrasound devices (Hologic Sahara, Hologic UBA575+, Osteometer DTUone). A subgroup of 102 (76 on the DTUone) healthy women aged 20–40 years was used to estimate the young adult mean and SD for each QUS and DXA measurement parameter to calculate *T*-scores. The age-related decline in *T*-scores for QUS measurement parameters was half the rate observed for the bone mineral density (BMD) measurements. The average *T*-score for a woman aged 65 years was –1.2 for QUS measurements and –1.75 for the BMD measurements. When osteoporosis was defined by a *T*-score $\leqslant$–2.5 the prevalence of osteoporosis in healthy postmenopausal women was 17%, 16% and 12% for lumbar spine, femoral neck and total hip BMD respectively. When the same definition was used for QUS measurements the prevalence of osteoporosis ranged from 2% to 8% depending on which ultrasound device and measurement parameter was used. Four different approaches, based on DXA-equivalent prevalence rates of osteoporosis, were utilized to examine which *T*-score threshold would be appropriate for identifying postmenopausal women at risk of osteoporosis using QUS measurements. These ranged from –1.05 to –2.12 depending upon the approach used to estimate the threshold and on which QUS device the measurements were performed, but all were significantly lower than the threshold of –2.5 used for BMD measurements. In conclusion, the WHO threshold of $T = -2.5$ for diagnosing osteoporosis requires modification when using QUS to assess skeletal status. For the three QUS devices used in this study, a *T*-score threshold of –1.80 would result in the same percentage of postmenopausal women classified as osteoporotic as the WHO threshold for BMD measurements. Corresponding *T*-score thresholds for individual measurement parameters on the two commercially available devices were –1.61, –1.94 and –1.90 for Sahara BUA, SOS and estimated heel BMD respectively and –1.45 and –2.10 for DTU BUA and SOS respectively Additional studies are needed to determine suitable *T*-score thresholds for other commercial QUS devices.

**Keywords:** Calcaneus; Osteoporosis; Quantitative ultrasound; *T*-scores

*Correspondence and offprint requests to:* Ms M. L. Frost, Osteoporosis Unit, Guy's Hospital, St Thomas Street, London SE1 9RT, UK. Fax: +44(0)207 955 8883. e-mail: michelle.frost@kcl.ac.uk

## Introduction

In recent years there has been a large increase in the number of quantitative ultrasound (QUS) devices in use worldwide. QUS technology has the potential to meet the increased demand for bone densitometry services due

to the progressive aging of the world's population. Despite this, the widespread clinical application of QUS has been limited by the fact that dual-energy X-ray absorptiometry (DXA) remains the accepted gold standard for assessing skeletal status. However, DXA devices are relatively expensive and require patients to be referred to hospital-based facilities. QUS has several advantages including no patient or operator exposure to ionizing radiation, low cost and portability. Large prospective fracture studies have demonstrated that both broadband ultrasound attenuation (BUA) and speed of sound (SOS) at the calcaneus can predict osteoporotic fracture as well as can DXA at the spine and hip [1,2]. Recently the US Food and Drug Administration approved a number of commercial QUS systems. As a consequence, the number of QUS devices in clinical use is set to increase and it is important to develop improved strategies for the clinical use of QUS.

At present, there is a widely accepted convention of defining osteoporosis in terms of *T*-scores as recommended by the World Health Organization (WHO) Working Party [3]. The WHO definition of osteoporosis of a *T*-score value of less than –2.5 was developed for bone mineral density (BMD) measurements at the spine, hip and forearm. This threshold value of BMD greater than 2.5 SD below the mean for a young adult population identifies approximately 30% of all postmenopausal Caucasian women as having osteoporosis at either the spine, hip or forearm. This is similar to the lifetime risk of fracture at these sites [4]. It has been reported that this definition of osteoporosis may not be appropriate at other skeletal sites or for different technologies such as QUS [5,6]. These studies show that few patients have a QUS *T*-score value below –2.5 and suggest that it may be necessary to provide a *T*-score criterion specific to the measurement technology employed. In addition to this, there are many different QUS devices available worldwide, all using manufacturer-supplied reference databases, which will further increase the heterogeneity between devices.

The aim of this study was to examine the age-dependence of *T*-scores, and the estimated prevalence of osteoporosis as defined by the WHO, for two commercial QUS devices (Hologic Sahara, Osteometer DTUone) and the Hologic UBA575+ used in early QUS prospective fracture studies. We also examined the optimum *T*-score threshold that could be used to identify postmenopausal women at risk of sustaining a fragility fracture using calcaneal QUS.

## Subjects and Methods

### *Subjects*

The study population consisted of two groups: (i) 420 healthy premenopausal and postmenopausal Caucasian women; (ii) 97 postmenopausal women with vertebral fractures. The healthy women from group 1 were recruited from three sources: (i) patients referred by their general practitioner for routine bone density screening by DXA; (ii) young hospital personnel; (iii) women from the general population who volunteered to participate in clinical research. To obtain a group of healthy women which is a close representation of the general population, women were excluded if they had a history of low-trauma fracture, a menopause before the age of 40 years, a history of amenorrhea or any treatments or diseases known to affect bone metabolism. The women from group 2 were recruited from the Guy's Hospital metabolic bone clinic. Written informed consent was obtained from all study participants and the study was approved by the Guy's Hospital Research Ethics Committee.

### *Measurements*

BMD measurements of the lumbar spine (L1–4), femoral neck and total hip were performed using a Hologic QDR4500 (Hologic, Bedford, MA). All subjects had calcaneal QUS measurements on the Hologic Sahara and the UBA575+. Three hundred and thirty-eight of the 420 women also had calcaneal QUS measurements on the Osteometer DTUone.

