Optimization of Follow-up Measurements of Bone Mass

Luis F. Verheij, Jacobus A.K. Blokland, Socrates E. Papapoulos, Aeilko H. Zwinderman, and Ernest K.J. Pauwels

Division of Nuclear Medicine, Department of Diagnostic Radiology, and Clinical Investigation Unit, Department of Endocrinology, University Hospital, Leiden, The Netherlands, and Department of Medical Statistics, State University, Leiden, The Netherlands

In bone densitometry, the precision of the instrument, the number of measurements and the time-points of the measurements are important criteria for monitoring bone mass changes. The most appropriate follow-up procedure can be determined by numerical comparison of various combinations of these three criteria. This can be done by computing the confidence interval of changes in bone mass. We developed a model to estimate the length of a confidence interval for the observed changes in individual patients. With specific instrument precision, a specified number of measurements and, assuming a linear rate of bone mass changes, the best estimate of the actual changes in bone mass is obtained by measurements at the end of an observation period. With the current precision of bone densitometers, follow-up of patients with yearly duplicate measurements is recommended. A shorter scan time interval offers no additional information unless very rapid bone loss is expected.

J Nucl Med 1992; 33:1406–1410

Osteoporosis is a complex, multifactorial chronic disease that may progress silently for years until characteristic fractures appear late in life (*1*). It is characterized by low bone mass, microarchitectural deterioration of bone tissue leading to enhanced bone fragility and a consequent increase in fracture risk (*2*). Because the structure of the bone tissue cannot yet be determined in vivo and alternatives to bone mass measurements are inadequate (*3*), bone densitometry is at present the most important tool for the diagnosis and follow-up of patients at risk and for the evaluation of interventions. The instruments' accuracy and precision, critical in diagnosis and in monitoring changes, respectively (*4*), are of great importance for these purposes.

Both the accuracy and precision of bone mineral density (BMD) measurements have been improved with the introduction of dual-energy x-ray absorptiometry (DEXA). Its accuracy has now reached an acceptable level, whereas its precision needs to be further improved and is often overrated. The interpretation of follow-up measurements in individuals is difficult, because the average bone mass changes are small with respect to the uncertainty in the estimation of the actual changes. Reduction in this uncertainty will result in improved monitoring of bone mass changes.

Apart from the instrument's precision, the uncertainty in the observed BMD changes depends on two additional parameters: the number of measurements and the time-points at which these measurements are performed (*5,6*). Various combinations of the three parameters can be numerically compared if the magnitude of this uncertainty can be determined. In fact, the length of a confidence interval can serve as a measure of that. The shorter the confidence interval, the better the estimate of the rate of bone mass change.

The effect of short-term precision of current instruments on the length of the confidence interval should be considered in practice optimal, since it appears to have reached its limits (*7*). Attention should therefore focus on the remaining two parameters, which constitute the follow-up procedure. The optimal combination of number and time-points of measurements will yield the shortest confidence interval. We describe a model for the computation of the length of confidence intervals which allows for the numerical expression of the degree of uncertainty of observed bone mass changes. The model is flexible and enables clinicians to determine a follow-up strategy for individual patients which best suits their practice.

METHOD

For the development of the model, some generally accepted assumptions were used (*5–8*). It was assumed that during the observation period, the rate of BMD change is constant, that the reproducibility is the same in all patients, that measurements are independent and that the instrument is stable. We also assumed that BMD values are normally distributed. These assumptions allow the use of a linear regression model.

Let y_i be the BMD at time-point t_i and let the change in

Received Nov. 19, 1991; revision accepted Feb. 13, 1992.

For reprints contact: J.A.K. Blokland, PhD, Division of Nuclear Medicine, University Hospital Leiden, Building 1, C4-Q, PO Box 9600, 2300 RC Leiden, The Netherlands.

