Osteoporos Int (2005) 16: 1565–1575
DOI 10.1007/s00198-005-1864-x

ORIGINAL ARTICLE

Comparison of questionnaire and quantitative ultrasound techniques as screening tools for DXA

R.B. Cook · D. Collins · J. Tucker · P. Zioupos

Received: 3 September 2004 / Accepted: 31 January 2005 / Published online: 10 May 2005

Abstract The aim of the study is to assess the sensitivity and specificity of different techniques and their ability to act as screening tools in relation to dual energy X-ray absorptiometry (DXA) in a group of 208 postmenopausal women. In this study we examined eight screening systems for the diagnosis of osteoporosis, the osteoporosis self-assessment tool (OST), the osteoporosis risk assessment instrument (ORAI), the osteoporosis index of risk (OSIRIS), a risk index derived using data from the study of osteoporotic fractures (SOFSURF), the simple calculated osteoporosis risk estimation (SCORE), patient body weight (pBW), along with two ultrasound based systems, the Sunlight Omnisense (Sunlight Medical, Rehovot, Israel) and the CUBA Clinical (McCue plc, Winchester, UK). The sensitivity and specificity of the different techniques in relation to DXA were plotted as receiver-operating characteristic (ROC) curves at three different levels (DXA T-score −2.5 osteoporosis, −2 and −1 osteopenia). The areas under the curves (AUC) were calculated and showed broadband ultrasound attenuation (BUA) at the calcaneus to provide consistently the highest AUC (0.77–0.81). The velocity of sound (VOS) of the calcaneus (AUC = 0.72–0.76) was equally good, but was out-performed by some of the questionnaire systems (AUC = 0.66–0.79). Both the questionnaire systems and the CUBA Clinical out-perform the Sunlight Omnisense (AUC = 0.58–0.7), which showed comparable performance with body weight (AUC = 0.66–0.69). The results show that QUS is capable of selecting patients with low bone density as measured by DXA. A patient displaying a low QUS value should be followed up with a DXA scan to confirm the diagnosis.

Keywords DXA · Osteoporosis · Screening · Sensitivity · Specificity · Ultrasound scanners

R.B. Cook · P. Zioupos (✉)
Department of Materials and Medical Sciences,
Cranfield University Postgraduate Medical School,
Shrivenham, SN6 8LA, UK
E-mail: p.zioupos@cranfield.ac.uk
Tel.: +44-1793-785932
Fax: +44-1793-783076

D. Collins
Department of Rheumatology, Great Western Hospital,
Swindon, UK

J. Tucker
Department of Radiology, Great Western Hospital,
Swindon, UK

Introduction

With 1 in 3 women and 1 in 12 men over the age of 50 suffering from osteoporosis, and the resultant estimated cost to the NHS and the government being £1.7 billion/annum (UK figures), the diagnosis and management of osteoporosis is a very important issue. The 'gold-standard' method for osteoporosis diagnosis is by dual-energy X-ray absorptiometry (DXA), which relies on the measurement of bone mineral density (BMD) at various sites, notably the lumbar spine, hip or distal radius. DXA, however, is expensive, it uses ionizing radiation (albeit at safe levels), requires the supervision of a qualified radiographer and is, therefore, best suited for the precise/accurate diagnosis of osteoporosis, rather than for the screening of large populations. Large numbers of potential patients can be reached by questionnaire-based methods or the use of portable quantitative ultrasound (QUS) scanners.

A number of questionnaire-based systems have been devised in an attempt to produce a cost-effective method of screening for osteoporosis. Questionnaires focus on clinical risk factors for osteoporosis and use a varying number of them to produce quantitative scores. The scores are designed to give information on those patients at risk of having low bone mineral density, who need to undergo a full assessment of their bone status. Examples of previously examined questionnaires are the Osteoporosis Self-assessment Tool (OST, OSTA) [1, 2, 3, 4, 5, 6, 7], Osteoporosis Risk Assessment Instrument (ORAI)

[3, 8], Simple Calculated Osteoporosis Risk Estimation (SCORE) [3, 9, 10, 11, 12, 13], Osteoporosis Index of Risk (OSIRIS) [14, 15], the risk index derived using data from the study of osteoporotic fractures (SOFSURF) [3, 16], the Osteoporosis Assessment Questionnaire (OPAQ) [17] and three further systems, one based around nutritional assessment [18], another based on life-style factors [19] and the third based purely on the patient's body weight (pBW) [20].

Alternative approaches for mass screening advocate the use of quantitative ultrasound (QUS) systems. QUS has been utilized for the measurement of bone and provides information that is not only comparable to, but also independent of DXA. QUS has the benefits of reduced costs in comparison to DXA, is a portable and radiation-free system with shorter investigation times than DXA and is, therefore, a better proposition for performing screening of large populations.

In this study we intend to examine comparatively and assess six questionnaire-based screening systems (OST, ORAI, OSIRIS, SOFSURF, SCORE and pBW) and two ultrasound-based systems, the Sunlight Omnisense (Sunlight Medical, Rehovot, Israel) and the CUBA Clinical (McCue plc, Winchester, UK). The aim of the study is to assess the abilities of these alternative methods/systems as screening tools for DXA and to investigate the cut-off levels for the various techniques.

Materials and methods

The study group consisted of 208 women, considered to be postmenopausal through natural or unnatural causes, recruited with consent from the DXA scanning clinics at the Great Western Hospital, Swindon, UK. All subjects attended the clinic because of referral from GPs or hospital-based clinics as a consequence of the presence of one or more clinical risk factors for osteoporosis. There were no general exclusion criteria for the study. Table 1 shows the anthropometrical data for the study group.

Each patient received a DXA scan of their lumbar spine L1–L4, and four sites around the proximal femur using a Hologic QDR-4500C (Hologic Inc., Bedford, Mass.), answered a questionnaire designed to give information on the patient's nutritional, life-style and historical risk factors for osteoporosis, received paired scans on their distal radius, the proximal phalanx of the middle finger, and mid-shaft tibia using the Sunlight Omnisense and paired measurements of the calcaneus using the CUBA Clinical. The QUS machines underwent system quality verification tests each day prior to any measurements being performed using the phantoms supplied by the manufacturers. All ultrasound scans were performed on the patient's non-dominant side, using Parker ultrasound gel (Parker Laboratories Inc., Fairfield, N.J.) to provide the coupling between the ultrasound probes and skin surface.

