View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DSpace at VU

Reproducibility of bone mineral density measurement in daily practice

M C Lodder, W F Lems, H J Ader, A E Marthinsen, S C C M van Coeverden, P Lips, J C Netelenbos, B A C Dijkmans, J C Roos

Ann Rheum Dis 2004;**63**:285–289. doi: 10.1136/ard.2002.005678

See end of article for authors' affiliations

Correspondence to: Dr M C Lodder, Department of Rheumatology, VU University Medical Centre, Room 4A42, PO Box 7057, 1007 MB Amsterdam, The Netherlands; secr.reumatologie@vumc.nl

Accepted 28 April 2003

Background: Bone mineral density (BMD) measurements are frequently performed repeatedly for each patient. Subsequent BMD measurements allow reproducibility to be assessed.
Objective: To examine the reproducibility of BMD by dual energy *x* ray absorptiometry (DXA) and to investigate the practical value of different measures of reproducibility in a group of postmenopausal women.
Methods: Ninety five women, mean age 59.9 years, underwent two subsequent BMD measurements of spine and hip. Reproducibility was expressed as smallest detectable difference (SDD), coefficient of variation (CV), and intraclass correlation coefficient (ICC). Sources of variation were investigated by multilevel analysis.
Results: The median interval between measurements was 0 days (range 0–45). The mean difference (SD) between the measurements (g/cm^2) was −0.001 (0.02) and −0.0004 (0.02) at L1-4 and the total hip, respectively. At L1-4 and the total hip, SDD (g/cm^2) was ±0.05 and ±0.04 and CV (%) was 1.92 and 1.59, respectively. The ICC at spine and hip was 0.99.
Conclusions: Reproducibility in the postmenopausal women studied was good. In a repeated DXA scan a BMD change exceeding $2\sqrt{2}$CV (%), the least significant change (LSC), or the SDD should be regarded as significant. Use of the SDD is preferable to use of the CV and LSC (%) because of its independence from BMD and its expression in absolute units. Expressed as SDD, a BMD change of at least ±0.05 g/cm^2 at L1-4 and ±0.04 g/cm^2 at the total hip should be considered significant.

Bone densitometry was developed for diagnosis and treatment evaluation of osteoporosis. Dual energy *x* ray absorptiometry (DXA) is the most widely used modality for bone mineral density (BMD) measurement.[1 2] The definitions of osteopenia and osteoporosis, as proposed by the WHO, are based on results of DXA measurements.[3] Meanwhile, therapeutic options for treatment of osteoporosis have been developed which create possibilities of effective intervention. Therefore, screening for and treatment of osteoporosis are widely practised in postmenopausal women and in people with an increased risk of osteoporosis because of underlying diseases.[4 5] It has become more and more common practice to perform a second DXA measurement to monitor BMD status or the effect of therapeutic intervention, even though BMD is a surrogate for fracture risk. The reproducibility of DXA measurements—that is, the ability of an instrument to reproduce the same results in several measurements, is claimed to be good. Precision, expressed as percentages, of phantom measurements is around 0.5%. In in vivo studies in healthy young people, these figures range from 0.6 to 1.4, depending on machine type, study group, and measurement site.[6–8]

When a second measurement is performed on a patient, the BMD change is considered statistically relevant if it exceeds at least twice the precision error of the measurement.[9–11] Despite the abundance of publications on BMD variability in different patient groups, only limited data from postmenopausal women, the patients commonly considered for BMD measurement, are available on short term BMD variability.[6 12 13] Note that reproducibility studies established by pre- and post-measurements are different from repeated measurements studies of BMD where the prime focus is on *change* in BMD as opposed to measurement *variability*.