TABLE 1A

Length of 95% Confidence Intervals for Various Instrument Precisions or Standard Deviations of the Technique and Diverse Single, Equidistant Measurements

Measurements*	CV† = 1	CV† = 1.5	CV† = 2	s.d.‡ = 0.008	s.d.‡ = 0.016	s.d.‡ = 0.024
2	±2.77	±4.16	±5.54	±0.022	±0.044	±0.067
3	±2.77	±4.16	±5.54	±0.022	±0.044	±0.067
4	±2.63	±3.94	±5.26	±0.021	±0.042	±0.063
5	±2.48	±3.72	±4.96	±0.020	±0.040	±0.060
6	±2.34	±3.51	±4.69	±0.019	±0.037	±0.056

* Total number of measurements.
† Coefficient of variation in %.
‡ Standard deviation in g/cm².

BMD be linear in time; the regression equation is then equal to

$$y_i = \beta_0 + \beta_1 * (t_i - t_0) + \epsilon_i, \qquad \text{Eq. 1}$$

where β_0 is the BMD at the start (t_0) of the observation period, β_1 represents the rate of change (slope of the regression line), t_i is the time-point of the measurement and ϵ_i is the random error. The random error ϵ_i comprises all time-independent variables, such as variations due to counting statistics and machine variability.

The confidence interval for β_1 is defined as

$$(b_1 - z_\alpha * S_b) < \beta_1 < (b_1 + z_\alpha * S_b), \qquad \text{Eq. 2}$$

where b_1 represents the observed rate of change, z_α is the normalized standard deviate which is related to the confidence level (see Table 1) and S_b is the standard error of b_1:

$$S_b = \frac{1}{\sqrt{(n-1)}} * \frac{S_{y.t}}{S_t}. \qquad \text{Eq. 3}$$

In Equation 3, n is the number of measurements and $S_{y.t}$ the instrument's precision. The short-term precision can be obtained from duplicate measurements and is defined as the coefficient of variation (CV), which is the standard deviation (s.d.) divided by the mean and expressed as a percentage. Reproducibility is often used to describe the long-term precision of phantom measurements (*9*).

S_t is the s.d. of the time-points:

$$S_t = \sqrt{\frac{\sum_{i=1}^{n} (t_i - \bar{t})^2}{(n-1)}}, \qquad \text{Eq. 4}$$

where $\bar{t}$ is the mean of the time-points $\bar{t}_i$ of all measurements within the observation period. If we use Equation 4 to substitute S_b in Equation 3 we obtain:

$$S_b = \frac{S_{y.t}}{\sqrt{\sum_{i=1}^{n} (t_i - \bar{t})^2}}. \qquad \text{Eq. 5}$$

The length of the observation period, that is, the period over which the changes are evaluated, is arbitrary. It can be set equal to 1, and thus becomes the unit of time.

RESULTS

The lengths of confidence intervals for various possible measurement set-ups are computed on the basis of the presented model. Tables 1A and 1B shows the results with the measurement precision expressed in both CV (%) and s.d. (g/cm²). These tables may be of help in deciding which combination suits best a particular precision.

The confidence intervals computed from two and three measurements within the same period of time have equal lengths, provided the third observation is carried out exactly at the mean time-point. Extra equidistant measurements have practically no effect on this interval. Replicate measurements reduce the length of confidence intervals considerably. This reduction is inversely proportional to the square root of the number of replicate measurements per time point. There is a direct relation between the confidence level and the length of the confidence interval. The effect of various confidence levels can be computed using the numbers depicted in Table 2. A practical appli-

TABLE 1B

Length of 95% Confidence Intervals for Various Instrument Precisions or Standard Deviations of the Technique and Diverse, Replicate Measurements at Only Two Time-Points

Measurements*	CV† = 1	CV† = 1.5	CV† = 2	s.d.‡ = 0.008	s.d.‡ = 0.016	s.d.‡ = 0.024
2 × 2	±1.96	±2.94	±3.92	±0.016	±0.031	±0.047
2 × 3	±1.60	±2.40	±3.20	±0.013	±0.026	±0.038
2 × 4	±1.39	±2.08	±2.77	±0.011	±0.021	±0.032

* Total number of measurements.
† Coefficient of variation in percent.
‡ Standard deviation in g/cm².