Table 1 Anthropometric data for the subjects ($n = 208$) of the present study

	Range	Average
Age	29–87 years	59.7 years
Height	137–182 cm	161.1 cm
Weight	41.3–104.8 kg	65.6 kg
BMI	15.7–43	25.4
Years since menopause	0–54 years	15.4 years

The precision of the QUS machines was previously examined in a group of 16 normals (aged 25–58). Four repeated measurements were performed on each individual, with repositioning between scans, and gave a RMSCV of 0.29% for the distal radius, 0.55% for the proximal phalanx, 2.88% for BUA of the calcaneus and 0.31% for VOS of the calcaneus.

Calculation of risk indices

The questionnaire systems were calculated as follows: OST [1, 2, 3, 4, 5, 6, 7] is based on only two variables, body weight and age. Two very similar methods have been suggested for its calculation, both giving virtually identical results. The first method is calculated as: 0.2 [weight (kg)]–0.2 [age (years)]; the last digit is dropped from each to give an integer, and the resulting values added together. The second method is calculated as (weight in kilograms – age in years) ×0.2 and truncated to give an integer. The risk index for OST varies between populations and the site being assessed. For a Caucasian population [1], values $> +2$ signify low risk, $< +2$ to > -3 indicate moderate risk and < -3 denote high risk of low BMD.

ORAI [3, 8] is based on three variables, age, weight and estrogen usage. The ORAI index is calculated using the following scoring system: +2 points for non-current usage of estrogen; +9 points for a body weight of less than 60 kg or +3 points for a body weight between 60 and 70 kg and 0 points for weight above 70 kg; +15 points for ages 75 years or more; +9 points for ages between 65 and 74 years; +5 points for ages between 55 and 64; 0 points for 45 and 54. The risk index for ORAI stipulates that results < 9 are low risk of low BMD, between > 9 and < 17 is moderate risk and > 17 denotes high risk of low BMD.

OSIRIS [14, 15] is based on four variables, age, weight, HRT usage and history of low trauma fracture. The index is calculated as current age times –2 and truncated to an integer, weight in kg times 2 and truncated to an integer, +2 points if a current HRT user and –2 points for a history of prior low impact fracture. The risk index for OSIRIS shows that values $> +1$ indicate low risk of low BMD, $< +1$ and > -3 are the intermediate risk category, and values < -3 indicate a high risk of low BMD.

SOFSURF [3, 16] is also based on four variables, age, weight, smoking and history of postmenopausal

fracture. The index is calculated as +0.2 points for every year over 65 years of age and –0.2 points for every year under 65 years of age; +3 points for weight below 130 lb (59 kg) and +1 point for weight between 130 lb (59 kg) and 150 lb (68 kg); +1 point for a current smoker and +1 point for a history of post-menopausal fracture. The risk index for SOFSURF shows that values <0 indicate a low risk of low BMD; >0 and $<+4$ indicate an intermediate risk of low BMD, with values $>+4$ denoting those at high risk of low BMD.

SCORE [3, 9, 10, 11, 12, 13] uses six variables, race, rheumatoid arthritis, history of non-traumatic fracture, HRT usage, age and weight. The index is calculated as +5 points for a race other than Black, +4 points for rheumatoid arthritis sufferers, +4 points for non-traumatic fractures (wrist, hip and rib) over the age of 45, up to a total of 12 points; +1 if never used HRT, 3 times the first digit of the patients age, and –1 times the patients weight in pounds divided by ten and truncated to an integer. The risk index for SCORE based on a Caucasian population uses values $<+7$ to indicate low risk, $>+7$ and $<+15$ to show moderate risk and $>+15$ to denote high risk of low BMD.

pBW [20] relies on one variable, weight (kg). The recommended cut-off value is 70 kg, with women below 70 kg in weight being considered to be at risk of osteoporosis. The risk index states that pBW >70 kg indicates a low risk, between 57 kg and 70 kg a moderate risk and below 57 kg a high risk of the patient having low BMD.

Correlations

The correlations between the measurement sites were calculated using Minitab Statistical Software Release v.13.31.

Sensitivity and specificity calculation

Previous papers [21, 22, 23, 24] have reported the correct method for calculation of factors that allow for the comparison of the diagnostic ability of different techniques, in particular the sensitivity and specificity of a technique. The calculations are based around a 2×2 table as shown in Table 2. The sensitivity is a measure of how good a test is at picking up people who have the condition. The specificity is a measure of how well a test correctly excludes people without the condition. The positive and negative predictive values are measures of probability, with the positive referring to the probability of the individual having the condition if they test positive, and the negative to the probability of an individual not having a condition, should they test negative.

Three levels were assessed using the DXA T-scores: a T-score of −2.5, which denotes confirmed osteoporosis, a T-score of −2 and a T-score of −1, below which patients are considered to be osteopenic, but including the osteoporotic patients. The DXA results for the lumbar spine and the total hip were combined, so that a patient showing a T-score of –2.5 or below at only one of the two sites was classified osteoporotic, even if the other site was diagnosed as normal. A population-specific database was not used for calculation of the T-scores; all T-scores were computed using the databases supplied with the systems.

Different levels were then assessed using the diagnostic tools (questionnaires and QUS) to gain a range of sensitivity values with a corresponding specificity value. Using SigmaPlot version 8.0, the results were plotted as sensitivity vs. 1-specificity. This supplied a range of receiver operator characteristic curves (ROC), from which the area under the curve (AUC) could then be calculated.

Cut-off points

The purpose of a screening tool is to select correctly those patients that have, or are at risk of having, low BMD and to exclude those patients who are subsequently found to have normal BMD levels. The optimum screening tool would provide a cut-off point that could be used to provide the correct diagnosis of every individual's bone status and provide no false positives or false negatives. It is therefore important that a point be selected above which patients are considered to be normal, and below which they are deemed to need a further investigation. Previous studies performing validation [2, 8] of screening tools have used cut-offs, which supply a sensitivity of 90%, regardless of the specificity, to ensure that the percentage of patients with low BMD correctly selected is high. In addition to this method of cut-off selection, the best balance between the sensitivity and specificity was investigated. By combining the sensitivity

Table 2 Combination table of outcomes of alternative diagnostic test results with respect to the DXA outcome

Diagnostic tool test result	DXA test result		
	Positive	Negative	
Positive	True positive (TP)	False positive (FP)	TP + FP
Negative	False negative (FN)	True negative (TN)	FN + TN
	TP + FN	FP + TN	TP + FN + FP + TN

Sensitivity: TP/(TP + FN). Specificity: TN/(FP + TN); +ve predictive value: TP/(TP + FP); –ve predictive value: TN/(FN + TN)

and specificity scores, a value was supplied that varied between cut-off levels, and the highest combined value was used as the cut-off level.

