JOURNAL OF BONE AND MINERAL RESEARCH
Volume 23, Number 11, 2008
Published online on June 30, 2008; doi: 10.1359/JBMR.080602

Heteroscedastic Regression Analysis of Factors Affecting BMD Monitoring

Mohsen Sadatsafavi,[1] Alireza Moayyeri,[2] Liqun Wang,[3] and William D Leslie[3] for the Manitoba Bone Density Program

ABSTRACT: Identifying factors affecting BMD precision and interindividual heterogeneity in BMD change can help optimize BMD monitoring. BMD change for the lumbar spine and total hip for short-term reproducibility ($n = 328$) and long-term clinical monitoring ($n = 2720$) populations were analyzed with heteroscedastic regression using linear prediction for mean (monitoring population only) and log-linear prediction for SD (both populations). For clinical monitoring, male sex, baseline body mass index (BMI), and systemic corticosteroid use were associated with greater SD of BMD change. Weight gain was negatively associated with SD for the hip, whereas height change was positively associated with SD for the spine. Each additional year of monitoring increased the SD by 6.5–9.2%. Osteoporosis treatment affected mean change but did not increase dispersion. For short-term reproducibility, performing scans on a different day increased the SD of measurement error by 38–44%. Baseline BMD, difference in bone area, and a repeat scan performed by different technologists were associated with higher measurement error only for the hip. For both samples, heteroscedastic regression outperformed models that assumed homogeneous variance. Heteroscedastic regression techniques are powerful yet underused tools in analyzing longitudinal BMD data and can be used to generate individualized predictions of BMD change and measurement error.
J Bone Miner Res 2008;23:1842–1849. Published online on June 30, 2008; doi: 10.1359/JBMR.080602

Key words: bone densitometry, DXA, heteroscedastic regression, osteoporosis, precision

INTRODUCTION

Osteoporosis and its clinical expression, fragility fractures, have large public health implications. Worldwide, the number of fracture sufferers in 2000 was estimated at 56 million, with ~9 million new osteoporotic fractures each year.[1] This is projected to result in a loss of 5.8 million disability-adjusted life years. Osteoporosis costs $13.8 billion annually in the United States alone.[2]

Serial BMD measurement with DXA is widely used for monitoring change in patients susceptible to bone loss and for assessing the impact of osteoporosis treatment.[3] Direct BMD monitoring is necessary because the between-subject variability in BMD change makes the exact prediction of a future BMD impossible.[4] On the other hand, the BMD measurement itself is accompanied by measurement error; hence, proper interpretation of an observed change in BMD requires knowledge of the precision of the measurement technique.[5] Identifying factors that affect both the average rate of change and heterogeneity among subjects contributes to understanding mechanisms of BMD change. Furthermore, knowledge about the factors that affect measurement error will assist clinicians in interpreting observed BMD change in an individual and may help to optimize BMD monitoring in clinical practice.

There are several reports assessing factors affecting short-term and long-term measurement error of BMD monitoring.[4,6–18] Typically these are based on small sample sizes[6–9] and have often assessed factors separately.[7–13] Even when the joint effect of more than one factor is studied, the analysis has been based on stratifying patients on each combination of factors.[4,14,15] Although such stratification enables statistical inference about the effect of a desired set of covariates, the effect size attributable to each covariate is difficult to quantify, and statistical power is lost as the analysis is performed in each stratum separately. Hence, only a few categorical variables can be studied in this way. On the other hand, there are statistically rigorous methods, called heteroscedastic regression or functional variance models, for the analysis of dispersion that provide important advantages over previous methods.[16,17] Because data are not divided into separate strata, the power for statistical inference is higher, and the estimated regression coefficients have straightforward interpretations.

In this paper, we analyze both the short-term precision and long-term BMD change using heteroscedastic regression analysis. Although the primary aim of this analysis was to identify factors affecting short-term precision and

The authors state that they have no conflicts of interest.

[1]Center for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Institute, Vancouver, British Columbia, Canada; [2]Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; [3]University of Manitoba, Winnipeg, Manitoba, Canada.

long-term variability of BMD change, we also provide equations for predicting individualized short-term precision and the mean and SD for long-term BMD change based on patient-specific covariates.

MATERIALS AND METHODS

Study populations

Anonymized data from the Manitoba Bone Density Program were used for the analyses. The study was approved by the Research Ethics Board for the University of Manitoba and the Health Information Privacy Committee of Manitoba Health. The program was established in 1997 and provides all bone density services to the population of Province of Manitoba, Canada.[18] An electronic database collects data on all DXA bone density tests performed in Manitoba and is >99% complete and accurate as judged by chart audit.[19] All equipment and technologist performance is subject to a rigorous quality assurance program developed from published models.

