Development and Validation of a Simple Questionnaire to Facilitate Identification of Women Likely to Have Low Bone Density

Eva Lydick, PhD; Karen Cook, PhD; Jennifer Turpin, MS; Mary Melton, MD; Robert Stine, PhD; and Christine Byrnes, MD

Abstract

The relationship between low bone mass and risk of fracture is well documented. Although bone densitometry is the method of choice for detecting low bone mass, its use may be limited by the availability of equipment, cost, and reimbursement issues. Improved patient selection for bone densitometry might increase the cost-effectiveness of screening for osteoporosis, a goal we sought to achieve by developing and validating a questionnaire based solely on patient-derived data. Responses to the questionnaire were used to assign postmenopausal women to one of two groups: (1) those unlikely to have low bone mineral density (defined as 2 standard deviations or more below the mean bone mass at the femoral neck in young, healthy white women) and therefore probably not currently candidates for bone densitometry; and (2) those likely to have low bone mineral density and therefore probably candidates for bone densitometry. We asked community-dwelling perimenopausal and postmenopausal women attending one of 106 participating multispecialty centers (both academic and community based) to complete a self-administered questionnaire and undergo bone density measurement using dual x-ray absorptiometry. We used regression modeling to identify factors most predictive of low bone density at the femoral neck in the postmenopausal group. A simple additive scoring system was developed based on the regression model. Results were validated in a separate cohort of postmenopausal women. Data were collected from 1279 postmenopausal women in the development cohort. Using only six questions (age, weight, race, fracture history, rheumatoid arthritis history, and estrogen use), we achieved a target of 89% sensitivity and 50% specificity. The likelihood ratio was 1.78. Validation in a separate group of 207 postmenopausal women yielded 91% sensitivity and 40% specificity. Assuming population characteristics similar to those of our development cohort, use of our questionnaire could decrease the use of bone densitometry by approximately 30%. Sensitivity and specificity can be varied by changing the level for referral for densitometry to provide the most cost-effective use within a particular healthcare setting. Thus use of our questionnaire, an inexpensive prescreening tool, in conjunction with physician assessment can optimize the use of bone densitometry and may lead to substantial savings in many healthcare settings where large numbers of women require evaluation for low bone mass.

(Am J Man Care 1998;4:37-48)

The strong relationship between low bone mass and risk of fracture has been well documented in several large, prospective epidemiologic studies.[1-6] Bone mass can now be measured safely and accurately, allowing confirmation of a diagnosis of osteoporosis through quantitative assessment of bone mineral content. Increased recognition of osteoporosis as a preventable and treatable disease will likely increase use of bone densitometry for predicting fracture risk.

Based on estimates from epidemiologic data, screening for and estrogen treatment of osteoporosis is as cost-effective as established screening and treatment regimens for diseases such as hypertension and breast cancer.[7-9] Nonetheless, measuring bone mass in all postmenopausal women to detect those who are likely to develop or who already have osteoporosis can be costly. Use of an inexpensive patient-administered

questionnaire with adequate sensitivity in conjunction with physician assessment to identify women at risk for osteoporosis, followed by a test with higher specificity (ie, bone densitometry) may be the ideal screening strategy for osteoporosis, both on the basis of costs and the availability of densitometric equipment.

Our study objective was to develop a questionnaire based on easily obtained patient information that could be used to assign postmenopausal women to one of two groups: (1) those unlikely to have low bone mineral density (BMD) and therefore probably not candidates for bone densitometry at the present time; and (2) those likely to have low BMD and therefore should be referred for further evaluation. Because desirable levels of sensitivity and specificity vary based on costs of densitometry and other variables within a particular healthcare setting, our goal was to create an instrument with the flexibility of optimizing either sensitivity or specificity at several levels of bone density, thus maximizing generalizability and potential clinical utility. After we developed the questionnaire, we validated its properties in a separate cohort of postmenopausal women. We named the resulting computationally simple questionnaire SCORE*, for Simple Calculated Osteoporosis Risk Estimation.

··· PATIENTS AND METHODS ···

A total of 106 investigators specializing in family medicine, geriatrics, or general internal medicine (50% of sites), endocrinology (20%), rheumatology (20%), or gynecology (10%) participated in the study. Investigators at each site were required to obtain institutional review board approval and informed consent from each patient. Investigators were asked to recruit the first 10 to 15 women who were seen for routine check-up or follow-up of any medical condition and who met all of the inclusion and none of the exclusion criteria.

Women were enrolled in the study between October 1994 and February 1995. Community-dwelling perimenopausal and postmenopausal women aged 45 or older were eligible for the study. Study participants had to be able to read English and provide informed consent. They could be of any race or ethnic group. Women were excluded if they had significant scoliosis, trauma, or sequelae of orthopedic procedures prohibiting BMD measurements of the spine or hip using dual x-ray absorptiometry;

metabolic bone disease (other than osteoporosis); cancer with metastasis to bone; or renal impairment (serum creatinine > 2.5 mg/dL). Only women classified as postmenopausal (ie, amenorrheic for at least 6 months before study enrollment) were included in the analysis.

Development Cohort

Each woman in the development cohort was asked to complete a self-administered questionnaire composed of approximately 60 questions on factors possibly or probably associated with osteoporosis (Table 1). We selected the items included on our questionnaire based on an extensive review of the medical literature.[10-23] Questions elicited closed-ended responses whenever possible and were written using familiar or lay language. Site personnel did not help the study participants interpret or complete the questionnaire.

Blood and urine samples were collected for measuring serum bone-specific alkaline phosphatase[24] and urinary N-telopeptide of type I collagen[25,26] as markers of bone formation and resorption, respectively, and analyzed for the first 1000 women enrolled.