Results

Of the 208 women included in this study, the T-scores from DXA scans classified 21.6% (45) of the patients as being osteoporotic at the lumbar spine or hip, 47.6% (99) as being osteopenic and 30.8% (64) of the patients as being of normal bone density.

Table 3 shows the correlations between the different questionnaires, QUS techniques and DXA. The CUBA Clinical displayed the best correlation, in relation to DXA at the lumbar spine ($r = 0.568$), with OSIRIS correlating best with the total hip ($r = 0.658$). The Sunlight Omnisense correlated poorly with the DXA results ($r = 0.127$–0.340), having correlations below those for the two questionnaire systems ($r = -0.417$–0.658), also below the two CUBA Clinical measures: BUA and VOS ($r = 0.473$–0.650) and only comparable to pBW ($r = 0.33$–0.492). The mid-shaft tibia result was non-significant in five out of nine cases and marginal in one of the rest. This is neither surprising nor disappointing as this index/measurement method has not yet obtained FDA approval and is not recommended for clinical use. In all but a few cases the correlations showed a high level of statistical significance, with *P*-values generally ≤ 0.001.

Analysis of the AUC for the ROC curves as shown in Fig. 1 allows for comparisons between techniques, with the larger AUC value indicating a greater clinical utility. The results in Table 4 show BUA at the calcaneus to provide consistently the highest AUC (0.77–0.81), although the VOS of the calcaneus (AUC = 0.72–0.76) is outperformed by some of the questionnaire systems (AUC = 0.66–0.79). Both questionnaire systems and the CUBA Clinical outperform the Sunlight Omnisense (AUC = 0.58–0.7), which only showed comparable performance with pBW (AUC = 0.65–0.69).

Tables 5 and 6 show the potential cut-off values for the different techniques and their associated sensitivity, specificity, positive and negative predictive values, and percentage of incorrectly diagnosed patients. The sensitivity and specificity method incorrectly diagnoses a smaller percentage of the overall group in comparison to the 90% sensitivity method. The positive and negative predictive values provide important information on the techniques' abilities, with techniques providing probabilities of over 0.9 for both exclusion of patients because of negative test results and correct diagnosis of sufferers because of positive test results, depending on the DXA cut-off level assessed.

An effort was also made to explore the potential of using these QUS systems and questionnaires in combination. We performed stepwise regression analysis to predict the minimum of the two T-scores for hip and spine (worst case scenario) by using the raw data output

Table 3 Correlation coefficients of the various T-score results of the various physical (DXA, QUS) and empirical (questionnaire) methods

		Weight	SOFSURF	SCORE	OST	OSIRIS	ORAI	Distal radius	Proximal phalanx	Mid-shaft tibia	BUA calcaneus	VOS calcaneus	BMD L1-L4
Questionnaires	SOFSURF	–0.460, 0.000											
	SCORE	–0.536, 0.000	0.897, 0.000										
	OST	0.774, 0.000	–0.883, 0.000	–0.878, 0.000									
	OSIRIS	0.666, 0.000	–0.904, 0.000	–0.927, 0.000	0.945, 0.000								
	ORAI	–0.469, 0.000	0.880, 0.000	0.778, 0.000	–0.819, 0.000	–0.839, 0.000							
Sunlight omnisense	Distal radius	0.112, 0.113	–0.474, 0.000	–0.453, 0.000	0.400, 0.000	0.457, 0.000	–0.402, 0.000						
	Proximal phalanx	0.220, 0.001	–0.574, 0.000	–0.558, 0.000	0.512, 0.000	0.543, 0.000	–0.479, 0.000	0.485, 0.000					
	Mid-shaft tibia	–0.087, 0.219	–0.143, 0.042	–0.157, 0.025	0.092, 0.191	0.136, 0.053	–0.132, 0.060	0.262, 0.000	0.183, 0.009				
CUBA Clinical	BUA calcaneus	0.233, 0.001	–0.590, 0.000	–0.560, 0.000	0.515, 0.000	0.594, 0.000	–0.507, 0.000	0.362, 0.000	0.383, 0.000	0.244, 0.000			
	VOS calcaneus	–0.028, 0.689	–0.438, 0.000	–0.391, 0.000	0.299, 0.000	0.379, 0.000	–0.367, 0.000	0.312, 0.000	0.323, 0.000	0.294, 0.000	0.792, 0.000		
Hologic	BMD L1-L4	0.330, 0.000	–0.463, 0.000	–0.460, 0.000	0.451, 0.000	0.508, 0.000	–0.417, 0.000	0.295, 0.000	0.318, 0.000	0.225, 0.001	0.568, 0.000	0.473, 0.000	
QDR-4500C	BMD total hip	0.492, 0.000	–0.599, 0.000	–0.598, 0.000	0.633, 0.000	0.658, 0.000	–0.558, 0.000	0.275, 0.000	0.340, 0.000	0.127, 0.070	0.650, 0.000	0.519, 0.000	0.717, 0.000