The precision error is usually expressed as the coefficient of variation (CV),[14–16] although several other statistics to express reproducibility exist and other measures, such as the smallest detectable difference (SDD), may be preferable to the CV.[13 17] Data on potential sources of measurement variability, related to the device, technician, or patient, show conflicting results.[13 18 19] For example, BMD measurement error was independent of age in one study,[18] whereas others found greater measurement error in older osteoporotic subjects[13] or reproducibility dependent on age related factors other than BMD.[19] Therefore, we investigated short term variability, the practical significance of different measures of variability, and the sources of variability in postmenopausal women in a university hospital. In addition, short term variation in a limited number of children was investigated because measurements on bone density in children are more often performed than in the past and precision in this group is assumed to be better than in adults.[2 20]

METHODS

Subjects

The subjects were recruited in one centre from participants in four studies. The first three of these studies concerned randomised clinical trials on the effects of the selective oestrogen receptor modulator raloxifene.[21–23] During two months in 1998, consecutive postmenopausal women undergoing a BMD measurement of lumbar spine and femur who had had several BMD measurements in the past (that is, between July 1994 and June 1997) were recruited. The maximum

Abbreviations: BMD, bone mineral density; BMI, body mass index; CV, coefficient of variation; DXA, dual energy *x* ray absorptiometry; ICC, intraclass correlation coefficient; LSC, least significant change; SDD, smallest detectable difference

interval between the two DXA measurements was 45 days. A change in BMD was not expected during this interval.

At the same time girls participating in a local study on the effects of puberty on BMD had their BMD measured twice.[24]

Ethics

All four study protocols were approved by the local medical ethics committee.

Methodology of BMD measurement

All BMD measurements were performed on a Hologic QDR 2000 machine (Hologic Inc, Waltham, Massachusetts). The software version used was V4.7. The DXA scans were obtained by standard procedures supplied by the manufacturer for scanning and analysis. The compare feature was used for the second scan. No records were kept of difficulties observed in the positioning of patients. Plain *x* rays documenting the presence of arthritic changes were not used. Daily quality control was carried out by measurement of a Hologic anthropomorphic spine phantom. At the time of the duplo measurements, phantom measurements showed stable results. The phantom precision expressed as the CV (%) was 0.82. The BMD measurements were carried out by experienced technicians.

Patient BMD was measured at the lumbar spine (L1-4) (posteroanterior projection) and at the left femur. At the femur the following sites were studied: total hip, femoral neck, and trochanter. When the two measurements were made on the same day, the patient was completely repositioned after the initial measurement.

T and Z scores were calculated using the reference population provided by the manufacturer. In the T score, the patient's BMD value is expressed as SD as compared with the mean BMD of a reference population of young adults. For Z score calculation, the patient's BMD is compared with the mean BMD of people of the same sex and age and also expressed as SD.

ΔBMD and the ΔT score were calculated by subtracting the results of the second measurement from the results of the first. The range of the difference in BMD as a percentage was calculated by dividing the difference between the first (a) and the second (b) measurement by the mean of those two figures, giving the fraction of difference between the two measurements as compared with the mean of the two measurements. The normally distributed variables are presented as mean (SD).

Precision

The measurement error was calculated using Bland and Altman's 95% limits of agreement method. Other methods used to evaluate reliability and agreement are also described. These are the CV and the intraclass correlation coefficient (ICC).

Precision expressed according to the Bland and Altman's 95% limits of agreement method[25] gives an absolute and metric estimate of random measurement error, also called SDD. In this case, where there are two observations for each subject, the standard deviation of the differences (SD_{diff}) estimates the within variability of the measurements. Most disagreements between measurements are expected to be between limits called ''limits of agreement'' defined as $d \pm z_{(1-\alpha/2)}\ SD_{diff}$ where d is the mean difference between the pairs of measurements and $z_{(1-\alpha/2)}$ is the $100(1-\alpha/2)$th centile of the normal distribution.[25] The value d is an estimate of the mean systematic bias of measurement 1 to measurement 2. d is expected to be 0 because we do not assume a true change in BMD to occur during the interval between the two BMD measurements. Defining α to be 5%, the limits of agreement are $+1.96SD_{diff}$ and $-1.96SD_{diff}$. Thus, about twice the standard deviation (SD) of the difference scores gives the 95% limits of agreement for the two measurements by the machine. A test is considered to be capable of detecting a difference, in absolute units, of at least the magnitude of the limits of agreement.

