1. Introduction

Osteoporosis is a worldwide major public health problem. Bone densitometry has become the “gold standard” in its diagnosis and treatment evaluation (El Maghraoui and Roux 2008). With its advantages of high precision, short scan times, low radiation dose, and stable calibration, dual-energy x-ray absorptiometry (DXA) has been established by the World Health Organization (WHO) as the technique of reference for assessing bone mineral density (BMD) in postmenopausal women and based the definitions of osteopenia and osteoporosis on its results. Recently, efficient 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 (e.g. chronic rheumatic diseases especially when treated by corticosteroids) (Phillipov, Seaborn et al. 2001). Moreover, BMD measurement is needed to select patients for osteoporosis treatment, as there is no proof that drugs for osteoporosis (other than hormone replacement therapy [HRT]) are beneficial in women with clinical risk factors for fractures but normal BMD values.

It has also become more and more common to perform a second DXA measurement to monitor BMD status or the effect of therapeutic intervention. When a second measurement is performed on a patient, the clinician needs to distinguish between a true change in BMD and a random fluctuation related to variability in the measurement procedure. The reproducibility of DXA measurements is claimed to be good. Such variability is due to multiple causes, such as device errors, technician variability, patients’ movements, and variation due to other unpredictable sources (Nguyen, Sambrook et al. 1997; Lodder, Lems et al. 2004).

The precision error is usually expressed as the coefficient of variation (CV), which is the ratio of the standard deviation (SD) to the mean of the measurements, although several other statistics to express reproducibility exist such as the smallest detectable difference (SDD) or the least significant change (LSC). The SDD represents a cut-off that can be measured in an individual and is usually considered more useful than the CV in clinical practice (Fuleihan, Testa et al. 1995; Ravaud, Reny et al. 1999).
2. Methods of bone mineral density reproducibility measurement

Precision errors are evaluated by performing repeated scans on a representative set of individuals to characterize the reproducibility of the technique. Most published studies examine the short-term precision error, based on repeated measurements of each subject performed over a time period of no more than 2 weeks. Over such a short period, no true change in BMD is expected.

2.1 The coefficient of variation (CV)

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, is calculated as CV (%) = (√(Σ(a-b)²/2n))/((Ma+Mb)/2)x100 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.

Reproducibility is far better for BMD measurement than for most laboratory tests. Reproducibility expressed by the CV is usually 1–2% at the spine on anteroposterior images and 2–3% at the proximal femur in individuals with normal BMD values; the difference between the two sites is ascribable to greater difficulties with repositioning and examining the femur, as compared to the spine. However, these data obtained under nearly experimental conditions may not apply to everyday clinical practice. Reproducibility depends heavily on quality assurance factors, including tests to control the quality and performance of the machine, as well as the experience of the operator. Assessment of machine performance requires daily scanning of a phantom (which may be anthropomorphic or not), followed by calculation of the in vitro coefficient of variation (CV), which serves to evaluate short-term and longterm performance and to detect drift in measurement accuracy. These in vitro data, however, do not necessarily reflect in vivo reproducibility, which should be evaluated at each measurement centre. Measurements are obtained either three times in each of 15 patients or twice in each of 30 patients, and the CV (m/r) is calculated from the mean (m) and standard deviation (r) of these repeated measurements. The CV is expressed as a percentage and depends on mean BMD values(Phillipov, Seaborn et al. 2001). The standard deviation reflects measurement error, which is a characteristic of machine performance and is independent from the value measured.

2.2 The least significant change (LSC)

For two point measurements in time, a BMD change exceeding 2√2 times the precision error (PE) of a technique is considered a significant change (with 95% confidence): the corresponding change criterion has been termed "least significant change" or LSC. LSC = 2.8 x PE; where PE is the largest precision error of the technique used (or more easily the CV expressed in percentage). This smallest change that is considered statistically significant is also expressed in percentage(Gluer 1999).

