Evaluating the Value of Repeat Bone Mineral Density Measurement and Prediction of Fractures in Older Women

The Study of Osteoporotic Fractures

Teresa A. Hillier, MD, MS; Katie L. Stone, PhD; Doug C. Bauer, MD; Joanne H. Rizzo, MS; Kathryn L. Pedula, MS; Jane A. Cauley, DrPH; Kristine E. Ensrud, MD, MPH; Marc C. Hochberg, MD; Steve R. Cummings, MD

Background: Whether repeat bone mineral density (BMD) measurement adds benefit beyond the initial BMD measurement in predicting fractures in older women is unknown.

Methods: We prospectively measured total hip BMD in 4124 older women (mean±SD age, 72±4 years) from 1989 to 1990 and again 8 years later. Incident nontraumatic hip and nonspine fractures were validated by radiology reports (>95% follow-up). In addition, spine fractures were defined morphometrically in 2129 of these women by lateral spine x-ray films from 1991 to 1992 and then again 11.4 years later. Prediction of fracture risk was assessed with proportional hazards models and receiver operating characteristic curves for BMD measures.

Results: Over a mean of 5 years after the repeat BMD measure, 877 women experienced an incident nontraumatic nonspine fracture (275 hip fractures). In addition, 340 women developed a spine fracture. After adjustment for age and weight change, initial and repeat BMD measurements were similarly associated with fracture risk (per unit standard deviation lower in BMD) for nonspine (hazard ratio, 1.6), spine (odds ratio, 1.8-1.9), and hip (hazard ratio, 2.0-2.2) fractures (P<.001 for all models). Areas under the receiver operating characteristic curves (AUC) revealed no significant differences to discriminate nonspine (AUC, 0.65), spine (AUC, 0.67-0.68), or hip (AUC, 0.73-0.74) fractures between models with initial BMD, repeat BMD, or initial BMD plus change in BMD. Stratification by initial BMD t scores (normal, osteopenic, or osteoporotic), high bone loss, or hormone therapy did not alter results.

Conclusion: In healthy, older, postmenopausal women, repeating a measurement of BMD up to 8 years later provides little additional value besides the initial BMD measurement for predicting incident fractures.

Arch Intern Med. 2007;167:155-160
placements were excluded. All women provided written consent, and the Study of Osteoporotic Fractures was approved by each site’s institutional review board.

Between 1989 and 1990, 8141 women had a second clinic visit, which included the initial measurement of BMD of the proximal femur and subregions (intertrochanter, trochanter, femoral neck, and Ward triangle) using DXA (Hologic QDR 1000; Hologic Inc, Waltham, Mass). All women in the cohort were invited to participate in another Study of Osteoporotic Fractures visit a mean of 8 years later, beginning in 1997, and 7008 women (93.0% of survivors) completed this examination in the clinic or at home; of these women, 4124 underwent a repeat BMD measurement (on the same machine), and these women are the final study sample for this analysis.

**BMD AND OTHER MEASUREMENTS**

Measurement and quality control procedures were rigorous (detailed elsewhere). The DXA BMD measurement standards and precision have also been previously detailed. The initial total hip BMD measurement was performed from 1989 to 1990, and the repeat total hip BMD measurement was performed a mean of 8 years later (range, 6.3-9.8 years) (Figure 1). The rate of change in BMD was calculated using the 2 total hip BMD measurements, expressed as annualized percentage change and absolute change (measured in grams per centimeter squared per year). The t scores were calculated using the National Health and Nutrition Examination Survey as the reference, and were computed by World Health Organization criteria. Weight was measured in light clothing without shoes by a balance beam scale, and height by stadiometer, at both visits. We calculated percentage weight change between the BMD measures, because weight change was previously found to be a strong predictor of fracture risk in older women.

