

# Universal Standardization of Bone Density Measurements: A Method with Optimal Properties for Calibration Among Several Instruments

SIU L. HUI,<sup>1,2</sup> SUJUAN GAO,<sup>1</sup> XIAO-HUA ZHOU,<sup>1,2</sup> C. CONRAD JOHNSTON, JR.,<sup>1</sup> YING LU,<sup>3</sup>  
CLAUS C. GLÜER,<sup>4</sup> STEPHEN GRAMPP,<sup>3</sup> and HARRY GENANT<sup>3</sup>

## ABSTRACT

The International Dual-Photon X-Ray Absorptiometry (DXA) Standardization Committee (IDSC) conducted a cross-calibration study among three models of DXA machines from three different manufacturers. In that study, 100 subjects were scanned on all three machines. A set of equations were derived to convert bone mineral density (BMD) on each machine to a “standardized BMD” (sBMD) such that sBMD from the same subject derived from different machines would be approximately the same. In a reanalysis of the cross-calibration data, we showed that the conversion method used in the IDSC study did not achieve several optimal properties desirable in such conversions. We derived new conversion equations to sBMD based on minimizing differences among sBMD from the three machines. More important is that the new conversions have no residual bias that was present in the IDSC conversions. The performance of the methods were compared on the cross-calibration data as well as an external data set. We conclude that the IDSC conversions are adequate for clinical use on other machines worldwide, but that researchers should standardize their own machines in a laboratory using the new method. (*J Bone Miner Res* 1997;12:1463–1470)

## INTRODUCTION

**B**ONE MINERAL DENSITY (BMD) is the primary determinant of skeletal fragility, and, as such, plays a central role in the diagnosis of osteoporosis. It remains, however, somewhat difficult for clinicians to use BMD measurements as readily as would be desirable. There are a number of reasons for these difficulties, but primary among them is the systematic difference in reported BMD among the manufacturers of densitometers. While the reasons for the discrepancies are many, the goal of this paper is not to discuss the biological or technical contributors to the problem,<sup>(1–3)</sup> but rather to introduce an appropriate algorithm for converting measurements from different machines to a universal standard scale whereby the measurements on the same subject on different machines are comparable.

The first attempt at universal standardization of BMD was made on dual-photon X-ray absorptiometry (DXA) measurements. The International DXA Standardization Committee (IDSC) sponsored a cross-calibration study which measured 100 women on three DXA scanners made by three different manufacturers.<sup>(1,4)</sup> The data showed that the measurements on the three machines were highly correlated and linearly related to one another; hence, simple linear regression equations were derived for converting BMD measurements on any one machine to another. To avoid designating any of the machines as the “gold standard,” the IDSC study then went on to derive a universal standardized measurement called standardized bone mineral density (sBMD). The aim was to convert each manufacturer’s BMD to sBMD using a formula such that the sBMD would give “approximately the same value when

<sup>1</sup>Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, U.S.A.

<sup>2</sup>Regenstrief Institute for Health Care, Indianapolis, Indiana, U.S.A.

<sup>3</sup>Musculoskeletal Section and Osteoporosis Research Group, Department of Radiology, University of California, San Francisco, California, U.S.A.

<sup>4</sup>University of Kiel, Kiel, Germany.

scanning one patient on all machines” and to “peg” the values to the “true” density of a reference phantom.<sup>(1)</sup> Since no standard statistical procedure was readily available for deriving the universal standard, the investigators developed an ad hoc method, which, unfortunately, had several problems. In particular, systematic differences remained between the same patient’s sBMD on different machines. In this paper, we evaluate the extent of this bias in the original cross-calibration data and go on to show that the problem may be negligible for standardizing measurements made on machines other than those used in the cross-calibration study.

The IDSC conversion equations from spine BMD to sBMD have now been implemented in new DXA scanners.<sup>(5)</sup> The machine-generated sBMDs are intended primarily for clinical use worldwide. These conversions, no matter how good, were optimized only for the three specific machines used in the original cross-calibration study. Although we will show that clinical application of the IDSC conversions are appropriate, researchers who wish to standardize multiple machines in their own laboratories for research studies should derive their own conversions that are optimized for their own machines. To this end, we propose a new conversion algorithm that improves upon the IDSC algorithm by minimizing the differences in sBMD on the same subjects and removing all residual biases.

Our proposed algorithm is to be used only *after* standard regression analysis has established linearity between BMD measured on all possible pairs of machines, as was done in the IDSC study. The major steps of the proposed algorithm are: (1) subtract the mean BMD from the individual BMD measured on each machine; (2) multiply the mean-adjusted BMD by a factor specific to each machine such that the total squared difference among machines is minimized; (3) add a common constant to each multiple of mean-adjusted BMD to obtain sBMD such that the mean sBMD of the “pegging” phantom from all machines equals its theoretical “true” density.

We will compare the performance of the different algorithms by applying them to the data from the original IDSC study and to an external data set.

