Precision of Dual Photon Absorptiometry Measurements

Adrian D. LeBlanc, Harlan J. Evans, Cherri Marsh, Victor Schneider, Philip C. Johnson, and Satish G. Jhingran

Department of Medicine, Baylor College of Medicine; and Johnson Space Center, NASA, Houston, Texas

One of the important uses of bone absorptiometry is to examine the rate of bone mineral change in order to evaluate therapy and to identify individuals who need therapy. Generally, this involves comparing the difference between two scans obtained months to years apart. This study investigates the precision of dual photon absorptiometry using a human torso phantom, normal subjects, and abnormal patients. These studies showed that bone mineral calculated as g/cm^2 was more precise than g/cm. Reanalysis of the same scan by the same individual produced an average error equivalent to that produced by scanning and analyzing the same subject on multiple occasions. Interobserver analysis error was essentially equal to the intraobserver error. In order to obtain maximum precision, care must be taken that the integrated area of a repeat scan is identical to the previous scan. Our findings indicate that to be confident (95%) of a real change between two scans a difference of at least 5.6% must be measured.

J Nucl Med 27:1362–1365, 1986

Over the last 20 years, several techniques have been used to assess bone mass: simple radiography, neutron activation analysis, radiogammetry, and absorptiometry (*1–8*). In the early 1970s single photon absorptiometry gained popularity because of the good precision and accuracy (~3%), and the low radiation exposure (~5 mrem). However, single photon absorptiometry is limited to the peripheral skeleton which, although correlated with the clinically important axial skeletal regions, is not optimal for diagnosis of axial skeletal disease (*9–10*). Dual photon absorptiometry allows the assessment of bone mineral of the lumbar spine and femoral neck. These instruments are useful for establishing present or future fracture risk and to monitor change. Spinal fracture risk is believed to be highly correlated with skeletal bone mineral, and good accuracy is important. The accuracy is ~5% (*11*). Since fracture risk is not believed to be significantly increased until bone mineral has decreased by 20% (bone mass is 80% of normal), an accuracy of 5% seems adequate and is not likely to significantly affect diagnosis (*12,13*).

Precision is of primary importance in determining temporal changes in bone mineral to decide whether an observed difference is real or not. Temporal changes are monitored to evaluate therapy and to identify those who need therapy (fast vs. slow bone mineral loss). Precision needs to be known for initial design and analysis of results of longitudinal studies.

Our laboratory has been using commercial dual photon instruments for ~3.5 yr and has performed over 1,000 scans of patients and research subjects. It is the purpose of this paper to report on a study of long-term precision of two dual photon absorptiometers using scans of phantoms, normal, and abnormal subjects. We have investigated actual and potential inter- and intraobserver variability of bone mineral calculated by dividing grams by scanned area (BMD) or scanned length (BML).

MATERIALS AND METHODS

All studies were conducted using one of two available dual photon absorptiometers* employing gadolinium-153†. Both instruments were set at the two photon peaks daily and standard values were obtained weekly from an average of 40 scans of the bone mineral standard. All results were obtained using version 5 of the scanner software and an 8-mm collimator. Five observers took part in the scan analysis; four trained and one untrained individual given minimal (<30 min) training but who was knowledgeable with the computer instrumentation.

Received Sept. 5, 1985; revision accepted Feb. 19, 1986.

For reprints contact: Adrian LeBlanc, PhD, The Methodist Hospital, 6565 Fannin, M.S. S396, Houston, TX 77030.

A lower torso phantom (thigh through lumbar vertebra) was used in this study and consisted of tissue equivalent material and human bones similar to that normally used in radiotherapy departments except that the metal antimony was not added to this particular phantom because of the possibility of altering the absorption coefficients of gadolinium. A scan of this phantom was obtained on each of the two scanners once or twice per month for ~1 yr. During this time the gadolinium source was changed on one of the instruments. The results from the two instruments were evaluated by one observer for the entire period. In every case, the integrated area was L2 through L4 and the results from the automatic and manual methods were recorded. The automatic method used the computer determined values for bone edges and baseline while the manual method involved a line by line inspection of edges and baseline. The bone mineral is reported both as BMD (g/cm^2) and BML (g/cm). The BMD is obtained by dividing bone mineral in grams by the integrated areas in cm^2 while the BML is obtained by dividing the bone mineral in grams by the length in cm of the scanned area. In each case the mean and standard deviation (s.d.) and coefficient of variation were calculated. These measurements allow a comparison of the two instruments directly in terms of accuracy and reproducibility as well as comparing the automatic and manual methods for calculating bone mineral.