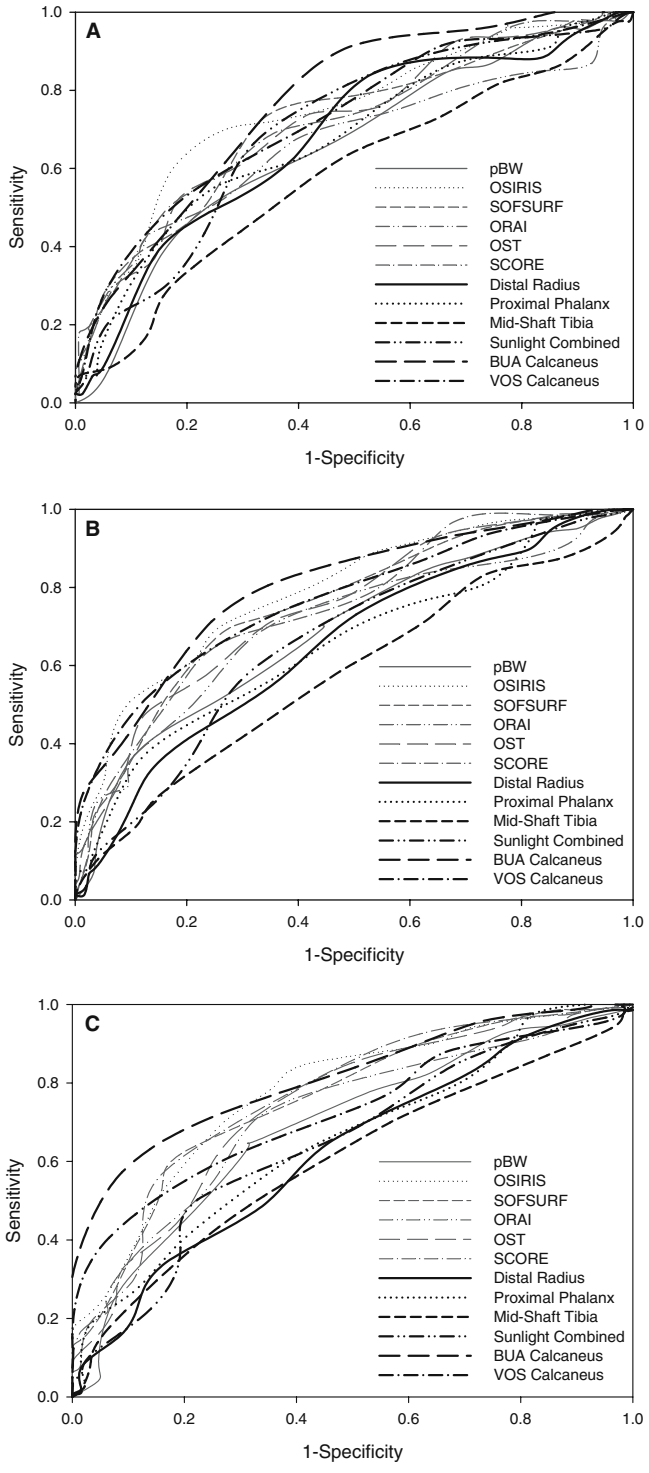

Fig. 1 ROC curves for the questionnaire and QUS systems in comparison to combined T hip and L1-L4 spine DXA. **A** Prediction of combined DXA T-score –2.5. **B** Prediction of combined DXA T-score –2. **C** Prediction of combined DXA T-score –1

of the QUS systems and questionnaires. We considered three likely situations: (1) clinicians who possess no instruments and would be able to only apply questionnaires (with knowledge of weight and age of course), (2) a case where one possesses the instruments, but chooses not to use the questionnaires and (3) the case where both QUS scanners and questionnaires are available for full use by the clinician. The results of the best stepwise regression are shown in Table 7. In this analysis, the use of scanners was shown to be more effective than the use of questionnaires. Taking account of questionnaire values in addition to QUS scanner values only improved the R^2 of the relationship marginally from $R^2 = 44.6$ to 46.8. Fig. 2 depicts graphically the result for the cases of predictive equations 1 and 10 (Table 7).

Discussion

This study comparatively examined and assessed six questionnaire-based screening systems (OST, ORAI, OSIRIS, SOFSURF, SCORE and pBW) and two ultrasound based systems (Sunlight Omnisense, CUBA Clinical) as potential screening tools for DXA. The study assessed the abilities of these alternative methods/systems and investigated the cut-off levels for the various techniques. The aim was not necessarily to replace DXA, but to explore various strategies and approaches by which the demand on DXA services could be reduced by, for instance, screening large sections of the population for the exclusion of individuals who upon DXA examination would have shown themselves as normal.

The decision to investigate the systems in comparison to low DXA, rather than fracture risk, was due to the nature of the DXA clinics from which the subjects were recruited. The DXA clinic aimed to supply a T-score value that would enable a clinician to assess potential courses of medical treatment or the cessation of currently undertaken treatments. Our approach reflected current recommendation standards for clinical practice as outlined in the official positions of the International Society of Clinical Densitometry [25]. The alternative method would be to supply a parameter related to risk of fracture or the probability of fracture. The outputs are essentially only numerically different, the basic essence of the course of action or the implications for having, or being likely to have, a future fracture are basically very similar.

One factor that may influence the robustness of the conclusions and results is the make up of the cohort of subjects that participates in a study. Our volunteers showed that they were either osteoporotic or osteopenic (judged by DXA) at 69% of the total. This percentage was higher than would be found if a database like cross section of the population had been recruited. Our intention, however, was not to seek specifically normals, but to apply our methodology and approach to the most relevant cross section of subjects. Hence, volunteers were recruited from a DXA scanning clinic, where all patients were referred because of the presence of at least one risk factor.

Only one previous study [5] has provided any correlation between a questionnaire system and DXA, and

Table 4 Area under the curve values (*AUC*) of the various diagnostic methods in order to predict the DXA outcome at *t* -score levels between −2.5 to −1. The AUC allows for a direct single value comparison of the different techniques' clinical accuracy for the prediction of DXA. The provided 95% CI also allow for a fast visual inspection for statistical differences

Investigation	AUC DXA T-score –2.5	95% CI	AUC DXA T-score −2	95% CI	AUC DXA T-score −1	95% CI
OSIRIS	0.747	0.805–0.702	0.788	0.813–0.788	0.773	0.758–0.800
SOFSURF	0.717	0.777–0.670	0.754	0.773–0.746	0.756	0.739–0.785
ORAI	0.664	0.739–0.595	0.703	0.733–0.679	0.718	0.704–0.742
OST	0.716	0.775–0.669	0.746	0.767–0.737	0.740	0.717–0.775
SCORE	0.720	0.779–0.674	0.751	0.772–0.743	0.764	0.741–0.796
Distal radius	0.676	0.731–0.628	0.651	0.670–0.640	0.624	0.593–0.659
Proximal phalanx	0.678	0.737–0.629	0.657	0.685–0.639	0.652	0.630–0.685
Mid-shaft tibia	0.582	0.645–0.521	0.578	0.606–0.553	0.600	0.580–0.622
Sunlight combined	0.698	0.751–0.654	0.661	0.679–0.650	0.643	0.614–0.678
BUA calcaneus	0.766	0.805–0.743	0.794	0.825–0.776	0.808	0.814–0.817
VOS calcaneus	0.723	0.781–0.676	0.760	0.793–0.739	0.717	0.720–0.724
pBW	0.655	0.708–0.607	0.684	0.707–0.667	0.687	0.660–0.717

this showed moderate correlation coefficients between OST and BMD of the femoral neck and lumbar spine of 0.62 and 0.49, respectively ($P < 0.0001$). In the present study the correlation between OST and BMD was closely comparable to this with 0.633 and 0.451 for total hip and lumbar spine, respectively.