TABLE 2
A Number of Frequently Used Confidence Levels and Their Related Values for the Normalized Standard Deviate, z_α

Confidence level (%)	Normalized standard deviate
99	2.58
95	1.96
90	1.65
80	1.28
68.2	1

cation of the model is described in the Appendix. The confidence intervals in two examples with different rates of bone mass changes have been computed. Although completely different, they yield confidence intervals of similar lengths.

DISCUSSION

We developed a model to estimate the length of a confidence interval for measured BMD changes in individuals. To examine group responses, a different statistical consideration is needed (*10*).

The results indicate that even with the currently best achievable precision, monitoring of individual patients still includes a relatively high degree of uncertainty. The short-term precision is limited by the inhomogeneity of the soft-tissue layer (*11*) and can hardly be improved. Therefore, determination of an optimal follow-up procedure is necessary in order to minimize the uncertainty of the observed BMD changes.

We assumed that the rate of BMD change is linear. Actually, biologic variations resulting from factors such as menopause (*12*) or temporary disability may cause nonlinear changes (*7*). However, it must be taken into consideration that the exact effect of these variations in individuals is unknown and that bone mass changes may be considered linear during a short interval even if they are nonlinear over a longer period (*10*). We felt, therefore, that the assumption made is the best possible approximation for actual bone mass changes during an observation period, especially at the beginning of a follow-up study.

In the model, it was also assumed that the short-term precision is the same for all patients. In fact, the short-term precision is not exactly the same in all patients due to its relation to the BMD and to the composition and thickness of the soft-tissue layer (*11*). Therefore, relatively low BMD values and inhomogeneous or thick soft-tissue layers may lead to an underestimate of the confidence interval, or to an overestimate in the opposite case. To account for this problem, we have expressed the precision not only in percentages, as is usually done, but also in g/cm^2 (Table 1A and 1B).

In addition, it should be noted that the lengths of the confidence intervals are minimum values. The use of the standard normal (z) distribution is only allowed in case the reproducibility is known and has a constant value. The Student's t-distribution should be applied if the intervals are calculated from multiple measurements and the individual reproducibility is estimated (*6*). Thus, an underestimate by the ratio of z/t occurs.

The best approach would be to calculate each person's variability. This, however, cannot be done at the beginning of the study since it requires at least three or more measurements to obtain a reasonably accurate estimate of variability (*6*). Hence, applying the same precision for all patients is often a necessary assumption. Moreover, an estimate of only the average length of the confidence interval is needed as the purpose of the presented model is to determine a follow-up procedure suitable for the majority of patients.

Another assumption concerned the instrument's stability. The absence of instrument drift should be checked regularly by phantom measurements. This has to be accounted for if a drift is detected.

Because of the assumptions made, the presented model may not define the exact lengths of the confidence intervals, but provides fairly accurate, easily calculated estimates of the degree of uncertainty. The model can also be utilized for the computation of confidence intervals in other fields for which the assumptions made here are applicable.

It appeared that measurements at the ends of the observation period invoke a far greater reduction of the length of a confidence interval than measurements close to the mean time-point. In fact, confidence intervals do not change at all when measurements are performed exactly at the mean time-point. This was observed in earlier studies (*5,8*), but was never supported by mathematical evidence: observations at the mean time-point do not contribute to the summation in the denominator in Equation 5. The more an individual time-point differs from the mean time-point, the larger the contribution to this summation and hence the smaller the standard error in the observed rate of change. This is described in more detail in the Appendix.