Eleven normal male volunteers were scanned once a week for five consecutive weeks and the mean ± s.d. and percent coefficient of variation (% CV) of BMD and BML were calculated for each subject. The mean of the individual coefficient of variations was then obtained. The area of integration was L2 through L4. Since the auto method of analysis periodically gave aberrant values only the manual method was used for the subject and patient analysis. Since these studies were conducted in connection with a metabolic study, the subjects had similar activity during this period and were on a controlled diet containing 1,000 mg calcium per day. These studies represent the precision that can be expected with normal subjects with a single individual analyzing the data over a short period of time.

Six normal scans (BMD >1.0 g/cm^2) and six abnormal scans (BMD <0.96 g/cm^2) were evaluated as described above by four trained and one minimally trained observer. Several weeks afterward the same individuals were asked to reanalyze the same 12 scans after the edges and baselines were changed back to that originally established by the instrument. These intraobserver results are expressed as the mean ± s.d. of the percent differences between the two scans for each observer. The trained and untrained observers were evaluated separately. In addition, the values for BMD and BML were obtained from all four trained observers using the same region of interest to simulate the variability produced by different observers trying to repeat the same scan. This would occur, for example, when a patient received a repeat scan a year later but with the analysis being done by a different observer. These interobserver results are reported as the mean % eV and s.d. of normal and abnormal scans. The reported values are obtained by calculating the mean of the % CV of the four observers. In the case of the untrained observer the mean of the normal and abnormal scans are compared and results expressed as the percent difference from the mean of the four trained observers. It was clear from these studies that the integration area is crucial in order to obtain high precision, i.e., the region to be integrated must be matched directly with the current scan. In order to demonstrate this further, one observer analyzed all 12 scans integrating a region shifted three lines above the originally chosen lines and L1–L3. The BMD and BML values were compared with the original L2–L4 values.

RESULTS

Table 1 summarizes the torso phantom measurements from both dual photon instruments. Instrument No. 1 gave slightly (2–5%) lower values than instrument No. 2. The BMD percent coefficient of variation by manual calculation was 2.2 and 1.8% for the two instruments while the auto mode was 2.6 and 2.1%. There was a significant difference between manual and auto BMD for both instruments, i.e., 3.5 and 6.1% which was not the case for BML, i.e., 0.4 and 1.2%. In every case, BML gave smaller % CV than BMD indicating better reproducibility. This is not the case for patient scans as will be shown later. No effect from the source change could be distinguished from a plot of BMD with time.

Table 2 provides the BMD and BML results from 11 subjects each scanned once a week for five consecutive weeks. The mean % CV for BMD was 1.9 ± 0.6% while BML was somewhat larger, 2.5 ± 0.7%.

Table 3 gives the intraobserver error as the mean % difference and s.d. of duplicate analysis of normal and abnormal scans for all five observers. These results suggest that a second BMD analysis of a scan will be within 3.5% (95% confidence level) for either normal or abnormal bone mineral content. For an untrained observer, the error is larger especially with abnormal bone mineral scans. As the data showed in Table 2, BML gave less precise results than BMD.

The untrained observer, while performing less well than the trained observer, nevertheless did quite well reproducing the initial analysis. However, comparing the actual scan values to those of the trained observers, the error was significant (p

TABLE 1
BMD and BML of Torso Phantom from Two Densitometers over 1-year Period

	Instrument No. 1			Instrument No. 2		
	Manual	Auto	Δ%	Manual	Auto	Δ%
BMD ± s.d. (g/cm^2)	1.122 ± 0.025	1.083 ± 0.028	3.5	1.171 ± 0.021	1.100 ± 0.023	6.1
% CV	2.2	2.6		1.8	2.1	
BML ± s.d. (g/cm)	4.78 ± 0.09	4.80 ± 0.09	0.4	4.99 ± 0.06	5.05 ± 0.06	1.2
% CV	1.9	1.9		1.2	1.2	
Number of scans	26	16		20	17	

TABLE 2
BMD and BML of 11 Normal Subjects Measured Each Week for Five Consecutive Weeks (n = 5)

Subject	BMD (g/cm²) $\bar{X}$ ± s.d.	% CV	BML (g/cm) $\bar{X}$ ± s.d.	% CV
1	1.135 ± 0.018	1.6	5.20 ± 0.09	1.7
2	1.175 ± 0.014	1.2	4.76 ± 0.09	1.9
3	1.176 ± 0.027	2.3	4.59 ± 0.14	3.0
4	1.544 ± 0.027	1.7	6.44 ± 0.11	1.7
5	1.071 ± 0.032	3.0	4.56 ± 0.16	3.5
6	1.252 ± 0.019	1.5	5.98 ± 0.23	3.8
7	1.239 ± 0.013	1.0	5.43 ± 0.12	2.2
8	1.137 ± 0.029	2.6	4.67 ± 0.15	3.2
9	1.372 ± 0.022	1.6	5.67 ± 0.13	2.3
10	0.974 ± 0.023	2.4	4.53 ± 0.12	2.6
11	0.985 ± 0.016	1.6	4.30 ± 0.09	2.1
$\bar{X}$ ± s.d.		1.9 ± 0.6		2.5 ± 0.7

< 0.001), i.e., normal scan BMD was 6.5 ± 1.6%, and abnormal scan BMD was 9.6 ± 2.2%.