2.3 The smallest detectable difference (SDD)

The measurement error can be calculated using Bland and Altman’s 95% limits of agreement method(Bland and Altman 1986). Precision expressed by this method 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\text{diff}) estimates the within variability of the measurements. Most disagreements between measurements are expected to be between limits called “limits of agreement” defined as \(d \pm z_{(1-a/2)} \cdot SD_{\text{diff}}\) where \(d\) is the mean difference between the pairs of measurements and \(z_{(1-a/2)}\) is the 100(1-a/2)th centile of the normal distribution. The value \(d\) is an estimate of the mean systematic bias of measurement 1 to measurement 2. \(d\) is expected to be 0 because a true change in BMD is not assumed to occur during the interval between the two BMD measurements. Defining \(a\) to be 5%, the limits of agreement are +1.96SD\text{diff} and -1.96SD\text{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.

3. Clinical implications of bone mineral density reproducibility measurement

In clinical practice, two absolute values (g/cm\(^2\)) have to be compared, rather than two percentages (T-scores). When serial measurements are obtained in a patient, only changes greater than the LSC (in %) or the SDD (in g/cm\(^2\)) can be ascribed to treatment effects. Smaller changes may be related to measurement error.

We studied recently the in vivo short term variability of BMD measurement by DXA in three groups of subjects with a wide range of BMD values: healthy young volunteers, postmenopausal women and patients with chronic rheumatic diseases (most of them taking corticosteroids). In all studied subjects, reproducibility expressed by different means was good and independent from clinical and BMD status. Thus, the clinician interpreting a repeated DXA scan of a subject should be aware that a BMD change exceeding the LSC is significant, in our centre arising from a BMD change of at least 3.56% at the total hip and 5.60% at the spine. Expressed as SDD, a BMD change should exceed 0.02 g/cm\(^2\) at the total hip and 0.04 g/cm\(^2\) at the spine before it can be considered a significant change (El Maghraoui, Do Santos Zounon et al. 2005). Indeed, it has become usual to perform repeated DXA measurement: in postmenopausal women to monitor efficacy of treatment and in patients with chronic rheumatic diseases where high prevalence of bone loss has been demonstrated (Maillefert, Aho et al. 2001; Johnson, Petkov et al. 2005) especially when long term corticosteroid therapy is used. In the reports published, variability is usually expressed as CV and the figures for short term variability are lower than the ones we found [7-9]. However, two studies showed variability data more in line with our results. In Ravaud et al. (Ravaud, Reny et al. 1999) study, 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. Eastell showed an LSC (%) of 5.4 at the lumbar spine and 8 at the total hip, respectively, in osteoporotic postmenopausal women (Eastell 1996). It has been suggested that 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 (Gluer, Blake et al. 1995). However, our study failed to show better variability, expressed as CV (%), in young healthy volunteers (El Maghraoui, Do Santos Zounon et al. 2005). Another reason advocated was that osteoarthritis in postmenopausal women may contribute to poorer variability than found in healthy young subjects. The SDD values found in our study were comparable to the figures presented by Ravaud et al. (Ravaud, Reny et al. 1999). In the first group of postmenopausal
women (mean age 53 years) they describe, the SDD was 0.02 (g/cm²) 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. In Lodder et al. (Lodder, Lems et al. 2004) study (Ninety five women, mean age 59.9 years), the SDD was 0.04 (g/cm²) at the total hip and 0.05 at the lumbar spine. The SDD values of the children studied in this study tended to be lower than the values in the postmenopausal women (table 1). 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 (figure 1).