Attention must be paid to the fact that biologic variations, as mentioned earlier, restrict the effect of replicate measurements in improving the estimate of the *rate* of bone mass change but not its effect on the estimate of the *actual* loss. Davis et al. (*7*) have used longitudinal data to assess the contribution of biologic nonlinearity to the uncertainty in the computation of the rate of bone loss. Their data suggest that the shorter the confidence interval, the relatively higher this contribution. This is explained by the fact that biologic variations generate a patient-related, and hence nonstatistical, factor.

Regarding the frequency of bone mass measurements, Heaney (*8*) recommends multiple, evenly-spaced observations to avoid the potential problem of instrument's instability. However, we feel that this factor should not influence the follow-up strategy as long as the instrument's stability is checked daily. With the same number of meas-

urements, replicate measurements yield the optimal reduction of the confidence interval for the BMD change.

Care should be taken that measurements are repeated in the same season since seasonal variations in bone mineral mass may mask real bone mass changes (*13,14*). Yearly measurements avoid this problem. Shorter scan time intervals yield confidence intervals that are large with respect to the average patients' bone loss during this period. These are only required if rapid losses are expected, e.g., in case of ovariectomy or immobilization. Rapid bone loss can double the patient's fracture risk over two years (*7*), which necessitates early detection. Longer scan time intervals are preferable if the precision is low, or if the patient is not treated and has a bone mass value higher than the average for his/her age and sex. In the other cases, such intervals should be avoided.

Further scan time intervals will be decided after the first follow-up period and should depend on the actual level of bone mass and the rate of its change. The initial scan time interval can be maintained if the observed loss is within the range of the average loss for the patient's age and sex, plus or minus half the length of the computed confidence interval (as depicted in Table 1A and 1B). In case the detected loss is larger, the scan time interval must be shortened; in case of smaller losses (or gains), the interval can be extended. Apart from these considerations, the choice of the most appropriate follow-up strategy may also depend on other important factors such as costs, patient's discomfort, and availability of personnel and equipment. Adequate interventions should always be available.

In conclusion, we found that even with the current precision of DEXA instruments, confidence intervals are large with respect to the observed changes in bone mass. Therefore, an optimal follow-up procedure is required to minimize the uncertainty in the estimation of bone mass changes. Replicate measurements yield the optimal estimate of the actual bone mass changes with respect to a specified number of measurements. Examinations should be repeated during the same season. A scan time interval of less than a year offers no advantages in a clinical setting unless rapid changes in bone mass are expected.

APPENDIX

As an example, two different follow-up procedures are presented. One with equidistant measurements, the other with measurements at the start and end of the observation period only.

In the first case, the instrument precision is 2%, as can be obtained with DPA in patient studies (*15*). A single measurement is performed every 6 mo during a 3-yr period. Hence, the number of measurements, n, is equal to 7 and, as the observation period is set equal to 1, $t = 1/2$. The standard error of the observed change can now be computed:

$$S_b = \frac{S_{y.t}}{\sqrt{\sum_{i=1}^{n} (t_i - \bar{t})^2}}$$

$$= \frac{2}{\sqrt{\left[\left(0-\frac{1}{2}\right)^2 + \left(\frac{1}{6}-\frac{1}{2}\right)^2 + \left(\frac{2}{6}-\frac{1}{2}\right)^2 + \left(\frac{3}{6}-\frac{1}{2}\right)^2 + \left(\frac{4}{6}-\frac{1}{2}\right)^2 + \left(\frac{5}{6}-\frac{1}{2}\right)^2 + \left(1-\frac{1}{2}\right)^2\right]}}$$

$$= 2.27\%.$$

If an 80% confidence level is chosen, $z_\alpha = 1.28$ (Table 2) and hence the length of the confidence interval $z_\alpha * S_b = \pm 2.91\%$. Suppose the observed rate of BMD change equals +1% (in 3 yr time); then the confidence interval for β_1 is equal to

$$(b_1 - z_\alpha * S_b) < \beta_1 < (b_1 + z_\alpha * S_b)$$

$$(1 - 2.91\%) < \beta_1 < (1 + 2.91\%)$$

$$-1.91\% < \beta_1 < +3.91\%.$$

So, with 80% certainty (and under the given assumptions), the real change in BMD will be between −1.91% and +3.91%.