From the BMD database, we identified all individuals who had at least two measurements (baseline and follow-up) for total spine (L_1–L_4) and the total hip performed on the same instrument before March 31, 2007. For subjects who had more than two measurements, only the first and second measurements were used in the analysis. We excluded cases where scanning was performed on different instruments, where the lumbar spine and/or hip were not scanned or were unsuitable for clinical reporting, and where there were vertebral exclusions for focal structural defects using conventional criteria.[20] Data from the included patients were linked to the provincial Drug Program Information Network (DPIN) database using an anonymous numeric identifier.[21] DPIN captures information about pharmaceutical dispensations in real time for all Manitoba residents. Based on these data, total duration (in years) of treatment with an antiresorptive treatment (systemic estrogen, oral bisphosphonate, raloxifene, or parenteral calcitonin) between the two scans was calculated for each subject. Individuals with a single dispensation were considered as not receiving treatment as single prescription usually indicates nonpersistence with therapy. No adjustment was made for the bioequivalency of the antiresorptive treatments. Systemic corticosteroid use was defined as dispensation for 90-day duration or longer after baseline BMD measurement at a mean daily dose of prednisone (or equivalent) of 7.5 mg or greater. For each subject, total dose (in milligrams) of prednisone-equivalent intake between the two scans was calculated. Overall, 2726 scan-pairs with complete data were available for the final "clinical monitoring sample."

Short-term replicate measurements of the total spine and total hip were obtained from a convenience sample of individuals referred for bone density testing who were agreeable to undergoing a repeat assessment. These scans were acquired as part of the Manitoba Bone Density Program's ongoing clinical quality assurance program and are therefore representative of BMD test precision during the years of clinical monitoring. During the course of this study, 20 technologists were involved in the precision assessments. No individual outliers or temporal variation was identified in terms of technologist performance. For comparability with the clinical monitoring population, individuals with severe focal structural defects in the lumbar spine were not included in the reproducibility assessment. The only three men in the reproducibility sample were excluded, limiting the reproducibility sample to women. The final "reproducibility sample" consisted of 331 total spine and 328 total hip scan-pairs.

BMD measurements

DXA scans were performed and analyzed in accordance with manufacturer recommendations. Before 2000, DXA measurements were performed with a pencil-beam instrument (Lunar DPX; GE Lunar, Madison, WI, USA), and after this date, a fan-beam instrument was used (Lunar Prodigy; GE Lunar). Instruments were cross-calibrated using 59 volunteers and anthropomorphic phantoms. No clinically significant differences were identified (T-score differences < 0.2). Therefore, all analyses are based on the unadjusted numerical results provided by the instrument. Densitometers showed stable long-term performance (CV < 0.5%).

Statistical analysis

We used a joint heteroscedastic regression model with a linear prediction equation for the mean change of BMD and a log-linear prediction equation for the SD of BMD change:

$$\Delta BMD_i \sim Normal(\mu_i, \sigma_i)$$

$$\mu_i = \beta_0 + \beta_1 \,.\, X_1 + \beta_2 \,.\, X_2 + \ldots + \beta_n \,.\, X_n$$

$$\log(\sigma_i) = \beta'_0 + \beta'_1 \,.\, X_1 + \beta'_2 \,.\, X_2 + \ldots + \beta'_n \,.\, X_n$$

Here the ΔBMD_i is the dependent variable (BMD change), and X_i are the explanatory covariates. β and β' are the estimated coefficients for mean and logarithm of SD, respectively. The difference between an ordinary (homoscedastic) regression and the above regression model should be contrasted. Both models assume that the true (unobserved) BMD change is a linear function of covariates, confounded by noise that has a zero mean. On the other hand, in ordinary linear regression, the variance of the residual errors is assumed to be equal across the whole range of the dependent variable. In a heteroscedastic model, on the other hand, the variance (or SD) of the residual error is no longer considered constant; rather, it is assumed to be a particular function of the explanatory variables. In our model, we assumed the logarithm of SD to be a linear function of covariates. The interpretation of the coefficients for the mean is the same as in the ordinary linear regression. For the SD, the exponent of the coefficients can be interpreted in a multiplicative manner. For example, a coefficient of 0.15 indicates that one unit increase in the value of the independent covariate is accompanied by $exp(0.15) = 1.16$ times (or 16% increase) in SD.

Baseline age, body mass index (BMI), BMD, and bone area were taken from the time of the first scan. Changes in weight, height, and bone area were defined as their difference between the second and first visits. Our database does not capture menopausal status or age of menopause; therefore, this variable was not entered into the model. Observations with missing values in the dependent variable or any of the covariates were excluded from the analysis. $p \leq 0.05$ was considered significant. For the clinical monitoring sample, both mean and log(SD) were regressed on the same set of covariates. For the short-term reproducibility sample, given the short interval between the two scans and that all scan-pairs were done on the same device, the only variable that could affect the mean change in BMD was difference in bone area. A preliminary analysis showed that this covariate had no effect on BMD change. Consequently, only the log(SD) was regressed on covariates for the reproducibility sample and the mean was assumed to be zero.