Hip and posterior-anterior lumbar spine BMD were measured using dual x-ray absorptiometry. To ensure generalizability of study findings, use of Lunar (Madison, WI), Hologic (Waltham, MA), and Norland (Fort Atkinson, WI) densitometers was permitted. Each site was required to provide its own long-term quality control data using hydroxyapatite phantoms and any additional devices as recommended by the manufacturer. All BMD measurements were expressed as t scores based on the mean and SD of the manufacturer's reference population at the particular anatomic site. Bone mass in the same person may be considered low as determined by one instrument and its associated normative values, yet not considered low when measured using an instrument from a different manufacturer with a different reference population.[27] Given the lack of consensus on reference populations, we used values recommended by the manufacturer of the specific instrument. For this analysis, low BMD was defined as 2 SD or more below the mean bone mass at the femoral neck in young, healthy white women. Although no single level of BMD should be used as the sole basis for treatment, this level of bone mass is associated with a significant risk of fracture and recognized as the "fracture threshold."[28-31]

Model Development

The initial pool of potential factors associated with osteoporosis contained more than 350 variables. To reduce the pool size to a more manageable number of

---

factors for modeling purposes, we performed univariate analyses on each potential factor to explore the correlation of that variable with each of the following two measures of bone density: (1) the dichotomous outcome of low versus not low BMD at the femoral neck; and (2) the actual t scores for BMD at the femoral neck. When bone density was expressed as low or not low, we used Fisher's exact test for binary response factors to determine whether the pattern of responses for the potential factor differed by whether a woman's BMD was considered low or not low. We used the extended Mantel-Haenszel test for ordered categorical factors and the chi-square test for nominal factors.[32,33] We also used univariate analyses of variance to assess the correlation of each potential factor with the actual BMD t scores. Factors having at least a marginal relationship with one of the two measures of BMD ($P \leq 0.2$) were retained in the variable pool as candidate factors. In addition, because of a purported relationship to bone mass, several factors with strong clinical interest were retained in the pool despite P values greater than 0.2 (eg, smoking, alcohol abuse, and family history of fracture). Through this process, we identified 123 potential factors. Items in Table 1 are general categories covered on the questionnaire; within these categories were many variables.

The next step was to assess the amount of missing data for each of these factors. Some answers were left blank on the questionnaire. Generally, any item with more than 4% missing data was excluded from the pool. Among the factors excluded because of excessive missing data were several factors examining family history of fracture (between 12% and 20% missing data). In an attempt to make the family history questions usable, we ran the analyses again assuming that if there was a family history of fracture or humped posture, the women completing the development questionnaire probably would have been aware of it. Therefore, the missing or unknown responses were probably representative of women who would have answered "no" to the family history questions. However, when the analyses were rerun, all variables remained nonsignificant ($P > 0.20$). Mean BMD t scores were smaller, although not significantly, for women answering "yes" to each history question compared with women who responded "no" to these questions. Mean BMD t scores for women who could not answer the family history questions consistently fell between the mean t scores of women answering "no" and those answering "yes." These mean values suggested that the

nonresponders were probably a mixture of women for whom a family history was present and for whom it was not. This does not imply that family history is

Table 1. Items on Questionnaire Used in the Development Phase

Demographics	Race/ethnic group
	Hair color
	Eye color
	Sunburn tendency
Body Measurements	Weight
	Weight pattern
	Height
	Height loss
	Posture
Lifestyle Data	Education
	Recreational activities
	Weight-bearing exercise
	Walking
	Lifting
	History of physical activity
	Exposure to sunlight
	Smoking
	Alcohol consumption
	Caffeine consumption
	Childhood nutrition
	Milk consumption
	Yogurt consumption
	Calcium supplementation
Reproductive History	Menarche
	Menopause
	Pregnancy
	Oral contraceptive use
	Hysterectomy
	Ovariectomy
Other Medical History	General health
	Bedridden in past
	Diabetes
	Eating disorder
	Arthritis
	Fracture history after age 45
	Tooth loss
	Fracture and height loss in family
Current, and Past Medication	Estrogen
	Progestins
	Antacids containing aluminum
	Steroids
	Thyroid medication
	Thiazide diuretic
	Sedatives
	Medication for epilepsy
	Anti-estrogens
	Heparin

not a risk factor, but rather that women in this study could not reliably provide the information, unassisted, in an office setting.

After we excluded factors with too much missing data, 101 variables remained in the candidate pool for potential inclusion in a model predictive of BMD. The modeling process was to be repeated twice, both with and without markers of bone turnover (bone-specific alkaline and N-telopeptide of type I collagen) in the candidate factor pool to determine whether information on the rate of bone formation and resorption increased the predictive ability of the model. Using the reduced pool of candidate factors, we performed a multivariate linear regression to model actual t scores and a multivariate logistic regression to model risk of low bone density. We used three methods of variable selection: forward selection, backward selection, and stepwise selection. The forward and backward selection procedures were useful in further reducing the candidate pool to a size more manageable with the computer-intensive stepwise selection procedure. All three variable selection techniques gave similar results. Because the stepwise procedure is more flexible in its selection algorithm, we used this method to finalize the prediction model.

We evaluated the screening characteristics of the candidate linear and logistic regression models by calculating the sensitivity and specificity for the cutpoint value of probability that gave 90% sensitivity. The linear model slightly outperformed the other candidate models and was chosen for adaptation to a simplified, additive scoring system. Only results of the linear model and the scoring system are reported in this article.

Before finalizing the model, we performed residual analyses to test for normality, heterogeneity of variance, model inadequacies (such as a missing term), outliers, and missing higher-degree polynomial terms.[34] Factors retained in the final model were significant at an alpha level of 0.05.

Model Evaluation

We estimated a woman's BMD t score by substituting her responses to the development questionnaire for the factors in the final linear model equation. The resulting prediction was then used to estimate the probability of the woman having low BMD. Using the predicted t score and the SD of the predicted t score for an individual observation, and assuming the t scores were normally distributed, the probability of a woman having low BMD was the probability that the estimated t score was -2.0 or less. A woman would be classified as having low BMD if this probability exceeded a pre-

specified probability, called the cutpoint. The cutpoint chosen was the probability that gave 90% sensitivity for the women in the development cohort.