*Sahara Clinical Bone Sonometer (SAH).* The Sahara Clinical Bone Sonometer (Hologic, Bedford, MA) consists of two unfocused transducers mounted coaxially on a motorized calliper. One transducer acts as a transmitter and the other as a receiver. The transducers are acoustically coupled to the heel using soft rubber pads and an oil-based coupling gel. The Sahara device measures both BUA and SOS at a fixed region of interest in the mid-calcaneus and the results are combined to provide an estimate of heel BMD (Est.heel BMD) with units of grams per square centimeter using the following equation:

$$\text{Estimated heel BMD} = 0.002592 \times (\text{BUA} + \text{SOS}) - 3.687(\text{g/cm}^2)$$

It is important to note that estimated heel BMD is inferred from a linear combination of BUA and SOS and is not an actual measurement of calcaneal BMD.

*Ultrasonic Bone Analyzer 575+ (UBA).* The UBA575+ (Hologic, Bedford, MA) consists of two unfocused transducers mounted coaxially in a water bath containing a surfactant. The heel is positioned in the water bath and a rectilinear scan is performed with measurements taken in a $3 \times 3$ grid located in the mid-calcaneus. Both BUA and SOS are calculated. A pulse echo technique is utilized to provide an estimate of bone thickness and a third parameter is calculated termed bone velocity (VB) which is the velocity of sound through bone only.

*DTUone (DTU).* The DTUone (Osteometer Meditech, California, USA) consists of two focused transducers

mounted coaxially in a water bath containing a surfactant. A rectilinear scan of the calcaneus is performed yielding an image size of approximately $60 \times 80$ mm and a pixel size of 0.5 mm. Both BUA and SOS are calculated at each pixel and an automatic region of interest is selected in an area with a local minimum of attenuation, located in the posterior tuberosity of the calcaneus.

### *Data Analysis*

A subgroup of 102 (76 for the DTUone) healthy women aged 20–40 years were selected to estimate the young normal mean and SD for each QUS and DXA measurement parameter for the purpose of calculating *T*-scores:

$$T\text{-}score = \frac{\text{Measurement value} - \text{Young adult mean}}{\text{Young adult population SD}}$$

Women aged 50+ years were then classified into three groups according to their *T*-scores as defined by the WHO:

*Normal*: a *T*-score greater than or equal to –1.
*Low bone mass (osteopenia)*: a *T*-score less than –1 but greater than –2.5.
*Osteoporosis*: a *T*-score less than or equal to –2.5.

The proportions of postmenopausal women aged over 50 years in each WHO diagnostic category were expressed as percentages. To examine the age-related decrease in *T*-scores for each measurement parameter women were placed into 5 year age groups (20–24 years, 25–29 years, etc.) and the mean *T*-score was then calculated for each age group. To investigate the optimum threshold for identifying a high-risk group using QUS four different approaches were compared:

*Approach 1.* Simple linear regression was performed between the age-related decline in each of the QUS *T*-score parameters and the age-related decline in total hip *T*-scores (forcing the line through the origin) for healthy women in group 1. The regression coefficient was then multiplied by –2.5 to estimate the equivalent *T*-score threshold for QUS.

*Approach 2.* A threshold for QUS was estimated by taking the *T*-score threshold that would diagnose 15% of the healthy women in group 1 as osteoporotic. This figure of 15% was chosen as it was the average prevalence of osteoporosis in healthy postmenopausal women using the spine (17%), femoral neck (16%) and total hip (12%) BMD data individually (see Results).

*Approach 3.* The percentage of women with vertebral fractures in group 2 with a total hip *T*-score equal to or less than –2.5 was calculated. The *T*-score threshold for QUS was chosen by estimating the *T*-score required to detect the same percentage of women with vertebral fractures as identified by total hip BMD. This method was identical to that applied by Hans et al. [6] to the EPIDOS study QUS data in hip fracture patients.

*Approach 4a.* This approach is similar to that applied by Hans et al. [6] (approach 3) but it was applied to the healthy women in group 1 identified as osteoporotic on the basis of their lumbar spine *T*-score. The *T*-score threshold for QUS was chosen by estimating the *T*-score required to detect the same percentage of women with a lumbar spine *T*-score equal to or less than –2.5 as identified by total hip BMD.

*Approach 4b.* This approach is the same as approach 4a but a *T*-score threshold was chosen by estimating the *T*-score required to detect the same percentage of women with a total hip *T*-score equal to or less than –2.5 as identified by lumbar spine BMD.

The index of positive agreement ($P_{pos}$) was calculated to assess the agreement between the different QUS devices and between QUS and DXA for identifying postmenopausal women with osteoporosis. This index was proposed by Cicchetti and Feinstein [7] to estimate the proportion of agreement between two techniques when neither is regarded as the 'gold standard'. In the present study, $P_{pos}$ represents the proportion of women that are classified by both devices as osteoporotic and is expressed as a fraction of the mean number identified by the two devices separately.

## Results

### *Calculation of* T-*scores*

The young adult mean and SD for each measurement parameter are shown in Table 1. Also shown in Table 1 are the manufacturer values for the young adult mean and SD. Although all three QUS devices measure BUA and SOS in the mid-calcaneus there is wide variation in both the mean and SD values. The young adult mean for BUA as measured on the DTUone was approximately 30% lower than that seen for the other QUS devices. The BUA young adult SD was also significantly lower for the DTUone imaging device. The bone velocity (VB) parameter displayed a larger young adult mean and SD compared with the other velocity measurements. The mean and SD obtained using our reference population for each measurement parameter were similar to those used by the manufacturer, although the manufacturer young adult mean tended to be slightly lower (except for DTU SOS). The manufacturer SD was higher than that obtained for our reference population for DTU BUA, DTU SOS and total hip BMD but lower for estimated heel and lumbar spine BMD. The reference data used in this study for femoral neck and total hip BMD were closer to the Hologic manufacturer's reference data than reference data from the NHANES study [8].