Unlike in the reproducibility sample, the observed BMD change in the clinical monitoring sample is a combination of the real change in BMD and change introduced by the imprecision of the measurement technique. It is therefore imperative to appreciate the difference in the effect of these two sources of variation. For the factors that might affect the true BMD change, the effect size is a function of the interval between two measurements, whereas factors that introduce bias in the measurement have a one-time, absolute effect. To account for this difference, baseline covariates (sex, age, BMI, and BMD) were modeled as having an interaction with interval in their effect on mean BMD change. The coefficients for these covariates can be interpreted as their effect on the "annual rate of change." Similarly, the regression intercept for the mean BMD change predictor was forced to be zero to prevent an implausible prediction of nonzero change for those with (hypothetically) zero interval between the scans, and the coefficient for interval between scans can be considered as an interval-dependent residual error. On the other hand, change in height, weight, and bone area between the two scans was modeled as having a one-time effect regardless of the interval between scans; thus, their coefficients should be interpreted as the absolute impact of the covariate on the long-term bias of the measurement. Total years of antiresorptive treatment and total dose of steroid use were entered into the model without interaction because they naturally depend on the interval between scans. The coefficient for the anti-osteoporosis treatment represents BMD change caused by 1 yr of treatment, whereas for the steroid use, it is the change in BMD associated with the intake of 1000 mg of prednisone. For models of SD of change, no interaction between the interval between scans and covariate effects was included.

To obtain an optimal model specification, we compared three competing models using the corrected Akaike Information Criterion (AICc).[22,23] The AIC is a popular model selection criterion in statistics, based on the adjusted log likelihood penalized for the complexity of the model (because the model with more covariates always fits the data better). According to this criterion, the model with smallest AICc value is preferred. For the clinical monitoring sample, the performance of three models was compared using the AICc: simple (assuming that change in BMD is not a function of any covariate and is equal in all subgroups), homoscedastic (ordinary linear regression model with a homoscedastic variance), and heteroscedastic. Because for the short-term reproducibility sample the mean change is not regressed on covariates, the AICc for the heteroscedastic regression model was only compared with that of a model that estimates a homogenous SD for the whole population. All statistical computations were performed using SAS (version 9.1.3; SAS Institute, Cary, NC, USA).

Table 1. Characteristics of Subjects in the Clinical Monitoring and Reproducibility Samples

	Clinical monitoring (n = 2720)	*Reproducibility (n = 328)*
Female [*n* (%)]	2539 (93.3)	328 (100.0)
Age at first BMD (yr)	57.5 (11.1)	55.4 (11.1)
Interval between scans	3.4 (1.3) yr	4.6 (6.3) days
Treatment with antiresorptive agents between the two scans [*n* (%)]	1742 (64.0)	N/A
Systemic corticosteroid use between the two scans [*n* (%)]	504 (18.5)	N/A
Baseline BMI (kg/m^2)	25.09 (4.63)	26.66 (5.36)
Baseline weight (kg)	66.0 (13.4)	68.4 (13.8)
Weight change between scans	0.18 (5.32)	N/A
Baseline BMD		
Lumbar spine (g/cm^2)	0.998 (0.162)	1.074 (0.168)
Total hip (g/cm^2)	0.860 (0.131)	0.924 (0.129)
Change in BMD		
Lumbar spine (g/cm^2)	0.000 (0.063)	0.001 (0.023)
Total hip (g/cm^2)	–0.003 (0.045)	–0.001 (0.014)
Scanner type		
Fan-beam [*n* (%)]	2527 (92.9)	241 (73.6)
Pencil-beam [*n* (%)]	193 (7.1)	87 (26.5)
Scanner timing		
Same day	N/A	165 (50.3)
Different day		163 (49.7)
Technologist		
Same technologist	N/A	105 (32.0)
Different technologists		223 (68.0)

Data are mean (SD) unless otherwise specified.
N/A, not applicable.

RESULTS

Baseline characteristics of patients

The subject characteristics for the two samples are presented in Table 1. From the original 2727 observations in the clinical monitoring sample, 7 were excluded because of missing values in one or more covariates. For the reproducibility sample, five scan-pairs for the lumbar spine and seven scan-pairs for the total hip were excluded because of missing values for covariates. Females comprised 93% of the clinical monitoring sample and 100% of the reproducibility sample. Mean ages were similar in the two samples (57.5 ± 11.1 versus 55.4 ± 11.1 [SD] yr).