## MATERIALS AND METHODS

### *The IDSC Cross-Calibration Study*

A cross-calibration study of three DXA densitometers was conducted at University of California at San Francisco under the auspices of the IDSC. Details of the IDSC study have been published.<sup>(1,4)</sup> Briefly, 100 healthy, nonpregnant women evenly distributed over the age 20–80 years were recruited. Posteroanterior (PA) lumbar spine and hip measurements were made on each subject on all three scanners—a Norland XR26 Mark II (Norland Corp., Fort Atkinson, WI, U.S.A.), a Lunar DPX-L (Lunar Corp., Madison, WI, U.S.A.), and a Hologic QDR 2000 (Hologic Inc., Waltham, MA, U.S.A.). In addition, a number of phantoms were measured multiple times on the same scanners. Cross-calibration equations were derived for both human and phantom BMD. Since the phantom data devi-

ated systematically from the human cross-calibration equations, BMD data on the 100 women were used to derive conversion equations for sBMD, with only one phantom’s measurements used for “pegging” the sBMD values.

### *An external data set*

To compare the IDSC and our new conversion algorithms on some machines other than those used in the cross-calibration study, we gathered a set of data with 56 normal subjects who previously had their spine BMD measured on both a Lunar DPXL and a Hologic 1000W at Indiana University. Other than a few healthy employees associated with the bone studies, these were subjects who participated in multiple study protocols that used those two different scanners. The primary studies were observational studies designed to investigate various factors related to BMD at different ages; these protocols enrolled only healthy subjects who had no metabolic bone disease and had not taken medication that affected bone metabolism. All subjects were white and all but one were women, with mean age of 47 years (range 24–84). The measurements of any subject on the two machines were no more than 1.2 years apart (50% were within 0.2 year). Although this was a convenience sample, neither the IDSC nor the new algorithm assumed any distribution of BMD in the cross-calibration sample; hence, no bias could have been introduced by any sampling scheme.

### *The IDSC algorithm and its problems*

Only the cross-calibration of spine BMD is used for illustration in this article. The published conversions equations for spine sBMD are:

$$\text{sBMD} = 1.0755 \text{ BMD}, \quad \text{for Hologic}$$

$$\text{sBMD} = 0.9522 \text{ BMD}, \quad \text{for Lunar}$$

$$\text{sBMD} = 1.0761 \text{ BMD}, \quad \text{for Norland}$$

These were derived by first fitting six no-intercept regressions through all possible pairs of spine BMD on the 100 subjects. For any given pair of scanners, they “normalized” the regression by averaging the slope of one regression, say  $y$  on  $x$ , and the inverse of the slope of  $x$  on  $y$ . The ratios of these “normalized” slopes were used to solve for the conversion parameters for sBMD, with the constraint that the mean sBMD of the midvertebra of the European spine phantom (ESP) was equal to its true density of 1.0 g/cm<sup>2</sup>. This particular phantom was chosen because its BMD was closest to the human regression lines. The major problems of the IDSC algorithm are described in order of severity as follows:

(1) The derivation method is not internally consistent, i.e., applying the method to a given set of cross-calibration data does not lead to a unique set of conversion equations from BMD to sBMD. Let  $X$ ,  $Y$ , and  $Z$  denote BMD measurements of the same individuals on three machines. The first inconsistency in the IDSC approach results from averaging the regression slope  $b_1$  of  $Y$  on  $X$  and the reciprocal of the slope  $b_2$  of  $X$  on  $Y$  to get a linear relationship between

$X$  and  $Y$ ; the result would not be the same if one had chosen to use the average of  $1/b_1$  and  $b_2$ . The second inconsistency is that solutions to the conversion equations are dependent on which two of the three linear relationships between machines are used. This lack of internal consistency allows one to obtain different sets of conversion equations from analyzing the same set of data using the same general algorithm.

(2) The IDSC algorithm makes no attempt to minimize differences in sBMD of the same subjects measured on different machines even though it is the primary purpose of creating sBMD.

(3) The algorithm forces all linear relationships of sBMD between machines through the origin even though regression analysis on BMD has shown statistically significant intercepts in those relationships. Two types of bias may result in the sBMD. First is that the mean sBMD of the cross-calibration study subjects measured on different machines may not be equal. Second, the derived sBMD from any given subject may be systematically higher or lower on one machine than another depending on the value of the individual's BMD. The magnitude of the latter bias increases with the size of the nonzero intercept in the regression of BMD between machines, as will be illustrated in the results.

#### *Derivation of the new algorithm*

The method developed here retains all of the desirable properties of the IDSC algorithm while avoiding its shortcomings. This method can be applied to any set of cross-calibration data from two or more scanners. For simplicity, we describe the method as applied to spine BMD measured on three scanners in the IDSC cross-calibration study.<sup>(1)</sup> Both the IDSC study and a subsequent study<sup>(2)</sup> found that regression lines based on phantom data differed systematically from regression lines based on human measurements; hence, we derived the conversions from BMD to sBMD based on the measurements of the 100 women in the cross-calibration study. As in the IDSC study, the midvertebra ESP was used as the "pegging" phantom mainly because it was near the center of the distribution of BMD in the study.

Details of the algorithm are given in the Appendix. The proposed method is summarized in the following steps:

(1) Start with standard regression analyses to fit the relationships between BMD on pairs of scanners. For each pair of scanners, regress BMD of machine 1 on machine 2, and vice versa. Add quadratic or higher order terms to the model to test whether each relationship deviates from linearity. This step provides interconversion equations from any machine to another. Standard errors provide measures of reliability of the coefficients of conversion equations.