It has been shown that reproducibility expressed using the SDD is independent of the BMD value whereas reproducibility expressed using the CV or the derived LSC depend on the BMD value. Ravaud et al. (Ravaud, Reny et al. 1999) reported that using SD, the values of the cut-offs are 0.024, 0.030, 0.020, and 0.021 g/cm² for postmenopausal women aged ≤70 years and 0.040, 0.033, 0.033, and 0.038 g/cm² for postmenopausal osteoporotic women aged >70 years at the spine, femoral neck, greater trochanter, and total hip, respectively. Using CV, cut-offs vary depending on the BMD level. In postmenopausal women aged >70 years, for a BMD level between 0.600 g/cm² and 1.000 g/cm², the cut-offs derived from CV vary between 0.015 g/cm² and 0.024 g/cm², 0.024 g/cm² and 0.041 g/cm², 0.018 g/cm² and 0.030 g/cm², 0.015 g/cm² and 0.025 g/cm² for the spine femoral neck, greater trochanter, and total hip, respectively. In postmenopausal osteoporotic women aged >70 years, for the same range of BMD level, cut-offs vary between 0.031 g/cm² and 0.051 g/cm², 0.038 g/cm² and 0.063 g/cm², 0.043 g/cm² and 0.071 g/cm², 0.038 g/cm² and 0.063 g/cm² for the same bone sites. Consequently, to express variability on a percentage basis using CV leads to underestimate variability in patients with low BMD and to overestimate variability in patients with high BMD. Previous reports in the literature, as well as Ravaud, Lodder’s data and our data (table 1) demonstrate that absolute precision errors derived from SD are constant across a wide range of BMD values and independent of the level of BMD. Because of therapeutic consequences, the clinician should be especially careful in judging an apparent BMD change in patients with osteoporosis. Influence of age on BMD reproducibility is controversial. Previous studies have suggested that BMD measurement errors were independent of age even some studies suggested that SDD may vary in extreme ages (children and elderly) probably because of age-related factors other than BMD. However, a few data exist for reproducibility of DXA in women over 70. Ravaud et al. data, as well as those of Fuleihan (Fuleihan, Testa et al. 1995), and Maggio et al. (Maggio, McCloskey et al. 1998) show that the measurement error is greater in older osteoporotic subjects. Several factors such as difficulties in repositioning could explain the increase of measurement error in this kind of patients. Therefore, 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.

In contrast with all previous publications about DXA reproducibility, we found in our centre better results for the hip BMD variability than the lumbar spine. This is due to the fact that our study was the first to use the mean measure of the two femurs (dual femur). In this study, we showed in a group of young healthy volunteers that the SDD was ±0.0218 g/cm² when both femurs were measured whereas it was ±0.0339 g/cm² when only one femur was
measured. Thus, these results enhance to encourage the use of the measurement of both hips to improve the reproducibility of DXA at this site.

Fig. 1. Graph of the difference score against the mean score of the total hip (a) and the lumbar spine (b) BMD measurements (g/cm²) in our centre [17] using Bland and Altman method. The outermost (solid) lines represent the SDD.
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/cm²); CV, coefficient of variation (%); LSC, least significant change (%); ICC, intraclass correlation coefficient.

Table 1. Comparison of BMD measurement reproducibility evaluation in two studies including various groups of patients.
Although the variability as expressed by the CV, 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 $\Delta T$ 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.

In summary, reproducibility of BMD measurement by DXA in different kinds of patients (postmenopausal women, patients with chronic rheumatic diseases, elderly...) 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. At each measurement centre, the SDD should be calculated from in vivo reproducibility data. In clinical practice, the SDD should be used to estimate the significance of observed changes, in absolute values.

5. Other factors influencing DXA monitoring

The first factor is the time interval between two measurements in the same patient which must be long enough to allow occurrence of a change greater than the SDD or the LSC. Therefore, it depends on the expected rate of change in BMD measurement (which varies according to whether the measurement site is composed predominantly of trabecular or of cortical bone) and the reproducibility of BMD measurement at that site. Thus, in clinical practice, a treatment-induced BMD increase can only be detected in general after 2 years. However, in patients receiving long term steroid therapy, the changes in BMD may be so important that they can be detected at 1 year. Thus, although the spine may not be the best site for the diagnosis of osteoporosis given the high prevalence of spinal degenerative disease, it is the most sensitive site for detecting changes over time. However, our study showed that measurement of both femurs (called “dual femur” in Lunar machines) increases the reproducibility at this site.

In another side, the changes in BMD measurements are influenced by the ability of osteoporosis treatments to increase the BMD at the different skeletal sites. For some treatments such as teriparatide and the more potent bisphosphonates, statistically significant changes in spine BMD occur on time scales of 1 to 2 years in the majority of patients, although for other treatments, such as raloxifene, the changes are often not large enough to be statistically significant. Recently, the strontium ranelate trials show BMD increases at the spine and hip 2-5 times larger than those for BPs and SERMs, and comparable or greater than with teriparatide. However, it is important to appreciate that much of this effect is due to the higher atomic number of strontium compared with calcium. Thus, strontium in bone attenuates X-rays much more strongly than calcium, and BMD is overestimated compared with the true mass of bone mineral present. This effect may persist for many years after the patient stops treatment and may affect the relationship between BMD and fracture risk.