TABLE 2. RESULTS OF THE REGRESSION ANALYSIS ON THE CLINICAL MONITORING POPULATION

	Spine L_1–L_4 (n = 2717)					*Total hip (n = 2719)*				
	Mean		*SD*			*Mean*		*SD*		
Covariates	*β*	*95% CI*	*β*	*95% CI*	*Percent change*	*β*	*95% CI*	*β*	*95% CI*	*Percent change*
Male gender (vs. female)†	0.012*	0.009, 0.016	0.210*	0.097, 0.323	23.4	0.006*	0.002, 0.009	0.413*	0.294, 0.533	51.2
Age (per decade)†	0.001	0.000, 0.001	0.002	−0.024, 0.029	0.2	0.000	−0.001, 0.000	−0.020	−0.046, 0.006	−2.0
Baseline BMD (per 0.1 g/cm^2)†	−0.003*	−0.003, −0.002	0.005	−0.012, 0.022	0.5	−0.002*	−0.003, −0.002	−0.005	−0.026, 0.016	−0.5
Baseline BMI (per 5 kg/m^2)†	0.003*	0.002, 0.004	0.042*	0.013, 0.070	4.3	0.002*	0.002, 0.003	0.047*	0.019, 0.075	4.8
Weight change (per kg)‡	0.001*	0.001, 0.002	0.005*	0.001, 0.009	0.5	0.002*	0.001, 0.003	−0.001	−0.005, 0.003	−0.1
Height change (per cm)‡	−0.001	−0.002, 0.000	0.011	0.000, 0.023	1.2	0.000	0.000, 0.001	−0.029*	−0.039, −0.020	−2.9
Cumulative corticosteroids (per 1000 mg prednisone)‡	0.001	0.000, 0.002	0.018*	0.006, 0.030	1.8	−0.001*	−0.002, 0.000	0.024*	0.011, 0.037	2.4
Change in the bone area (per cm^2)‡	0.005*	0.004, 0.006	−0.001	−0.011, 0.010	−0.1	0.002*	0.001, 0.004	0.027*	0.006, 0.049	2.8
Pencil beam scanner (vs. fan beam)‡	0.003	−0.003, 0.009	−0.003	−0.126, 0.119	−0.3	0.007*	0.003, 0.011	−0.024	−0.141, 0.093	−2.3
Osteoporosis treatment duration (yr)‡	0.012*	0.010, 0.014	−0.009	−0.033, 0.014	−0.9	0.008*	0.007, 0.010	−0.005	−0.029, 0.018	−0.5
Monitoring interval (yr)	0.002	−0.005, 0.009	0.088*	0.065, 0.112	9.2	0.004	−0.001, 0.009	0.063*	0.040, 0.087	6.5
Regression intercept (β_0)	0§	0§	−3.529	−3.811, −3.247	N/A	0§	0§	−3.592	−3.846, −3.337	N/A

* Significant at 0.05 level.
† Effect on the mean annualized rate of change.
‡ Effect on the mean absolute change.
§ Forced to be zero.
N/A, not applicable.

TABLE 3. RESULTS OF THE REGRESSION ANALYSIS ON THE SHORT-TERM REPRODUCIBILITY SAMPLE

	Spine L1–L4 (n = 326)			*Total hip (n = 322)*		
	SD			*SD*		
Covariates	*β*	*95% CI*	*Percent change*	*β*	*95% CI*	*Percent change*
Age (per decade)	0.011	–0.070, 0.092	1.1%	0.018	–0.062, 0.098	1.8%
Baseline BMD (per 0.1 g/cm^2)	0.035	–0.014, 0.084	3.6%	0.070*	0.003, 0.138	7.3%
Baseline BMI (per 5 kg/m^2)	0.044	–0.032, 0.121	4.5%	0.050	–0.029, 0.129	5.1%
Change in the bone area (per cm^2)	0.021	–0.046, 0.088	2.1%	0.168*	0.028, 0.308	18.3%
Pencil beam scanner (vs. fan beam)	0.003	–0.197, 0.203	0.3%	–0.130	–0.332, 0.072	–12.2%
Different day (vs. same day)	0.321*	0.164, 0.478	37.9%	0.365*	0.200, 0.529	44.1%
Different technologist (vs. same technologist)	–0.068	–0.251, 0.115	–6.6%	0.211*	0.024, 0.398	23.5%
Regression intercept (β_0)	–4.653	–5.543, –3.763	N/A	–5.648	–6.462, –4.833	N/A

* Significant at 0.05 level.
N/A, not applicable.

Clinical monitoring sample

Results of the heteroscedastic regression analysis on the clinical monitoring sample are presented in Table 2. Mean change in BMD for the sample was 0.000 ± 0.063 g/cm^2 for the lumbar spine and –0.003 ± 0.045 g/cm^2 for the total hip after a mean interval of 3.4 ± 1.3 yr. Several covariates had a significant impact on the mean change. Men gained slightly more BMD between the two visits at both the lumbar spine and total hip ($p < 0.001$). Weight gain between the scans ($p = 0.010$) and greater difference in bone area between the two scans were all associated with increasing BMD at both sites ($p = 0.001$). As expected, anti-osteoporotic medication use was significantly associated with gain in BMD at both sites ($p < 0.001$). Use of a pencil-beam scanner (versus fan-beam) was positively associated with increase in BMD only at the hip ($p = 0.002$). Likewise, systemic steroid use was associated with BMD decline only at hip ($p = 0.021$). Baseline BMD was negatively associated with BMD change, an expected observation caused by the regression-to-the-mean phenomenon.[24]

Several variables also had a significant impact on the dispersion of BMD change in the clinical monitoring sample. Male sex had a large impact on the dispersion, increasing SD of change by 23% for total spine and 51% in total hip (both $p < 0.001$). Baseline BMI showed a positive association with SD of change at both sites (SD increase 4–5% per each five-point increase in BMI, $p < 0.05$ at both sites). Also systemic corticosteroid use had a small positive association with the dispersion of BMD change ($p = 0.04$ at both sites). Weight gain was associated with a larger SD of change at the lumbar spine ($p = 0.020$), whereas height change was associated with a smaller SD of change at the hip ($p < 0.0001$). The interval between the two scans had a great impact on the SD of change. Each additional year of monitoring increased the SD by 9.2% (95% CI, 6.7; 11.8%) for the lumbar spine and 6.5% (95% CI, 4.1, 9.1%) for the total hip (both $p < 0.0001$). Osteoporosis treatment duration did not increase BMD dispersion ($p > 0.1$).

