Universal Standardization of Forearm Bone Densitometry

JOHN A. SHEPHERD,[1] XIAO GUANG CHENG,[1] YING LU,[1] CHRIS NJEH,[1] JORG TOSCHKE,[1]
KLAUS ENGELKE,[2] MICHAEL GRIGORIAN,[1] and HARRY K. GENANT[1]

ABSTRACT

As part of an effort to quantify device-dependent differences in forearm bone density, 101 women, aged 20–80 years (~16 women in each age decade), were scanned on six forearm bone densitometers: the Aloka DCS-600EX, the Hologic QDR-4500A, the Lunar PIXI, the Norland pDEXA, the Osteometer DTX-200, and the Pronosco X-posure System. Regression statistics are reported for all similar regions of interest (ROIs). However, comparisons were confounded because of large differences in the ROI size and placement. The number of ROIs reported for a single scan by each device varied from 1 to 12. The correlation coefficients ranged from $0.7 < r < 0.97$, with the highest correlation coefficients and lowest SEs for comparisons between the most similar ROIs. Standardized units of bone mineral density are derived for distal (sdBMD), mid-(smBMD), and proximal (spBMD) ROIs that allow for comparable mean bone densities to be derived for patient populations. Five phantoms were scanned and characterized on five of the devices and the precision and mean values were reported. These phantom values will aid in the in vitro cross-calibration between manufacturers to recreate the presented in vivo relationships. Care should be exercised when using these equations for cross-calibrating patient databases or pooling clinical data from different devices because the least significant differences detectable from measurements taken on two different machines can be increased substantially. (J Bone Miner Res 2002;17:734–745)

Key words: dual-energy X-ray absorptiometry, digital X-ray radiogrammetry, cross-calibration, osteoporosis, peripheral

INTRODUCTION

Bone density is the single-most predictive measure of fracture risk. Bone density at virtually any available measurement site (spine, proximal femur, forearm, whole body, calcaneus, and tibia) can be used to predict risks of all fractures. Of those bone density measurement sites, in clinical practice, the forearm has been used the longest, applying quantitative methods such as dual-energy X-ray absorptiometry (DXA), single-energy X-ray absorptiometry (SXA),

or the radioisotope equivalents dual photon absorptiometry (DPA) and single photon absorptiometry (SPA). Düppe et al. showed that a single SPA measurement predicted hip fractures 25 years later with a relative risk of 1.66/SD decrease in bone mineral density (BMD).[1]

The purpose of this study is to quantify the interrelationships of currently marketed devices that measure forearm bone density. The International Committee of Standards in Bone Measurement (ICSBM) commissioned the study. This committee was formed to address issues of accuracy, standardization, and comparability of densitometers from different manufacturers. All manufacturers were invited to participate with the committee, as were representatives from

[1]Osteoporosis and Arthritis Research Group, Department of Radiology, University of California San Francisco, San Francisco, California, USA.
[2]Institute of Medical Physics, University of Erlangen/Nürnberg, Erlangen, Germany.

academic institutions. The ICSBM has commissioned two previous studies, one for the standardization of the antero-posterior (AP) spine BMD as measured by DXA[2] and the second for the standardization of the proximal femur BMD.[3] The standardization equations were reported as letters to the editor of selected journals[4–7] and standardize BMD units are now available on the applicable machines. The committee also has recommended standardized projectional density units of milligrams per squared centimeter (sometimes referred to as areal density) to distinguish it from the manufacture-specific BMD units typically reported in grams per squared centimeter.

In the previous spine standardization study, regions of interest (ROIs) were found to be very similar. The agreement between the three participating manufacturers (Hologic, Inc. Bedford, MA, USA; and Lunar, Madison, WI, USA; and Norland Medical Systems, White Plains, NY, USA) was better than $r = 0.99$ with only slight differences attributable to bone edge detection. The T scores compared well also. In the proximal femur, the intermanufacturer definitions of the neck and total femur ROIs are larger than in the spine. The BMD agreement between manufacturers ranged from $0.97 < r < 0.98$ for the total femur and $0.94 < r < 0.96$ for the femur neck.[3] In addition, the femur reference data used by the three manufacturers differed considerably.[8] It was the suggestion of the ICSBM to define one reference data set, the Third National Health and Nutrition Examination Survey (NHANES III) total hip BMD,[9] to be provided in standardized BMD by all manufacturers to remove the diagnostic differences.[6] To date, reference data have not been standardized for any sites other than the femur.

Different forearm densitometers from a single manufacturer tend to agree well by design with a unity regression slope, no significant mean value offsets, and a small amount of scatter around the fit line. Intergenerational comparisons have been published for Hologic[10,11] forearm devices. Faulkner et al.[12] found that although there was high correlation ($r \sim 0.99$) between the Hologic QDR-1000 and Lunar DPX-L at the ultradistal and $\frac{1}{3}$ distal sites, the regression slope was up to 12% different from unity with offsets of 0.07–0.08 g/cm^2. There are no universal standards for the measurement of forearm bone density. Consequently, each of the commercially available densitometers has different ROIs (size, location, and number), different calibration standards, and unique racial and ethnic reference populations. Yet, practitioners commonly apply the standardized World Health Organization (WHO) or National Osteoporosis Foundation (NOF) diagnostic criteria equally to all devices. Thus, there is a strong possibility of treatment disparities and disagreement in clinical studies. We report on the absolute BMD differences, regression relationships, and define standardized equations to compare BMD results of similar ROIs between most of the commercially available forearm densitometers.