Overall, there were excellent ($r = 0.95$) to moderate ($r = 0.46$) correlations seen between the various questionnaires. The variability may, to a certain degree, have been due to the method of questionnaire data collection whereby questionnaires relied on patient's self-reported data. Bearing in mind that the various questionnaire systems follow different designs and philosophies of approach (focusing on different risk factors) and that their implementation also differs considerably may explain the variability of performance between them. In correlation to DXA, questionnaires performed less effectively ($r = 0.417$–0.658), but this is to no extent disappointing, and if anything considering that the population had a bias to the lower BMD end of the population and that the questionnaires were not designed to be replacement measurements for BMD but an indicator of a patient's bone status, it is very encouraging.

The results for the CUBA Clinical and Sunlight Omnisense within this study showed close comparison with those of the previous studies. Correlations between measurement sites using the ultrasound are affected by the physical principle of the ultrasound application and the kind of bone matrix that is scanned (cortical or cancellous). Hence the CUBA Clinical, which assesses a load-bearing cancellous bone site that matches the make up and loading of the hip and spine, correlates better with the axial skeleton. The correlations from this study of ($r = 0.473$–0.650) for BUA and VOS were in close agreement with previous studies ($r = 0.20$–0.64) [26, 27, 28, 29, 30, 31], being higher than the correlations for the Sunlight Omnisense in both this study ($r = 0.127$–0.340) and previous studies [32, 33] ($r = 0.21$–0.48). The moderate correlations for the CUBA Clinical can be further explained; the attenuation and velocity of an ultrasound pulse will be affected by the structure of the material it is passing through [34], with strong and complex trabecular structure affecting the ultrasound in different ways in fragile and broken trabeculae, a factor not taken into consideration by the measurement of density alone.

Calculation of the AUC showed BUA of the calcaneus to have the highest value for each DXA cut-off level. The questionnaire systems and VOS of the calcaneus obtained the next highest values with the Sunlight Omnisense, and pBW had the lowest values. The AUC values for the questionnaire systems ranged from 0.66–0.75 for osteoporosis (T-score –2.5), within the range (0.594–0.87) of previous studies [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. At a DXA T-score of –2, the results from this study were lower than previous studies.

The AUC results showed that the CUBA Clinical, measuring both BUA and VOS, demonstrated a better diagnostic ability in comparison to the Sunlight Omnisense. Previous studies [30, 35] looking at the AUC for the CUBA Clinical prediction of osteoporosis at the hip and spine achieved AUC of 0.72–0.856 and 0.75–0.816 for BUA, and 0.68–0.871 and 0.72–0.820 for VOS, respectively. The results from the present study are in agreement with these previous results for the CUBA Clinical, and also with other calcaneal devices, which ranged from 0.64–0.888 for BUA and 0.68–0.871 for VOS [35, 36, 37, 38, 39], depending on the measurement site and device.

J. Damilakis et al. [40] reported AUC for the phalanx of 0.709, which was slightly higher than achieved here, and for the radius 0.659, which was slightly lower than in the present study. However, both results were within the same range.

The AUC for the different techniques also supplied information on the diagnostic accuracy of the different technique. J. Swets [24] supplied different areas in terms of their diagnostic accuracy with AUC between 0.50–0.70 representing low accuracy, between 0.70–0.90 showing moderate diagnostic accuracy and any technique showing an AUC of greater the 0.90 considered to

Table 5 The suggested cut-off points that allow for the best sensitivity and specificity balance

Site	Combined DXA cut-off level	Sensitivity + specificity cut-off value	Sensitivity	Specificity	+ve predictive value	–ve predictive value	% of group (total FN + FP)
OSIRIS	–2.5	0	0.7	0.73	0.42	0.89	27.9% (58)
SOFSURF	–2.5	1	0.72	0.67	0.38	0.89	32.2% (67)
ORAI	–2.5	14	0.43	0.86	0.48	0.84	23.1% (48)
OST	–2.5	–1	0.52	0.82	0.44	0.56	25% (52)
SCORE	–2.5	12	0.5	0.83	0.46	0.85	24% (50)
Distal radius	–2.5	–0.5	0.84	0.47	0.30	0.91	43.3% (90)
Proximal phalanx	–2.5	–0.5	0.56	0.75	0.38	0.86	29.3% (61)
Mid-shaft tibia	–2.5	–0.5	0.64	0.51	0.26	0.84	45.2% (94)
Sun combined	–2.5	–1.5	0.69	0.66	0.35	0.89	32.2% (67)
BUA calcaneus	–2.5	–1.5	0.91	0.51	0.34	0.95	40.4% (84)
VOS calcaneus	–2.5	–3.5	0.53	0.8	0.42	0.86	26% (54)
Weight	–2.5	60kg	0.56	0.7	0.34	0.85	33.2% (69)
OSIRIS	–2	1	0.7	0.74	0.63	0.80	27.4% (57)
SOFSURF	–2	1	0.68	0.74	0.62	0.79	28.4% (59)
ORAI	–2	10	0.68	0.68	0.57	0.77	32.2% (67)
OST	–2	1	0.71	0.65	0.56	0.79	32.7% (68)
SCORE	–2	8	0.68	0.71	0.59	0.78	30.3% (63)
Distal radius	–2	–0.5	0.73	0.5	0.48	0.74	39.9% (83)
Proximal phalanx	–2	–1	0.37	0.88	0.64	0.70	31.3% (65)
Mid-shaft tibia	–2	–1.5	0.27	0.85	0.53	0.65	36.5% (76)
Sun combined	–2	–1.5	0.59	0.68	0.53	0.73	33.2% (69)
BUA calcaneus	–2	–2	0.71	0.76	0.61	0.83	26% (54)
VOS calcaneus	–2	–3.5	0.51	0.88	0.72	0.74	26.4% (55)
Weight	–2	55kg	0.36	0.9	0.69	0.69	30.8% (64)
OSIRIS	–1	2	0.7	0.73	0.86	0.52	28.8% (60)
SOFSURF	–1	0	0.63	0.8	0.87	0.49	32.2% (67)
ORAI	–1	8	0.71	0.67	0.83	0.51	30.3% (63)
OST	–1	2	0.72	0.67	0.83	0.51	29.8% (62)
SCORE	–1	8	0.56	0.84	0.98	0.46	35.1% (73)
Distal radius	–1	–0.5	0.65	0.54	0.77	0.39	37% (77)
Proximal phalanx	–1	0	0.51	0.7	0.79	0.39	42.3% (88)
Mid-shaft tibia	–1	–1.5	0.25	0.93	0.90	0.35	52.9% (110)
Sun combined	–1	–1.5	0.51	0.77	0.84	0.39	38.9% (81)
BUA calcaneus	–1	–2	0.56	0.92	0.94	0.48	33.2% (69)
VOS calcaneus	–1	–3	0.61	0.72	0.85	0.45	35.6% (74)
Weight	–1	65 kg	0.71	0.55	0.82	0.46	34.1% (71)