The CV, the most commonly presented measure for BMD variability, is the SD corrected for the mean of paired measurements. CV, expressed as a percentage, was calculated as $CV\% = (\sqrt{((\sum(a-b)^2)/2n)})/((Ma+Mb)/2) \times 100$ where a and b are the first and the second measurement, Ma and Mb are the mean values for the two groups, and n is the number of paired observations.[26] For two point measurements in time, a BMD change exceeding $2\sqrt{2}$ times the precision error of the technique is considered a significant change (with 95% confidence).[9] Gluer *et al* called this smallest change that is considered statistically significant, the least significant change (LSC).[11] In the current study, the LSC (%) was computed for the different BMD measurement sites. In these calculations the precision error is expressed as the CV (%).

The ICC equals variance between patients divided by variance between patients plus variance between measurements. The value of the ICC ranges from 0 to 1, 1 representing perfect reliability of the measurement.

Multilevel analysis of variability

Because multiple linear regression analysis does not allow us to discern whether the observed variation in BMD is attributable to individual differences between patients or to the influence of interval length, multilevel analysis was used.[27] Multilevel analysis can separate and quantify individual variability in contrast with measurement of interval length variability. The models used were two level variance component models, with measurement interval at the first level and patients at the second level. Demographic variables—for example, age (years) and body mass index (BMI) (kg/m^2), and BMD variables, such as area (cm^2) of BMD measurement and technician identity, were included as sources of BMD variability—that is, possible confounders that need correction. Separate models were used for different measurement sites. The fixed parameters in the final models describe the average contribution of each confounder. Patients with their corresponding interval between measurements will vary around the predictive value of the model. The degree of variation at patient level and at interval level is estimated in the random part of the model.

Statistical analysis was carried out using SPSS, version 9.0 (SPSS, Chicago, Illinois) and MLwiN for multilevel analysis.[28]

RESULTS

Postmenopausal women

Patient characteristics

The BMD measurements of 95 postmenopausal women were collected during the recruitment period. The mean (SD) age of the women was 59.9 (8.1) years. Their mean (SD) height was 163.7 (6.6) cm and their mean (SD) weight 68.8 (10.5) kg. The mean (SD) BMI was 25.7 (3.8) (kg/m^2). The interval between the first and the second spine and hip DXA ranged between 0 and 45 days, with a median of 0 days.

Table 1 shows the BMD data and the derived T and Z score data for each measurement site. The mean (SD) difference between the first and the second measurement (g/cm^2) was −0.001 (0.02) at L1-4 and −0.0004 (0.02) at the total hip. The mean (SD) T scores of the first measurement were −1.59 (1.50) and −1.34 (1.26) at L1-4 and total hip, respectively. Some patients had T scores far below −2.5 SD. The range of the difference between the T score of the first and the second BMD measurement was −0.42 to +1.30 at L1-4 and −1.84 to +0.60 at the total hip.

Table 1 BMD, T, and Z scores in postmenopausal women (n = 95)

	L1-4	Total hip	Femoral neck	Trochanter
Mean BMD (g/cm^2) (a)	0.87 (0.17)	0.82 (0.15)	0.69 (0.13)	0.63 (0.13)
Range BMD (g/cm^2) (a)	0.49 to 1.22	0.48 to 1.20	0.37 to 1.12	0.33 to 0.97
Mean ΔBMD (g/cm^2) (a−b)	−0.001 (0.02)	−0.0004 (0.02)	−0.0008 (0.02)	−0.0006 (0.01)
Range ΔBMD (g/cm^2) (a−b)	−0.05 to 0.14	−0.04 to 0.07	−0.06 to 0.09	−0.03 to 0.05
Range ΔBMD (%)*	−6.6 to 14.9	−6.2 to 7.6	−12.7 to 13.1	−7.6 to 6.1
Mean T score (a)	−1.59 (1.50)	−1.34 (1.26)	−2.06 (1.33)	−1.07 (1.40)
Range T score (a)	−5.08 to 1.54	−4.09 to 1.86	−5.27 to 2.26	−4.32 to 2.71
Mean ΔT score (a−b)	−0.01 (0.22)	−0.02 (0.24)	−0.007 (0.22)	−0.006 (0.15)
Range ΔT score (a−b)	−0.42 to 1.30	−1.84 to 0.60	−0.60 to 0.91	−0.35 to 0.51
Mean Z score (a)	−0.19 (1.35)	−0.27 (1.10)	−0.44 (1.18)	−0.04 (1.28)
Range Z score (a)	−3.40 to 2.47	−2.81 to 2.54	−3.40 to 3.66	−3.46 to 3.32

Mean (SD) for normally distributed variables. (a) First BMD measurement; (b) second BMD measurement.
*Formula: (a−b)/(a+b/2)×100.