MATERIALS AND METHODS

Subjects

The study goal was to recruit 100 women, aged 20–80 years with ~16 women in each age decade. One hundred and one subjects were actually recruited with 13–19 subjects in each 10-year period. This population was chosen to provide a broad range of clinically observed BMD values. The self-declared racial breakdown was 74% white/Hispanic, 18% Asian, 4% black, 4% Pacific Islander/Indian. The following exclusion criteria were used: women known to be pregnant; women with a history of fracture at the distal radius of either arm; the known presence of generalized bone diseases of bone other than osteoporosis including hyperparathyroidism, hypoparathyroidism, Paget's disease, renal osteodystrophy, Cushing's disease, and steroid-induced osteoporosis or other metabolic diseases; a history of malignant diseases localized to bone or treatment by local resection; the presence of rheumatoid arthritis or other arthritic processes that severely limit patient mobility; and the presence of senile dementia severe enough to hinder adequate compliance and understanding of the study. The University of California San Francisco's (UCSF's) local institutional review board approved the study protocol.
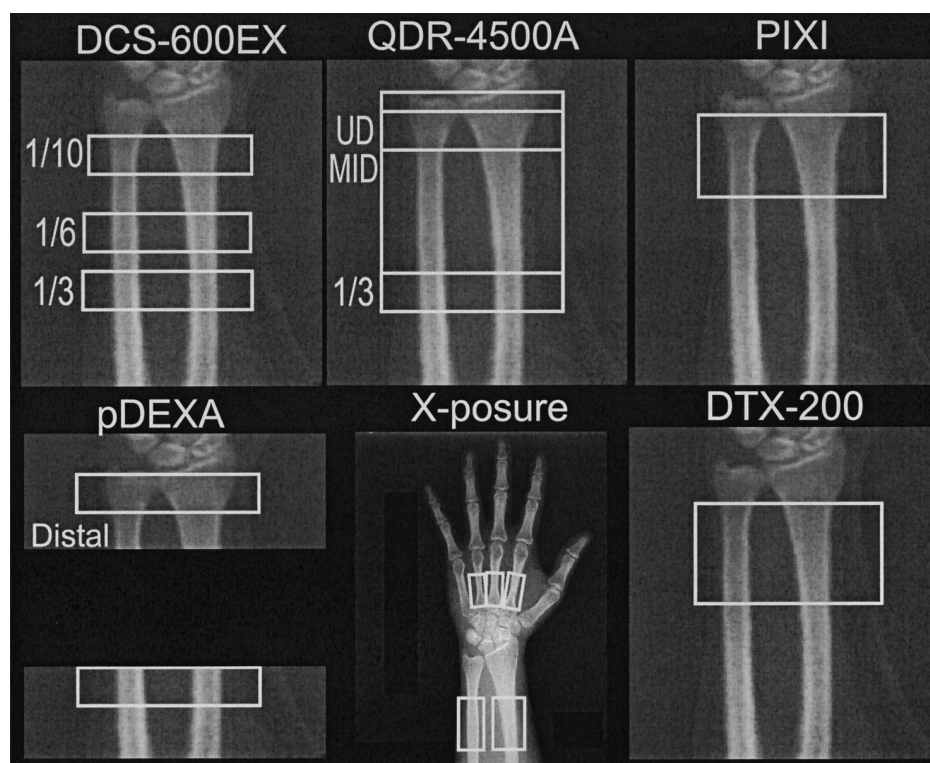
Devices

Each manufacturer was asked to identify the device they would like to include in the study and how each device was to be used. Some of the manufacturers produce several devices that measure forearm densitometry. However, it was out of the scope of this study to include all of a manufacturer's offerings or previous scanner generations. Table 1 is a list of the devices used in this study along with the associated manufacturer. The Hologic QDR-4500A and the Osteometer DTX-200 (Osteometer MediTech, Hawthorne, CA, USA) were already on site. The other devices were obtained on loan for the course of the study from the manufacturers. Each device acquired an image of both the radius and ulna. All devices except the Pronosco (Vedbaek, Denmark) collected a DXA image from the carpals to a specified proximal distance on the forearm. The Pronosco X-posure acquires a single-energy radiograph of the hand and proximal forearm. When a forearm length was needed for the calculation of ROI placement, consistent lengths were used for all devices. ROIs as configured by each manufacturer for their specific device were used. These ROIs are shown graphically in Fig. 1. No attempt was made in this study to alter the default ROIs because this would amount to nonstandard analysis and the reference data would not apply.

The Aloka DCS-600EX (Aloka Co., Ltd., Tokyo, Japan) algorithm automatically placed three 10-mm-long regions on the radius and ulna centered at $\frac{1}{10}$, $\frac{1}{6}$, and $\frac{1}{3}$ of the forearm length measured from the styloid process on the ulna.[13] Both the radius and ulna were included in each of the BMD calculations.

In the Hologic QDR-4500A, first the user had to place a global ROI in which three individual regions were defined: $\frac{1}{3}$ distal, middistal, and ultradistal ROIs.[14] The $\frac{1}{3}$ distal ROI was defined as 20 mm long centered at a distance equal to $\frac{1}{3}$ of the forearm length measured from the distal tip of the ulna. The ultradistal ROI is a region nominally 15 mm in length positioned proximal to the end plate of the radius. The middistal is the region between the $\frac{1}{3}$ and the ultradis-

TABLE 1. LIST OF DEVICES WITH VERSION OF THE SCAN PROTOCOL INCLUDED IN THIS STUDY

Make	Model	Technique	Scan protocol
Aloka Co., Ltd.	DCS-600EX	DXA	Forearm V 1.10
Hologic, Inc.	QDR-4500A	DXA	Forearm V 8.24a:3
General Electric Medical Systems Lunar	PIXI	DXA	Forearm V 1.42
Norland Medical Systems, Inc.	pDexa	DXA	Forearm V 3.8.0/1.1.1
Pronosco	X-posure	DXR	Forearm V 1.1.2
Osteometer MediTech	DTX-200	DXA	Forearm V 1.63

FIG. 1. Graphical representation of the ROIs for each device: Aloka DCS-600EX, the Hologic QDR-4500A, GE Lunar PIXI, the Norland Medical Systems pDEXA, Pronosco X-posure System, and the Osteometer DTX-200. Masking added to X-posure image to cover labeling. Note that the six manufacturers use different placement criteria for all ROIs.

tal regions. These regions were defined for both the radius and the ulna. A total ulna, total radius, and total forearm BMD is reported that is a sum of the radius and ulna for each of the three ROIs.

The Lunar PIXI forearm ROI was positioned automatically to span 30 mm from the radius and ulna radiographic junction proximally up the forearm.[8,15] Both the radius and the ulna were analyzed and single BMD, bone mineral content (BMC), and projected bone area values were reported.

The Norland pDEXA has a global distal site (radius + ulna), global ⅓ proximal site (radius + ulna), and a ⅓ proximal radius site.[16] The distal ROI spans 10 mm of the

lowest BMD region in the distal forearm and is found using an automated search routine. The proximal site spans 10 mm starting at the ⅓ forearm length and continuing proximally.