**FOLLOW-UP FOR ASCERTAINMENT OF FRACTURES AND MORTALITY**

Participants were contacted every 4 months by postcard (with telephone follow-up for nonresponders) to ascertain incident hip and nonspine fractures; more than 95% of these contacts were completed. Incident nonspine fractures were physician adjudicated from radiology reports. To compare the overall sensitivity and specificity of the 2 BMD measurements in predicting incident fractures, it is necessary to compare the same fracture outcomes (ie, only incident fractures after the second repeat measurement). Therefore, we excluded the 513 incident nonspine and the 72 hip fractures that occurred between the initial and repeat BMD measurement, and compared the prediction of incident nonspine and hip fractures after the second BMD measurement in all models (Figure 1). Spine fractures were ascertained by morphometric analysis of lateral thoracic and lumbar spine x-ray films, measured in 2129 women from 1991 to 1992 and at an examination a mean of 11.4 years later (Figure 1). An incident fracture on the second x-ray film was defined as greater than 20% and at least a 4-mm decrease in vertebral height at any level.

**STATISTICAL ANALYSES**

To test the association between BMD and fracture risk, we used Cox proportional hazards models for nontraumatic hip and nonspine fractures. Logistic regression was used to examine the association between BMD and spine fracture because the date of the actual fracture was not known.

We then used logistic regression to examine receiver operating characteristic (ROC) curves and area under the curve (AUC) for BMD to discriminate fracture vs no fracture, with BMD assessed 4 ways: (1) initial BMD alone, (2) repeat BMD alone (a mean of 8 years later), (3) change in BMD between the 2 examinations, and (4) a combined model of initial BMD plus the change in BMD between the 2 examinations.

We evaluated unadjusted models and models adjusted for age and weight change. We also evaluated additional models adjusted for age and baseline weight, to confirm that results were similar to the age- and weight change–adjusted models. There was no evidence of multicollinearity between predictor variables (r < 0.4). The goal of this analysis was to assess the overall utility of each BMD measure to predict fractures, and not to explore the etiology of other potential risk factors for osteoporosis (or to calculate the adjusted hazards ratio of BMD), so we did not adjust for other covariates in these BMD models.

We tested differences in the AUC between the final 4 BMD models with a Wilcoxon-based statistic that adjusts for the between-area correlation of the ROC curves for the entire population. Additional stratified ROC analyses were also evaluated, looking for a potential subgroup in which the ROC curves might diverge more than in the entire population (ie, where there may be a particular benefit for a repeat BMD measurement in fracture risk prediction). We performed analyses stratified by initial total hip BMD t score (normal, >-1.0; osteopenic, -1.0 to -2.5; and osteoporotic, <-2.5). We also stratified by those with a high rate of BMD loss vs other levels of BMD change, and examined 3 different thresholds of high BMD loss (where high BMD loss was defined as the highest 5th, 10th, and 25th percentile).

Because estrogen use—particularly a change in hormone therapy status between BMD measures—could confound this relationship, we also evaluated the ROC curves for each of the BMD measures stratified by women taking estrogen at either BMD examination compared with women not taking estrogen at either examination. Moreover, although the first bisphosphonate (alendronate sodium) was only beginning to be used in clinical practice at the follow-up BMD examination, we did separate analyses, excluding the 8% of women taking alendronate at the repeat BMD examination to rule out any potential confounding.
The characteristics of the 4124 older women, who were a mean age of 72 years at the initial BMD measurement, are shown in Table 1. On average, the participants' BMD at the initial examination was in the low bone mass range (mean initial t score, −1.37), and participants experienced an average BMD loss of −0.59% per year (range, 4.1% gain per year to 6.6% loss per year), which resulted in an average repeat BMD t score 8 years later of −1.64. The cut point values for the highest bone loss groups at the 5, 10, and 25 percentile levels were −2.1%, −1.7%, and −1.1% per year, respectively. During a mean 5 years of follow-up after the repeat BMD measurement, 877 women experienced an incident nontraumatic nonspine fracture, 275 of which were hip fractures. In addition, 340 women developed a morphometric spine fracture during a mean 11.4 years between lateral x-ray examinations.