Table 3 also gives the interobserver results where different observers (experienced only) are compared against each other. The percent coefficient of variation for each patient is calculated and the mean presented for the six normals and six abnormals. Comparison of interobserver error with Table 2 indicates that experienced observers analyzing the same scan do not increase the expected error compared to a single individual analyzing all scans. A repeat scan performed a year after the first does not require the same individual for analysis to preserve maximum precision.

Table 4 shows the error produced by integrating an area shifted three lines or one vertebra for both BMD and BML. This table shows the large error that can result if a different area is chosen in an otherwise identical scan and indicates the need to carefully compare new scans with old in order to obtain maximum precision.

TABLE 3
Intra- and Interobserver Variability—Mean Percent (± s.d.)

	Intraobserver*			
	Normals (n = 6)		Abnormals (n = 6)	
	BMD	BML	BMD	BML
Trained observers (n = 4)	0.3 ± 1.6	0.6 ± 2.5	0.0 ± 1.8	−1.3 ± 4.1
Untrained observers (n = 1)	0.9 ± 1.4	0.1 ± 1.2	−3.2 ± 2.7	−3.1 ± 3.9
	Interobserver†			
Trained observers (n = 4)	1.2 ± 0.6	1.2 ± 0.2	1.9 ± 0.3	1.6 ± 0.8

*Each observer (n = 4) analyzed each scan twice several weeks apart. Mean and s.d. of percent differences were then calculated.

†Each scan was analyzed by all four trained observers and percent coefficient of variation calculated. Data were separated into normals and abnormals and mean and s.d. of percent coefficient of variations calculated.

TABLE 4
Effect of Integration Area on BMD and BML—Percent Change ± s.d. from Initial (L2–L4) Values

	Normals (n = 6)		Abnormals (n = 6)	
	Shifted 3 lines	L1–L3	Shifted 3 lines	L1–L3
BMD (g/cm²)	−1.8 ± 1.5	−3.4 ± 2.5	−1.8 ± 1.1	−3.7 ± 2.9
BML (g/cm)	−4.9 ± 0.9	−8.2 ± 1.8	−7.2 ± 4.4	−6.9 ± 3.9

DISCUSSION

Obviously, if an absorptiometer is improperly peaked or calibrated large errors are possible. Other than these obvious sources of error we experienced two instances of potential error attributable to instrumentation. Although counting error is not significant, if the true background changes and/or is otherwise different from that in the computer software, significant error (10%) can result especially when the instrument's radiation source is weak and/or the patient is large. Therefore, background should be monitored routinely to insure that the actual background value corresponds with that in the computer software. Analyzer window size or detector resolution changes can occur which will affect deadtime, energy window spillover correction, and the air values. As a continuous check for this, we have modified the air value printout to compare current air values to the decay corrected original values. A difference of greater than a few percent is a signal to investigate the system. Long-term drifts in the equipment is another factor which can affect precision but over which the operator has little control. There is no other equipment problem that we have encountered that could cause errors without a clear indication of a malfunction.

There are three parameters in the scan analysis over which the operator has control that can affect precision and, to a lesser extent, accuracy. These are adjustments to the baseline and edges and selection of integrated area. It was the purpose of this paper to analyze these effects with the aim of assessing the minimum change in an observed measurement that can be ascribed as real. We determined this under a number of expected situations, i.e., abnormal vs. normal values, single vs. multiple observers, manual vs. auto analysis, integration area selection, phantom vs. human scans.

We showed using a torso phantom that the auto and manual methods gave significantly different BMD values with the manual method having better precision. The phantom BML gave better precision than BMD, and the auto or manual methods had the same precision and essentially the same mean values. This would indicate that the manual/auto BMD differences result

from a difference in edge selection. This study was carried out using version 5 software. Since this study began, new analysis software, version 7, has been released by the manufacturer. The principle difference between version 5 and 7 software is that version 7 has a more compressed data storage but there are also differences in the edge detection and baseline algorithms. These changes might improve the automatic edge detection and therefore improve precision in this mode. Since most of the scans were analyzed using the manual mode the conclusions and maximum attainable precision should not be affected by this change in software. The two instruments differed significantly by 2–5% depending on whether the phantom analysis was for BMD or BML using the manual or auto method of analysis. No explanation for this difference is known at this time.