(2) After step 1 has established linearity between all pairs of machines, use the new algorithm to derive sBMD. First, to remove the problem of nonzero intercepts, subtract the sample mean from the individual BMD. If  $X$ ,  $Y$ , and  $Z$  denote BMD measurements on Hologic, Lunar, and Norland scanners, respectively, we obtain the following variables:

$$\begin{aligned}x &= X - \bar{X} \\y &= Y - \bar{Y} \\z &= Z - \bar{Z}\end{aligned}\quad (1)$$

(3) Multiply  $x$ ,  $y$ , and  $z$  from step 2 by different factors,  $a$ ,  $b$ , and  $c$  to obtain  $ax$ ,  $by$ , and  $cz$ , respectively. The multipliers  $a$ ,  $b$ , and  $c$  are chosen to minimize

$$E^2 = (ax - by)^2 + (by - cz)^2 + (cz - ax)^2 \quad (2)$$

over the entire sample, subject to the constraint  $a^2 + b^2 + c^2 = L$ , a norming constant.

(4) Add a common constant  $K$  to  $ax$ ,  $by$ , and  $cz$  to obtain sBMD:

$$\begin{aligned}\text{sBMD}_x &= ax + K \\ \text{sBMD}_y &= by + K \\ \text{sBMD}_z &= cz + K\end{aligned}\quad (3)$$

for the Hologic, Lunar, and Norland scanners, respectively. The constant  $K$  is chosen such that the mean sBMD of the "pegging" phantom from the three machines is equal to the phantom's "true" density.

Several desirable properties result from this algorithm. Step 2 ensures that the linear relationships between scanners all pass through the sample means, as any unbiased linear relation should. Step 3 ensures that, among all linear conversions, this conversion produces sBMDs that are closest between machines by the least-squares criterion. Step 4 "pegs" the sBMD to the theoretical density. The Appendix contains an exact solution for deriving sBMD as outlined in steps 2–4. We call it the "optimal" algorithm. Note that the least-squares criterion is only optimal if the measurement errors are the same across machines. Otherwise, it should be modified to a weighted least-squares criterion, with the more precise machines getting greater weights. Since the exact solution requires a symbolic programming language that may not be widely available, we also developed an approximate method that uses more commonly available statistical packages. Both of these numerical methods are internally consistent, so that there is a unique solution set for any given set of data. A bootstrap procedure is also described briefly in the Appendix that could be used to obtain variance estimates for the conversion parameter estimates, but this procedure is generally too computing intensive for its worth unless comparison between parameters is necessary.

## RESULTS

The universal standardization method was applied to the spine (L2–L4) BMD of the 100 subjects measured on the three scanners. The measurements from each pair of scanners are plotted in Fig. 1. Standard regression analyses between pairs of scanners produced the same results as previously reported<sup>(1)</sup> so they are not repeated here. Assuming true linear relationships of BMD between scanners, we applied the new conversion method from BMD to sBMD on the IDSC cross-calibration data.

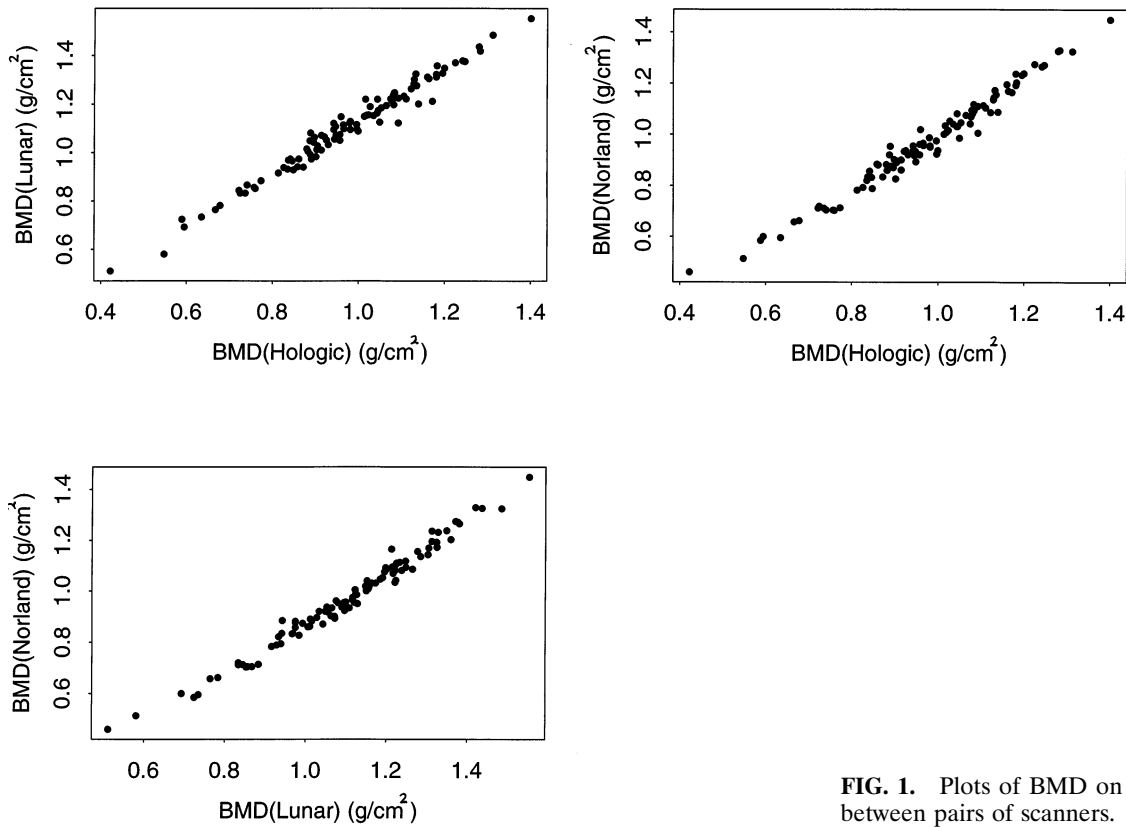


FIG. 1. Plots of BMD on 100 subjects between pairs of scanners.