The Osteometer DTX-200 ROI was positioned automatically 24 mm proximal to the position where the radius and ulna are separated by 8 mm.[17]

The Pronosco X-posure System automatically defines a single BMD calculated from five ROIs.[18] Three of the ROIs are centered on the shafts of the three inner metacarpals, one ROI is on the distal ulna cortical shaft, and the last ROI on the distal shaft of the radius. Unlike the other five devices, the X-posure derives a BMD estimate from radio-

TABLE 2. VARIABLE NAMES FOR THE SCAN PROTOCOL DEFINED IN THE TEXT

Device	Ultradistal ROI	Mid-ROI	Proximal ROI	Total	No. of regions
Aloka DCS-600EX	ARU10BMD	ARU6BMD	ARU3BMD		3
GE/Lunar PIXI		PBMD			1
Hologic QDR-4500A	HUUBMD	HUMBMD	HU13BMD	HUTOTBMD	12
	HRUBMD	HRMBMD	HR13BMD	HRTOTBMD	
	HRUUBMD	HRUMBMD	HRU13BMD	HRUTOTBMD	
Norland pDEXA	NDBMD		NPBMD		3
			NPRBMD		
Osteometer DTX-200		OBMD			1
Pronosco X-posure		DXR-BMD			1

The Aloka DCS-600EX reports the results for the combined radius and ulna (ARU. . .). Hologic reports a radius (HR. . .), ulna (HU. . .), and radius + ulna (HRU . . .) BMD for each of three ROIs and a global ROI. The Norland distal (NDBMD) is the result for both the radius and the ulna while the proximal ROI reports both a radius and ulna (NPBMD) or a radius only (NPRBMD). Lunar and Osteometer reported a single BMD value that included both the radius and ulna. Pronosco reported a single BMD estimate from three metacarpals as well as the radius and ulna.

TABLE 3. FOREARM PHANTOMS IDENTIFIED FOR USE AS EITHER QUALITY CONTROL OR CROSS-CALIBRATION STUDIES

Make	Model	Description
CIRS	CIRS 400 mg	400 mg/cm^3 volumetric density, forearm mold of entire radius, ulna, and water-equivalent soft tissue (no carpals)
CIRS	CIRS 600 mg	Same as CIRS 400 mg but with 600 mg/cm^3
CIRS	CIRS 800 mg	Same as CIRS 400 mg but with 800 mg/cm^3
Kalender/Rüegsegger	European forearm Phantom I (EFPI)	Segmented cylindrical phantom, orthoradial representation of approximately last ⅙ forearm length
Kalender/Engelke	European forearm Phantom II (EFPII)	Same as EFPI but with conical shape to distal radius and longer forearm length (~½ distal forearm length.)
Lunar	PIXI QC Forearm	Metal stamp of distal forearm projection potted in acrylic, carpals to ~⅙ distal forearm length
Norland	Forearm QC	HA anthropomorphic projection potted in acrylic, from carpals to ½ distal forearm
Osteometer	DTX-200	HA anthropomorphic projection potted in acrylic, from carpals to ½ distal forearm

Note that none of the phantoms were scannable on all devices.
HA, hydroxiapatite.

grammetry applied to the cortical part of the bone (i.e., no reference bone material is used.) The BMD estimate is called digital X-ray radiogrammetry–BMD (DXR).

Unique BMD variable names were defined for the ROIs described previously and listed in Table 2. Each scanner was maintained per manufacturer's instructions. Data were collected over a 4-month period beginning in February 1999. There were no signs of calibration drifts of >0.5% for any of the devices during the study period.

Interdevice BMD comparison statistics

Because there are a total of 21 reported BMD values, reporting the correlation relationships between all ROIs was not practical for this publication. We chose to present the BMD correlations in three groupings most commonly used for diagnostic and clinical trial purposes: distal (i.e., the ultradistal), mid- (i.e., middistal), and proximal (i.e., ⅓

distal) ROIs. The Lunar PIXI, Osteometer DTX-200, and Pronosco X-posure each report one BMD value exclusively and will be compared in all three groups. The Norland pDEXA scans and reports a BMD for distal and proximal ROIs. Thus, the distal ROI for the Norland pDEXA is included in the distal as well as the mid-ROI analysis. Simple linear regression using SAS's General Linear Model procedure, procGLM, (SAS Institute, Cary, NC, USA) was used to determine a correlation coefficient, slope, intercept, and an SE of the estimate (SEE) for reported combinations. The significance of the intercept was determined using a *p* value.

Universal standardization equations

Universal standardization equations were calculated using the "optimal conversion" method by Hui et al.[19] This method minimizes a common entropy equation [Eq. (2) in

TABLE 4. TABLE OF POPULATION STATISTICS FOR EACH
DEVICE AND SELECTED ROIs

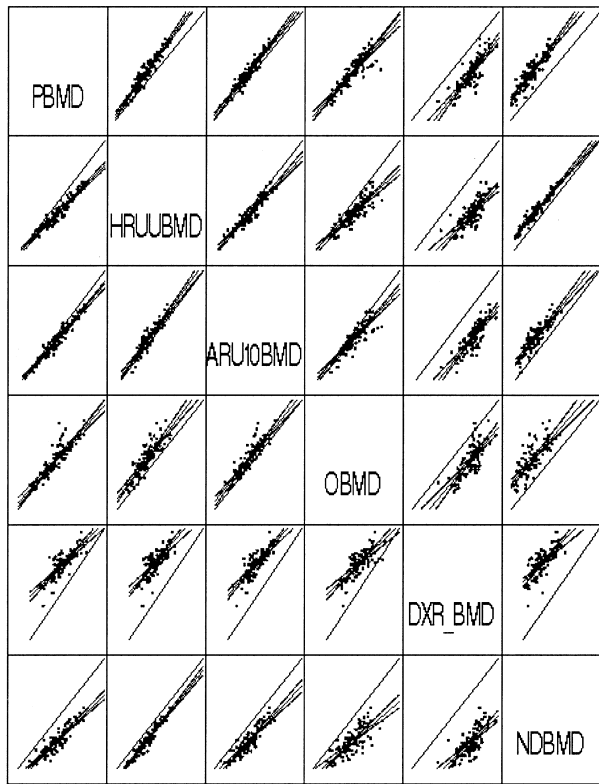| ROI	Study population mean (g/cm²)	Study population CV (%)	Study population STD (g/cm²)
ARU10BMD	0.394	15.6	0.061
AR6BMD	0.491	13.7	0.067
ARU3BMD	0.651	13.5	0.088
HRUUBMD	0.369	15.5	0.057
HRUMBMD	0.546	12.8	0.070
HRU13BMD	0.648	12.2	0.079
HRUTOTBMD	0.523	12.5	0.065
PBMD	0.420	16.5	0.069
NDBMD	0.319	18.6	0.059
NPBMD	0.777	13.7	0.106
OBMD	0.425	16.4	0.070
DXR-BMD	0.544	10.511	0.057

FIG. 2. Scatterplots of the distal BMD ROIs. The range of all x and y axes are the same and extend from 0.2 to 0.7 g/cm². The diagonal line represents the line of identity.

FIG. 3. Scatterplots of the middistal ROIs. The range of both the x and y variables extend from 0.2 to 0.8 g/cm². The diagonal line represents the line of identity.