Using the estimated probabilities of low BMD obtained from the linear model, we assessed the goodness of fit using the Hosmer-Lemeshow test.[35] Receiver operating characteristics (ROC) curves for the linear model were obtained by calculating sensitivity and specificity for different cutpoint values and their areas under the curve (AUC) were measured.[36]

Development of SCORE

Using those factors selected by the linear regression model as input, we developed an equation that preserved the predictive ability of the model but had a simpler form. To increase ease of use, we modified the regression coefficients for variables to yield integer values. The integer values derived for each variable were then summed to give the SCORE, or Simple Calculated of Osteoporosis Risk Estimation.

Validation Cohort

A shorter questionnaire incorporating the questions most predictive of BMD was tested in a second cohort of perimenopausal and postmenopausal women at a subset of the sites that participated in the development phase of the study. Study participants were enrolled between September and October 1995. Inclusion and exclusion criteria and study procedures were identical to those for the development phase. To evaluate the reliability of findings, women completed the shortened self-administered questionnaire on two different occasions, separated by no less than 2 days and no more than 30 days. Based on results seen during model development, we did not assess markers of bone turnover in the validation cohort.

Overall assessments of reproducibility for each candidate prediction model were calculated using the intraclass correlation coefficient[37] and the concordance correlation coefficient.[38] The ROC curves and their AUCs[39] were calculated and compared for both the development and the validation cohorts. The Hosmer-Lemeshow test was used to assess the fit of the models.

⋯ RESULTS ⋯

Approximately 1600 women were asked to participate in the development phase of the study. A total of 1424 women from 106 physician practices participated; 1279 were considered to be postmenopausal. Femoral neck BMD measurements were available for 1246 women, of whom 473 (38%) were classified as having

low BMD. Demographic characteristics for the development cohort are shown in Table 2.

Modeling

Modeling efforts focused on the subset of postmenopausal women, using the femoral neck measurement as the most clinically relevant site. Adding data on markers of bone turnover did not improve model predictions (R^2 of 0.44 with markers and 0.45 without). The final linear model had an R^2 of 0.40, had no significant lack of fit ($P > 0.2$, Hosmer-Lemeshow test), and included eight questionnaire items. The variables selected for the final model and parameter estimates are summarized in Table 3.

Model Evaluation

The models were evaluated by calculating the sensitivity and specificity for various cutpoints. Sensitivity is the proportion of women with true low bone mass classified correctly by the model. Specificity is the proportion of women who do not have low bone mass who are classified correctly. The area under the ROC curve gives an overall measure of joint sensitivity and specificity of the model. This area measures the probability that, in randomly paired low BMD and non-low BMD women, the predicted probability of low BMD is larger for women with true low BMD. Values of the AUC in the range of 0.80 or greater are considered acceptable.[36] The ROC curve for the linear model had an AUC of 0.811 (Figure 1). Selecting a cutpoint for the linear model that gave 90% sensitivity had a corresponding specificity of 47%.

Development of SCORE

In developing a simplified scoring system, we removed height loss as a variable because of missing data (14% of all responses). Years postmenopause was eliminated because of the amount of missing data and a significant change in the responses from the first and second administration of the questionnaire. To simplify patient recall, years taking estrogen was dropped and converted to "never" or

Table 2. Demographic Characteristics of Development and Validation Cohorts

	Development Cohort (n = 1424)	Validation Cohort (n = 259)
Postmenopause	1279 (90%)	208 (80%)
Perimenopause	145 (10%)	31 (12%)
Unknown	0	20 (8%)
Postmenopausal Women		
Age (y) (mean ± SD)	61.5 ± 9.6	63.1 ± 9.5
Race (%)		
White	89	94
Black	6	1
Hispanic	3	2
Other or missing	3	3
Estrogen (%)		
Current users	45	54
Past users	12	23
Rheumatoid Arthritis (%)	5	24
Densitometer Used (%)		
Lunar	54	55
Hologic	41	45
Norland	3	0
Unknown	2	<1
Low BMD* (%)		
Hip	38	44
Spine	24	21
Both hip and spine	18	17
Low BMD* (%) at the Hip for:		
Estrogen		
Current users	183/552 (33%)	36/111 (32%)
Never/past users	280/669 (42%)	55/93 (59%)
Rheumatoid Arthritis		
No	432/1166 (37%)	65/157 (41%)
Yes	37/66 (56%)	26/49 (53%)
Race		
White	437/1101 (40%)	84/194 (43%)
Black	10/65 (15%)	1/2 (50%)
Hispanic	15/38 (40%)	1/4 (25%)
Other or missing	7/28 (25%)	5/7 (71%)
Hip Fracture		
No	434/1185 (37%)	89/202 (44%)
Yes	19/21 (90%)	0/2 (0%)
Rib Fracture		
No	419/1158 (36%)	81/194 (42%)
Yes	32/45 (71%)	6/7 (86%)
Wrist Fracture		
No	413/1137 (36%)	77/180 (43%)
Yes	44/71 (62%)	10/20 (50%)

SD = standard deviation; BMD = bone mineral density
*Low BMD was defined as 2 SD or more below the mean bone mass for young, healthy white women based on the manufacturer's reference database.