The mean spine BMDs (in  $\text{g/cm}^2$ ) for the sample were  $\bar{X} = 0.972$ ,  $\bar{Y} = 1.100$ , and  $\bar{Z} = 0.969$  for the Hologic, Lunar, and Norland scanners, respectively. These means were subtracted from the individual measurements to remove the intercept problem initially, i.e., Eqs. (1) became:

$$\begin{aligned} x &= X - 0.972 \\ y &= Y - 1.1 \\ z &= Z - 0.969 \end{aligned} \tag{1'}$$

We then minimized expression (2) (pairwise squared differences among  $ax$ ,  $by$ , and  $cz$ ) subject to the constraint that  $a^2 + b^2 + c^2 = 3 \times 1^2$ . Using the numerical method described in the Appendix, we obtained the scale parameters

$$\begin{aligned} a &= 1.0550 \\ b &= 0.9683 \\ c &= 0.9743 \end{aligned}$$

The final calibration step was based on the midvertebra ESP, which had BMD (in  $\text{g/cm}^2$ ) measured at 0.916, 1.074, and 0.922 on the Hologic, Lunar, and Norland scanners, respectively. Substituting the estimates of  $a$ ,  $b$ , and  $c$  into Eqs. (3) and equating the mean sBMD from the three scanners to the true density of the phantom,  $1.0 \text{ g/cm}^2$ , we obtained  $K = 1.0436$ . Based on these estimates, the optimal universal standardized measurements were given by:

$$\begin{aligned} \text{sBMD}_H(O) &= 1.0550(X - 0.972) + 1.0436, && \text{for Hologic} \\ \text{sBMD}_L(O) &= 0.9683(Y - 1.100) + 1.0436, && \text{for Lunar} \\ \text{sBMD}_N(O) &= 0.9743(Z - 0.969) + 1.0436, && \text{for Norland} \end{aligned} \tag{4}$$

We then applied the approximate method described in the Appendix to the same data and obtained the following approximate results for universal standardization:

$$\begin{aligned} \text{sBMD}_H(A) &= 1.0546(X - 0.972) + 1.0433, && \text{for Hologic} \\ \text{sBMD}_L(A) &= 0.9686(Y - 1.100) + 1.0433, && \text{for Lunar} \\ \text{sBMD}_N(A) &= 0.9745(Z - 0.969) + 1.0433, && \text{for Norland} \end{aligned} \tag{5}$$

Note that the approximate estimates are very close to the optimal estimates because the measurements of the 100 subjects on the three scanners were very close to a straight line (Fig. 1).

Table 1 presents descriptive statistics for the sBMD derived from the IDSC study (sBMD(I)), as well as the exact solution for the optimal procedure (sBMD(O)) and the approximate conversion method (sBMD(A)) described above. We show that using either the optimal or the approximate method the mean sBMD of the 100 subjects are identical for all three scanners, whereas this condition is neither imposed nor achieved in the IDSC method. Furthermore, the standard deviations of sBMD using either the optimal or the approximate method are nearly identical across the three scanners so that one unit difference in

TABLE 1. MEAN AND STANDARD DEVIATION (IN PARENTHESES) FOR ORIGINAL SPINE BMD MEASURED ON THREE SCANNERS AND STANDARDIZED BMD FROM THE IDSC STUDY (sBMD(I)), THE OPTIMAL METHOD (sBMD(O)), AND THE APPROXIMATE METHOD (sBMD(A)) FOR 100 SUBJECTS IN THE IDSC STUDY

Bone density measurements	Hologic	Lunar	Norland
BMD	0.9725 (0.1762)	1.1000 (0.1919)	0.9692 (0.1907)
sBMD(I)	1045.9 (189.6)	1047.4 (182.7)	1043.0 (205.3)
sBMD(O)	1043.6 (185.9)	1043.6 (185.8)	1043.6 (185.9)
sBMD(A)	1043.3 (185.9)	1043.3 (185.9)	1043.3 (185.9)

BMD measurements are in g/cm<sup>2</sup> and sBMD in milligram per square centimeter.

sBMD always means the same magnitude of difference regardless of the original scanner. In contrast, the IDSC's sBMD from different machines have standard deviations ranging from 0.183 to 0.205 g/cm<sup>2</sup>, a difference of over 10% between scanners. All three methods guarantee that the average sBMD on the reference phantom is equal to the "true" density.

We compared the performance of the three sBMD algorithms in several ways. First, Pearson correlation between sBMD from each pair of scanners was calculated for each algorithm. The three algorithms were indistinguishable by this criterion because all of the correlations were extremely high (>0.987). Therefore, we assessed the conversions by the mean squared differences in sBMD on the same individuals since sBMD is meant to be approximately the same for the same subject measured on different scanners. This is similar to comparing the mean squared error between different regression models. In Table 2, the overall difference for the IDSC method was 18% larger than for the optimal sBMD, which by definition should have the smallest overall mean squared difference. Indeed, the optimal method had smaller mean squared difference for every pair of scanners. Surprisingly, the approximate method had almost identical performance to that of the optimal method.

