A Comparison of Prediction Models for Fractures in Older Women

Is More Better?

Kristine E. Ensrud, MD, MPH; Li-Yung Lui, MA, MS; Brent C. Taylor, PhD, MPH; John T. Schousboe, MD, MS; Meghan G. Donaldson, PhD; Howard A. Fink, MD, MPH; Jane A. Cauley, DrPH; Teresa A. Hillier, MD, MS; Warren S. Browner, MD, MPH; Steven R. Cummings, MD; for the Study of Osteoporotic Fractures Research Group

Background: A Web-based risk assessment tool (FRAX) using clinical risk factors with and without femoral neck bone mineral density (BMD) has been incorporated into clinical guidelines regarding treatment to prevent fractures. However, it is uncertain whether prediction with FRAX models is superior to that based on parsimonious models.

Methods: We conducted a prospective cohort study in 6252 women 65 years or older to compare the value of FRAX models that include BMD with that of parsimonious models based on age and BMD alone for prediction of fractures. We also compared FRAX models without BMD with simple models based on age and fracture history alone. Fractures (hip, major osteoporotic [hip, clinical vertebral, wrist, or humerus], and any clinical fracture) were ascertained during 10 years of follow-up. Area under the curve (AUC) statistics from receiver operating characteristic curve analysis were compared between FRAX models and simple models.

Results: The AUC comparisons showed no differences between FRAX models with BMD and simple models with age and BMD alone in discriminating hip (AUC, 0.75 for the FRAX model and 0.76 for the simple model; \( P = .26 \)), major osteoporotic (AUC, 0.68 for the FRAX model and 0.69 for the simple model; \( P = .51 \)), and clinical fracture (AUC, 0.64 for the FRAX model and 0.63 for the simple model; \( P = .16 \)). Similarly, performance of parsimonious models containing age and fracture history alone was nearly identical to that of FRAX models without BMD. The proportion of women in each quartile of predicted risk who actually experienced a fracture outcome did not differ between FRAX and simple models (\( P \approx .16 \)).

Conclusion: Simple models based on age and BMD alone or age and fracture history alone predicted 10-year risk of hip, major osteoporotic, and clinical fracture as well as more complex FRAX models.

Arch Intern Med. 2009;169(22):2087-2094

Advanced age and low bone mineral density (BMD) are strongly associated with higher fracture risk in older women.\(^1\) Data from randomized trials support the recommendation to initiate pharmacologic therapy to lower fracture risk among older women with osteoporosis as defined by BMD T scores of −2.5 or less (at least 2.5 SDs below average for healthy young women). However, despite a lower risk of fracture, more fractures occur in the much larger group of older women without osteoporosis (BMD T scores above −2.5) in whom there is much less certainty about the efficacy of drug therapy.

In addition to low BMD and greater age, many other independent risk factors for fractures, in particular hip fracture, have been identified.\(^2,3\) Thus, recent efforts by the World Health Organization Metabolic Bone Disease Group have focused on developing a risk assessment tool (FRAX) using clinical risk factors with and without femoral neck BMD to enhance fracture prediction.\(^4\) To develop the Web-based FRAX tool, Kanis and colleagues\(^5-13\) used data from 9 epidemiologic cohorts on baseline BMD and common clinical risk factors easily determined by primary care clinicians that were identified from previous meta-analyses. The performance characteristics of the FRAX tool were then validated in 11 independent population-based cohorts.\(^14\) The FRAX algorithm\(^15\) is country specific and uses clinical risk factor data, with or without consideration of femoral neck BMD measurement, to calculate an individual patient’s 10-year probability of hip fracture and 10-year probability of major osteoporotic (hip, clinical vertebral, wrist, or humerus) fracture.

See Invited Commentary at end of article
The development of the FRAX tool has been supported by organizations including the International Osteoporosis Foundation and the National Osteoporosis Foundation (NOF) in the United States, which have strongly advocated its use in clinical decision making. On the basis of the results of a US-specific cost-effective analysis, the NOF recently modified its treatment guidelines to recommend pharmacologic therapy for adults 50 years or older meeting specific criteria, including those based on the presence of osteopenia (BMD T score between −1 and −2.5) and level of 10-year absolute probability of hip (3% or higher) or major osteoporotic (20% or higher) fractures as calculated by the FRAX tool. As noted by the NOF, widespread adoption of these new guidelines would shift the treatment approach from one based primarily on BMD measurement to a new approach based on absolute risk of fracture. This shift in approach may have major implications for the proportion of individuals for whom pharmacologic treatment is recommended to lower their risk of fracture. For example, it has been estimated that application of the new NOF guidelines would result in recommending pharmacologic treatment to a very large proportion of older white women in the United States (at least 72% of those ≥65 years and 93% of those ≥75 years).