Table 3. Linear Regression Models for Bone Mineral Density (*t* Score): Variables and Parameter Estimates

Variable	Initial Linear Regression Model		Simplified Linear Regression Model	
	Estimate	Standard Error	Estimate	Standard Error
Intercept	-1.9193	0.3684	-0.3686	0.3090
Weight	0.0153	0.0011	0.0144	0.0009
Age	-0.0279	0.0058	-0.0424	0.0034
Years Postmenopause	-0.0144	0.0049	Dropped	
Years Taking Estrogen	0.0155	0.0048	Dropped	
Never Used Estrogen			-0.1278	0.0635
Rheumatoid Arthritis	-0.5434	0.1591	-0.5736	0.1415
Race				
Black	0.7093	0.1753		
Hispanic	0.2015	0.1886		
Asian	0.3120	0.3287		
Native American	-0.0253	0.1372		
Other	-0.1083	0.4230		
Race other than Black			-0.7002	0.1447
Height Loss	-0.2503	0.0821	Dropped	
Fracture*				
Rib	-0.6307	0.2049		
Hip	-0.9168	0.3027		
Wrist	-0.4416	0.1557		
Fracture				
One of the above			-0.5619	0.1098
Two of the above			-1.1507	0.3113

*Self-reported fracture after age 45

Figure 1. Receiver Operating Characteristic Curve Based on a Linear Model for the Development Cohort
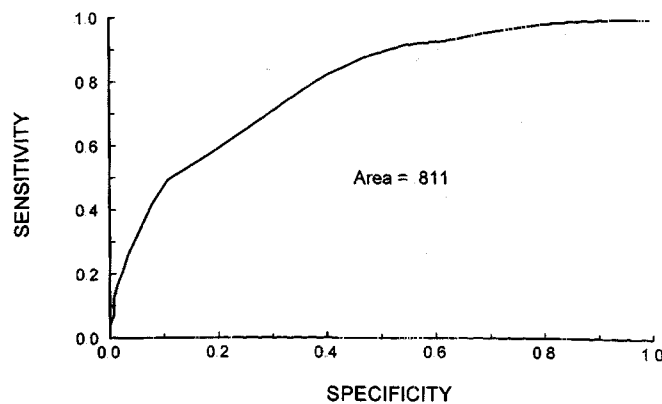

Area = 811

"currently past use" in SCORE. Because being black was the only racial variable that was a significant predictor of bone density, race was re-expressed as 0 if black and 1 if other. Regressing the remaining variables on bone density *t* scores resulted in the parameter estimates shown in Table 3.

We developed a scoring system using these new parameter estimates. SCORE expresses other coefficients in the regression model relative to the coefficient for weight. Age and weight were rescaled for multiples of 10, allowing the use of smaller integers. Weight and age are not rounded to 10, but rather truncated for ease of administration and calculation so that, for example, 67 becomes 6. Because the coefficient for a history of wrist fracture was approximately the same as that for rib or hip fracture, we combined fracture history into a single variable from 0 (no history of wrist, rib, or hip fracture) to 3 (positive history of all three fracture types).

Final variables included in SCORE and a worked example are shown in Table 4. The threshold score closest to the target of 90% sensitivity was a value of 6; this threshold resulted in a specificity of 50%. The relationship between measured BMD and SCORE in the development cohort is shown in Figure 2. Nearly all women with a SCORE of 12 or more were in the osteoporotic or osteopenic range of BMD (ie, BMD ≤ -1 SD) and those with a SCORE above 20 were all in the osteoporotic range (BMD ≤ -2 SD). The positive predictive value for SCORE using a threshold value of 6 was 52%; the nega-

tive predictive value was 89%. The likelihood ratio[40,41] at this threshold was 1.78.

When various subsets of women in the development cohort were examined, the predictive properties of SCORE did not differ substantially from the overall findings. For example, sensitivity and specificity in women whose BMD was measured using Hologic devices was 87% and 58%, respectively; for BMD measured using Lunar devices, comparable values were 93% and 45%, respectively. When current estrogen users were removed from the development cohort, sensitivity and specificity were 90% and 47%, respectively. When black women and women with rheumatoid arthritis also were removed, sensitivity and specificity were 90% and 43%, respectively. When data from the development cohort were examined by age decade, sensitivity of SCORE ranged from 65% in women aged 45 to 54 to 100% in women aged 75 or older.

These findings apply to the predictive ability of SCORE for bone mass at the femoral neck when low BMD is defined as 2 SD or more below the mean bone mass in young, healthy white women. If low BMD is defined as 2.5 SD or more below the mean bone mass, approximately 90% sensitivity is attained at a threshold score of 7, with 52% specificity (Table 5). When low BMD is defined as 1 SD or more below the mean, similar results are seen at a threshold SCORE of 4. Sensitivity and specificity of SCORE for identifying women with low bone mass (t score $\leq$ -2.0) at the spine was 88% and 43%, respectively.

Validation Sample

Two hundred fifty-nine women participated in the validation phase of the study. Of these, 208 classified themselves as postmenopausal (ie, amenorrheic for at least 6 months before study entry). Bone mineral density measurements were available for 207 patients, of whom 44% had BMD readings at the femoral neck that were 2 SD or more below the mean bone mass in young, healthy white women (Table 2).

Validation of Model and SCORE

The reliability of model scores obtained during the first and second times the women answered the questionnaire was estimated by substituting values of the relevant variables into each of the models, multiplying by the

corresponding parameter estimates, and summing across variables. All models had excellent reliability, with both the intraclass correlation coefficient and the concordance correlation coefficient exceeding 0.96 (values > 0.75 indicate excellent reliability).[42]

The linear regression model and its scored application performed equivalently in both the development and validation cohorts. Using the cutpoint for linear regression from the development phase for 90% sensitivity, the validation data gave values for sensitivity of 97% and for specificity of 39%. The AUC for the validation phase results was 0.75 (Table 6). Despite significant lack of fit with the validation phase data (P = 0.03, Hosmer-Lemeshow test), the regression model demonstrated excellent discriminatory power in terms of sensitivity and specificity. When examining the contribution of the individual deciles to the overall Hosmer-Lemeshow test results, lack of fit primarily occurred in two deciles. Although the predicted probabilities differed from what was expected, in the one decile the model had 100% sensitivity and 71% specificity. The other decile corresponded to the tenth decile of women having the largest estimated probabilities. All these women were predicted to have low BMD. The sensitivity was 100%, but the specificity was 0% (0 out of 5). Using a threshold SCORE of 6, sensitivity was virtually identical for both the development and the validation phases (0.89 versus 0.91); specificity was 0.50 and 0.40, respectively.