**Table 1.** The young adult[a] mean and SD for QUS and BMD parameters used for the calculation of *T*-scores

| Measurement parameter | Local reference population | | Manufacturers' values | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| SAH BUA (dB/MHz) | 77.7 | 13.5 | n/a | n/a |
| SAH SOS (m/s) | 1560.7 | 25.1 | n/a | n/a |
| SAH Est.heel BMD (g/cm$^2$) | 0.561 | 0.10 | 0.537 | 0.08 |
| UBA BUA (dB/MHz) | 80.2 | 15.5 | n/a | n/a |
| UBA VB (m/s) | 1639.0 | 46.4 | n/a | n/a |
| UBA SOS (m/s) | 1507.5 | 6.6 | n/a | n/a |
| DTU BUA (dB/MHz) | 54.0 | 6.0 | 51.3 | 6.4 |
| DTU SOS (m/s) | 1553.5 | 8.4 | 1557.8 | 10.2 |
| DXA lumbar spine (g/cm$^2$) | 1.068 | 0.12 | 1.047 | 0.11 |
| DXA femoral neck (g/cm$^2$) | 0.892 | 0.10 | 0.895 | 0.1 |
| DXA total hip (g/cm$^2$) | 0.988 | 0.10 | 0.975 | 0.12 |

[a]102 (76 for the DTU) healthy women aged 20–40 years.
SAH, Hologic Sahara; UBA, Hologic UBA575+; DTU, Osteometer DTUone; n/a, not available.

## *Age-Related Decline in* T-*scores*

The age-related decline in *T*-scores for both BMD and QUS measurements is shown in Fig. 1. The *T*-scores calculated for BMD measurements begin to fall at approximately age 40 years and decline thereafter (Fig. 1a). By age 65 years the average lumbar spine and hip BMD *T*-score was approximately –1.75. The mean *T*-scores for the lumbar spine display an apparent increase after the age of 65 years that probably reflects degenerative changes in the spine, which is a problem when measuring elderly subjects. As a comparison, the estimated heel BMD parameter measured by the Sahara device has also been plotted with the spine and hip BMD parameters (Fig. 1a). *T*-scores for estimated heel BMD fell at approximately half the rate of the spine and hip BMD parameters and at age 65 years the average *T*-score was –1.10. When the three BUA measurements were compared, the Sahara and UBA *T*-scores decline at a similar rate while the DTU displays a rather slower age-related decline in *T*-scores (Fig. 1b). The BUA *T*-scores decline to about –1 at age 65 years, which is significantly higher than the mean *T*-score seen for the BMD measurements at the same age. The age-related decline

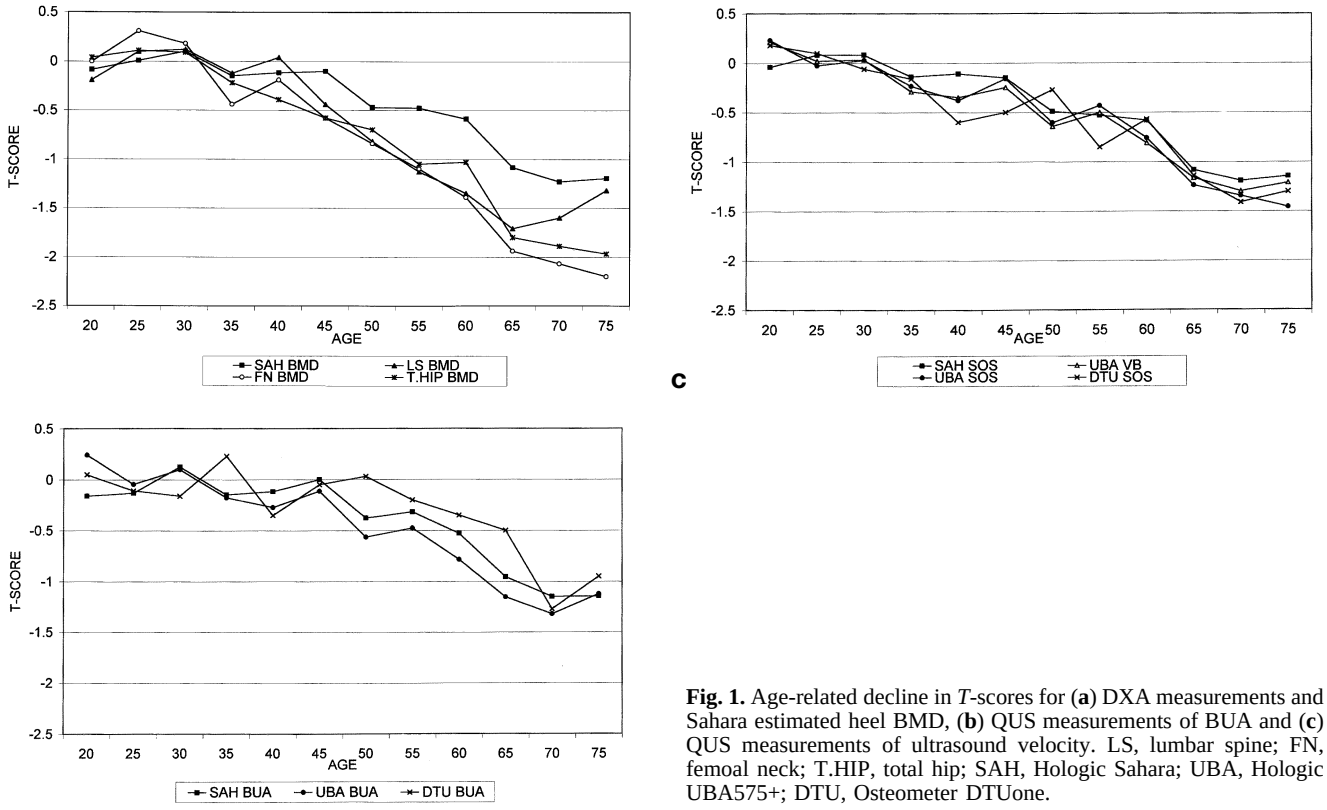


**Fig. 1.** Age-related decline in *T*-scores for (**a**) DXA measurements and Sahara estimated heel BMD, (**b**) QUS measurements of BUA and (**c**) QUS measurements of ultrasound velocity. LS, lumbar spine; FN, femoal neck; T.HIP, total hip; SAH, Hologic Sahara; UBA, Hologic UBA575+; DTU, Osteometer DTUone.

in *T*-scores for the velocity measurements displays a similar pattern to BUA (Fig. 1c). All four velocity measurements follow a similar age-related decline in *T*-scores, falling to approximately –1.20 by age 65 years.