In the second case, the instrument's precision is equal to 1.1%, close to the present in-vivo short-term precision of lumbar spine measurements using DEXA (*16*). Duplicate measurements are performed at both the start and end of a 1-yr observation period, and the observed BMD change is equal to −4%. Again $\bar{t} = 1/2$ and, using the same equation for S_b, we obtain

$$S_b = \frac{1.1}{\sqrt{\left\{\left(0-\frac{1}{2}\right)^2 + \left(0-\frac{1}{2}\right)^2 + \left(1-\frac{1}{2}\right)^2 + \left(1-\frac{1}{2}\right)^2\right\}}}$$

$$= 1.1\%.$$

by choosing a 99% confidence level, $z_\alpha = 2.58\%$, hence $z_\alpha * S_b = 2.84\%$, so that β_1 ranges from

$$(-4 - 2.84\%) < \beta_1 < (-4 + 2.84\%)$$

$$-6.84\% < \beta_1 < -1.16\%.$$

In these two specific cases, the equations regarding S_b can be simplified. For equidistant measurements we can rewrite the equation to:

$$S_b = \frac{S_{y.t}}{\sqrt{\sum_{i=1}^{n} \left\{\frac{(i-1)}{(n-1)} - \frac{1}{2}\right\}^2}}.$$

$$S_b = \frac{S_{y.t}}{\sqrt{\left\{\frac{1}{(n-1)^2} * \sum_{i=1}^{n} (i-1)^2 - \frac{1}{(n-1)} * \sum_{i=1}^{n} (i-1) + \frac{n}{4}\right\}}}.$$

With the mathematical series for $(i-1)$ and $(i-1)^2$:

$$S_b = \frac{S_{y.t}}{\sqrt{\frac{n(n+1)}{12(n-1)}}}.$$

By substituting the numbers from the first example, i.e., $S_{y.t}$ = 2%, and $n = 7$:

$$S_b = \frac{2}{\sqrt{\frac{7(7+1)}{12*6}}} = 2.27\%.$$

For replicate measurements at the start and end of the observation period only, it can be derived that

$$S_b = \frac{S_{y.t}}{\sqrt{\left\{\frac{n}{2} * \left(0 - \frac{1}{2}\right)^2 + \frac{n}{2} * \left(1 - \frac{1}{2}\right)^2\right\}}} = \frac{2*S_{y.t}}{\sqrt{n}}.$$

by substituting the numbers from the second example, i.e., $S_{y.t}$ = 1.1% and $n = 4$:

$$S_b = \frac{2*1.1}{\sqrt{4}} = 1.1\%.$$

It can be shown that with any given number of measurements, measurements at the ends of an observation period yield smaller confidence intervals than equidistant measurements. If the ratio of the standard errors of both measurement set-ups, indexed as $S_{b_{ends}}$ and $S_{b_{eq}}$, is computed, it follows that:

$$\frac{S_{b_{ends}}}{S_{b_{eq}}} = \frac{(2*S_{y.t})/(\sqrt{n})}{(2*S_{y.t})/\left\{\frac{1}{3}\sqrt{3}*\sqrt{n}*\sqrt{\frac{(n+1)}{(n-1)}}\right\}} = \frac{1}{3}\sqrt{3}*\sqrt{\frac{(n+1)}{(n-1)}}.$$

The set-ups are exactly the same for $n = 2$. With $n = 4$, it follows:

$$\frac{1}{3}\sqrt{3}*\sqrt{\frac{(4+1)}{(4-1)}} = \frac{1}{3}\sqrt{5} < 1.$$

It can be shown inductively that the ratio is always less than 1 for all other even values of $n > 4$. So, with the assumptions made for the presented model, n replicate measurements at both ends of the observation period result in a better estimate of the rate of BMD change than $2n$ equidistant measurements.