Short-term reproducibility sample

Results of the heteroscedastic regression analysis on the short-term reproducibility sample are presented in Table 3. The only covariate that was significantly associated with SD of change at both sites was whether the scan-pairs were performed on the same day or not ($p < 0.0001$). Performing the replicate scan in a different day (compared with the same-day scan) increased the SD of measurement error by 37.9% (95% CI, 17.8, 61.3%) for lumbar spine and by 44.1% (95% CI, 22.2, 69.8%) for total hip (both $p < 0.001$). Baseline BMD (7.3% per 0.1 g/cm^2, $p = 0.03$), change in bone area (18.3% per cm^2, $p = 0.009$), and whether the repeat scan was performed by another technologist (23.5% as opposed to the same technologist, $p < 0.001$) were all associated with larger measurement error at the hip site only.

Model selection

For the clinical monitoring population, AICc for the simple, homoscedastic, and heteroscedastic models for lumbar spine was –7480, –8161, and –8260, respectively. For total hip, the corresponding numbers were –9221, –9825, and –10,005, respectively. These numbers indicate that, for both sites, the heteroscedastic regression was the best among the three models. For the short-term reproducibility sample, the AICc of the simple and heteroscedastic models for lumbar spine was –1583 and –1591, respectively. For total hip, the AICc was –1852 for the simple model and –1866 for the heteroscedastic model. Again, the lower AICc indicated the superiority of the heteroscedastic regression for both sites over the model that assumes precision is homoscedastic.

An example of how the results of the heteroscedastic regression analysis can be used in both short-term reproducibility and long-term monitoring for individualizing BMD monitoring parameters is provided in the Appendix.

DISCUSSION

If there was no error involved in measuring BMD in a subject, the interpretation of an observed BMD change would be purely clinical. In the presence of measurement error, however, the interpretation of an observed change requires statistical inference to distinguish between the signal and noise, where the signal is the true BMD change

and noise is the measurement error. In this paper, we show the impact of several factors on the rate and heterogeneity of BMD change and measurement error. The novel aspect of our analysis is reporting the adjusted effect sizes of multiple covariates on the SD of change in a way that lends itself to a straightforward interpretation. Our results not only shed light on the factors contributing to BMD change and errors of measurement but also potentially enable clinicians to apply their knowledge of the signal-versus-noise relationships to make better patient decisions.

Other groups have studied factors affecting BMD precision. Tothill[25] has reported on the long-term bias of measurement over a 6-mo period. This author reported a positive correlation between bone area and BMD. A positive relation between weight change and BMD was also reported for the Lunar Prodigy instrument (the device used in >90% of our BMD measurements), which is again in accordance with our findings. In another study, Tothill and Hannan[9] reported the short- and long-term precision in different groups of patients. For long-term precision, they fitted a different regression model (negative exponential or linear) and derived the short-term and long-term LSC from the SD of the regression residuals. A similar approach has been used by Patel et al.,[8] who fitted a linear regression model and assumed that the residual errors are caused by the long-term measurement error. We have concerns regarding the validity of such assumptions. It is very unlikely that a complex biological system like bone turnover follows a perfectly linear or exponential change in all individuals over time. Therefore, the regression residuals in such analyses include some elements from the true BMD change. In addition, whereas such models enable statistical inference for the effect of several covariates on the mean change in BMD, statistical inference cannot be easily performed on the residual SD, and hence judgment on factors affecting precision remains qualitative. A different approach has been taken to analyze precision by Phillipov et al.[26] They performed a nonparametric linear regression on 12 serial BMD measurements in 24 subjects. They were able to perform statistical inference (e.g., confidence interval) on the components of the overall variance using bootstrap techniques and analysis of covariance (ANCOVA). ANCOVA also enables testing for the effect of covariates. For example, a significant effect of body fat on BMD precision was found on total spine and total femur, although the absolute effect size could not be reported in such analysis.

We have previously reported a univariate analysis on the same short-term reproducibility sample.[14] In that study, among the several covariates studied, the only significant factor affecting the short-term precision was whether the replicate measure is performed in the different day versus same day, whose effect size was similar to our present estimate. Other covariates in our previous analysis were not found to consistently affect short-term precision. In another analysis, we found that the difference in bone area (categorized as <2% or >2%) adversely affected the precision in lumbar spine and femoral neck, but not total hip,[12] whereas in this analysis, a significant affect was found for hip area but not the lumbar spine. Such differences might be explained by the lack of adjustment for other covariates and/or the effect of using continuous covariates versus dichotomized measures.

The combined results on short-term reproducibility and long-term clinical monitoring samples allow for a coherent application in the recently developed Bayesian models for the interpretation of BMD change.[27,28] Such Bayesian methods combine a priori knowledge of the BMD change with the information from the current experiment (which is defined by the observed BMD change and the magnitude of the measurement error) to construct a posteriori knowledge about the true BMD change. The predicted mean and SD of the change from the clinical monitoring population can be used to construct a patient-specific prior distribution, whereas the individualized SD of measurement error can be used to represent the uncertainty around the observed BMD change. In essence, such an approach will enable the incorporation of the patient- and setting-specific characteristics into the decision making process, resulting in a more valid representation of our confidence on the true BMD change in a subject. The absolute impact of such a process on the output of a BMD monitoring program could be the focus of future research.