Variability

Table 2 presents the results of the various methods of calculating variability for the three most frequently used measurement sites. Figures 1 and 2 show the scatter plots of the difference between the two measurements against their mean, for lumbar spine and total hip. The horizontal lines in these graphs show the mean of the differences and the limits of agreement. When Bland and Altman's 95% limits of agreement method was used, the mean of the difference scores approached zero, reflecting no systematic bias between measurements (the 95% CI included zero difference). In this method, random measurement error, is expressed as SD of the difference scores. Twice this value approaches the 95% limits of agreement. Thus, for the total hip the SDD in BMD measurements based on two BMD values with a short interval is 0.04 g/cm^2. The SDD at the spine was 0.05 g/cm^2.

The CV (%) was 1.59 at the total hip and 1.92 at the spine. The LSC (%) was 4.50 and 5.43 at the total hip and at the spine, respectively. Therefore, in an individual subject a BMD change at the total hip can be considered significant if the change between the measurements exceeds the SDD (expressed in absolute units) of 0.04 g/cm^2 or the LSC (%) of 4.50%.

Reliability expressed by ICC was 0.99 with narrow 95% confidence intervals at all measurement sites.

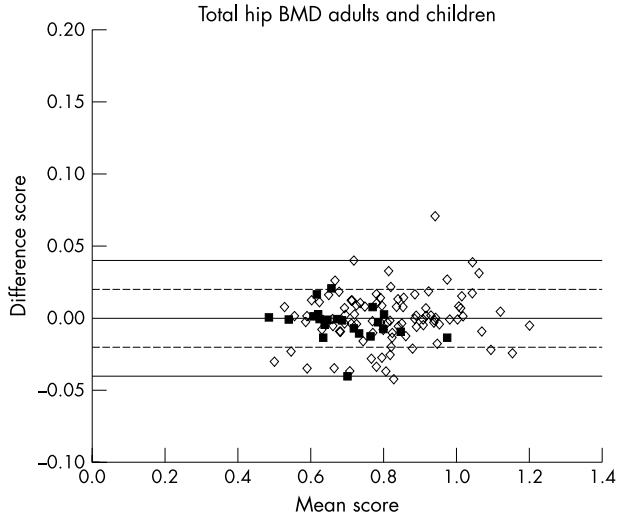

Figure 1 Graph of the difference score against the mean score of the two total hip BMD measurements (g/cm^2) in postmenopausal women (open diamonds) and children (closed squares). The outermost (solid) lines represent the 95% limits of agreement for postmenopausal women. The inner (dashed) lines represent the 95% limits of agreement for children.

Multilevel analysis

The fixed part of the final model for the spine (L1-4), apart from the constant, contained the variables technician, age, and area of BMD measurement. Both age and area predict BMD, though these variables do not influence BMD variability. The random part of the model shows that most of the variance in BMD (98.6%) can be attributed to the patients, whereas only 0.7% can be attributed to interval length. The percentage variance attributed to the technician is also 0.7%. In the final model for the trochanter, age and BMI are predictors of BMD. For the femoral neck, predictors were age and area. Finally, the model for total hip included age, BMI, and technician. Age was the only negatively correlated variable in the fixed part of each model. At all

Table 2 Reproducibility of BMD (g/cm^2) measurement in postmenopausal women (n = 95) and children (n = 23)