ACKNOWLEDGMENTS

This work was supported by the Dutch Praeventiefonds under grant 28-1474. Presented in part at the Society of Nuclear Medicine Annual Meeting, June 1991, Cincinnati, OH, and at the European Association of Nuclear Medicine Congress, Vienna, Austria, September 1991.

REFERENCES

1. Melton LJ, Wahner HW, O'Fallon WM, Riggs BL. Epidemiology of vertebral fractures in women. *Am J Epidemiol* 1989;129:1000–1011.
2. Anonymous. Consensus development conference: prophylaxis and treatment of osteoporosis. *Osteoporosis Int* 1991;1:114–117.
3. Johnston CC, Melton LJ, Lindsay R, Eddy DM. Clinical indications for bone mass measurements. A report from the scientific advisory board of the National Osteoporosis Foundation. *J Bone Min Res* 1989;4:1–28.
4. Mazess RB. Measurement of skeletal status by noninvasive methods. [Editorial] *Calcif Tissue Int* 1979;28:89–92.
5. Wahner HW. Can the rate of bone loss be measured with repeated bone mineral measurements. In: Dequeker J, Geusens P, Wahner HW, eds. *Bone mineral measurement by photon absorptiometry.* Leuven: Leuven University Press; 1988:301–304.
6. Cummings SR, Black D. Should perimenopausal women be screened for osteoporosis? *Ann Intern Med* 1986;104:817–823.
7. Ross PD, Davis JW, Wasnich RD, Vogel JM. The clinical application of serial bone mass measurements. *Bone Min* 1991;12:189–199.
8. Heaney PR. En recherche de la différence ($p < 0.05$). *Bone Min* 1986;1:99–114.
9. Appledorn CR, Appledorn PS, Wellman HN, Witt RM, Johnston CC, Hanish BL. Accuracy, precision, and long-term reproducibility of dual photon bone mineral absorptiometric measurements. In: Schmidt HAE, Emrich D, eds. *Clinical demands on nuclear medicine.* Stuttgart, New York: FK Schattauer Verlag; 1986:315–317.
10. Davis JW, Ross PD, Wasnich RD, MacLean CJ, Vogel JM. Long-term precision of bone loss rate measurements among postmenopausal women. *Calcif Tissue Int* 1991;48:311–318.
11. Valkema R, Verheij LF, Blokland JAK, et al. Limited precision of dual photon absorptiometry by variations in the soft tissue background. *J Nucl Med* 1990;31:1774–1781.
12. Elders PJM, Netelenbos JC, Lips P, van Ginkel FC, van der Stelt PF. Accelerated vertebral bone loss in relation to the menopause: a cross-sectional study on lumbar bone density in 286 women of 46 to 55 years of age. *Bone Min* 1988;5:11–19.
13. Elders PJM, Netelenbos JC. Machine drift, source replacement, seasonal variation. Does it matter? In: Dequeker J, Geusens P, Wahner HW, eds. *Bone mineral measurement by photon absorptiometry.* Leuven: Leuven University Press; 1988:165–169.
14. Bergstralh EJ, Sinaki M, Offord KP, Wahner HW, Melton III LJ. Effect of season on physical activity score, back extensor muscle strength, and lumbar bone mineral density. *J Bone Min Res* 1990;5:371–377.
15. Verheij LF, Blokland JAK, Papapoulos SE, Pauwels EKJ. Automated comparison of dual-photon absorptiometric studies of the lumbar spine. *J Bone Min Res* 1991;6:575–581.
16. Kelly TL, Slovik DM, Schoenfield DA, Neer RM. Quantitative digital radiography versus dual photon absorptiometry of the lumbar spine. *J Clin Endocrinol Metab* 1988;67:839–844.