Finally, we compared the conversion algorithms by examining whether there were any systematic differences in sBMD between scanners in different ranges of measurements. To look for patterns of such biases, we calculated the mean and the difference for each subject's sBMD on each pair of scanners and plotted the difference against the mean sBMD (Fig. 2). Since some linear trends were apparent, we estimated the correlations between the differences and the means. In Table 3, the sBMDs derived from either the optimal or approximate method showed no correlation in any case, but the IDSC-derived between-scanner difference in sBMD was significantly correlated with the mean sBMD for all three pairs of scanners. The smallest of the significant correlations occurred between Hologic and Lunar be-

TABLE 2. COMPARISONS OF BETWEEN-SCANNER DIFFERENCES IN sBMD ACROSS THREE METHODS OF CONVERSION—ORIGINAL IDSC METHOD (sBMD(I)), OPTIMAL METHOD (sBMD(O)) AND APPROXIMATE METHOD (sBMD(A))

	sBMD(I)	sBMD(O)	sBMD(A)
Mean difference among scanners			
Hologic-Lunar	1.529	0.0	0.0
Lunar-Norland	4.464	0.0	0.0
Norland-Hologic	2.935	0.0	0.0
Root mean squared difference among scanners			
Hologic & Lunar	30.34	29.49	29.49
Lunar & Norland	36.08	26.79	26.79
Norland & Hologic	34.75	29.12	29.12
Overall	33.81	28.49	28.49

All values are given in milligrams per square centimeter.

cause the regression of Hologic on Lunar BMD had an almost zero intercept.<sup>(1)</sup> The strongest correlation was between Lunar and Norland, as can be seen in the top, middle panel of Fig. 2; it shows that the IDSC-derived sBMD is systematically higher for Lunar than Norland in the lower range of sBMD, while the opposite is true in the higher range. The systematic differences are about  $\pm 50$  mg/cm<sup>2</sup> at the high and low ends.

Table 4 compares the performance of the three conversion methods between a Lunar DPXL and a Hologic 1000W for a different group of 56 subjects in Indiana. The mean difference between scanners is marginally larger by the IDSC algorithm, but the magnitude of this overall bias of 0.6–0.7% is negligible in all cases. The root mean squared differences are approximately the same, about 3.5%, across the three methods. The correlation between individuals' mean sBMD and their differences on the two scanners are marginally stronger for the IDSC method but of no consequence in any case. Thus, the performances of the three algorithms are comparable for these two external machines.

## DISCUSSION

We have shown that the sBMD derived by the IDSC from three scanners, though highly correlated, had several methodologic problems. Fortunately, even the most severe problem of residual bias became negligible when the conversion formulas were applied to two other scanners made by Hologic and Lunar. This leads us to believe that the IDSC conversion formulas, which have already been implemented in the recently manufactured scanners, are very satisfactory for standardizing between Hologic and Lunar scanners. Even though some of the Indiana subjects were not measured on both machines on the same day, the individuals' differences in sBMD between Hologic and Lunar were of the order of 3.5%, which would rarely affect clinical decisions for individuals. To support the worldwide adoption of the IDSC-derived sBMD for clinical use, our findings