	Postmenopausal women			Children		
	L1–4	Femoral neck	Total hip	L1–4	Femoral neck	Total hip
Bland and Altman's methods						
Mean $_{difference}$ (95% CI) (systematic bias)	−0.001 (−0.05 to 0.05)	−0.0008 (−0.04 to 0.04)	−0.0004 (−0.04 to 0.04)	−0.0009 (−0.02 to 0.02)	−0.004 (−0.05 to 0.04)	−0.003 (−0.03 to 0.02)
SD $_{difference}$ (random measurement error)	0.0238	0.0220	0.0184	0.0088	0.0122	0.0116
SDD ~95% limits of agreement (assuming no systematic bias)	±0.0466	±0.0432	±0.0361	±0.0172	±0.0238	±0.0228
CV (%)	1.92	2.25	1.59	0.84	1.34	1.19
LSC (%)	5.43	6.36	4.50	2.38	3.79	3.37
Random effects ICC (95% CI)	0.99 (0.98 to 0.99)	0.99 (0.98 to 0.99)	0.99 (0.99 to 0.99)	0.99 (0.99 to 1.00)	0.99 (0.97 to 1.00)	0.99 (0.99 to 1.00)

Mean $_{difference}$, mean of the difference between the first and the second BMD measurement; SD $_{difference}$, SD of the difference between the first and the second BMD measurement; SDD, smallest detectable difference (g/cm^2); CV, coefficient of variation (%); LSC, least significant change (%); ICC, intraclass correlation coefficient. SD $_{difference}$ and SDD are rounded off to two decimal places in the text of the article.

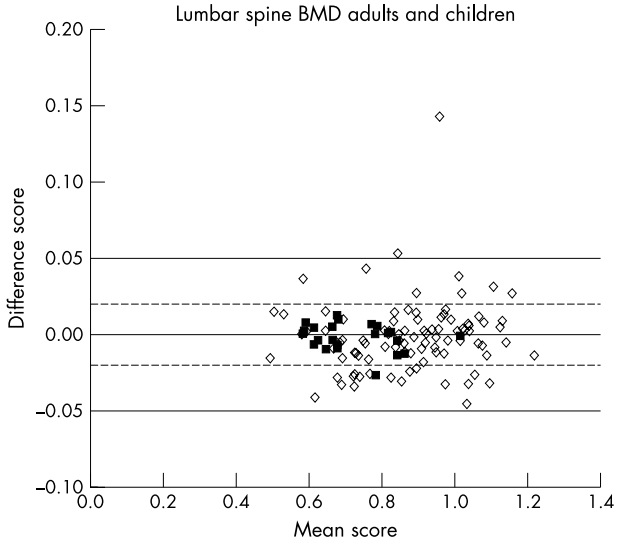

Figure 2 Graph of the difference score against the mean score of the two lumbar spine BMD measurements (g/cm^2) in postmenopausal women (open diamonds) and children (closed squares). The outermost (solid) lines represent the 95% limits of agreement for postmenopausal women. The inner (dashed) lines represent the 95% limits of agreement for children.

measurement sites in the hip, BMD variation is mainly explained by patient variability in BMD; the interval between measurements and technician variation in the case of the total hip explain only a small part of BMD variation (figures not shown).

Children

Patient characteristics

The mean (SD) age of the 23 girls investigated was 11.2 (1.3) years. In each girl, the first and second BMD measurements were carried out on the same day, with complete repositioning between the two measurements. The mean (SD) of the first BMD measurement (g/cm^2) was 0.72 (0.11) at the spine, 0.66 (0.08) at the femoral neck, and 0.70 (0.11) at the total hip. Although the mean of the difference in BMD values for the children was of the same order as for the postmenopausal women, the SD and the range of the difference were smaller in children than in the postmenopausal women. At the lumbar spine, for example, the mean (SD) of the difference between the two measurements was −0.0009 (0.009) (range −0.03 to +0.01) in children, whereas these figures were −0.001 (0.02) (range −0.05 to +0.14) in the women. At the total hip these figures were −0.003 (0.01) (range −0.04 to +0.02) in children and −0.0004 (0.02) (range −0.04 to +0.07) in the postmenopausal women, respectively.

Variability

Table 2 and figs 1 and 2 show that the SDD tended to be smaller in the children than in the women.

In the children CV% at L1-4 and the total hip were 0.84 and 1.19, whereas in the postmenopausal women these figures were 1.92 and 1.59, respectively. Consequently, the LSC in children is smaller. Hence, as compared with the women, a smaller change in BMD can be regarded as a significant change.

The ICC was as high in the children as in the postmenopausal women (table 2).

Multilevel analysis

The group of children described was considered too small for multilevel analysis to be performed.