Hui et al.] to simultaneously solve for the best-fit solution of all devices at once. The result is a set of equations that converts each device's BMD into a BMD in standardized units. In addition, the optimal conversion method differs from simple linear regressions in that it creates standardized relationships that are invertible and it assumes that errors exist in all v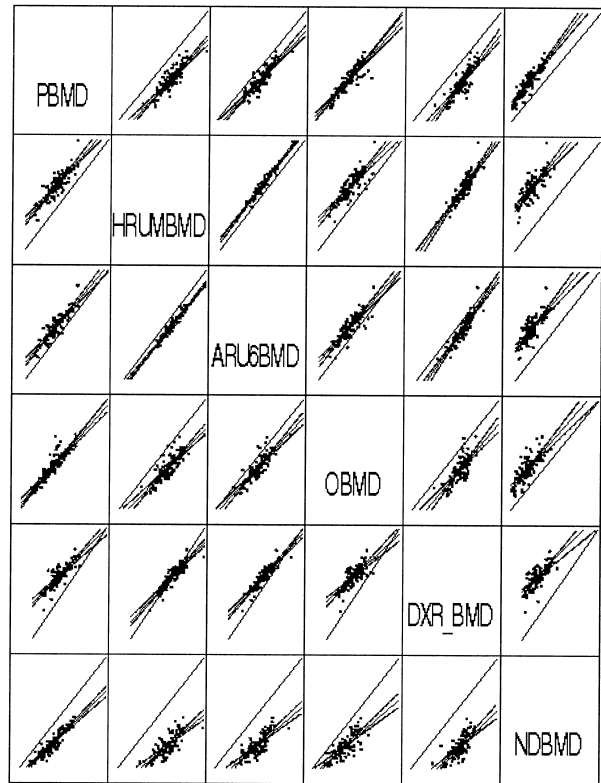ariables. For this reason, the standardized equations are more appropriate for cross-calibrating devices than the linear regressions. The optimal conversion standardization equations were solved using the SAS proc NONLIN. Because it was not possible to scan a single phantom on all devices, a common average BMD was picked by averaging the population means for each device. Standardization equations were calculated for the distal, mid-, and the proximal ROIs with the equations referred to as the sdBMD, smBMD, and the spBMD equations, respectively. As in the linear regression, devices with only one ROI were included in all three standardization equations. After applying the standardized BMD equations to the study subject data, we compared each device using the Bland-Altman analysis to look for systematic residual differences.[20]

Relative BMD value

To look at the equivalence of the standardized BMD as a function of age, we plotted standardized BMD values versus age for each ROI grouping. The SD for each variable was calculated independently for each decade. A p value was used to test for significant differences between similar ROIs for different age decades. In addition, the manufacturer-specific peak reference BMDs and population SDs were converted to standardized BMD units. T-score values of

−2.5 were calculated and compared to determine diagnostic equivalence.

Phantom measurements

The phantoms in Table 3 were identified for potential use in quality control and cross-calibration procedures. Two additional phantoms were made available and considered: the Aloka PHA-8158 and the Leeds phantom used by the Pronosco System. The Aloka and Pronosco phantoms were designed specifically for the manufacturer-specific machine and could not be analyzed using any of the other manufacturer's forearm algorithms. The European Forearm Phantom (EFP) I is not anthropomorphic and could not be analyzed using the default analysis instructions provided by the manufacturers and would have required special analysis. The EFP II could not be scanned in the default mode on the DCS-600EX. The three CIRS phantoms were designed specially and built for this study. The volumetric bone density of 400, 600, and 800 mg/cm^3 were chosen by CIRS. Because of their anthropomorphic shape and full forearm length, we had to prop the CIRS phantoms with foam pads to keep the phantoms from rotating during the scan. None of the phantoms could be analyzed by the X-posure in vivo algorithm. Each phantom in Table 3 was scanned on the remaining five devices with repositioning between each scan. The Aloka DCS-600EX scans were analyzed for radial BMD only. The Aloka device was shipped with radius-only analysis software and upgraded only after the phantom scans were acquired.

FIG. 4. Scatterplots of the proximal ROIs. The range of both the x and y variables extend from 0.2 to 1.1 g/cm^2. The diagonal line represents the line of identity.

RESULTS

Intermanufacturer BMD correlations

The population mean, SD, and CV for each of the chosen BMD variables is shown in Table 4. For qualitative visualization, Figs. 2–4 are compilations of the scatterplots for all the distal, mid-, and proximal ROIs, respectively. Each plot contains the scatter data, the least square fit line (solid) with 95% confidence limits (dotted), and a line of identity (diagonal). The x variables are in columns and the y variables are in rows. To simplify visualization, we used the same x and y axis scales for each plot compilation. The x and y axes in Fig. 2 range from 0.2 to 0.7 g/cm^2. For Figs. 3 and 4, the axes range from 0.2 to 0.8 and 0.2 to 1.1, respectively. Tables 5–7 show the linear regression statistics for the distal, mid-, and proximal sites, respectively. The r values ranged from 0.70 to 0.97. The SEE ranged from 2.9% to 13.1% of the mean.

Standardization relationships between devices

Table 8 shows the standardized BMD equations for each of the devices for the ultradistal, mid-, and proximal ROIs. Regression of the difference in sBMD values to the mean sBMD values for each device (Bland-Altman analysis) showed no systematic differences between mean values.
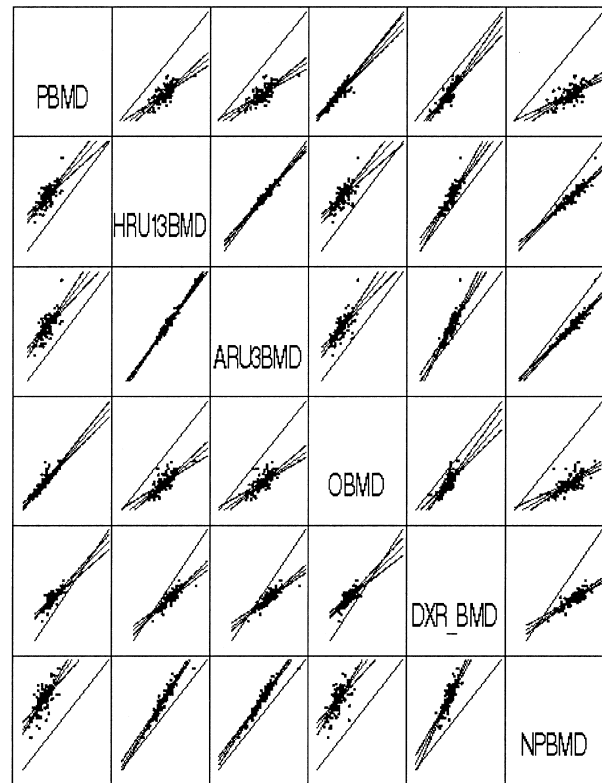
Standardization BMD values versus age

The standardized BMD versus the average decade age for similar ROIs across manufacturers is plotted in Figs. 5–7. The error bars show the SD of the mean for each decade mean value independently. For this study population and presentation method, there was a trend for all BMD values to increase relative to the young BMD values before a perimenopausal drop, with this trend becoming more pronounced for the more distal ROIs. In general, there was a similar age-related change in BMD for the ultradistal, mid-, and ⅓ proximal sites. For the proximal ROIs, none of the spBMD values of the young (20- to 30-year-old) subjects were significantly different ($p > 0.05$) except those between PBMD and DXR-BMD, which differed by 30 mg/cm^2 ($p = 0.03$). For the mid-ROIs, all smBMDs for the 20- to 30-year-old subjects were equivalent except for a 30-mg/cm^2 ($p = 0.04$) difference between ARU6BMD and NDBMD. For the distal ROIs, no age decade was significantly different except for a 0.01-g/cm^2 ($p = 0.044$) difference between PBMD and ARU3BMD. However, none of the 20- to 30-year-old sdBMD values were significantly different for $p < 0.01$.