The initial and repeat BMD measurements were highly correlated (r = 0.92, P < .0001). However, when expressed as change in BMD, the correlation with initial BMD was only r = 0.14 (P < .0001), so we were able to analyze initial BMD plus change in BMD in the same model without an issue of collinearity.

In each of the 4 BMD models (initial BMD, repeat BMD, change in BMD between the 2 examinations, and initial BMD plus change in BMD), BMD was a significant predictor of incident nonspine and hip fracture risk, and was associated with morphometric spine fractures (Table 2). Each standard deviation lower in either initial or repeat total hip BMD was associated with a 55% to 61% increased risk of incident nonspine fracture, a 102% to 121% increased risk of incident hip fractures, and a 75% to 86% increased risk of spine fractures (P < .0001 for all models).

Although change in BMD was an independent predictor of all fracture types, it was a weaker predictor than either initial or follow-up BMD measure alone (Table 2). Each standard deviation loss in BMD was associated with a 26% to 45% increase in fracture risk among the 3 fracture types (P < .0004 for all models), which was less than half the risk associated with 1 SD lower in either initial or repeat BMD measures for each fracture type (Table 2). Moreover, when we included initial BMD with BMD change, the effect of change in BMD was attenuated (Table 2).

In ROC analyses, the only model that consistently performed worse was BMD change alone (P < .05 vs initial BMD). However, there were no statistical differences between the AUC for the other 2 models (repeat BMD or initial plus change in BMD) compared with initial BMD. In fact, the AUC was essentially identical regardless of which of these 3 BMD models was used (Table 2). That is, across the range of sensitivities and specificities, the ROC curves were essentially superimposed for initial BMD, repeat BMD, and initial plus change in BMD to discriminate all fracture types (Figure 2).

The ROC curves and AUCs were similar when results were stratified by initial t score group (normal t score, >−1.0 [n = 1348]; osteopenic t score, −1.0 to −2.5 [n = 2314]; and osteoporotic t score, <−2.5 [n = 621]). We also stratified by those with highest BMD loss vs other BMD change, and results were similar whether the threshold of high BMD loss was the highest 25th, 10th, or 5th percentile (data not shown). Finally, we stratified by any use of estrogen at either BMD measurement (1021 women [24.8%]) vs no estrogen use, and found the results were consistent with the analyses for the full population (eg, the AUC for hip fracture with any estrogen use was 0.74 for initial BMD, 0.76 for repeat BMD, and 0.76 for initial plus change in BMD; P > .40 for both models vs initial BMD). Analyses excluding the 346 women (8.4%) who were using a bisphosphonate at the repeat BMD measurement (this drug was not available at the initial BMD measurement) were also similar to the full population results.

We also assessed ROC curves, with adjustment only for age, and the AUC results were similar. Finally, unadjusted results were similar, and there was not a significant interaction between age and BMD.

In this large, prospective, cohort study of 4124 community-dwelling women 65 years and older, we did not find any improvement in the overall predictive value, or AUC, in a second measure of BMD, obtained a mean of 8 years later, in prediction of hip, spine, or overall nonspine fracture risk. In other words, the initial BMD was highly, and similarly, predictive of fracture risk in our population. The initial and repeat BMD measurements were highly correlated in our population (r = 0.92), consistent with our results showing that each BMD measure was independently predictive of fracture risk, but with a similar AUC.
We found that rate of bone loss was an independent risk factor—albeit weaker than initial BMD—for fracture risk; however, this association was attenuated in models that included the initial BMD measurement (Table 2). Several earlier studies, evaluating BMD change as an independent predictor were limited by small sample sizes, and they measured forearm BMD by single-photon absorptiometry. A recent French study, of 671 postmenopausal women, with annually measured forearm DXA over a median of 11 years before fracture, found that the highest tertile of bone loss compared with the lowest tertile of bone loss predicted fracture risk independent of baseline BMD (hazard ratio, 1.45-1.70), although the point estimates for bone loss were consistently lower than baseline BMD dichotomized at a t score of less than −2.5 (hazard ratio, 2.0-2.5). Nguyen and colleagues observed 966 postmenopausal women 60 years or older for more than 10 years; they found that femoral neck bone loss by DXA was independently associated with fracture risk (hazard ratio, 1.4 per 5% loss), as was baseline femoral neck BMD (hazard ratio, 2.0 per standard deviation). The findings of these recent studies are consistent with our findings that BMD change is an independent, but relatively weaker, predictor of fracture risk compared with baseline BMD.