### *Prevalence of Osteoporosis*

Figure 2 displays the percentage of postmenopausal women aged 50+ years classified as normal, osteopenic and osteoporotic according to the WHO criteria. The proportion of women classified as normal ranges from 40% for the BMD measurements to almost 60% for the QUS measurements. The prevalence of osteopenia is approximately 40% for all measurement parameters. The prevalence of osteoporosis for the BMD measurements is 17% at the lumbar spine, 16% at the femoral neck and 12% for the total hip site. The mean prevalence of osteoporosis for all three skeletal sites was 15%. When defined as a *T*-score $\leqslant$ –2.5 the prevalence of osteoporosis for QUS measurements is significantly lower compared with the lumbar spine and hip, ranging from just 2% for the UBA bone velocity parameter to 8% for the DTU SOS measurement. The average for all the QUS measurements was 3.8%.

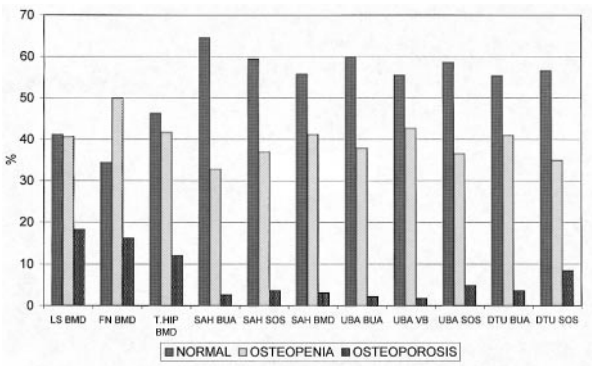


**Fig. 2.** The proportion of postmenopausal women classified as normal, osteopenic and osteoporotic according to the WHO criteria. VB, velocity in bone; BUA, broadband ultrasound attenuation; SOS, speed of sound; other abbreviations as in Fig. 1.

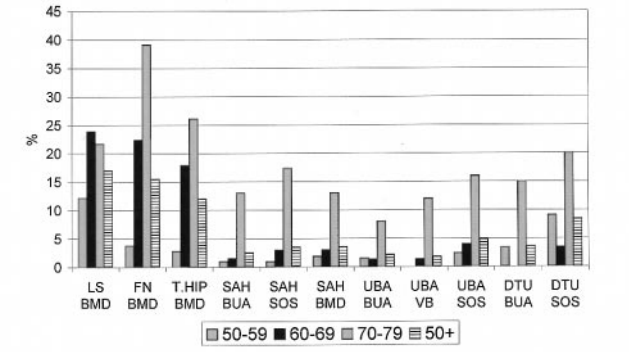


**Fig. 3.** The prevalence of osteoporosis as defined by the WHO criteria in postmenopausal women aged 50–80 years for BMD and QUS parameters. Abbreviations as in Fig. 2.

To examine this further, the prevalence of osteoporosis according to the WHO criteria was calculated for each decade for postmenopausal women aged 50–80 years for both the BMD and QUS parameters (Fig. 3). As expected, the prevalence of osteoporosis increases as the population gets older. By ages 70–79 years, the prevalence of osteoporosis for the BMD measurements ranges from 22% at the spine to almost 40% at the femoral neck. The prevalence of osteoporosis at the spine is lower because of the increased occurrence of osteoarthritis seen in women of this age, which will artificially increase BMD values. If we examine postmenopausal women aged 50 years and over, the prevalence varies considerably according to the measurement site and measurement technique. The prevalence of osteoporosis is significantly higher if a BMD measurement is taken, with the prevalence being on average 15%. In comparison the QUS measurements classify a much lower number of women as osteoporotic for every age decade.

### *The Optimum* T*-score Threshold for Interpretation of QUS Results*

The four different approaches used to estimate the optimum threshold for QUS measurements to identify postmenopausal women at risk of osteoporosis yielded a range of different *T*-score thresholds. However, in all cases the calculated *T*-score threshold was significantly less negative than the –2.5 currently used for BMD measurements (Table 2). If linear regression was performed (approach 1) to estimate the difference in slopes between the age-related decline in total hip BMD and the QUS parameters, the average *T*-score threshold was –1.50 (range –1.05 to –1.73). The *T*-score thresholds that captured 15% of postmenopausal women as osteoporotic (approach 2) had a mean of –1.80 (range –1.45 to –2.10). Fifty-one (55%) of the women with vertebral fractures had a total hip *T*-score equal to or less than –2.5. The *T*-score threshold that identified 55% of the vertebral fracture group (approach 3) yielded the lowest *T*-score thresholds for the QUS parameters, with a mean of –1.89 (range –1.39 to –2.12). The sensitivity of total hip BMD for identifying spinal osteoporosis was 41% and the sensitivity of lumbar spine for identifying osteoporosis at the hip was 50%. These figures were used to estimate *T*-score thresholds for QUS in approaches 4a and 4b. Approach 4 yielded *T*-score thresholds for QUS ranging from –1.32 to –2.19, averaging –1.79 and –1.59 for approaches 4a and 4b, respectively. Regardless of which approach was utilized, the BUA parameter tended to have slightly less negative *T*-score thresholds than the SOS measurement parameters. The individual results of these different approaches are summarized in Fig. 4.