Figure 8 is a comparison of the peak reference values and population SDs used to calculate T scores after being converted to sBMD. The sBMD values are shown for each ROI grouping. Note that some of the devices are represented in

TABLE 5. LINEAR REGRESSION STATISTICS FOR THE DISTAL SITES

Independent variable	Dependent variable				
	DXR-BMD	HRUUBMD	NDBMD	OBMD	PBMD
r Value					
ARU10BMD	0.82	0.92	0.88	0.89	0.95
DXR-BMD		0.76	0.72	0.76	0.83
HRUUBMD			0.96	0.86	0.94
NDBMD				0.78	0.90
OBMD					0.89

Independent variable	Dependent variable					
	ARU10BMD	DXRBMD	HRUUBMD	NDBMD	OBMD	PBMD
Intercept						
ARU10BMD		0.252	0.044	**−0.005**	**0.039**	**0.017**
DXR-BMD	−0.102		**−0.044**	−0.088	**−0.078**	−0.128
HRUUBMD	**0.014**	0.266		−0.049	**0.037**	**−0.002**
NDBMD	0.095	0.325	0.075		0.132	0.084
OBMD	0.053	0.280	0.069	**0.035**		0.046
PBMD	**0.025**	0.258	0.043	**−0.006**	0.051	
Slope						
ARU10BMD		0.747	0.831	0.827	0.988	1.031
DXR-BMD	0.909		0.760	0.747	0.927	1.007
HRUUBMD	1.023	0.754		0.997	1.051	1.144
NDBMD	0.931	0.687	0.923		0.919	1.055
OBMD	0.798	0.620	0.707	0.668		0.881
PBMD	0.872	0.680	0.778	0.774	0.890	
SEE						
ARU10BMD		0.032	0.022	0.028	0.032	0.021
DXR-BMD	0.035		0.038	0.042	0.046	0.039
HRUUBMD	0.024	0.038		0.017	0.036	0.023
NDBMD	0.030	0.040	0.016		0.044	0.030
OBMD	0.028	0.038	0.029	0.037		0.032
PBMD	0.020	0.032	0.019	0.026	0.033	

Bold represents not significantly different from 0 ($p > 0.05$).

each ROI grouping and thus the peak sBMD values change relative to the standardization grouping used. T scores calculated using sBMD values can now be compared and are shown in Fig. 8. Note that although there is no relative difference between the sBMD values in each grouping, the peaks, SDs, and thus the T score are different most likely because of the use of different reference populations.

Phantom results

As stated in the methods, six of the available phantoms could be scanned on five machines in the manufacturer's default scan mode. A forearm length of 25 cm was defined when the software asked for it. The mean values and CVs are given for the remaining phantoms on each device in Table 9. One Osteometer phantom scan was excluded from the DCS-600EX data because of poor positioning. The PIXI phantom was too short to have a distal site available for the

QDR-4500A, the pDEXA, and the DCS-600EX devices. The CIRS and Lunar phantoms did not contain a density gradient in the ultradistal region for the pDEXA to use to find the NDBMD ROI. The Osteometer phantom's handle interfered with its ability to lay flat on the horizontal scanners like the QDR-4500A, PIXI, DCS-600EX, and the pDEXA. All Osteometer phantom scans were done with the phantom propped and a horizontal bone plane.

DISCUSSION

Several principal points can be derived from this study. First, baseline and follow-up examinations must be acquired on the same make and model of densitometer. The monitoring of the same patient on a different machine comes up in a variety of situations including device upgrades and when a patient moves and/or changes primary care giver.

TABLE 6. LINEAR REGRESSION STATISTICS FOR THE MIDSITES

Independent variable	Dependent variable				
	DXR-BMD	HRUMBMD	NDBMD	OBMD	PBMD
			r Value		
ARU6BMD	0.90	0.97	0.78	0.85	0.88
DXR-BMD		0.89	0.72	0.76	0.83
HRUMBMD			0.78	0.83	0.87
NDBMD				0.78	0.90
OBMD					0.89

Independent variable	Dependent variable					
	ARU6BMD	DXR BMD	HRUMBMD	NDBMD	OBMD	PBMD
			y Intercept			
ARU6BMD		0.178	0.055	**−0.009**	**0.003**	**−0.007**
DXR-BMD	−0.102		**−0.049**	−0.088	−0.078	−0.128
HRUMBMD	−0.027	0.148		**−0.042**	**−0.024**	**−0.049**
NDBMD	0.200	0.325	0.253		0.132	0.084
OBMD	0.134	0.280	0.193	**0.035**		0.046
PBMD	0.117	0.258	0.177	**−0.006**	0.051	
			Slope			
ARU6BMD		0.748	1.004	0.673	0.864	0.875
DXR-BMD	1.086		1.095	0.747	0.927	1.007
HRUMBMD	0.945	0.724		0.661	0.822	0.860
NDBMD	0.906	0.687	0.920		0.919	1.055
OBMD	0.835	0.620	0.830	0.668		0.881
PBMD	0.885	0.680	0.877	0.774	0.890	
			SEE			
ARU6BMD		0.024	0.016	0.036	0.036	0.032
DXR-BMD	0.029		0.032	0.042	0.046	0.039
HRUMBMD	0.015	0.026		0.037	0.040	0.035
NDBMD	0.042	0.040	0.044		0.044	0.030
OBMD	0.036	0.038	0.040	0.037		0.032
PBMD	0.032	0.032	0.035	0.026	0.033	

Bold represents not significantly different from 0 ($p > 0.05$).