Table 6 The suggested cut-off points that allow for a guaranteed 90% sensitivity level

Site	Combined DXA cut-off level	90% sensitivity cut-off	Sensitivity	Specificity	+ve predictive value	–ve predictive value	% of group (total FN + FP)
OSIRIS	–2.5	5	0.96	0.22	0.26	0.95	61.5% (128)
SOFSURF	–2.5	–2	0.91	0.22	0.25	0.9	63% (131)
ORAI	–2.5	0	0.96	0.056	0.22	0.82	74.5% (155)
OST	–2.5	3	0.91	0.33	0.28	0.93	53.8% (112)
SCORE	–2.5	4	0.93	0.26	0.26	0.93	59.1% (123)
Distal radius	–2.5	1	0.93	0.12	0.22	0.86	68.3% (142)
Proximal phalanx	–2.5	2	0.91	0.15	0.23	0.86	67.8% (141)
Mid-shaft tibia	–2.5	1.25	0.93	0.1	0.22	0.84	70.2% (146)
Sun combined	–2.5	–0.5	0.93	0.24	0.25	0.92	57.7% (120)
BUA calcaneus	–2.5	–1.5	0.91	0.51	0.34	0.95	40.4% (84)
VOS calcaneus	–2.5	–2.5	0.91	0.35	0.28	0.93	52.9% (110)
Weight	–2.5	80kg	0.96	0.15	0.24	0.92	67.8% (141)
OSIRIS	–2	4	0.91	0.41	0.49	0.88	39.9% (83)
SOFSURF	–2	–2	0.95	0.27	0.46	0.9	46.6% (97)
ORAI	–2	2	0.925	0.1	0.39	0.68	58.2% (121)
OST	–2	3	0.9	0.39	0.48	0.86	41.3% (86)
SCORE	–2	4	0.96	0.33	0.47	0.93	42.8% (89)
Distal radius	–2	0.5	0.9	0.19	0.41	0.74	51.9% (108)
Proximal phalanx	–2	2	0.94	0.18	0.41	0.82	52.9% (110)
Mid-shaft tibia	–2	1.25	0.92	0.1	0.39	0.68	56.7% (118)
Sun combined	–2	0	0.97	0.091	0.4	0.85	53.8% (112)
BUA calcaneus	–2	–1	0.93	0.33	0.43	0.90	33.2% (96)
VOS calcaneus	–2	–2	0.94	0.26	0.44	0.87	48.1% (100)
Weight	–2	80kg	0.94	0.16	0.41	0.81	53.8% (112)
OSIRIS	–1	5	0.9	0.36	0.76	0.61	26.9% (56)
SOFSURF	–1	–2	0.9	0.38	0.76	0.62	26.4% (55)
ORAI	–1	2	0.93	0.14	0.71	0.47	31.3% (65)
OST	–1	5	0.94	0.25	0.74	0.64	27.4% (57)
SCORE	–1	3	0.91	0.36	0.76	0.64	26% (54)
Distal radius	–1	1	0.92	0.19	0.73	0.5	28.4% (59)
Proximal phalanx	–1	2	0.9	0.22	0.72	0.5	30.8% (64)
Mid-shaft tibia	–1	1.5	0.94	0.033	0.69	0.2	31.7% (66)
Sun combined	–1	0	0.95	0.11	0.72	0.46	27.9% (58)
BUA calcaneus	–1	–0.5	0.96	0.27	0.75	0.74	25.5% (53)
VOS calcaneus	–1	–1.5	0.95	0.063	0.695	0.36	32.2% (67)
Weight	–1	80 kg	0.92	0.23	0.73	0.58	28.8% (60)

Table 7 Stepwise regression equations for three scenarios, where the introduced parameters are shown in column 1, the best equations produced in sequence are in column 2 and the r^2 values in column 3. Note that equations 5–7 are identical to equations 2–4 even in the presence of the questionnaires as additional factors

Parameters	Equation	r^2	Eq. no.
Weight, OSIRIS, OST, SOFSURF, SCORE, ORAI	Min DXA T-score = 0.178 OSIRIS –1.915	31.0	1
Weight, age, distal radius, proximal phalanx, BUA Calcaneus, VOS calcaneus	Min DXA T-score = 0.0472 BUA –4.471	37.7	2
	Min DXA T-score = 0.0431 BUA + 0.0249 weight (kg) –5.841	43.0	3
	Min DXA T-score = 0.0259 BUA + 0.0308 weight (kg) + 0.0085 VOS –18.505	44.6	4
Weight, OSIRIS, OST, SOFSURF, SCORE, ORAI, distal radius, proximal phalanx, BUA calcaneus, VOS calcaneus	Min DXA T-score = 0.0472 BUA –4.471	37.7	5
	Min DXA T-score = 0.0431 BUA + 0.0249 weight (kg) –5.841	43.0	6
	Min DXA T-score = 0.0259 BUA + 0.0308 weight (kg) + 0.0085 VOS –18.505	44.5	7
	Min DXA T-score = 0.0215 BUA + 0.0220 weight (kg) + 0.0080 VOS + 0.047 OSIRIS –17.001	45.3	8
	Min DXA T-score = 0.0208 BUA + 0.0304 weight (kg) + 0.0085 VOS + 0.120 OSIRIS –0.112 OST –18.287	46.0	9
	Min DXA T-score = 0.0198 BUA + 0.0460 weight (kg) + 0.0087 VOS + 0.088 OSIRIS –0.220 OST –0.144 SOFSURF –19.289	46.8	10