Approach 2 yielded a mean *T*-score threshold of –1.80 for the three QUS devices. To assess the agreement between the different QUS devices and between QUS and DXA in identifying postmenopausal women with osteoporosis, the index of positive agreement ($P_{pos}$) was

**Table 2.** *T*-score thresholds for identifying women at risk of osteoporosis using QUS measurements, estimated using the four approaches described in Subjects and Methods

| | Approach 1 | Approach 2 | Approach 3 | Approach 4a | Approach 4b | Average *T*-score threshold |
|---|---|---|---|---|---|---|
| | Linear regression | 15% POST women | Vertebral fracture | LS *T*-score <–2.5 | T.Hip *T*-score <–2.5 | |
| *Sajara* | | | | | | |
| BUA (dB/MHz) | –1.33 | –1.61 | –1.81 | –1.50 | –1.50 | –1.55 |
| SOS (m/s) | –1.46 | –1.94 | –2.03 | –1.83 | –1.32 | –1.72 |
| Heel BMD (g/cm$^2$) | –1.48 | –1.90 | –2.11 | –1.63 | –1.45 | –1.71 |
| *UBA 575+* | | | | | | |
| BUA (dB/MHz) | –1.57 | –1.75 | –1.39 | –1.75 | –1.69 | –1.63 |
| VB (m/s) | –1.64 | –1.77 | –1.83 | –2.00 | –1.77 | –1.80 |
| SOS (m/s) | –1.73 | –1.89 | –2.12 | –2.19 | –1.74 | –1.93 |
| *DTUone* | | | | | | |
| BUA (dB/MHz) | –1.05 | –1.45 | –1.91 | –1.55 | –1.43 | –1.48 |
| SOS (m/s) | –1.71 | –2.10 | –1.91 | –1.88 | –1.82 | –1.88 |
| Average *T*-score threshold | –1.50 | –1.80 | –1.89 | –1.79 | –1.59 | |

LS, lumbar spine; T.Hip, total hip; POST, postmenopausal.

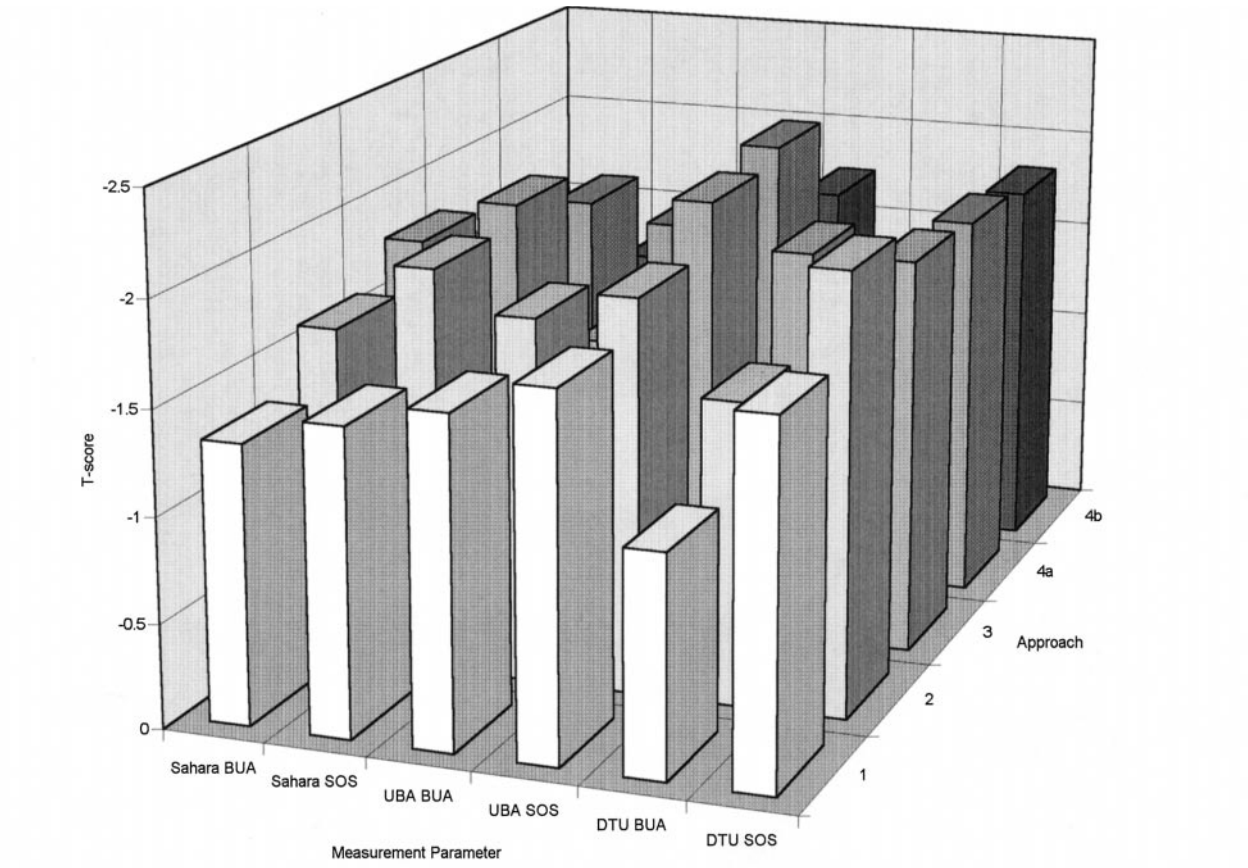


**Fig. 4.** The average *T*-score thresholds for QUS measurement parameters using approaches 1 to 4 described in Subjects and Methods. Abbreviations as in Fig. 2.

calculated (Table 3). Postmenopausal women with a *T*-score threshold of $\leqslant$ –1.80 for either BUA or SOS on each QUS device, or $\leqslant$ –2.50 at either the lumbar spine or total hip, were considered osteoporotic. When the three QUS devices were compared, $P_{pos}$ ranged from 0.68 to 0.76, indicating that approximately 72% of women were classified similarly using different QUS devices. $P_{pos}$ ranged from 0.43 to 0.46 when QUS and DXA were compared, indicating that approximately 45% of women were classified similarly as osteoporotic. As a

**Table 3.** The index of positive agreement between each of the QUS devices and between QUS and BMD for diagnosing osteoporosis

| | UBA | SAH | DTU |
|---|---|---|---|
| SAH | 0.76 | – | – |
| DTU | 0.68 | 0.75 | – |
| BMD | 0.44 | 0.46 | 0.43 |

UBA, UBA575+; SAH, Sahara; DTU, DTUone.
Women were classified as osteoporotic if they had a *T*-score $< -2.5$ at either the lumbar spine or total hip for DXA measurements or $< -1.80$ for either BUA or SOS for QUS measurements.

comparison, 45% ($P_{pos} = 0.45$) of women were classified similarly using lumbar spine and total hip BMD measurements.