Table 4. Coefficients for Calculating SCORE (Simple Calculated Osteoporosis Risk Estimation)

| Variable	Score	If Woman
Race	5	is NOT black
Rheumatoid Arthritis	4	HAS rheumatoid arthritis
History of Fractures	4	for EACH TYPE (wrist, rib, hip) of nontraumatic fracture after age 45 (maximum score = 12)
Age	3	times first digit of age in years
Estrogen	1	if NEVER received estrogen therapy
Weight	-1	times weight divided by 10 and truncated to integer

SCORE equals sum of above.
EXAMPLE: A 126-pound, 67-year-old white woman with a history of rheumatoid arthritis, no history of fractures, and a history of estrogen therapy would have a SCORE of 15, or 5 (for race) + 4 (for rheumatoid arthritis) + (4 x 0) (for no history of fracture) + (3 x 6) (for age) + 0 (for previous estrogen therapy) - (1 x 12) (for weight). Since 15 is greater than the threshold score of 6, this woman should be referred for bone densitometry.

Figure 2a. Relationship Between Bone Density (t Score) and SCORE (Simple Calculated Osteoporosis Risk Estimation) in Postmenopausal Women in the Development Cohort
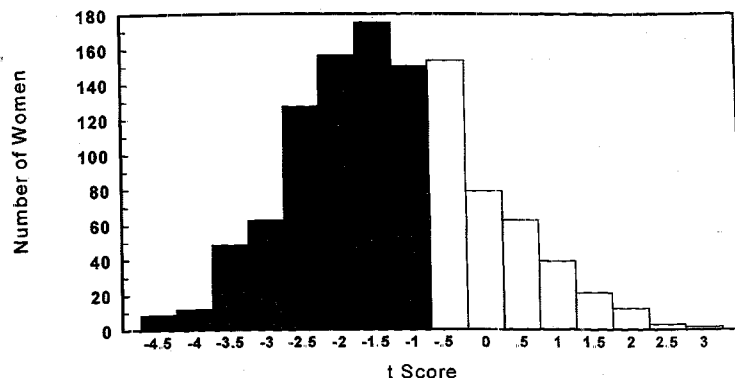

Distribution of t scores for bone mineral density of the femoral neck

Figure 2b. Distribution of Estimated SCORE for the Same Women
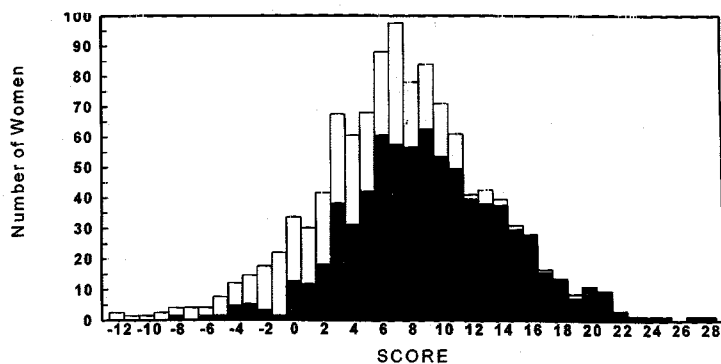
For both histograms, darkest shading indicates bone mass 2 standard deviations or more below the mean for young, healthy white women; lighter shading indicates bone mass between 1 and 2 standard deviations below the mean for young, healthy white women.

Figure 2c. Scatter Plot of t Score Distribution by SCORE
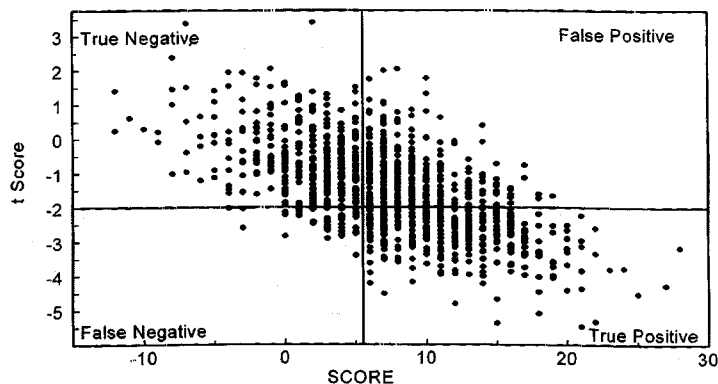
Horizontal line denotes t score of 2 standard deviations below the mean for young, healthy white women; vertical line denotes SCORE of 6.

··· DISCUSSION ···

Other researchers have previously examined the relationship between clinical variables and bone mass,[10-14,16-20,22,23] and their efforts have ben reviewed by Ribot et al.[43] Most of these studies focused on understanding the relationship between various factors and bone mass to identify possible causes of osteoporosis and define subgroups of patients at high risk for this disease. Some investigators used multivariate modeling techniques to determine which variables account for most of the variance in bone mass. Based on their findings, some authors suggested that prediction models might be useful in reducing the number of women needing BMD measurement. However, most researchers have been disappointed with their results, largely because their models lacked adequate sensitivity, misclassifying a large percentage of women with low bone mass.

We designed a model capable of predicting BMD better than most models previously developed, probably because of the large number of women used in the development phase, their greater demographic diversity, and a larger pool of potential factors than that used in most other studies. Our initial model was based on the most predictive factors from a pool of more than 100 variables among 1246 postmenopausal women, and results were validated in a separate cohort of 207 postmenopausal women. This model and a simplified scoring system (SCORE) derived from the model identified women with low BMD at the hip with approximately 90% sensitivity and at least 40% specificity in both the development and the validation cohorts. Both the model and the scoring system thus correctly classified approximately 62% of the women in the validation cohort (90% of the women with low BMD and 40% of the women without low BMD).