Fig. 2 Predictions of the minimum hip or spine DXA T-score by using equation 1 (*open symbols* and *dotted lines*) and equation 10 (*solid symbols* and *lines*) of Table 7. The regression lines with their 95% prediction intervals are shown in both cases. The 0, –1, and –2.5 score levels are also highlighted

be of high accuracy. The AUC results for the different techniques in this study showed the majority of the methods being considered to have moderate diagnostic accuracy, with the measurements from the Sunlight Omnisense, ORAI and pBW showing low diagnostic accuracy.

The two different methods for the selection of cut-off values both have utility within the clinical environment, but depend on the chosen DXA cut-off level that is deemed to be desirable for predictions. For the prediction of osteoporosis, neither the 90% sensitivity, nor the sensitivity and specificity cut-off levels, will provide a high probability of a positive test result, meaning the individual has the condition. However, by using the suggested cut-offs, a clinician could confidently exclude an individual with a negative test result from any further diagnosis. Should the clinician wish to ensure the correct inclusion of individuals with osteoporosis or osteopenia (DXA–The cut off values for the different questionnaire systems vary from those previously published for the techniques [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 20]; the variation in the cut-off values can be explained by the different demographics of the study populations and the methods of selection for the cut-off values.To the best of our knowledge only one previous paper has offered potential cut-off values for the Sunlight Omnisense [41] for the prediction of both osteoporosis and osteopenia. The cut-off values were much lower than those used within this study and not suitable for use as screening cut-offs, as only 10 patients out of 45 would be correctly diagnosed as osteoporotic using the distal radius with only 3 out of 45 correctly diagnosed at

the proximal phalanx and the mid-shaft tibia. The same problem was demonstrated using the cut-off levels for osteopenia with only 57, 21 and 14 out of 144 patients being correctly diagnosed at the distal radius, proximal phalanx and mid-shaft tibia, respectively.There are five previous studies [42] using QUS of the calcaneus as an investigation. Two studies recommend a T-score cut-off of –2, lower than found in the present study. The third recommends –1.3 for BUA and –1.5 for VOS, while the fourth using the CUBA Clinical recommends a cut-off of 63 dB MHz^{-1}, which equates to a T-score of between –1.58 and –1.64; these values show a close agreement with the cut-off of –1.5 recommended in this study. The final study by D. Hans et al. [43] sets out a full screening strategy based on the use of clinical risk factors and QUS of the calcaneus. There were discernable differences between the studies in terms of study population, QUS systems, the prediction of BMD of the femoral neck and not a combination of total hip and spine, and additional risk factors not considered within the previously devised questionnaire systems. Despite this, the resultant cut-off values supplied by the study for the QUS were similar to those within this study.There are limitations to the present study that one may wish to consider before using the cut-off values suggested. The population was relatively small in numbers in comparison to the validation cohorts used for the development of these systems. The population was also somewhat biased to the lower bone density due to the collection of data from a DXA scanning clinic, where clinical referral criteria have already been applied to the participants to ensure they all display at least one risk factor for osteoporosis. For the purposes of this study, this was intentional and constitutes a strong positive factor, but in a more general situation researchers are reminded of these simple facts of our study.With these considerations in mind, we believe that the analysis of the present study, especially based on ROC curves, allowed for a robust comparison of the different scanning techniques and an assessment of their clinical utility. The CUBA Clinical through the measurement of BUA appears to provide a useful tool in the screening of this population, with the various questionnaire systems proving that they have good diagnostic accuracy. The Sunlight Omnisense on the other hand showed overall less promise as a screening tool for DXA.

Acknowledgements Support has been provided by the UK Department of Transport under the BOSCOS project, which allowed us to gain access and make use of the Sunlight Omnisense and CUBA Clinical systems. Special thanks are due to all members of the Department of Radiology of GWH in Swindon for their generous help with this study.