This study had several limitations. The whole sample for this study came from a single densitometry program and thus the external validity of our results could be questioned. Our database does not capture the menopausal status of participants, a factor that is known to affect the rate[29,30] and possibly the heterogeneity of BMD change and measurement error. We did not adjust for the type and dosage of anti-osteoporosis treatments. In addition, our reproducibility sample consisted only of women, and the effect of sex on short-term precision could not be quantified. The reproducibly sample for this study was gathered gradually as part of the ongoing quality control, and it is possible that the precision had changed because of a variety of factors such as changes in the personnel and equipment. Finally, we did not perform empirical model comparison to elucidate the predictive power of the heteroscedastic regression compared with simpler models like ordinary regression. Such an analysis requires dividing the data into testing and training sets, which results in lower statistical power, especially for the short-term reproducibility sample.

The observed mean and SD in the clinical monitoring sample is made up of two components: the true changes in subjects and potential measurement error in the measurement technique. We were unable to separate these two effects in our analysis. For the mean change, we modeled some covariates as affecting the rate of change, whereas the others as affecting the absolute effect on BMD change, but this had been chiefly imposed by the nature of the model rather than a priori hypothesis. For example, it was inappropriate to model baseline covariates (e.g., sex and baseline BMD) and residual error as not having interaction with interval between the two scans; otherwise, the model would predict an implausible BMD change in subjects with zero interval because of the coefficients estimated for these covariates.

On the other hand, the effect of covariates on the SD of change was modeled multiplicatively (additive on the log

scale), so there is an implied interaction between covariates and the interval between scans. Whereas this seems to be a valid assumption for covariates that affect the between-subject heterogeneity in the rate of change, it becomes problematic for the factors affecting long-term precision. Nevertheless, a comparison of the SD of change between the reproducibility and clinical monitoring samples showed that the between-observer heterogeneity in true rate of change is the dominant component of the observed variance in the clinical monitoring sample. Thus, for a factor that only affects long-term precision, the effect size would need to be very large to overcome the noise caused by true BMD change to make it statistically significant. Given this, we believe the observed association between some covariates and long-term dispersion of BMD change in this analysis is likely because of their effect on the heterogeneity in the true rate of change rather than on the long-term precision.

To our knowledge, this study is the first analysis of BMD change using heteroscedastic regression models. Precision of BMD monitoring has a direct impact on the interpreting an observed change in BMD. Precision of a measurement technique often affects the dispersion of the observed parameter; thus, functional variance models are a natural choice in the study of factors affecting the precision. This work could pave the ground for further research using these powerful techniques because there is plenty of room for future studies on both short- and long- term precision and the between-subject heterogeneity of BMD change. Similar analyses from other centers are needed to corroborate our results. Theoretically, when more than two longitudinal observations per subject are available, more complex designs such as random-effect methods that model patient-specific rate of change could be used, which may lead to more valid and informative results. Alternative regression functions for dispersion could also be evaluated with an eye on the underlying biological and mechanical processes. Finally, further research is needed to assess the feasibility of incorporating the results of such analysis in routine practice. The regression equations that predict mean and SD of change are simple enough to run on a spreadsheet or a hand calculator, and many densitometry centers routinely record the variables that we used for our analysis. Thus, it might be the case that the benefit of implementing such methods will outweigh the additional complexities involved.

ACKNOWLEDGMENTS

We are indebted to Manitoba Health for providing data. The results and conclusions are those of the authors, and no official endorsement by Manitoba Health is intended or should be inferred. This article has been reviewed and approved by the members of the Manitoba Bone Density Program Committee.