## Discussion

As the use of QUS for clinical studies increases there is a need for guidelines on how to interpret the results obtained with these devices. At present, the most commonly accepted guidelines are those proposed by the WHO Study Group for the interpretation of BMD measurements at the lumbar spine, femoral neck and distal forearm [3]. However, it is becoming increasingly clear that these criteria cannot be applied to QUS measurements. The aim of this study was to examine the consequences of using the WHO criteria for the interpretation of QUS measurements and to derive an improved threshold for QUS which will be of practical use in a clinical setting.

The age-related decline in *T*-scores for QUS measurements was almost half that seen for the BMD measurements performed by DXA. By age 75 years, the average *T*-score for the QUS measurements was just –1.25. Similar findings have been reported for another study, which examined the age-related decrease in *T*-scores using manufacturer normative databases [5]. There could be two explanations for this: (i) the calcaneus displays a slower rate of age-related bone loss, or (ii) QUS has a higher population SD compared with DXA. The first explanation is unlikely as studies have revealed that the calcaneus displays similar rates of bone loss to the spine and hip [9–11]. One study showed that the annual changes in QUS measurement parameters were comparable to BMD measurements of the spine and hip [12]. The second explanation seems to be the more plausible. Random accuracy errors inherent in the QUS measurement technique may increase the young adult SD over and above the true variation due to BMD differences. The effects of accuracy errors for DXA of the lumbar spine caused by osteoarthritis, aortic calcification, etc., have been well documented but these problems are common in the elderly and so do not affect the accuracy of the young adult measurements which are used to calculate *T*-scores. One possible way in which the young adult SD could be artificially increased using QUS is a process called phase cancellation [13]. As the calcaneus is highly inhomogeneous in terms of bone tissue, when a broadband ultrasound signal propagates through the calcaneus different parts of the wavefront will travel at different speeds due to the dependence of ultrasound velocity on density (i.e., the ultrasound signal will travel faster through denser areas of bone). This means that the emerging wavefronts will arrive at the transducer face at different phases and will tend to cancel, leading to an underestimation of the true intensity of the signal and subsequently the attenuation will be overestimated. This phenomenon is larger at high BUA as the ultrasound signal will undergo more modification as it traverses the calcaneus and thus this could have a significant effect on the young adult SD. On examination of the young adult data for this study it would appear that this might be the explanation. When the young adult SD is expressed as a percentage of the young adult mean for BUA it is almost 20% while the corresponding value for BMD is just 10%. The young adult SD for SOS measurements are also high if they are compared with the clinical range observed for these measurements. The Sahara device combines both BUA and SOS to provide an estimate of heel BMD. If we compare the young adult SD obtained for this estimated heel parameter with the young adult SD estimated for DXA at the calcaneus observed in a recent study, the Sahara-estimated heel BMD SD is 16% higher [14].

The slower rate of age-related decline in *T*-scores observed for QUS compared with DXA has implications when using the WHO criteria to diagnose osteoporosis. The number of subjects identified as osteoporotic will vary according to the site and technique used as well as the reference population. The WHO report states that 30% of all postmenopausal Caucasian women will be identified as having osteoporosis based on BMD measurements at the spine, hip and forearm [3]. If results of a single technology such as DXA are compared at different skeletal sites, the prevalence of osteoporosis will vary [15,16]. In one study 45% of postmenopausal women were diagnosed as having osteoporosis at the lumbar spine, femoral neck or forearm [4] while in another study the prevalence of osteoporosis ranged from 10% to 45% depending on the site measured by DXA [17]. In the present study, the prevalence of osteoporosis was much lower when the WHO criteria were applied to QUS measurements compared with BMD. The prevalence of osteoporosis was approximately 3–4% for women aged 50 years and over. A similar prevalence rate has been reported in the literature for QUS [5]. This has implications when using QUS in a clinical setting, as very few women would be found to be osteoporotic and therefore recommended for preventive treatment if the current WHO definition of osteoporosis were to be applied uncritically to QUS.

In the present study we calculated *T*-scores using our own population of young adults to estimate the young adult mean and SD. This consistency in using the same young adult population to calculate *T*-scores is advantageous as any differences observed between QUS and

BMD measurements is due to the different technologies and not discrepancies in manufacturer reference ranges. If manufacturer-based reference data were used different proportions of women, as well as different individuals, would have been identified as osteoporotic. Significant discrepancies in diagnosing osteoporosis have been found when using different DXA systems with incompatible reference data [18–20]. In one study, which compared the Hologic femoral neck reference values with the NHANES reference data for the femoral neck, the prevalence of osteoporosis fell from 49% to 28% when the latter reference data were used [21]. This can increase the apparent heterogeneity between anatomic regions and measurement techniques. There are now many different QUS devices in use worldwide, all using different reference data to calculate *T*-scores, so one would expect differences in the way individuals are classified depending on which ultrasound device they were measured on.