Our study population was fairly diverse and meant to be representative of community-dwelling women older than age 45 seen in outpatient practices. In terms of distribution of BMD, our study population compared well with both the general population and the non-Hispanic white population in the United States. Based on data from the Third National Health and Nutrition Examination Survey (NHANES III), Looker et al[44] estimated that 20% of the US population of noninstitutionalized women older than age 50 would have a BMD measurement at the femoral neck of 2.5 SD or more below the mean bone mass for young, healthy women. This percentage increases to 22% when only non-Hispanic white women are included. Within our development population, 24% of all women over age 50 and 25% of white women over age 50 had BMD measurements at the femoral neck that were 2.5 SD or more below that of young, healthy women.

Because of our large sample size, outlying observations probably had little effect on the choice of factors that entered the model. Nevertheless, the following caveats should be noted. The model was developed

Table 5. Sensitivity and Specificity of SCORE (Simple Calculated Osteoporosis Risk Estimation) for Various Threshold Values and Levels of Bone Mineral Density (BMD)

Threshold SCORE	Prediction of BMD Sensitivity	≤ -2 SD Specificity	Prediction of BMD Sensitivity	≤ -2.5 SD Specificity	Predicition of BMD Sensitivity	≤ -1 SD Specificity
0	.993	.124	.996	.103	.980	.199
1	.986	.167	.992	.140	.963	.256
2	.976	.203	.992	.176	.948	.305
3	.966	.257	.992	.223	.924	.368
4	.942	.340	.974	.294	.872	.447
5	.925	.417	.954	.361	.831	.529
6	.894	.497	.936	.433	.776	.602
7	.816	.578	.883	.520	.694	.678
8	.737	.672	.822	.614	.616	.787
9	.674	.746	.769	.692	.540	.845
10	.604	.824	.708	.772	.456	.902

SD = standard deviation.

Table 6. Summary of Results in Development and Validation Cohorts

Method	Cohort	n*	Sensitivity	95% Confidence Interval[†]	Specificity	95% Confidence Interval[†]	AUC
Linear Model	Development	816	0.90	84 - 94	0.47	.42 - .51	0.81
	Validation	142	0.97	78 - 1.00	0.39	26 - 55	0.75
SCORE	Development	1102	0.89	86 - 92	0.50	47 - 52	0.77
	Validation	185	0.91	81 - 96	0.40	30 - 52	0.72

AUC = area under the curve; SCORE = simple calculated osteoporosis risk estimation.
*Number of women with nonmissing responses on all variables used in the predictor.
†Confidence intervals obtained using jackknife procedure on the logits

for predicting bone mass at the femoral neck only. Despite this, the predictive ability of SCORE for the spine was comparable to that for the femoral neck in our development population. However, findings may differ in other populations. In addition, factors other than those included in SCORE may be more predictive of bone mass at the spine. Separate models designed specifically for predicting spinal bone mass were not included in our analyses. Sensitivity and specificity of the current model also may vary depending on other characteristics of the cohort of women tested. For example, lower sensitivity was noted in postmenopausal women younger than age 50 in the development cohort. Thus alternative thresholds should be considered when applying SCORE to women in this age group. Additionally, since our study included very few nonwhite women, it is possible that further research may identify alternative thresholds more appropriate for women of other racial/ethnic groups.

A potential limitation of SCORE may be its development on the basis of BMD measurements made by using different densitometers with differing reference populations. However, develop- ing the instrument in this way increased its generalizability and potential clinical usefulness. In fact, when data derived from Hologic versus Lunar densitometers were compared, the sensitivity and specificity of SCORE did not differ substantially from the overall results.

The model and SCORE provide the flexibility to choose the sensitivity or specificity desired for a particular BMD level. For this report, we defined low BMD as 2 SD or more below the mean bone mass in young, healthy women. SCORE can also be used to identify postmenopausal women at risk for low bone mass at a different level of BMD (eg, 1 SD below the mean bone mass in young, healthy women), which may be useful when considering treatment of osteopenia and prevention of osteoporosis. We estimate that in a cohort of 1000 women similar to those in our study (assuming 38% prevalence of low bone mass), using SCORE as a pretest set at 90% sensitivity and 40% specificity, densitometry would not be indicated for approximately 30% of the women, with only 10% of those with low bone density missed during screening. In addition, we also would have identified 62% (200/321) of the women in our development cohort with osteopenia (BMD between 1 and 2 SD below the mean bone mass for young, healthy women) (Figure 2). The savings from the 30% of women not sent for densitometry could be substantial in areas where densitometry costs are high or access is limited.

Our model is based solely on six self-reported items, and the responses are scored in a simple manner, resulting in an inexpensive pretesting tool. Adding data on markers of bone turnover did not enhance predictive ability. This simple model cannot, of course, be used in place of densitometry. However, because of the great differential in cost in time and dollars, this instrument could be useful in settings in which large numbers of women require evaluation for low bone mass. Desirable levels of sensitivity and specificity will depend on both healthcare costs and disease-related variables.

With the exception of assessing the effects in clinical use, SCORE meets all the published criteria for clinical prediction models.[45,46] Application of the instrument needs to be explored with further validation studies and in populations with differing characteristics. The final value of screening for low bone mass, either with densitometry or SCORE followed by selective densitometry, will depend on the relative costs of treating hip fracture and osteoporosis.[7-9,47] However, based on its performance in this study and existing epidemiologic data, we expect that SCORE as a pretest for osteoporosis and in conjunction with physician assessment can optimize the use of densitometry and may have a place in many healthcare settings today.