One must be certain that the two measurements are comparable and how the use of the two different devices change affects the least significant change (LSC)[21] in BMD. When no change of devices has occurred, the LSC is defined as

$$LSC = 2.8 \times PE,$$

where PE is the largest precision error of the technique for a 95% statistical confidence that the change in BMD is significant. However, if two devices are used, the LSC has to be modified to include the SE of the estimate, the precision of the two devices, and the uncertainty of the slope and intercept relating the two devices. If precision drove the SEE, we would expect the SEE around the fit line to be on the order of the root mean square combination of their precision values. For equal precision on both devices, the SEE would be $\sqrt{2}$ times the precision. If the precision on two devices to be compared is determined by patient posi-

tioning and is ~2%, the expected SEE would be ~3%. The actual SEEs varied from ~3% to over 13% with the ROIs, most closely matched in position, having the lowest SEE. This implies that biological variations are mostly responsible for the large SEEs on dissimilar ROIs. Unfortunately, the manufacturers do not reveal their specific algorithms, for example, their method of bone segmentation. Thus, the quantification of the impact on ROI and scanner variations is difficult. A better comparability of forearm densitometry would require in particular a standardization of the ROIs to be used. In addition, the ratio of cortical to trabecular bone mass is known to increase when moving proximally up the forearm.[22] So, one would expect mismatched ROIs to have different rates of loss or gain. Needless to say, if one is following a subject of a clinical trial over time, a change of devices during the trial should be strongly avoided.

Another principal point is that monitoring changes in BMD for populations may be possible on cross calibrated devices for

TABLE 7. LINEAR REGRESSION STATISTICS FOR THE PROXIMAL SITES

Independent variable	Dependent variable				
	DXR-BMD	HRU13BMD	NPBMD	OBMD	PBMD
			r Value		
ARU3BMD	0.83	0.97	0.95	0.72	0.74
DXR-BMD		0.82	0.83	0.76	0.83
HRU13BMD			0.94	0.70	0.73
NPBMD				0.69	0.71
OBMD					0.89

Independent variable	Dependent variable					
	ARU3BMD	DXR BMD	HRU13BMD	NPBMD	OBMD	PBMD
			Intercept			
ARU3BMD		0.204	0.086	**0.047**	**0.067**	**0.061**
DXR-BMD	**−0.070**		**0.026**	**−0.067**	**−0.078**	−0.128
HRU13BMD	−0.056	0.162		**−0.037**	**0.030**	**0.007**
NPBMD	**0.030**	0.199	0.106		0.075	**0.060**
OBMD	0.256	0.280	0.311	0.329		0.046
PBMD	0.240	0.258	0.297	0.318	0.051	
			Slope			
ARU3BMD		0.525	0.865	1.127	0.553	0.556
DXR-BMD	1.321		1.142	1.553	0.927	1.007
HRU13BMD	1.089	0.589		1.257	0.611	0.638
NPBMD	0.796	0.444	0.697		0.451	0.463
OBMD	0.924	0.620	0.791	1.053		0.881
PBMD	0.972	0.680	0.834	1.093	0.890	
			SEE			
ARU3BMD		0.031	0.019	0.034	0.048	0.045
DXR-BMD	0.049		0.046	0.060	0.046	0.039
HRU13BMD	0.021	0.033		0.038	0.050	0.048
NPBMD	0.028	0.032	0.028		0.051	0.049
OBMD	0.062	0.038	0.057	0.077		0.032
PBMD	0.060	0.032	0.055	0.075	0.033	

Bold represents not significantly different from 0 ($p > 0.05$).

similar ROIs and some devices. This is a common practice to increase recruitment numbers in clinical trials by drawing subjects from more than one clinical site. To pool the results from different machines to look for treatment effects and/or describing patient populations, it is imperative to know their in vivo cross-calibration relationships. All measures would be converted into the measure from one device.

There are no guidelines on what r value and SE is appropriate when converting one device's BMD into another's for a clinical trial. This situation can be contrasted to predicting the total femur BMD from the posteroanterior (PA) spine BMD. The r value between comparative spine and hip measures on the same subjects is ~0.7 caused by biological variation and it is never recommended to omit the hip and use only the spine to predict hip BMD. To the contrary, it has been shown that one can increase the ability to predict fracture risk by using a combination of hip and spine BMD in determining fracture risk.[23] Thus, we con-

clude that an r value of 0.70 between devices suggests too large of an uncertainty between devices to predict data of one device from another for individual patients. One the other hand, generally, it is accepted that devices that have been compared with r values above 0.97 using a clinically broad range of values can be pooled reasonably and cross-calibrated. An example of this is the pooling of spine and hip BMD values from across manufacturers. What is not clear is the threshold of acceptability for correlation relationships between the 0.70 and the 0.97 values. One could argue that an r value of 0.90 could be used as a threshold because this would mean 80% of the data supports the linear relationship between the two scanners. In addition, an SEE similar to the LSC of a single device could be used as an additional criterion. In this study, ~20% of all correlations resulted in r values of >0.9 and 50% in r values of >0.85. Therefore, excellent comparability exists for some devices and for certain ROIs.

TABLE 8. STANDARDIZATION EQUATIONS FOR THE SUBMD, SMBMD, AND SPBMD ROIs

Standardization equations

Ultradistal

$$sdBMD = 0.945 * PBMD + 0.015$$
$$sdBMD = 1.158 * HRUUBMD - 0.019$$
$$sdBMD = 1.059 * ARU10BMD - 0.005$$
$$sdBMD = 0.802 * OBMD + 0.071$$
$$sdBMD = 0.960 * DXR\_BMD - 0.110$$
$$sdBMD = 1.027 * NDBMD + 0.084$$

Middistal

$$smBMD = 1.011 * PBMD + 0.033$$
$$smBMD = 0.894 * HRUMBMD - 0.030$$
$$smBMD = 0.964 * ARU6BMD - 0.016$$
$$smBMD = 0.856 * OBMD + 0.094$$
$$smBMD = 1.137 * DXR\_BMD - 0.161$$
$$smBMD = 1.106 * NDBMD + 0.105$$

Proximal

$$spBMD = 1.091 * PBMD + 0.119$$
$$spBMD = 0.861 * HRU13BMD + 0.020$$
$$spBMD = 0.712 * ARU3BMD + 0.114$$
$$spBMD = 0.917 * OBMD + 0.188$$
$$spBMD = 1.538 * DXR\_BMD - 0.259$$
$$spBMD = 0.596 * NPBMD + 0.114$$

suBMD, standardization for ultradistal BMD ROIs.
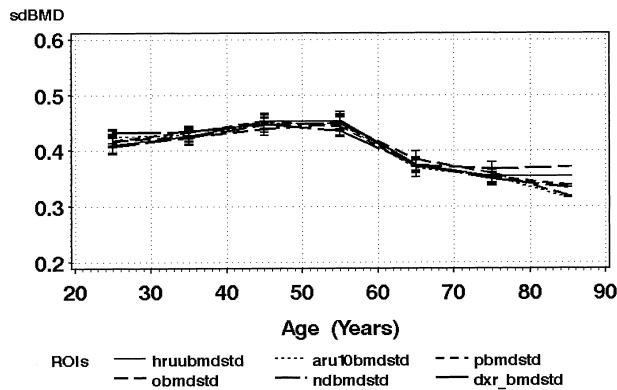
FIG. 5. Age-related change for the distal ROIs. Each point shows the decade average for age and sdBMD values. The sdBMD variables are represented as the variable name with a std suffix for standardized. The error bars show the ± 1 SD range around the mean values.