In the present study, four different approaches were used to derive a suitable *T*-score threshold for QUS measurements. For all four approaches the threshold for QUS was significantly less negative than the *T*-score threshold of –2.5 recommended for BMD measurements at the spine, hip or forearm, with QUS thresholds ranging from –1.05 to –2.19. The only other study to examine this to date revealed that a *T*-score of –1.5 for BUA and –2.3 for SOS detected 76% of hip fracture patients as well as did DXA [6]. The method used in the latter study by Hans et al. [6] was similar to approach 3 used in this study, although women with prevalent vertebral fractures were used in this case. If the *T*-score thresholds obtained in the present study were compared to those by Hans et al, they were comparable for BUA but the thresholds for SOS were higher than those obtained by Hans et al. [6]. However, the QUS device used by Hans et al. was different to those used in this study and this may indicate that the *T*-score thresholds are device-specific, with different devices having different optimum *T*-score thresholds. In addition to this, BUA measurements tend to have different *T*-score thresholds to velocity measurement parameters (Table 3). Therefore, at present it may not be possible to recommend a single *T*-score threshold which would be appropriate for all QUS devices and all QUS measurement parameters for identifying women at risk of osteoporosis. The diversity of its technology has been identified as a challenge for the advancement of QUS [22]. The International QUS Consensus Group recognizes that the standardization of methods of calibration and expression of measurement results would increase the clinical utility of QUS [22]. It was expected that the four approaches used in the present study would yield different *T*-score thresholds for QUS. The approach used in the present study that was most similar to that used by the WHO Working Party to define a threshold for diagnosing osteoporosis [3] was approach 2, which was based on classifying as osteoporotic a similar percentage of healthy postmenopausal women aged 50+ years as do BMD scans. The *T*-score thresholds obtained using this approach were relatively consistent for the three QUS devices, averaging –1.80, and therefore this *T*-score threshold is probably the optimum for use in identifying postmenopausal women at risk of osteoporosis using the three QUS devices employed in this study. It is important to note that, with any measurement technique, a patient's *T*-score result should be examined in conjunction with clinical risk factors such as history of fracture and age, so that an individuals risk of fracture can be fully assessed.

Using a *T*-score threshold of $\leqslant -1.80$ for QUS, the agreement between the different QUS devices in classifying individuals as osteoporotic was assessed, as well as the agreement between QUS and DXA measurements at the lumbar spine or total hip. The index of positive agreement ($P_{pos}$) was consistent among the three QUS devices, with approximately 72% of the women classified similarly. The agreement between QUS and DXA was less, as expected, with approximately 45% of women classified similarly (Table 3). This value of 45% was identical to that obtained when comparing lumbar spine and total hip BMD measurements. Although QUS and DXA identified different groups of women as osteoporotic, there was overlap between techniques and, as stated by Gluer and Hans [24], as long as the predictive capabilities are of similar magnitude, the two approaches are of equal value.

One of the reasons for the introduction of QUS was to increase the availability of bone densitometry services. It has been estimated that only 25% of Caucasian postmenopausal women in the US have access to bone densitometry services [25]. The introduction and acceptance of QUS and other peripheral devices has given more people the opportunity to have an assessment of their skeletal status. It is important to provide guidelines on how to interpret QUS scans, especially in view of the fact that QUS devices are more likely to be used away from specialist centers. Several recommendations have been published in recent years. Baran et al. [25] recommended that if a *T*-score lies between –2 and 1 for women under 65 years or between –2 and 0 for women over 65 years then further investigation should be performed by DXA at the spine and hip to exclude a false negative. The United Kingdom National Osteoporosis Society recommended that if a patient has a low QUS measurement then they should be referred on for axial bone density measurement [26]. Miller et al. [27] proposed that if an individual has a peripheral bone mass *T*-score of less than –1 they would not need additional central measurements. It is clear from these recommendations that a large number of women will continue to be referred for axial measurements at specialist centers after having a low 'QUS or other peripheral measurement. This may be impractical when we consider that the main reason for the increasing number of peripheral devices in use in recent years has been the lack of availability of resources. Therefore, demand for axial measurements may still exceed supply if individuals with low peripheral measurements are referred for additional testing. Suitable guidelines need to be established so

that QUS can be used effectively in a clinical setting without the need for referring a large proportion of women for additional testing. However, it must be noted that at present there is a lack of agreement on whether QUS can be used to monitor disease progression or treatment efficacy; therefore patients may still need to be monitored using DXA or other established bone densitometry techniques.

There are several limitations to this study. There are many different commercially available QUS devices but only three were assessed. In addition, the UBA is no longer commercially available and the Sahara was developed to match the performance of the UBA. Other QUS devices may yield different optimum *T*-score thresholds for diagnosing osteoporosis, especially those that measure QUS parameters at skeletal sites other than the calcaneus. The *T*-score thresholds have been derived using our own reference young adult population and therefore may not be applicable to the manufacturer-supplied reference data, although our reference data were similar to those used by the manufacturer (Table 1). Although the *T*-scores were calculated using our own reference population, a large range of *T*-score thresholds was obtained using the four approaches described here. Different methods of deriving equivalent thresholds for QUS may yield other optimum thresholds. The majority of the study population consisted of women who had volunteered for bone density studies and women were excluded from analysis if they had any risk factors associated with low BMD; consequently there may be a degree of bias towards a more 'healthy' study population. Large population-based sampling studies do not exclude women because of specific risk factors and so these studies may best address the issue of how the current WHO criteria can be adapted for QUS and other bone densitometry technologies. DXA-equivalent prevalence rates of osteoporosis were used to determine suitable *T*-score thresholds for identifying high-risk individuals using QUS. Other approaches, such as lifetime risk of fracture estimates [28–30], can also be used, although at present these have only been developed for hip fracture and have not been widely adopted in clinical practice. Other approaches for developing equivalent diagnostic criteria for QUS include using prospective fracture studies to determine fracture thresholds. Due to the diversity of QUS technology, which is expected to increase further, additional studies are required to determine suitable *T*-score thresholds for other QUS devices.