References

1. Geusens P, Hochberg MC, van der Voort DJM, et al (2002) Performance of risk indicies for identifying low bone density in postmenopausal women. Mayo Clin Proc 77:629–637
2. Koh LKH, Sedrine WB, Torralba TP, et al (2001) A simple tool to identify asian women at increased risk of osteoporosis. Osteoporos Int 12:699–705
3. Fujiwara S, Masunari N, Suzuki G, et al (2001) Performance of osteoporosis risk indices in a Japanese population. 62:586–594
4. Adler RA, Tran MT, Petkov VI (2003) Performance of the Osteoporosis Self-assessment Screening Tool for osteoporosis in American men. Mayo Clin Proc 78:723–727
5. Kung AWC, Ho AYY, Sedrine WB, et al (2003) Comparison of a simple clinical risk index and quantitative bone ultrasound for identifying women at increased risk of osteoporosis. Osteoporos Int 14:716–721
6. Richy F, Gourlay M, Ross PD, et al (2004) Validation and comparative evaluation of the Osteoporosis Self-Assessment Tool (OST) in a Caucasian population from Belgium. Q J Med 97:39–46
7. Park HM, Sedrine WB, Reginster JY, et al (2003) Korean experience with the OSTA Risk Index for Osteoporosis. J Clin Densitom 6:247–250
8. Cadarette SM, Jaglal SB, Kreiger N, et al (2000) Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry. Can Med Assoc J 162:1289–1294
9. Von Mühlen D, Lunde AV, Barrett-Connor E, et al (1999) Evaluation of the Simple Calculated Osteoporosis Risk Estimation (SCORE) in older Caucasian women: The Rancho Bernardo Study. Osteoporos Int 10:79–84
10. Sedrine WB, Devogelaer JP, Kaufman JM, et al (2001) Evaluation of the Simple Calculated Osteoporosis Risk Estimation (SCORE) in a sample of White women from Belgium. Bone 29:374–380
11. Cadarette SM, Jaglal SB, Murray TM (1999) Validation of the Simple Calculated Osteoporosis Risk Estimation (SCORE) for patient selection for bone densitometry. Osteoporos Int 10:85–90
12. Falasca GF, Dunston C, Banglawala YA (2003) Further validation of a questionnaire to identify women likely to have low bone density. J Clin Densitom 6:231–236
13. Ungar WJ, Josse R, Lee S, et al (2000) The Canadian Score Questionnaire. J Clin Densitom 3:269–280
14. Sedrine WB, Chevallier T, Zegels B, et al (2002) Development and assessment of the Osteoporosis Index of Risk (OSIRIS) to facilitate selection of women for bone densitometry. Gynecol Endocrinol 16:245–250
15. Reginster JY, Sedrine WB, Viethel P, et al (2004) Validation of OSIRIS, a prescreening tool for the identification of women with an increased risk of osteoporosis. Gynecol Endocrinol 18:3–8
16. Black DM, Steinbuch M, Palermo L, et al (2001) An assessment tool for predicting fracture risk in postmenopausal women. Osteoporos Int 12:519–528
17. Randell AG, Bhalerao N, Nguyen TV, et al (1998) Quality of life in osteoporosis: reliability, consistency, and validity of the Osteoporosis Assessment Questionnaire. J Rheumatol 25:1171–1179
18. Gerber V, Krieg MA, Cornuz J, et al (2003) Nutritional status using the Mini Nutritional Assessment Questionnaire and its relationship with bone quality in a population of institutionalized elderly women. J Nutr Health Aging 7:140–145
19. Goemaere S, Zegels B, Toye K, et al (1999) Limited clinical utility of a self-evaluating risk assessment scale for postmenopausal osteoporosis: lack of predictive value of lifestyle-related factors. Calcif Tissue Int 65:354–358
20. Michaëlsson K, Bergström R, Mallmin H, et al (1996) Screening for osteopenia and osteoporosis: selection by body composition. Osteoporos Int 6:120–126
21. Greenhalgh T (1997) How to read a paper: papers that report diagnostic or screening tests. BMJ 315:540–543
22. Glas AS, Lijmer JG, Prins MH, et al (2003) The diagnostic odd ratios: a single indicator of test performance. J Clin Epidemiol 56:1129–1135

23. Grimes DA, Schulz KF (2002) Uses and abuses of screening tests. Lancet 359:881–884
24. Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240:1285–1293
25. Leib ES, Lewiecki EM, Binkley N, et al (2004) Official positions of the international society for clinical densitometry. J Clin Densitom 7:1–6
26. Greenspan SL, Bouxsein ML, Melton ME, et al (1997) Precision and discriminatory ability of calcaneal bone assessment technologies. J Bone Miner Res 12:1303–1313
27. Brooke-Wavell K, Jones PR, Pye DW (1995) Ultrasound and dual X-ray absorptiometry measurement of the calcaneus: influence of region of interest location. Calcif Tissue Int 57:20–24
28. Johansen A, Evans W, Stone M (1999) Bone assessment in elderly women: what does a low bone ultrasound result tell us about bone mineral density? Arch Gerontol Geriatr 28:239–246
29. Njeh CF, Hans D, Li J, et al (2000) Comparison of six calcaneal quantitative ultrasound devices: precision and hip fracture discrimination. Osteoporos Int 11:1051–1062
30. Herd RJM, Blake GM, Miller CG, et al (1994) The ultrasonic assessment of osteopenia as defined by dual X-ray absorptiometry. Br J Radiol 67:631–635
31. Graafmans WC, Lingen AV, Ooms ME, et al (1996) Ultrasound measurements in the calcaneus: precision and its relation with bone mineral density of the heel and the lumber spine. Bone 19:97–100
32. Knapp KM, Blake GM, Spector TD, et al (2001) Multisite quantitative ultrasound: precision, age- and menopause- related changes, fracture discrimination, and *t* -score equivalence with dual-energy X-ray absorptiometry. Osteoporos Int 12:456–464
33. Damilakis JE, Papadokostakis G, Vrahoriti H, et al (2003) Ultrasound velocity through the cortex of phalanges, radius, and tibia in normal and osteoporotic postmenopausal women using a new multisite quantitative ultrasound device. Invest Radiol 38:207–211
34. Nicholson PHF, Strelitzki R, Cleveland RO, et al (2000) Scattering of ultrasound in cancellous bone: predictions from a theoretical model. J Biomech 33:503–506
35. Stewart A, Reid DM (2000) Quantitative ultrasound or clinical risk factors—which best identifies women at risk of osteoporosis? Br J Radiol 73:165–171
36. Adler RA, Funkhouser HL, Holt CM (2001) Utility of heel ultrasound bone density in men. J Clin Densitom 4:225–230
37. Ekman A, Michaëlsson K, Petrén-Mallmin M, et al (2002) Dual X-ray absorptiometry of hip, heel ultrasound, and densitometry of fingers can discriminate male patients with hip fractures from control subjects. J Clin Densitom 5:79–85
38. Sørensen HA, Jørgensen NR, Jensen J-EB, et al (2001) Comparison of quantitative ultrasound and dual X-ray absorptiometry in estrogen-treated early postmenopausal women. J Clin Densitom 4:97–104
39. Falgarone G, Porcher R, Duché A, et al (2004) Discrimination of osteoporotic patients with quantitative ultrasound using imaging or non-imaging device. Joint Bone Spine (corrected proof)
40. Damilakis JE, Papadokostakis G, Perisinakis K, et al (2003) Can Radial bone mineral density and quantitative ultrasound measurements reduce the number of women who need axial density skeletal assessment? Osteoporos Int 14:688–693
41. Knapp KM, Blake GM, Spector TD, et al (2004) Can the WHO definition of osteoporosis be applied to the multi-site axial transmission quantitative ultrasound? Osteoporos Int 15:367–374
42. Gambacciani M, de Aloysio D, Elia D, et al (2003) Quantitative ultrasound (qus) of bone in the management of postmenopausal women. Maturitas 47:139–149
43. Hans D, Hartl F, Krieg MA (2003) Device-specific weighted T-score for two quantitative ultrasounds: operational proportions for the management of osteoporosis for 65 years and older women in Switzerland. Osteoporos Int 14:251–258