The average BMD percent precision error of the two instruments was 4% at a 95% confidence level. Two scans having this scan error would have to differ by more than 5.6% ($\sqrt{(4)^2 + (4)^2}$) to be confident that a real change had occurred. Since phantoms may not vary to the same degree as human scans, a series of repeat scans (five) all analyzed by the same individual, are presented in Table 2. The BMD average % CV for 11 subjects was 1.9% or 3.1% at a 95% confidence level (2 s.d.). If only two measurements are made with this same average error then one would have to observe a >4.4% difference in BMD to have confidence (95%) that a real change had occurred. The BML value would need to be >5.5%. The precision of these short-term measurements (5 wk) are slightly better than the long-term (1 yr) phantom measurements.

Reanalyzing the same scan by an observer produces a BMD precision error essentially the same as when multiple scans of the same subject are analyzed. This would indicate that the scanning by itself does not introduce substantial error over that present in the analysis. Interobserver error was the same as intraobserver error indicating substantial error is not introduced by having different observers analyze a repeat scan. This was opposite to what we initially believed but it is important that all observers have similar training. Individuals with minimal training are able to be precise, but large errors can result if compared with a scan analyzed by a differently trained individual.

To obtain maximum precision, a previous scan must be compared directly with the current one. It is essential that the same area be compared or 2–4% additional error can result. It is usually more difficult to determine individual vertebra on patient scans and, therefore, it is more difficult to compare patient scans with previous scans. Table 4 demonstrated this as well as the greater error if BML is used. Overall BML proved to be less precise than BMD. The only exception to this was the phantom scans probably because the vertebrae in the phantom are further apart and more easily distinguished. We therefore conclude that BMD is the preferred parameter for examining repeat scans.

FOOTNOTES

* Lunar Radiation Corp. (Model DP3). (Both instruments are similar but not of identical design. Instrument No. 1 was the first commercial version built by this company.)

† Gulf Nuclear Inc., Webster, TX.

ACKNOWLEDGMENTS

The authors thank Jean Krebs and David Engelbretson for their assistance.

REFERENCES

1. Evans HJ, LeBlanc AD, Johnson PC: Feasibility study: In vivo neutron activation for regional measurement of calcium using Cf-252. *Med Phys* 3:148–152, 1976
2. Al-Hiti K, Thomas H, Altikrity SA, et al: Spinal calcium: Its in vivo measurement in man. *Int J Appl Rad Isot* 27:97–102, 1976
3. Barnett E, Nordin BEC: The radiological diagnosis of osteoporosis. A new approach. *Clin Radiol* 11:166–174, 1960
4. Virtama P, Mahonen H: Thickness of cortical layer as an estimate of mineral content of human finger bones. *Br J Radiol* 33:60–62, 1960
5. Cameron JR, Sorenson J: Measurement of bone mineral in vivo: An improved method. *Science* 142:230–236, 1963
6. Reed GW: The assessment of bone mineralization from the relative transmission of Am-241 and Cs-137 radiations. *Phys Med Biology* 11:174, 1966
7. Mazess RB, Ort M, Judy P, et al: Absorptiometric bone mineral determination using Gd-153. In *Proceedings of Bone Measurement Conference*, (Conf. 700515) Cameron JR, ed. U. S. Atomic Energy Commission, 1970, pp 308–312
8. Cann CE, Genant HK: Precise measurement of vertebral mineral content using computed tomography. *J Comp Assist Tomogr* 4:493–500, 1980
9. Krolner B, Nielson SP, Lund B, et al: Measurement of bone mineral content (BMC) of the lumbar spine. II. Correlation between forearm BMC and lumbar spine BMC. *Scand J Clin Lab Invest* 40:665–670, 1980
10. Mazess RB, Peppler WA, Chesney RW, et al: Does bone measurement on the radius indicate skeletal status? Concise communication. *J Nucl Med* 25:281–288, 1984
11. Wahner HW, Dunn WL, Mazess RB, et al: Dual-photon Gd-153 absorptiometry of bone. *Radiology* 156:203–206, 1985
12. Melton LJ, Wahner HW, Richelson LS, et al: Bone density-specific fracture risk: A population-based study of the relationship between osteoporosis and vertebral fractures. *J Nucl Med* 26:24, 1985
13. Riggs BL, Wahner HW, Dunn WL, et al: Differential changes in bone mineral density of the appendicular and axial skeleton with aging: Relationship to spinal osteoporosis. *J Clin Invest* 67:328–335, 1981