It is also important to distinguish between the information gleaned from a risk ratio, which is important in disease etiology, and the ROC curve, which is more useful in measuring the value of a screening test across a population—the aim of the present study. In fact, our results confirm the value of a BMD measurement in screening older women for fracture—with an AUC as high as 0.74 for hip fracture. However, for the overall population, we did not find an incremental increase in overall prediction of any fracture type with a repeat BMD test beyond the initial BMD measure. To our knowledge, this is the first study to evaluate the overall predictive value of a repeat BMD measurement compared with the initial BMD measurement to discriminate fractures vs no fractures.

What are the clinical implications of our findings in a population of older women? The ROC curves are a summary of the sensitivity and specificity for every woman’s measured BMD to predict her actual outcome of incident fracture. The lack of overall additional benefit we found for the repeat BMD measurement in women older than 65 years does not imply that a repeat BMD measurement may not be useful for some individual patients—particularly if intervening clinical factors are present that would likely accelerate BMD loss greater than average. It similarly does not contradict the importance of BMD, and BMD loss, in the disease etiology of osteoporosis. However, our results do suggest that, for the average healthy older woman 65 years or older, a repeat BMD measurement has little or no value in classifying risk for future fracture—even for the average older woman who has osteoporosis by initial BMD measure or high BMD loss.

Our study has several important strengths. It is a large prospective study of older women, with rigorous quality control of BMD measurements. Incident nonspine and hip fractures were physician adjudicated, and spine fractures were included only with a measured change by morphometric criteria. In addition, retention of survivors is excellent, as is more than 95% completion of follow-up information on triannual follow-up of fractures—including women who became too frail to attend subsequent visits.

Our study also has some limitations. Repeat BMD measurement is more commonly done every 2 to 5 years in clinical practice, not the 8-year interval present in this analysis. But, if anything, this longer interval should overestimate any potential benefit of more frequent repeat BMD testing. The baseline spine film was measured 2 years after the initial BMD measurement, and thus, we cannot capture incident spine fractures that occurred in those first 2 years. However, the spine results are consistent with the incident nonspine and hip results. Women who were able to attend the follow-up examination to have their BMD measurement were healthier than those who did not attend the follow-up examination. We did comparisons of the ROC curves of the initial BMD measurement in predicting fracture risk in those who attended the follow-up examination (and had the second BMD measurement) vs those who did not, and their curves were superimposed (data not shown). Thus, a large bias is unlikely in our results using only clinic attendees. It is still possible that those women who became too frail to return for a visit (and, thus, a repeat BMD measurement) may have had greater BMD loss. Thus, our results are relevant to healthy older women—and in clinical settings.