REFERENCES

1. Johnell O, Kanis JA 2004 An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. Osteoporos Int **15:**897–902.
2. Cummings SR, Bates D, Black DM 2002 Clinical use of bone densitometry: Scientific review. JAMA **288:**1889–1897.
3. Lenchik L, Kiebzak GM, Blunt BA 2002 What is the role of serial bone mineral density measurements in patient management? J Clin Densitom **5**(Suppl)**:**S29–S38.
4. Nguyen TV, Sambrook PN, Eisman JA 1997 Sources of variability in bone mineral density measurements: Implications for study design and analysis of bone loss. J Bone Miner Res **12:**124–135.
5. Bonnick SL, Johnston CC Jr, Kleerekoper M, Lindsay R, Miller P, Sherwood L, Siris E 2001 Importance of precision in bone density measurements. J Clin Densitom **4:**105–110.
6. Engelke K, Gluer CC, Genant HK 1995 Factors influencing short-term precision of dual X-ray bone absorptiometry (DXA) of spine and femur. Calcif Tissue Int **56:**19–25.
7. Maggio D, McCloskey EV, Camilli L, Cenci S, Cherubini A, Kanis JA, Senin U 1998 Short-term reproducibility of proximal femur bone mineral density in the elderly. Calcif Tissue Int **63:**296–299.
8. Patel R, Blake GM, Rymer J, Fogelman I 2000 Long-term precision of DXA scanning assessed over seven years in forty postmenopausal women. Osteoporos Int **11:**68–75.
9. Tothill P, Hannan WJ 2007 Precision and accuracy of measuring changes in bone mineral density by dual-energy X-ray absorptiometry. Osteoporos Int **18:**1515–1523.
10. Shepherd JA, Fan B, Lu Y, Lewiecki EM, Miller P, Genant HK 2006 Comparison of BMD precision for Prodigy and Delphi spine and femur scans. Osteoporos Int **17:**1303–1308.
11. El Maghraoui A, Do Santos Zounon AA, Jroundi I, Nouijai A, Ghazi M, Achemlal L, Bezza A, Tazi MA, Abouqual R 2005 Reproducibility of bone mineral density measurements using dual X-ray absorptiometry in daily clinical practice. Osteoporos Int **16:**1742–1748.
12. Leslie WD 2006 The impact of bone area on short-term bone density precision. J Clin Densitom **9:**150–153.
13. Lilley J, Walters BG, Heath DA, Drolc Z 1991 In vivo and in vitro precision for bone density measured by dual-energy X-ray absorption. Osteoporos Int **1:**141–146.
14. Leslie WD 2007 Factors affecting short-term bone density precision assessment and the effect on patient monitoring. J Bone Miner Res **23:**199–204.
15. Patel R, Blake GM, Herd RJ, Fogelman I 1997 The effect of weight change on DXA scans in a 2-year trial of etidronate therapy. Calcif Tissue Int **61:**393–399.
16. Davidian M, Carroll RJ 1987 Variance function estimation. J Am Stat Assoc **82:**1079–1091.
17. Verbyla A 1993 Modelling variance heterogeneity: Residual maximum likelihood and diagnostics. J Roy Statist Soc Ser B Methodological **55:**493–505.
18. Leslie WD, Metge C 2003 Establishing a regional bone density program: Lessons from the Manitoba experience. J Clin Densitom **6:**275–282.
19. Leslie WD, Caetano PA, Macwilliam LR, Finlayson GS 2005 Construction and validation of a population-based bone densitometry database. J Clin Densitom **8:**25–30.
20. Hansen KE, Binkley N, Christian R, Vallarta-Ast N, Krueger D, Drezner MK, Blank RD 2005 Interobserver reproducibility of criteria for vertebral body exclusion. J Bone Miner Res **20:**501–508.
21. Manitoba Center for Health Policy website. Health Databases: Pharmaceutical (DPIN). Available at http://umanitoba.ca/faculties/medicine/units/mchp/resources/repository/3701.htm. Accessed May 29, 2008.
22. Akaike H 1974 A new look at the statistical model identification. IEEE Trans Automatic Control **19:**716–723.
23. Cavanaugh JE 1997 Unifying the derivations for the Akaike and corrected Akaike information criteria. Stat Probab Lett **33:**201–208.
24. Barnett AG, van der Pols JC, Dobson AJ 2005 Regression to the mean: What it is and how to deal with it. Int J Epidemiol **34:**215–220.
25. Tothill P 2005 Dual-energy x-ray absorptiometry measurements of total-body bone mineral during weight change. J Clin Densitom **8:**31–38.
26. Phillipov G, Seaborn CJ, Phillips PJ 2001 Reproducibility of

DXA: Potential impact on serial measurements and misclassification of osteoporosis. Osteoporos Int **12:**49–54.
27. Nguyen TV, Pocock N, Eisman JA 2000 Interpretation of bone mineral density measurement and its change. J Clin Densitom **3:**107–119.
28. Sadatsafavi M, Moayyeri A, Wang L, Leslie WD 2008 Optimal decision criterion for detecting change in bone mineral density during serial monitoring: A Bayesian approach. Osteoporos Int (in press).
29. McClung MR, Wasnich RD, Hosking DJ, Christiansen C, Ravn P, Wu M, Mantz AM, Yates J, Ross PD, Santora AC II 2004 Prevention of postmenopausal bone loss: Six-year results from the Early Postmenopausal Intervention Cohort Study. J Clin Endocrinol Metab **89:**4879–4885.
30. Sirola J, Kroger H, Honkanen R, Sandini L, Tuppurainen M, Jurvelin JS, Saarikoski S 2003 Risk factors associated with peri- and postmenopausal bone loss: Does HRT prevent weight loss-related bone loss? Osteoporos Int **14:**27–33.

Address reprint requests to:
William D Leslie, MD, MSc
Department of Medicine (C5121)
409 Tache Avenue
Winnipeg, Manitoba R2H 2A6, Canada
E-mail: bleslie@sbgh.mb.ca

Received in original form March 6, 2008; revised form May 29, 2008; accepted June 27, 2008.