FIG. 6. Age-related change for the midforearm ROIs. Each point shows the decade average for age and smBMD values. The smBMD variables are represented as the variable name with a std suffix for standardized. The error bars show the ± 1 SD range around the mean values.

FIG. 7. Age-related change for the proximal forearm ROIs. Each point shows the decade average for age and spBMD values. The spBMD variables are represented as the variable name with a std suffix for standardized. The error bars show the ± 1 SD range around the mean values.

It is interesting to look at mistaken conclusions one could draw from uncalibrated BMD comparisons. For example, nonunity slopes between devices could cause confounding conclusions to be drawn depending on the device used. One would expect that a patient scanned on the DTX-200 (OBMD), for example, with a change between baseline and follow-up measurements of say 0.10 g/cm^2, would see only a 0.06-g/cm^2 change on the Norland pDEXA (NDBMD). However, even when the sBMD units are used, there still exist differences in the sBMD value for a T score of -2.5
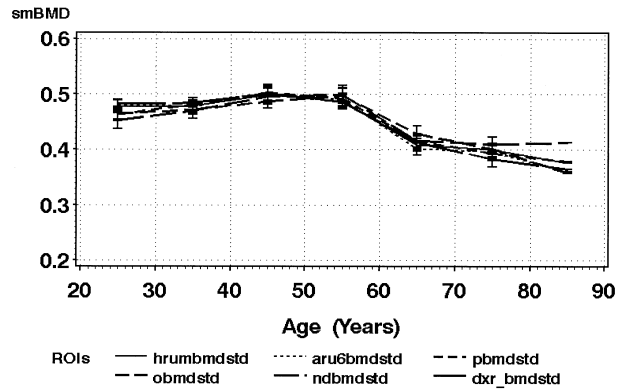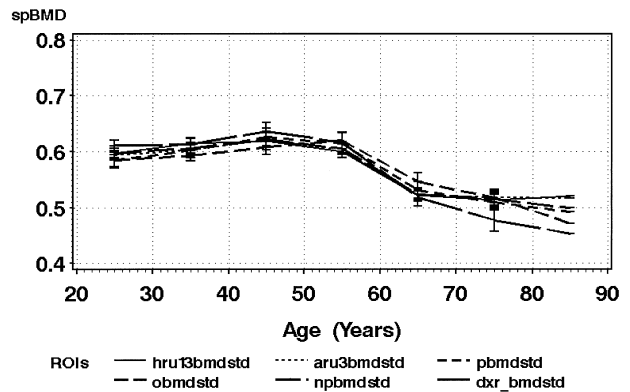
because different manufacturers use different reference populations as shown in Fig. 8.

Standard ROIs would simplify interpretation of forearm results. Early postmenopausal bone loss is monitored most sensitively using sites high in trabecular bone, like the ultradistal forearm. Global loss in skeletal mass is best monitored with cortical bone sites like the distal proximal forearm. Standardized ROIs would eliminate biological variation between measures on different devices, reduce the SEE, and increase confidence in cross-calibrating devices, pooling measurements, and monitoring drug therapy.

Because the phantoms are distributed through the clinical range, it may be useful to use two or three together when cross-calibrating scanners to get a reasonable slope. The CIRS phantoms were designed to give three reference points for all ROIs but were limited by their difficult positioning and high densities. Based on the experience of this study, an ideal cross-calibration phantom set could be con-

TABLE 9. PHANTOM MEAN BMD VALUES IN (g/cm^2) FOR EACH ROI

ROI	Phantom Mean BMD in g/cm^2 (% CV)						
	CIRS 400	CIRS 600	CIRS 800	Osteometer	EFPII	Norland	Lunar
AR3BMD	0.268 (2.2)	0.391 (1.7)	0.442 (0.6)	0.381 (0.2)	NA	0.794 (0.7)	NA
AR6BMD	0.261 (2.1)	0.380 (0.8)	0.421 (0.4)	0.386 (1.2)	NA	0.574 (1.6)	0.412 (0.2)
AR10BMD	0.276 (2.3)	0.419 (3.1)	0.467 (1.7)	0.377 (1.2)	NA	0.456 (1.2)	0.408 (1.0)
HRU3BMD	0.316 (1.3)	0.434 (1.6)	0.563 (1.9)	0.409 (1.4)	0.519 (2.6)	0.825 (0.8)	NA
HRUMBMD	0.309 (0.8)	0.421 (1.2)	0.541 (1.0)	0.409 (1.5)	0.325 (0.8)	0.610 (1.1)	0.450 (1.1)
HRUUBMD	0.400 (2.1)	0.600 (1.3)	0.777 (0.7)	0.416 (1.5)	0.425 (2.6)	0.414 (2.2)	0.464 (1.2)
OBMD	0.316 (2.2)	0.447 (1.2)	0.498 (0.9)	0.388 (0.6)	0.264 (0.5)	0.413 (0.7)	0.416 (0.6)
NDBMD	0.337 (2.4)	0.454 (1.1)	0.467 (2.1)	0.421 (0.7)	0.598 (1.1)	0.324 (0.9)	0.451 (1.1)
NPBMD	0.316 (2.8)	0.472 (1.8)	0.499 (2.7)	0.410 (0.4)	0.292 (0.7)	0.909 (0.5)	NA
PBMD	0.41 (1.8)	0.583 (1.9)	0.679 (1.1)	0.393 (0.5)	0.336 (1.0)	0.388 (1.1)	0.411 (0.3)

The phantom CV values with repositioning between scans is shown in parentheses as a percentage.
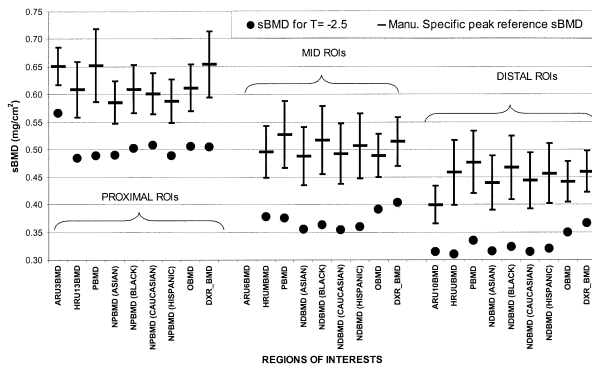
FIG. 8. Comparison of the manufacturer-specific peak reference values with ± 1 population SD in units of sBMD by grouping. These peaks and SDs are used to calculate the sBMD for a T score of -2.5 (circles). Thus, the T scores used for diagnoses still vary because different manufacturers use different reference populations. (ARU6BMD has no reference values.)