In conclusion, this study indicates that the current WHO criteria for the diagnosis of osteoporosis in postmenopausal women cannot be applied to calcaneal QUS measurements. It appears that different QUS devices have different optimum *T*-score thresholds for diagnosing osteoporosis. However, for the three QUS devices used in this study a *T*-score threshold of –1.80 may be appropriate for identifying postmenopausal women at risk of osteoporosis using ultrasound attenuation and velocity measurements at the calcaneus. Additional studies are required to determine suitable *T*-score thresholds for other commercial QUS systems. As the number of QUS devices in use worldwide increases and as bone densitometry services move away from specialist centers into primary care, it is important that a different clinical strategy be implemented for QUS so it can be used with confidence as an alternative diagnostic tool in the field of osteoporosis.

*Acknowledgements.* We are very grateful to the women who participated in this study and we also thank the DXA technicians A. Jefferies and F. Crane.

## References

1. Bauer DC, Gluer CC, Cauley JA, et al. Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women. Arch Intern Med 1997;157:629–34.
2. Hans D, Dargent-Molina P, Schott AM, et al. Ultrasonographic heel measurements to predict hip fracture in elderly women: the EPIDOS prospective study. Lancet 1996;348:511–4.
3. The WHO Study Group. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Technical report series 843. Geneva: WHO. 1994.
4. Melton LJ, Chrischilles EA, Cooper C, et al. How many women have osteoporosis now? J Bone Miner Res 1992;7:1005–10.
5. Faulkner KG, von Stetten E, Steiger P, et al. Discrepancies in osteoporosis prevalence at different skeletal sites: impact on the WHO criteria. Bone 1998;5:s194.
6. Hans D, Schott AM, Dargent-Molina P, et al. Is the WHO criteria applicable to quantitative ultrasound measurement? The EPIDOS prospective study. Bone 1998;5:s286.
7. Cicchetti DV, Feinstein AR. High agreement but low kappa. II. Resolving the paradoxes. J Clin Epidemiol 1990;43:551–8.
8. Looker AC, Johnston CC, Wahner HW, et al. Prevalence of low femoral bone density in older US women from NHANES III. J Bone Miner Res 1995;10:796–802.
9. Steiger P, Cummings SR, Black DM, et al. Age-related decrements in bone mineral density in women over 65. J Bone Miner Res 1992;7:625–32.
10. Vogel JM, Wasnich RD, Ross PD. The clinical relevance of calcaneal bone mineral measurements: a review. Bone Miner 1988;5:35–58.
11. Yamada M, Ito M, Hayashi K, et al. Dual energy x-ray absorptiometry of the calcaneus: comparison with other techniques to assess bone density and value in predicting risk of spine fracture. AJR 1994;163:1435–40.
12. Grampp S, Genant HK, Mathur A, et al. Comparisons of noninvasive bone mineral measurements in assessing age-related loss, fracture discrimination, and diagnostic classification. J Bone Miner Res 1997;12:697–711.
13. Petley GW, Robins PA, Aindow JD. Broadband ultrasonic attenuation: are current techniques inherently inaccurate? Br J Radiol 1995;68:1212–4.
14. Greenspan SL, Bouxsein ML, Melton ME, et al. Precision and discriminatory ability of calcaneal bone assessment technologies. J Bone Miner Res 1997;12:1303–13.
15. Greenspan SL, Maitland-Ramsey L, Myers E. Classification of osteoporosis in the elderly is dependent on site-specific analysis. Calcif Tissue Int 1996;58:409–14.
16. Lai K, Rencken M, Drinkwater BL, et al. Site of bone density measurement may affect therapy decision. Calcif Tissue Int 1993;53:225–8.
17. Arlot ME, Sornay-Rendu E, Garnero P, et al. Apparent pre- and postmenopausal bone loss evaluated by DXA at different skeletal sites in women: the OFELY cohort. J Bone Miner Res 1997; 12:683–90.
18. Ahmed AIH, Blake GM, Rymer JM, et al. Screening for osteopenia and osteoporosis: do the accepted normal ranges lead to overdiagnosis? Osteoporos Int 1997;7:432–8.

19. Faulkner KG, Roberts LA, McClung MR. Discrepancies in normative data between Lunar and Hologic DXA systems. Osteoporos Int 1996;6:432–6.
20. Patel R, Blake GM, Jefferies A, et al. A comparison of a peripheral DXA system with conventional densitometry of the spine and femur. Clin Densitometry 1998;1:235–44.
21. Chen Z, Maricic M, Lund P, et al. How the new Hologic hip normal reference values affect the densitometric diagnosis of osteoporosis. Osteoporos Int 1998;8:423–7.
22. Gluer CC. Quantitative ultrasound techniques for the assessment of osteoporosis: expert agreement on current status. J Bone Miner Res 1997;12:1280–8.
23. Melton LJ. How many women have osteoporosis now? J Bone Miner Res 1995;10:175–7.
24. Gluer CC, Hans D. How to use ultrasound for risk assessment: a need for defining strategies. Osteoporos Int 1999;9:193–5.
25. Baran DT, Faulkner KG, Genant HK, et al. Diagnosis and management of osteoporosis: guidelines for the utilization of bone densitometry. Calcif Tissue Int 1997;61:433–40.
26. National Osteoporosis Society. The use of quantitative ultrasound in the management of osteoporosis in primary or secondary care. NOS, 1998.
27. Miller PD, Bonnick SL, Johnston C, et al. The challenges of peripheral bone density testing: which patients need additional central density skeletal measurements? Clin Densitometry 1998;1:211–7.
28. Black D, Cummings SR, Melton LJ. Appendicular bone mineral and a woman's lifetime risk of hip fracture. J Bone Miner Res 1992;7:639–46.
29. Melton LJ, Kan SH, Wahner HW, et al. Lifetime fracture risk and approach to hip fracture risk assessment based on bone mineral density and age. J Clin Epidemiol 1988;41:985–94.
30. Suman VJ, Atkinson EJ, Wm OF, et al. A nomogram for predicting lifetime hip fracture risk from radius bone mineral density and age. Bone 1993;14:843–6.

*Received for publication 25 June 1999*
*Accepted in revised form 29 September 1999*