DISCUSSION

This study shows the in vivo short term variability of BMD measurement by DXA in a group of postmenopausal women with a wide range of BMD values. In the group of women studied, reproducibility expressed by different means is good. The clinician interpreting a repeated DXA scan of a subject should be aware that a BMD change exceeding the LSC is significant, here arising from a BMD change of at least 4.5% at the total hip and 5.4% at the spine. Expressed as SDD, a BMD change should exceed 0.04 g/cm^2 at the total hip and 0.05 g/cm^2 at the spine before it can be considered a significant change.

Despite the many publications on BMD variability in different patient groups, only limited data from a large number of postmenopausal women are available on short term BMD variability.[6 12 13] In the reports published, variability is usually expressed as CV and the figures for short term variability are lower than the ones we found.[6–8] Two studies showed variability data more in line with our results.[13 29] Two samples of healthy (n = 70) and elderly (n = 57) postmenopausal women showed a CV (%) of 0.9 and 1.8, respectively, at the spine, and of 0.9 and 2.3, respectively, at the total hip.[13] Eastell showed an LSC (%) of 5.4 at the lumbar spine and 8 at the femoral neck, respectively, in osteoporotic postmenopausal women.[29] The varying results of reproducibility studies might be explained by the ''population'' investigated; a phantom and healthy young subjects are likely to show more favourable variability than postmenopausal women, possibly in part because of easier positioning for measurement. The current study also shows better variability, expressed as CV (%), in children. Secondly, osteoarthritis in postmenopausal women may contribute to poorer variability than found in healthy young subjects. Besides, the majority of the studies mentioned had small patient samples, giving less precise results.

Alternative measures of variability are the SDD and ICC. The Bland and Altman plots visualise between measurement differences. The scatter plots of the current data show a random distribution of values, indicating the absence of a relationship between the measurement error and the true BMD value, as estimated by the mean of the two measurements. The SDD values found in the adult patients were slightly higher than the figures presented by Ravaud *et al*.[13] In the first group of postmenopausal women (mean age 53 years) they describe, the SDD was 0.02 (g/cm^2) at the total hip and 0.02 at the lumbar spine. In the second group described, women with a mean age of 80 years, these figures were 0.04 and 0.04, respectively. The SDD values of the children studied tended to be lower than the values in the postmenopausal women. Using the SDD one can state that a (BMD) change larger than the figure found is a true (BMD) change in 95% of the cases. The characteristics of the Bland and Altman method thus allow direct insight into the variability of the measurement under study. Previously published reports,[13 30] as well the current data, show that reproducibility expressed in absolute units (SDD) is independent of the BMD value. Reproducibility expressed as a percentage (CV) and the derived LSC, however, depend on the BMD value. Because of therapeutic consequences, the clinician should be especially careful in judging an apparent BMD change in patients with osteoporosis. The use of the SDD in the evaluation of an apparent BMD change gives a more conservative approach than the use of the CV at low BMD. Because of its independence from the BMD level and its expression in absolute units, the SDD is a preferable measure for use in daily clinical practice as compared with the CV and the derived LSC.

The ICC found in postmenopausal women and children was high, indicating good overall reproducibility of BMD measured by DXA. However, it is important to note that a

large variability between patients automatically increases the ICC. The ICC and the Bland and Altman method yield complementary information; the presence of systematic bias cannot be found by estimating ICC.

Although the variability as expressed by the ICC, and especially the SDD, is reassuring, showing good short term variability at group level, the wide range of the differences in BMD and the derived T scores indicates considerable individual differences between two consecutive BMD measurements in some patients. The range in ΔT scores, for example, indicates that in some patients the diagnosis, based on the diagnostic thresholds of the WHO, would change owing to the measurement variability. Measured as percentages, differences between BMD measurements in patients (table 1) should be interpreted with care because the difference as a percentage depends on the mean BMD of the two measurements. Similar BMD differences in the numerator of the fraction can yield a different percentage BMD change depending on the mean value of the BMD in the denominator of the fraction.