APPENDIX: A CLINICAL EXAMPLE USING THE POTENTIAL APPLICATION OF THE REGRESSION EQUATIONS FOR INDIVIDUALIZING BMD MONITORING

Suppose that the baseline total hip BMD for a 65-yr-old woman is 0.860 g/cm^2, with height 162 cm and weight 67 kg. We are interested in estimating an individualized least significant change (LSC) for this patient and in making inferences about her BMD 1 yr from now. The patient will not receive systemic corticosteroids but will be treated with an osteoporosis medication for the intervening period. For simplicity, we assume that the same technologist will perform both scans using the same pencil-beam device and that there is no change in height, weight, or scan area between the two visits.

The final model for the short-term reproducibility sample is

$$\begin{aligned}\mathrm{SD}(\Delta\mathrm{BMD}) = \mathrm{EXP}(&\beta_1 \times \mathrm{age\_per\_decade} + \beta_2 \times \mathrm{BMD\_per\_0.1} \\ &+ \beta_3 \times \mathrm{BMI\_per\_5} + \beta_4 \times \Delta\mathrm{bone\_area} + \beta_5 \\ &\times \mathrm{scanner\_type} + \beta_6 \times \mathrm{diff\_day} + \beta_7 \times \mathrm{diff\_tech} \\ &+ \beta_{\mathrm{intercept}})\end{aligned}$$

with coefficients and covariates in the same order as in Table 3. To calculate the SD of measurement error for this woman, we write the patient's covariates in the same order:

$$\begin{aligned}&\mathrm{age\ (per\_decade)} = 65/10 = 6.5 \\ &\mathrm{BMD\ (per\ 0.1\ g/cm^2)} = 0.860/0.1 = 8.6 \\ &\mathrm{BMI(per\ 5\ kg/m^2)} = 67/(1.62)^2/5 = 5.1 \\ &\mathrm{change\ in\ bone\ area\ (\Delta bone\_area)} = 0 \\ &\mathrm{pencil\text{-}beam\ scanner\ (scanner\_type)} = 1 \\ &\mathrm{different\ day\ scanning\ (diff\_day)} = 1 \\ &\mathrm{different\ technologist\ (diff\_tech)} = 0\end{aligned}$$

and take the β coefficients coming from Table 3. We will have:

$$\begin{aligned}\mathrm{SD}(\Delta\mathrm{BMD}) = \mathrm{EXP}(&0.018 \times 6.5 + 0.070 \times 8.6 + 0.050 \times 5.1 + 0.168 \\ &\times 0 - 0.130 \times 1 + 0.365 \times 1 + 0.211 \times 0 - 6.462) \\ = &\ 0.012\ \mathrm{g/cm^2}\end{aligned}$$

Compared with the SD of 0.014 g/cm^2 for the whole reproducibility sample (Table 1), the short-term SD of measurement (and hence LSC) is ~15% lower for this subject.

For the long-term mean change, the equation is:

$$\begin{aligned}\mathrm{mean}(\Delta\mathrm{BMD}) = (&\beta_1 \times \mathrm{male\_sex} \times \mathrm{interval} + \beta_2 \times \mathrm{age\_per\_decade} \\ &\times \mathrm{interval} + \beta_3 \times \mathrm{BMD\_per\_0.1} \times \mathrm{interval} + \beta_4 \\ &\times \mathrm{BMI\_per\_5} \times \mathrm{interval} + \beta_5 \times \Delta\mathrm{weight} + \beta_6 \\ &\times \Delta\mathrm{height} + \beta_7 \times \mathrm{corticosteroid} + \beta_8 \\ &\times \Delta\mathrm{bone\_area} + \beta_9 \times \mathrm{scanner\_type} + \beta_{10} \\ &\times \mathrm{osteo\_treatment} + \beta_{11} \times \mathrm{interval})\end{aligned}$$

Note that the first four terms are also multiplied by the interval between scans (1 yr in this example) as they are modeled as having an interaction with interval and that there is no intercept for this equation (see Materials and Methods section).

For the long-term SD of change, the equation is:

$$\begin{aligned}\mathrm{SD}(\Delta\mathrm{BMD}) = \mathrm{EXP}\ (&\beta_1 \times \mathrm{male\_sex} + \beta_2 \times \mathrm{age\_per\_decade} + \beta_3 \\ &\times \mathrm{BMD\_per\_0.1} + \beta_4 \times \mathrm{BMI\_per\_5} + \beta_5 \times \Delta\mathrm{weight} \\ &+ \beta_6 \times \Delta\mathrm{height} + \beta_7 \times \mathrm{corticosteroid} + \beta_8 \\ &\times \Delta\mathrm{bone\_area} + \beta_9 \times \mathrm{scanner\_type} + \beta_{10} \\ &\times \mathrm{osteo\_treatment} + \beta_{11} \times \mathrm{interval} + \beta_{\mathrm{intercept}})\end{aligned}$$

with all β coefficients coming from Table 2. Putting the covariates for our patients in the above equations, the mean and SD of change will be 0.008 and 0.031 g/cm^2, respectively. These figures can be used to construct a 95% CI for the future BMD of the patient:

$$\begin{aligned}&95\%\ \text{interval for the predicted}\ \Delta\mathrm{BMD}\ (\mathrm{g/cm^2}) = \\ &\quad [0.008 - 1.96 \times 0.031,\ 0.008 + 1.96 \times 0.031] = [-0.052,\ 0.068].\end{aligned}$$