### Table 2. Prediction of Fracture Risk With Initial and Repeat DXA Total Hip BMD Measurements

<table>
<thead>
<tr>
<th></th>
<th>Incident Nonspine Fractures (n = 4124)</th>
<th>Incident Hip Fractures (n = 4124)</th>
<th>Morphometric Spine Fractures (n = 2193)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)††</td>
<td>AUC ‡</td>
<td>HR (95% CI)††</td>
</tr>
<tr>
<td>1: Initial BMD</td>
<td>1.55 (1.43-1.67)</td>
<td>0.65</td>
<td>2.02 (1.76-2.33)</td>
</tr>
<tr>
<td>2: Repeat BMD</td>
<td>1.61 (1.49-1.74)</td>
<td>0.62</td>
<td>2.21 (1.91-2.54)</td>
</tr>
<tr>
<td>3: Change in BMD</td>
<td>1.26 (1.17-1.35)</td>
<td>0.61†</td>
<td>1.45 (1.29-1.62)</td>
</tr>
<tr>
<td>Initial BMD</td>
<td>1.51 (1.40-1.63)</td>
<td>0.65‡</td>
<td>1.94 (1.68-2.23)</td>
</tr>
<tr>
<td>Change in BMD</td>
<td>1.18 (1.10-1.26)</td>
<td>0.74‡</td>
<td>1.29 (1.15-1.45)</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under the receiver operating characteristic curve; BMD, bone mineral density; CI, confidence interval; DXA, dual-energy x-ray absorptiometry; HR, hazard ratio; NA, data not applicable; OR, odds ratio.

*Per standard deviation lower in BMD measurement; all models adjusted for age and weight change.

†P<.05 compared with the AUC percent for initial BMD.

‡Model 4 AUC percent is for Initial BMD plus Change in BMD.
in which prevention of the first fracture in a healthy older woman is the reason for the repeat BMD test. Repeat BMD testing to monitor response to treatment for diagnosed osteoporosis is a different issue that is not addressed by this study. Finally, our results were done in postmenopausal women 65 years or older, and may not be generalizable to men, other ethnic groups, and, in particular, younger women, who are transitioning through menopause at differing rates.

Our results indicate that, although BMD is highly predictive of fracture risk, for the average postmenopausal woman 65 years or older who has not yet developed a fracture, repeat BMD measurement provides little additional benefit as a screening tool. However, for individual patients with intervening illness, marked weight loss, or new medication use, such as glucocorticoids, a repeat BMD measurement could be important for identifying new rates of BMD change. Moreover, the value of a repeat BMD measurement in a perimenopausal or an early postmenopausal woman—in whom rates of BMD change are most unpredictable—is likely to be greater. More research is needed in other populations, including younger women, men, and other ethnic groups, to confirm our results.

Accepted for Publication: October 6, 2006.

Correspondence: Teresa A. Hillier, MD, MS, Center for Health Research, Kaiser Permanente Northwest/Hawaii, 800 N Interstate Ave, Portland, OR 97227 (Teresa .Hillier@kp.org).

Author Contributions: Dr Hillier had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Hillier, Stone, and Pedula. Acquisition of data: Hillier, Stone, Cauley, Ensrud, Hochberg, and Cummings. Analysis and interpretation of data: Hillier, Stone, Bauer, Rizzo, Pedula, and Ensrud. Drafting of the manuscript: Hillier. Critical revision of the manuscript for important intellectual content: Hillier, Stone, Bauer, Rizzo, Pedula, Cauley, Ensrud, Hochberg, and Cummings. Statistical analysis: Stone, Rizzo, and Pedula. Obtained funding: Hillier, Stone, Ensrud, and Cummings. Administrative, technical, and material support: Cauley. Study supervision: Stone and Cummings.

Financial Disclosure: Dr Cauley has received research grants from Merck & Co, Inc, Eli Lilly and Company, Pfizer Global Pharmaceuticals, and Novartis Pharmaceuticals; has received honoraria from Merck & Co, Inc, Novartis Pharmaceuticals, and Eli Lilly and Company; and has been on the speaker’s bureau of Merck & Co, Inc.

Funding/Support: This study was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases; and by Public Health Service grants AR35583, AG05407, AR35582, AG05394, and AR35584 from the National Institute on Aging.

Role of the Sponsor: The funding bodies had no role in data extraction and analyses, in the writing of the manuscript, or in the decision to submit the manuscript for publication.

Previous Presentation: This study was presented in part at the American Society for Bone and Mineral Research meetings; September 23-27, 2005; Nashville, Tenn.

Acknowledgment: We thank Diann Triebwasser for her technical assistance, and Martie Sucec for her editorial review.
REFERENCES