Jacksonville, FL; Harry Collins, Edison, NJ; Edmond Confino, Chicago, IL; Fiona J. Cook, Greenville, NC; Jahangir Cyrus, Louisville, KY; William Delaney, Danbury, CT; Jerry M. Earll, Washington, DC; Gary W. Edelson, Detroit, MI; W. Travis Ellison, Greer, SC; Mark Epstein, Encino, CA; Susan Erlenborn, Berwyn, IL; Richard Esham, Mobile, AL; Susan Fox, New York, NY; Karen Freund, Boston, MA; David R. Gelbart, Santa Cruz, CA; F. Wilton Germino, Orlando Park, IL; Husam Ghusn, Houston, TX; Joe Gladden, Kansas City, KS; F. Michael Gloth, III, Baltimore, MD; David Goff, Raleigh, NC; G.M. Gollapudi, Midland, TX; Elizabeth Grace, Denver, CO; Maria Greenwald, Palm Springs, CA; William Griffin, Corpus Christi, TX; Gordon Guthrie, Lexington, KY; M.J. Haller, Omaha, NE; Donna Heinemann, Stonybrook, NY; Mona Horwitz, Stony Brook, NY; Thomas Hughes, Memphis, TN; Alan Jacobs, Lincoln, NE; Jolene Johnson, Baton Rouge, LA; Fran E. Kaiser, St. Louis, MO; Avedis K. Khachadurian, New Brunswick, NJ; Romesh Khardori, Springfield, IL; Insung Kim, Tulsa, OK; Diana Koster, Albuquerque, NM; Leonard Kraut, Lakewood, NJ; June La Valleur, Minneapolis, MN; Susan Langbehn, Iowa City, IA; Matthew Leinung, Albany, NY; David Levenson, Miami, FL; Mitchell Lowenstein, Clearwater, FL; David Lurie, Johnson City, TN; Alan Marcus, Laguna Hills, CA; William R. Markel, Broomfield, CO; Raj Marwah, El Paso, TX; A. Wayne Meikle, Salt Lake City, UT; William Mitchell, Albuquerque, NM; Mack C. Mitchell, Baltimore, MD; Dianne Petrone, Dallas, TX; Stephen R. Plymate, Tacoma, WA; David Podlecki, Longmont, CO; Joe W. Ramsdell, San Diego, CA; Veronica Ravnikar, Worcester, MA; Lorie Reder, Summit, NJ; John Reed, Worcester, MA; Richard Reese, Lancaster, PA; B.A. Rimer, Amarillo, TX; Theodore Rooney, Des Moines, IA; Sidney Rosenblatt, Irvine, CA; Henry Rothschild, New Orleans, LA; Nicholas Salerno, Marlton, NJ; Raja Sayegh, Boston, MA; Catherine Scheraldi-Nunez, Englewood, NJ; Elliott Schwartz, Oakland, CA; John Seibel, Albuquerque, NM; Mona Shangold, Philadelphia, PA; Tammi Shlotzhauer, Rochester, NY; James Shwayder, Newport News, VA; Richard Sievers, II, Dayton, OH; Christine Simonelli, St. Paul, MN; David Small, Syracuse, NY; Richard F. Spark, Boston, MA; Shawn Stinson, Baltimore, MD; Richard M. Stockwell, Farmington, CT; John Stokes, Palm Harbor, FL; Carol Teutsch, Atlanta, GA; Robert Thoburn, Gainesville, FL; Angelina Trujillo, Sioux Falls, SD; Alan A. Wanderer, Englewood, CO; Richard Weinstein, Walnut Creek, CA; Margaret Wierman, Denver, CO; Donna Winn, Knoxville, TN; Dennis Yee, Mesa, AZ; Kevin Yingling, Huntington, WV; and Carol Young, Escondido, CA.

··· REFERENCES ···

1. Cummings SR, Black DM, Nevitt MC, et al. Bone density at various sites for prediction of hip fractures. *Lancet* 1993;341:72-75.

2. Gardsell P, Johnell O, Nilsson BE, Gullberg B. Predicting various fragility fractures in women by forearm bone densitometry: A follow-up study. *Calcif Tissue Int* 1993;52:348-353.

3. Melton LJ, Atkinson EJ, O'Fallon WM, Wahner HW, Riggs BL. Long-term fracture prediction by bone mineral assessed at different skeletal sites. *J Bone Miner Res* 1993;8:1227-1233.

4. Nguyen T, Sambrook P, Kelly P. Prediction of osteoporotic fractures by postural instability and bone density. *Br Med J* 1993;307:1111-1115.

5. Ross PD, Davis JW, Epstein RS, Wasnich RD. Pre-existing fractures and bone mass predict vertebral fracture incidence in women. *Ann Intern Med* 1991;114:919-923.

6. Hui SL, Slemenda CW, Johnston CC Jr. Baseline measurement of bone mass predicts fracture in white women. *Ann Intern Med* 1989;111:355-361.

7. Clark AP, Schuttinga JA. Targeted estrogen/progestogen replacement therapy for osteoporosis: Calculation of health care cost savings. *Osteoporos Int* 1992;2:195-200.

8. Tosteson ANA, Rosenthal DI, Melton LJ III, Weinstein MC. Cost effectiveness of screening perimenopausal white women for osteoporosis: Bone densitometry and hormone replacement therapy. *Ann Intern Med* 1990;113:594-603.

9. Whittington R, Faulds D. Hormone replacement therapy, II: A pharmacoeconomic appraisal of its role in the prevention of postmenopausal osteoporosis and ischaemic heart disease. *Pharmacoeconomics* 1994;5:513-554.

10. Allaway SL, Robinson D, Bailey AR, Hale AC. Bone density, biochemistry and life-style. *Methods Inform Med* 1993;32:233-236.

11. Bauer DC, Browner WS, Cauley JA, et al. Factors associated with appendicular bone mass in older women. *Ann Intern Med* 1993;118:657-665.