Lunar makes four different devices for forearm densitometry: Prodigy, DPX-IQ, Expert, and PIXI. The Prodigy, Expert, and IQ have the same X-ray technique but differ from the PIXI. The Prodigy, IQ, and Expert have multiple ROIs and the PIXI has only one. Thus, it will be difficult to apply the results of this study to those other devices. Hologic makes pencil and fan beam devices that have identically defined ROIs but different X-ray techniques. Osteometer has several algorithms for analyzing their scans that differ significantly from the analysis we used. Also, no peripheral quantitative computed tomography (pQCT) devices were included. Because in vivo precision was not quantified on this patient population, we could not use regression models that use weighted error corrections such as the model proposed by Mandel.[24] Instead, we assumed that the base precision of each device was similar. Although the precision reported in the literature for each device varies, this is a reasonable estimate because patient repositioning typically limits the precision of most clinical BMD measurements.

structed in several ways. The CIRS phantoms could be improved by shortening to half the forearm length, given a square shape on the proximal end to limit rotation and add stability, eliminate the 800-mg phantom, and provide a 200-mg phantom. Alternatively, two additional Norland or Lunar phantoms could be made of half and double their present densities to cover the clinical range. The EFP II covers all densities required in a single phantom but is not anthropomorphic enough to be analyzed on all devices using the default mode. A more complete anthropomorphic phantom of the hand as well as the forearm that included the medullar canals would be needed for Pronosco X-posure longitudinal QC. Because the Pronosco X-posure uses radiogrammetry instead of mass attenuation to derive bone density, it is not obvious how one could create a phantom set that could cross-calibrate the X-posure to the DXA devices. The fact that none of the phantoms could be scanned on all devices stresses the challenges in cross-calibrating different technologies to each other.

This study has several shortcomings. First, not all devices from all manufacturers were included in this study. GE/

CONCLUSIONS

This study describes the regression relationship between six forearm densitometers. In general, the agreement between forearm densitometers is high. Standardized units have been defined such that the mean BMD values for each device would, on average, agree with one another. Further standardization is necessary in the ROIs to improve the agreement and reduce the SE of the estimate.

ACKNOWLEDGMENTS

REFERENCES

1. Duppe H, Gardsell P, Nilsson B, Johnell O 1997 A single bone density measurement can predict fractures over 25 years. Calcif Tissue Int **60:**171–174.
2. Genant HK, Grampp S, Gluer CC, Faulkner KG, Jergas M, Engelke K, Hagiwara S, Van Kuijk C 1994 Universal standardization for dual X-ray absorptiometry: Patient and phantom cross-calibration results. J Bone Miner Res **9:**1503–1514.
3. Lu Y, Fuerst T, Hui S, Genant HK 2001 Standardization of bone mineral density at femoral neck, trochanter and Ward's triangle. Osteoporos Int **12:**438–444.
4. Steiger P 1995 Standardization of measurements for assessing BMD by DXA. Calcif Tissue Int **57:**469 (letter).
5. Steiger P 1995 Standardization of postero-anterior (PA) spine BMD measurements by DXA. Committee for Standards in DXA. Bone **17:**435 (letter).
6. Hanson J 1997 Standardization of proximal femur bone mineral density (BMD) measurements by DXA. Bone **21:**369–370 (letter).
7. Hanson 1997 Letter to the editor. Osteoporos Int **7:**500–501.
8. Nord R 1999 E-mail communication regarding PIXI ROIs. GE Lunar, Madison, WI, USA.
9. Looker AC, Wahner HW, Dunn WL, Calvo MS, Harris TB, Heyse SP, Johnston CC Jr, Lindsay R 1998 Updated data on proximal femur bone mineral levels of US adults. Osteoporos Int **8:**468–489.
10. Hagiwara S, Engelke K, Yang SO, Dhillon MS, Guglielmi G, Nelson DL, Genant HK 1994 Dual x-ray absorptiometry forearm software: Accuracy and intermachine relationship. J Bone Miner Res **9 :**1425–1427.
11. Bouyoucef SE, Cullum ID, Ell PJ 1996 Cross-calibration of a fan-beam X-ray densitometer with a pencil-beam system. Br J Radiol **69:**522–531.
12. Faulkner KG, McClung MR, Schmeer MS, Roberts LA, Gaither KW 1994 Densitometry of the radius using single and dual energy absorptiometry. Calcif Tissue Int **54:**208–211.
13. Aloka DCS-600EX Owner's Manual. Aloka Co., Ltd., Toyko, Japan, p. 9.6.
14. Hologic QDR-4500 Elite User's Guide, Forearm Option. Hologic, Inc., Bedford, MA, USA, pp. 7–9.
15. Lunar PIXI Owner's Manual. GE Lunar, Madison WI, USA, pp. 32–33.
16. Norland Medical Systems pDEXA Owner's Manual. Norland Medical Systems, Madison WI, USA, pp. viii.
17. Svendsen OL, Hassager C, Skodt V, Christiansen C 1995 Impact of soft tissue on in vivo accuracy of bone mineral measurements in the spine, hip, and forearm: A human cadaver study. J Bone Miner Res **10:**868–873.
18. Pronosco X-posure User Manual. Pronosco, Inc., Vebaek, Denmark.
19. Hui SL, Gao S, Zhou XH, Johnston CC Jr, Lu Y, Gluer CC, Grampp S, Genant H 1997 Universal standardization of bone density measurements: A method with optimal properties for calibration among several instruments. J Bone Miner Res **12:**1463–1470.
20. Bland JM, Altman DG 1986 Statistical methods for assessing agreement between two methods of clinical measurement. Lancet **1:**307–310.
21. Gluer CC 1999 Monitoring skeletal changes by radiological techniques. J Bone Miner Res **14:**1952–1962.
22. Nilas L, Norgaard H, Podenphant J, Gotfredsen A, Christiansen C 1987 Bone composition in the distal forearm. Scand J Clin Lab Invest **47:**41–46.
23. Lu Y, Genant HK, Shepherd J, Zhao S, Mathur A, Fuerst TP, Cummings SR 2001 Classification of osteoporosis based on bone mineral densities. J Bone Miner Res **16:**901–910.
24. Mandel J 1984 Fitting straight lines when both variables are subject to error. J Qual Technol **16:**1–15.