Before investigating sources of raw BMD variation we considered examining sources of ΔBMD variation. However, technician variability in ΔBMD could not be assessed because in some cases different technicians carried out the two consecutive DXA scans. To our knowledge, this is the first study using multilevel analysis to investigate the sources of BMD variation. It is difficult to interpret why the variable technician predicts BMD only at the spine and the total hip. Why area predicts BMD at the spine and femoral neck and not at the other measurement sites is also hard to understand. The well known determinants of BMD, age and BMI, are also found in this study. BMD variation was mainly explained by patient variability, while interval length and technician in the case of the spine and total hip only explain a small part of the BMD variation. The local technicians thus have negligible influence on BMD variation. Whether this applies to other centres remains a question to be answered at the individual centre.

It should be noted that the presented variability figures hold exclusively for this patient group and this particular DXA device in our hands. The figures show that at our centre the favourable variability values presented in the literature cannot be reproduced in daily practice in postmenopausal women. Thus, for optimal clinical decision making, individual centres should establish the reproducibility figures based on routine DXA measurements in different patient groups.

In conclusion, reproducibility of BMD measurement by DXA in postmenopausal women expressed by different means is good at a group level. However, the clinician must remain aware that an apparent BMD change in an individual patient may represent a precision error. In daily practice, the use of the SDD is preferable to the use of the CV and LSC because of its independence from BMD level and its expression in absolute units.

Authors' affiliations

M C Lodder, W F Lems, A E Marthinsen, B A C Dijkmans, Department of Rheumatology, VU University Medical Centre, Amsterdam, The Netherlands
H J Ader, Department of Clinical Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, The Netherlands
S C C M van Coeverden, Department of Paediatrics, VU University Medical Centre, Amsterdam, The Netherlands
P Lips, J C Netelenbos, Department of Endocrinology, VU University Medical Centre, Amsterdam, The Netherlands
J C Roos, Department of Nuclear Medicine, VU University Medical Centre, Amsterdam, The Netherlands