12. Elliot JR, Gilchrist N, Wells J, et al. Historical assessment of risk factors in screening for osteopenia in a normal Caucasian population. *Aust N Z J Med* 1993;23:458-462.

13. Hernandez Avila M, Stampfer MJ, Ravnikar VA, et al. Caffeine and other predictors of bone density among pre- and perimenopausal women. *Epidemiology* 1993;4:128-134.

14. Kroger H, Tupparainen M, Honkanen R, Alhava E, Saarikoski S. Bone mineral density and risk factors for osteoporosis—a population-based study of 1600 perimenopausal women. *Calcif Tissue Int* 1994;55:1-7.

15. Lissner L, Bengtsson C, Hansson T. Bone mineral content in relation to lactation history in pre- and postmenopausal women. *Calcif Tissue Int* 1991;48:319-325.

16. Nguyen TV, Kelly PJ, Sambrook PN, et al. Lifestyle factors and bone density in the elderly: Implications for osteoporosis prevention. *J Bone Miner Res* 1994;9:1339-1346.

17. Ooms M, Lips P, Van Lingen A, Valkenburg HA. Determinants of bone mineral density and risk factors for osteoporosis in healthy elderly women. *J Bone Miner Res* 1993;8:669-675.

18. Pouilles JM, Tremollieres F, Bonneu M, Ribot C. Influence of early age at menopause on vertebral bone mass. *J Bone Miner Res* 1994;9:311-315.

19. Ribot C, Pouilles JM, Bonneu M, Tremollieres F. Assessment of the risk of post-menopausal osteoporosis using clinical factors. *Clin Endocrinol* 1992;36:225-228.

20. Slemenda CW, Hui SL, Longcope C, Wellman H, Johnston CC Jr. Predictors of bone mass in perimenopausal women: A prospective study of clinical data using photon absorptiometry. *Ann Intern Med* 1990;112:96-100.

21. Soroko SB, Barrett-Connor E, Edelstein SL, Kritz-Silverstein D Family history of osteoporosis and bone mineral density at the axial skeleton: The Rancho Bernardo Study. *J Bone Miner Res* 1994;9:761-769

22. Spector TD, Edwards AC, Thompson PW. Use of a risk factor and dietary calcium questionnaire in predicting bone density and subsequent bone loss at the menopause. *Ann Rheum Dis* 1992;51:1252-1253.

23. Stevenson JC, Lees B, Devenport M, Cust MP, Ganger KF. Determinants of bone density in normal women: Risk factors for future osteoporosis? *Br Med J* 1989;298:924-928

24. Delmas PD. Biochemical markers of bone turnover in osteoporosis. In: Riggs BL, Melton LJ III, eds. *Osteoporosis.* Raven Press, New York, NY; 1988:297-316

25. Gertz BJ, Shao P, Hanson DA, et al. Monitoring bone resorption in early postmenopausal women by an immunoassay for cross-linked collagen peptides in urine. *J Bone Miner Res* 1994;9:135-142.

26. Rosen HN, Dresner-Pollak R, Moses AC, et al. Specificity of urinary excretion of cross-linked n-telopeptides of type I collagen as a marker of bone turnover. *Calcif Tissue Int* 1994;54:26-29

27. Faulkner KG, Roberts LA, McClung MR. Discrepancies in normative data between Hologic and Lunar systems *J Bone Miner Res* 1995;10:S146. Abstract.

28. Ettinger B, Miller P, McClung MR. Use of bone densitometry results for decisions about therapy for osteoporosis. *Ann Intern Med* 1996;125:623. Letter.

29. Meema HR, Meema S. Postmenopausal osteoporosis: Simple screening method for diagnosis before structural failure. *Radiology* 1987;164:405-410.

30. Nordin BEC. The definition and diagnosis of osteoporosis. *Calcif Tiss Int* 1987;40:57-58.

31. Meema HE. Improved vertebral fracture threshold in postmenopausal osteoporosis by radiogrametric measurements: Its usefulness in selection of therapy. *J Bone Miner Res* 1991;6:9-14.

32. Fleiss JL. *Statistical Methods for Rates and Proportions.* New York, NY: John Wiley & Sons; 1973:39-43, 109-118, 143-147.

33. Stokes ME, Davis CS, Koch GG. *Categorical Data Analysis Using the SAS System.* Cary, NC: SAS Institute Inc.; 1995:19-140.

34. Rawlings JO *Applied Regression Analysis: A Research Tool.* Pacific Grove, CA: Wadsworth & Brooks/Cole; 1988:237-280.

35. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons; 1989:135-145

36. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.

37. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: Statistics and strategies for evaluation. *Control Clin Trials.* 1991;12: 142S-158S.

38. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* 1989;45:255-268.

39. Hanley JA, McNeal BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* 1983;148:839-843.

40. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: Sample size estimation for diagnostic test studies. *J Clin Epidemiol.* 1991;44:736-770.

41. Simel DL, Samsa GP, Matchar DB. Likelihood ratios for continuous test results—making the clinicians' job easier or harder? *J Clin Epidemiol.* 1993;46:85-93.

42. Fleiss JL. *The Design and Analysis of Clinical Experiments.* New York, NY: John Wiley & Sons; 1986: 5-8, 39-43.

43. Ribot C, Tremollieres F, Pouilles JM. Can we detect women with low bone mass using clinical risk factors? *Am J Med* 1995;98:52S-55S

44. Looker AC, Johnston CC Jr, Wahner HW, et al. Prevalence of low femoral bone density in older US women from NHANES III. *J Bone Miner Res* 1995;10:796-802.

45. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules: Applications and methodological standards. *N Engl J Med* 1985;313:793-799.

46. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: A review and suggested modifications of methodological standards. *JAMA* 1997;277:488-494.

47. Melton LJ III, Eddy DM, Johnston CC Jr. Screening for osteoporosis. *Ann Intern Med* 1990;112:516-528.