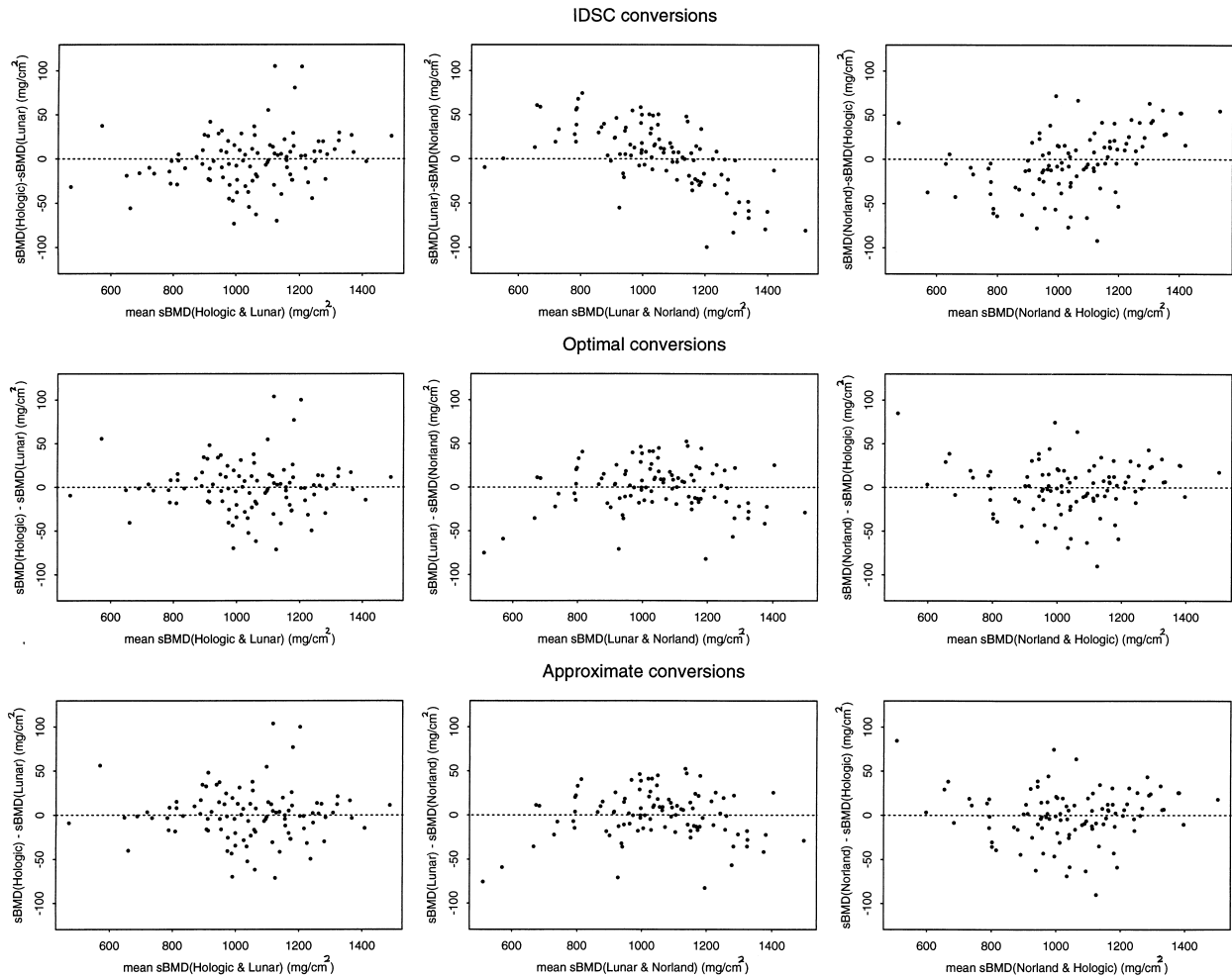


FIG. 2. Plots of the difference in sBMD versus the mean sBMD of 100 individuals, by pairs of scanners and by these conversion methods for sBMD.

TABLE 3. CORRELATIONS OF INDIVIDUAL DIFFERENCES AND MEANS OF sBMD FOR EACH PAIR OF SCANNERS—COMPARISONS ACROSS sBMD(I), sBMD(O), AND sBMD(A)

	sBMD (I)	sBMD (O)	sBMD (A)
Pairs of scanners			
Hologic-Lunar	0.23*	0.00	0.00
Lunar-Norland	-0.63†	-0.00	-0.00
Norland-Hologic	0.45†	-0.00	0.00

\*  $p < 0.05$ .

†  $p < 0.01$ .

should be corroborated with data from other cross-calibration studies based on different machines. In particular, the agreement between Norland sBMD and other manufacturers needs to be more broadly established since there were some systematic differences in the original calibration study.

TABLE 4. DIFFERENCES IN sBMD BETWEEN HOLOGIC AND LUNAR SCANNERS FOR 56 SUBJECTS MEASURED IN INDIANA COMPARISONS ACROSS THREE METHODS OF DERIVING sBMD

Difference in sBMD	sBMD (I)	sBMD (O)	sBMD (A)
Mean ( $\text{mg}/\text{cm}^2$ )	-7.02	-6.28	-6.30
Root mean squared difference ( $\text{mg}/\text{cm}^2$ )	34.67	34.35	34.37
Correlation with mean sBMD	0.14	-0.11	-0.11

In research studies, investigators always try to use the same instruments throughout a project. Over the years, however, bone laboratories need to update their scanners to keep up with technological advances. To explore certain research questions, sometimes it is expeditious to perform analyses on data that have been acquired on different scanners. Therefore, it makes sense to have conversions from the BMD measured on all the scanners in a laboratory to a common standard. The demand for precision and freedom from bias is more stringent in addressing research questions than for making clinical decisions for an individual.

We showed that optimal internal standardization could attain overall between-scanner differences in sBMD of under 3% while 3.5% difference was observed with one external application. Furthermore, a systematic bias of 50 mg/cm<sup>2</sup> for measurements of sBMD in the neighborhood of 600 mg/cm<sup>2</sup> would be unacceptable in research. Therefore, research laboratories should derive their own internal standardization that is optimized for their particular machines in the laboratory. Our proposed method should be used because it improves upon the IDSC method while retaining all of the desirable properties set forth by the IDSC. The improvement could be even more marked for standardizing BMD at other skeletal sites or between other machines if the linear relationships between BMDs have larger nonzero intercepts. The new method is also more flexible. For example, if one machine is known to have larger measurement error than the others, then the least-squares criterion can be modified to a weighted least-squares criterion whereby those differences with larger errors can be given smaller weights.

The proposed method is only appropriate after traditional regression analysis has first established the linear relationships between BMD measured on the machines to be standardized. If a simple linear relationship does not hold, then none of the existing conversion methods is appropriate. New methods will need to be developed for nonlinear conversions. Another situation that cannot be handled by available methods is the conversion of longitudinal data. When serial measurements have been made on an individual and a change in scanner is unavoidable, there is a need for a conversion method that optimizes the measurement of change in BMD.