REFERENCES

1 **Blake GM**, Fogelman I. Bone densitometry and the diagnosis of osteoporosis. *Semin Nucl Med* 2001;**31**:69–81.
2 **Genant HK**, Engelke K, Fuerst T, Gluer CC, Grampp S, Harris ST, *et al.* Noninvasive assessment of bone mineral and structure: state of the art. *J Bone Miner Res* 1996;**11**:707–30.
3 **Kanis JA**, Melton LJ III, Christiansen C, Johnston CC, Khaltaev N. The diagnosis of osteoporosis. *J Bone Miner Res* 1994;**9**:1137–41.
4 **Genant HK**, Cooper C, Poor G, Reid I, Ehrlich G, Kanis J, *et al.* Interim report and recommendations of the World Health Organization Task-Force for Osteoporosis. *Osteoporos Int* 1999;**10**:259–64.
5 **Miller PD**, Bonnick SL, Rosen CJ. Consensus of an international panel on the clinical utility of bone mass measurements in the detection of low bone mass in the adult population. *Calcif Tissue Int* 1996;**58**:207–14.
6 **Fuleihan GE**, Testa MA, Angell JE, Porrino N, LeBoff MS. Reproducibility of DXA absorptiometry: a model for bone loss estimates. *J Bone Miner Res* 1995;**10**:1004–14.
7 **Gluer CC**, Steiger P, Selvidge R, Elliesen-Kliefoth K, Hayashi C, Genant HK. Comparative assessment of dual-photon absorptiometry and dual-energy radiography. *Radiology* 1990;**174**:223–8.
8 **Pouilles JM**, Tremollieres F, Todorovsky N, Ribot C. Precision and sensitivity of dual-energy x-ray absorptiometry in spinal osteoporosis. *J Bone Miner Res* 1991;**6**:997–1002.
9 **Cummings SR**, Black D. Should perimenopausal women be screened for osteoporosis? *Ann Intern Med* 1986;**104**:817–23.
10 **Genant HK**, Block JE, Steiger P, Glueer CC, Ettinger B, Harris ST. Appropriate use of bone densitometry. *Radiology* 1989;**170**:817–22.
11 **Gluer CC**. Monitoring skeletal changes by radiological techniques. *J Bone Miner Res* 1999;**14**:1952–62.
12 **Lees B**, Stevenson JC. An evaluation of dual-energy X-ray absorptiometry and comparison with dual-photon absorptiometry. *Osteoporos Int* 1992;**2**:146–52.
13 **Ravaud P**, Reny JL, Giraudeau B, Porcher R, Dougados M, Roux C. Individual smallest detectable difference in bone mineral density measurements. *J Bone Miner Res* 1999;**14**:1449–56.
14 **Boutsen Y**, Jamart J, Esselinckx W, Devogelaer JP. Primary prevention of glucocorticoid-induced osteoporosis with intravenous pamidronate and calcium: a prospective controlled 1-year study comparing a single infusion, an infusion given once every 3 months, and calcium alone. *J Bone Miner Res* 2001;**16**:104–12.
15 **Cohen S**, Levy RM, Keller M, Boling E, Emkey RD, Greenwald M, *et al.* Risedronate therapy prevents corticosteroid-induced bone loss: a twelve-month, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum* 1999;**42**:2309–18.
16 **Diamond T**, McGuigan L, Barbagallo S, Bryant C. Cyclical etidronate plus ergocalciferol prevents glucocorticoid-induced bone loss in postmenopausal women. *Am J Med* 1995;**98**:459–63.
17 **Nguyen TV**, Eisman JA. Assessment of significant change in BMD: a new approach. *J Bone Miner Res* 2000;**15**:369–72.
18 **Kelly TL**, Slovik DM, Schoenfeld DA, Neer RM. Quantitative digital radiography versus dual photon absorptiometry of the lumbar spine. *J Clin Endocrinol Metab* 1988;**67**:839–44.
19 **Maggio D**, McCloskey EV, Camilli L, Cenci S, Cherubini A, Kanis JA, *et al.* Short-term reproducibility of proximal femur bone mineral density in the elderly. *Calcif Tissue Int* 1998;**63**:296–9.
20 **Slosman DO**, Rizzoli R, Bonjour JP. Bone absorptiometry: a critical appraisal of various methods. *Acta Paediatr Suppl* 1995;**411**:9–11.
21 **Neele SJ**, Marchien van Baal W, van der Mooren MJ, Kessel H, Netelenbos JC, Kenemans P. Ultrasound assessment of the endometrium in healthy, asymptomatic early post-menopausal women: saline infusion sonohysterography versus transvaginal ultrasound. *Ultrasound Obstet Gynecol* 2000;**16**:254–9.
22 **Oleksik A**, Ott SM, Vedi S, Bravenboer N, Compston J, Lips P. Bone structure in patients with low bone mineral density with or without vertebral fractures. *J Bone Miner Res* 2000;**15**:1368–75.
23 **Valk-de Roo GW**, Stehouwer CD, Lambert J, Schalkwijk CG, van der Mooren MJ, Kluft C, *et al.* Plasma homocysteine is weakly correlated with plasma endothelin and von Willebrand factor but not with endothelium-dependent vasodilatation in healthy postmenopausal women. *Clin Chem* 1999;**45**:1200–5.
24 **Van Coeverden SC**, De Ridder CM, Roos JC, van't Hof MA, Netelenbos JC, Delemarre-Van de Waal HA. Pubertal maturation characteristics and the rate of bone mass development longitudinally toward menarche. *J Bone Miner Res* 2001;**16**:774–81.
25 **Bland JM**, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**i**:307–10.
26 **Blake GM**, Wahner HW, Fogelman I. *Assessment of instrument performance; precision, installation of new equipment and radiation dose. the evaluation of osteoporosis: dual energy x-ray absorptiometry and ultrasound in clinical practice.* London: Martin Dunitz, 1999:147–72.
27 **Hox JJ**. *Applied multilevel analysis.* 1st ed. Amsterdam: TT-Publicaties, 1995.
28 **Goldstein H**, Rasbash J, Lewis IP, Draper D, Browne W, Yang M, *et al. A user's guide to MlwiN.* 1st ed. London: Institute of Education, 1998.
29 **Eastell R**. Assessment of bone density and bone loss. *Osteoporos Int* 1996;**6**(suppl 2):3–5.
30 **Gluer CC**, Blake G, Lu Y, Blunt BA, Jergas M, Genant HK. Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. *Osteoporos Int* 1995;**5**:262–70.