Thus, treatment dosages cannot be adjusted on the basis of BMD changes. Moreover, there is no proof that repeating BMD measurements improves compliance, as most patients discontinue antiresorptive medications after a few months because of administration constraints, side effects, cost of medications or lack of interest.
Above all, BMD is used as a surrogate marker for the fracture risk, yet BMD increases do not reliably reflect a reduction in the fracture risk. Although bisphosphonates, raloxifene, and strontium have not been compared in the same study, they seem to produce comparable reductions in the risk of vertebral fractures, of about 30–50%, whereas BMD changes differ markedly across medications. Studies have shown that BMD gains explain only a small proportion of the vertebral fracture risk reduction: 28% with risedronate (Li, Meredith et al. 2001), 16% with alendronate (Cummings, Karpf et al. 2002), and 4% with raloxifene (Sarkar, Mitlak et al. 2002). It has been suggested that the percentage of BMD change may be related to the change in the relative risk of fracture (Wasnich and Miller 2000). In one study, a linear relationship was found between these two parameters, but a 1% increase in spinal BMD was associated with an only 3% decrease in the relative risk of vertebral fracture (Cummings, Karpf et al. 2002). For peripheral fractures, in contrast, the risk reduction is clearly related to the BMD gain (Hochberg, Greenspan et al. 2002). Common sense indicates that a BMD increase during treatment should be preferable over a BMD decrease. However, data showing that the fracture risk may decrease despite a reduction in BMD have been reported (Watts, Geusens et al. 2005). It has also been shown that the fracture risk was more heavily dependent on BMD at baseline than on BMD changes during treatment (Hochberg, Ross et al. 1999).

6. Conclusion
Serial BMD measurements can be used to monitor current antiresorptive treatments (raloxifene, bisphosphonates or strontium ranelate). However, adequate quality-control procedures must be used (Roux, Garnero et al. 2005). Measurement error must be considered when evaluating serial assessments. A clear understanding of the interpretation of serial measurements and the statistical principles impacting upon their interpretation is necessary to determine whether a change is real and not simply random fluctuation. It is inadequate to simply use the manufacturer’s default precision error, which may underestimate the precision error in the clinical setting. Thus, every centre should calculate its own precision error from in vivo reproducibility data. International societies interested in osteoporosis diagnosis and management such as the International Society for Clinical Densitometry or the International Osteoporosis Foundation should add to their guidelines at least two recommendations about DXA monitoring highlighted in this paper: the measurement of both hips improves the reproducibility at this site and DXA measurement centres should determine and use the individual SDD. Indeed, 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. The exact definition and advices for the measure and use of these parameters in clinical practice should be clearly explained. It is clear that the choice of the optimum site for performing follow-up scans depends on the ratio of the BMD treatment effect to the precision of the measurements. The larger this ratio, the more statistically significant the observed changes are likely to be. Actually, all data agree in showing that the spine is the optimum site. In clinical practice, BMD measurements have to be spaced at least 2 years apart. The main goal of serial BMD measurement is to check that no further bone loss has occurred; estimation of BMD gains is the secondary objective. This should be explained to the patients, many of whom expect to recover normal BMD values.
7. References


The World Health Organization (WHO) has established dual-energy x-ray absorptiometry (DXA) as the best densitometric technique for assessing bone mineral density (BMD) in postmenopausal women and has based the definitions of osteopenia and osteoporosis on its results. DXA enables accurate diagnosis of osteoporosis, estimation of fracture risk and monitoring of patients undergoing treatment. Additional features of DXA include measurement of BMD at multiple skeletal sites, vertebral fracture assessment and body composition assessment, including fat mass and lean soft tissue mass of the whole body and the segments. This book contains reviews and original studies about DXA and its different uses in clinical practice (diagnosis of osteoporosis, monitoring of BMD measurement) and in medical research in several situations (e.g. assessment of morphological asymmetry in athletes, estimation of resting energy expenditure, assessment of vertebral strength and vertebral fracture risk, or study of dry bones such as the ulna).

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