One advantage of the proposed method in this paper is that one can sample the subjects in any manner in the cross-calibration study. For example, one can over-sample the two extremes on big men and children to add stability to the estimated relationships. Alternatively, one may desire higher precision for cross-calibration in the lower range of BMD since it is people with low BMD who are usually of primary concern in the field of osteoporosis. If so, one can over-sample the low end of BMD by measuring more frail elderly subjects in the cross-calibration study. This is the reason it was appropriate to evaluate the methods on the Indiana subjects even though they were not properly sampled. Consequently, however, a shortcoming shared by the IDSC and our proposed methods is that variance estimates for the conversion parameters can only be obtained by resampling methods such as bootstrapping. If one is willing to give up the choice of sampling schemes, and if the cross-calibration study subjects' BMD on different machines can be assumed to have a multivariate normal distribution, then one may choose to use the method by Lu et al.<sup>(6)</sup> The advantage of this method is that it is based on maximum likelihood estimation which produces asymptotic variance estimates for the conversion parameters, so statistical inference is more straightforward. Lu's parameter estimates from data from the IDSC cross-calibration study are very similar to the estimates presented in this paper. The new conversions are likely to be adopted by the manufacturers for future standardization of BMD at skeletal sites other than the spine.

In conclusion, the conversions of spine BMD to sBMD that are now available on DXA bone absorptiometers are adequate for clinical use. However, researchers who want to derive their own conversions should use the methods proposed in this article or the one by Lu.<sup>(6)</sup>

## APPENDIX

### *Derivation of optimal conversion equations*

After establishing linear relationships between the BMD from all pairs of scanners, we proceeded to derive conversion equations using the new method. First, we calculated the means of the 100 subjects' measurements on each scanner:  $\bar{X} = 0.972$ ,  $\bar{Y} = 1.100$ , and  $\bar{Z} = 0.969$ , and subtracted the respective mean from each individual BMD measurement to obtain:

$$x = X - 0.972$$

$$y = Y - 1.100$$

$$z = Z - 0.969$$

To minimize  $E^2$  in expression (3) subject to the constraint  $a^2 + b^2 + c^2 = L$ , an arbitrary constant, we first had to choose an appropriate  $L$ . If we want differences of 1 unit in sBMD to be similar to unit differences in BMD, then  $a$ ,  $b$ , and  $c$  need to be close to 1. A natural choice of  $L$  could be  $1^2 + 1^2 + 1^2 = 3$ , which preserves the overall size of the measurements, but other choices could be justified by other criteria. Having chosen  $L$ , we minimized  $E^2$  with an undetermined Lagrange multiplier  $\lambda$  for the constraint. That is, we minimized

$$\begin{aligned} & \Sigma(ax - by)^2 + \Sigma(by - cz)^2 + \Sigma(cz - ax)^2 \\ & + \lambda(a^2 + b^2 + c^2 - L) \end{aligned} \quad (6)$$

with respect to  $a$ ,  $b$ ,  $c$ , and  $\lambda$ . Differentiating expression (6) with respect to  $a$ ,  $b$ ,  $c$ , and  $\lambda$  and setting the differentials to zero, we obtained:

$$2a\Sigma x^2 - b\Sigma xy - c\Sigma zx + \lambda a = 0$$

$$2b\Sigma y^2 - c\Sigma yz - a\Sigma xy + \lambda b = 0$$

$$2c\Sigma z^2 - a\Sigma zx - b\Sigma yz + \lambda c = 0$$

$$a^2 + b^2 + c^2 - L = 0$$

This is a set of four simultaneous equations in four unknowns in second order. In general, no closed-form solution can be obtained, so we had to put in the data  $x$ ,  $y$ ,  $z$ , and the constant  $L$  and solved the equations numerically using Maple, a symbolic programming language.

After we obtained the solutions for  $a$ ,  $b$ , and  $c$ , constants were added to the scaled measurements to obtain sBMD for all scanners in the form:

$$\text{sBMD}_H = 1.0550x + K, \quad \text{for Hologic}$$

$$\text{sBMD}_L = 0.9683y + K, \quad \text{for Lunar}$$

$$\text{sBMD}_N = 0.9743z + K, \quad \text{for Norland}$$

Addition of a common  $K$  preserves the optimal criterion of least-squares and does not introduce any systematic bias in

any range of sBMD between scanners. Thus, the equality of mean sBMD is also preserved across all scanners. The estimate of  $K$  was used to calibrate the sBMD of the phantom, midvertebra ESP, to its theoretical density. The phantom BMD was 0.916 g/cm<sup>2</sup> on the Hologic, 1.074 g/cm<sup>2</sup> on the Lunar, and 0.922 g/cm<sup>2</sup> on the Norland. Equating the mean phantom sBMD from the three scanners to the true density of 1.0 g/cm<sup>2</sup>,  $[1.0550(0.916 - 0.972) + K + 0.9683(1.074 - 1.1) + K + 0.9743(0.922 - 0.969) + K]/3 = 1.0$  resulted in  $K = 1.0436$ . Thus, the conversion equations:

$$\text{sBMD}_H = 1.0550x + 1.0436, \quad \text{for Hologic}$$

$$\text{sBMD}_L = 0.9683y + 1.0436, \quad \text{for Lunar}$$

$$\text{sBMD}_N = 0.9743z + 1.0436, \quad \text{for Norland}$$

are presented as the optimal conversion Eqs. (4) in the text.

No closed-form solutions exist for variance estimates for the parameter estimates  $a$ ,  $b$ ,  $c$ , and  $K$ . One way to obtain such estimates is to use a bootstrap procedure as follows:

- (1) Randomly sample 100 subjects' measurements, with replacement, from the original sample to form a bootstrap sample.
- (2) Estimate  $a$ ,  $b$ ,  $c$ , and  $K$  from the bootstrap sample, and save the estimates.
- (3) Repeat steps 1 and 2 many times (usually thousands) to obtain a stable sampling distribution of the estimates.
- (4) The standard deviations of the estimates of  $a$ ,  $b$ ,  $c$ , and  $K$  in the sampling distribution give the standard errors of the estimates.

As one can tell from the steps described, it is a highly computing-intensive way to obtain these variance estimates. Unless one is interested in performing statistical inference, such as comparing two conversion parameters, the gain in information may not be worth the effort since linearity has previously been established and optimal properties of the sBMD conversion, such as least-squares, are guaranteed. Thus, no such variance estimates were calculated.

#### *Derivation of approximate conversion equations*

The exact solution for the optimal conversion requires a symbolic programming language such as Maple or *Mathematica*, which is not as widely available to researchers as some common statistical packages. We therefore developed the following approximate solution to Eqs. (3). The approximate method started with interconversion between scanners. Again, we removed the sample means from BMD initially and converted among  $x$ ,  $y$ , and  $z$  through  $x = k_1y$ ,  $y = k_2z$ , and  $z = k_3x$ , with  $k_1k_2k_3 = 1$ . Backward conversions through the same equations, e.g.,  $y = x/k_1$ , were also internally consistent. To estimate  $k_1$ , we first obtained two no-intercept regression lines of  $x$  on  $y$  and  $y$  on  $x$ . In the example, least-squares regression with no intercept resulted in:

$$x = 0.90678y$$

$$y = 1.07493x$$

$$y = 0.99549z$$

$$z = 0.98356y$$

$$z = 1.06882x$$

$$x = 0.91256z$$

The slope  $k_1$  was chosen to bisect the angle defined by the two regression lines between  $x$  and  $y$ . If the two regression lines are given by  $x = b_1y$  and  $y = b_2x$ , then  $k_1 = \tan [0.5(\arctan b_1 + \arctan (1/b_2))]$ . Similarly,  $k_2$  and  $k_3$  can be estimated and it can be shown that  $k_1k_2k_3 = 1$ . In the example,  $k_1 = 1.0888$ ,  $k_2 = 0.9940$ ,  $k_3 = 0.9240$ . We then estimated  $a$ ,  $b$ , and  $c$  for sBMD through  $k_1 = a/b$ ,  $k_2 = b/c$ , and  $k_3 = c/a$  with  $a^2 + b^2 + c^2 = 3$ . This can be done on a calculator by first letting, say,  $a' = 1$  and solving any two of the first three equations, e.g.,  $b' = 1/k_1 = 0.9185$  and  $c' = k_3 = 0.9940$ . Then we multiply  $a'$ ,  $b'$ , and  $c'$  by a constant,  $R$ , such that the last normalizing equation ( $a^2 + b^2 + c^2 = 3$ ) is satisfied. Solving  $R^2(a'^2 + b'^2 + c'^2) = 3$ , i.e.,  $R^2(1^2 + 0.9185^2 + 0.9940^2) = 3$ , gives  $R^2 = 1.1122$  and  $R = 1.0546$ . Therefore,  $a = Ra' = 1.0546$ ,  $b = Rb' = 0.9686$ , and  $c = Rc' = 0.9745$ . Substituting in the phantom data, we obtained  $K = 1.0433$ . Again bootstrap procedures could be used for obtaining variances for parameter estimates.

#### ACKNOWLEDGMENTS

This study was partially supported by National Institutes of Health grants AG04518 and AG05793. We thank Drs. C. Conrad Johnston, Jr. and Charles W. Slemenda for helpful discussions and Mrs. Crystal Wampler for secretarial assistance. The reviewers' many helpful suggestions are also gratefully acknowledged.

#### REFERENCES

1. 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.
2. Finkelstein JS, Butler JP, Cleary RL, Neer RM 1994 Comparison of four methods for cross-calibrating dual-energy X-ray absorptiometers to eliminate systematic errors when upgrading equipment. *J Bone Miner Res* **9**:1945-1952.
3. Peel NFA, Eastell R 1995 Comparison of rates of bone loss from the spine measured using two manufacturers' densitometers. *J Bone Miner Res* **10**:1796-1801.
4. Genant HK 1995 Universal standardization for dual X-ray absorptiometry: Patient and phantom cross-calibration results. Letter to the Editor. *J Bone Miner Res* **10**:997-998.
5. Steiger P 1995 Standardization of measurements for assessing BMD by DXA. *Calcif Tissue Int* **57**:469.
6. Lu Y, Ye K, Ashwini MK, Hui S, Fuerst TP, Genant 1997 Comparative calibration without a gold standard. *Stat Med* (in press).

Address reprint requests to:

*Siu L. Hui, Ph.D.*

*Professor of Medicine*

*Regenstrief Institute*

*1001 West 10th Street, RG6*

*Indianapolis, IN 46202 U.S.A.*

Received in original form June 6, 1996; in revised form March 31, 1997; accepted April 14, 1997.