Despite recommendations to incorporate fracture probabilities calculated by the FRAX model into clinical decision making, it is uncertain whether fracture prediction with FRAX models is superior to that based on parsimonious models. We used data collected in the Study of Osteoporotic Fractures to compare the value of FRAX models with BMD with that of simple models based on age and BMD alone for prediction of fractures in older women. In addition, because BMD is not available in all practice settings, we performed a secondary analysis in which we compared FRAX models without BMD with simple models based on age and fracture history alone.

METHODS

PARTICIPANTS

From 1986 to 1988, a total of 9704 women at least 65 years old were recruited for participation in the baseline examination of the prospective Study of Osteoporotic Fractures (SOF). Women were recruited from population-based listings in 4 areas of the United States. Women were recruited from population-based listings in 4 areas of the United States. Black women were originally excluded from SOF because of their low incidence of hip fracture. In addition, women were excluded if they were unable to walk without assistance or had a history of bilateral hip replacement. All surviving participants were invited to participate in a second examination between 1989 and 1990 that included measurement of femoral neck BMD. A total of 7963 (84.3% of survivors) had technically adequate femoral neck BMD measurements. Of these, 6252 women provided data for all FRAX clinical risk factors and femoral neck BMD and are the subject of this analysis. The institutional review board at each center approved the study protocol, and written informed consent was obtained from all participants.

CLINICAL RISK FACTORS AND BMD

Participants completed a questionnaire and were interviewed at the baseline examination and asked about race/ethnicity, history of fracture since the age of 50 years, physician diagnosis of rheumatoid arthritis, parental history of hip fracture, smoking status, alcohol intake, and use of oral glucocorticoids. Measurements of body weight and height were used to calculate a standard body mass index (weight in kilograms divided by height in meters squared). The BMD of the proximal femur including the femoral neck subregion was measured by means of dual-energy x-ray absorptiometry (QDR 1000; Hologic, Waltham, Massachusetts). Details of the measurement methods and densitometry quality control methods have been published elsewhere.

FRAX TOOL

This analysis used the FRAX tool (October 2008 version) developed for white subjects in the United States that included the following components: age, sex, fracture history, use of oral glucocorticoids, presence of rheumatoid arthritis, presence of specific conditions associated with secondary osteoporosis (type 1 diabetes mellitus, osteogenesis imperfecta in adults, untreated long-standing hyperthyroidism, premature menopause, chronic malnutrition or malabsorption, and chronic liver disease), parental history of hip fracture, smoking status (current smoker vs nonsmoker), alcohol intake (≥3 drinks per day vs <3 drinks per day), and body mass index. The algorithm calculated 4 fracture probabilities for each participant: the 10-year probability of hip fracture calculated with and without femoral neck BMD data and the 10-year probability of major osteoporotic fracture (defined as hip, clinical vertebral, wrist, or humerus fracture) calculated with and without femoral neck BMD data.

ASCERTAINMENT OF FRACTURES

After the baseline examination, we contacted (by postcard or telephone) participants every 4 months about fractures during the subsequent 10-year follow-up period. A 10-year time point was selected to match that specified by the FRAX tool. More than 98% of these follow-up contacts were completed. Fractures were confirmed by review of radiographic reports. Incident fracture outcomes for this analysis included hip fracture, major osteoporotic fracture (defined as hip, clinical vertebral, wrist, or humerus fracture), and any clinical fracture (nonvertebral and clinical vertebral fractures).

STATISTICAL ANALYSIS

We used logistic regression to examine receiver operating characteristic (ROC) curves for each model across a range of sensitivities and specificities. The model’s ability to discriminate between women who did sustain a fracture during the 10-year follow-up period and those who did not was assessed by the area under the curve (AUC) statistic. Models with AUC statistics of 0.50 do no better than chance alone, whereas models with substantially higher AUC statistics do much better than chance. We compared AUC statistics between FRAX models and simple models by means of the “roccomp” procedure in STATA, which uses a nonparametric approach (StataCorp, College Station, Texas).

In addition, we used the 10-fold cross-validation procedure to generate AUC statistics for FRAX models and simple models for each of the 3 fracture outcomes. The cross-validated AUCs were nearly identical in magnitude to the observed AUCs, and findings regarding the comparison of the ability of the FRAX models vs simple models for prediction of each of these fracture outcomes were unchanged (results not shown).

The clinical usefulness of a model is greater as the proportion of individuals at high risk who actually experience the out-
come increases and as the proportion of individuals at low risk who actually experience the outcome decreases. Thus, we also compared across models the proportion of women in each quartile of predicted risk who actually experienced a fracture outcome. To use a more stringent criterion for identification of high- and low-risk groups, we performed a secondary analysis comparing across models the proportion of women in each decile of predicted risk who actually experienced a fracture outcome. Because the results of this analysis were similar to those of the primary analysis using quartiles, findings from the primary analysis are presented in this article.

RESULTS

CHARACTERISTICS OF THE STUDY POPULATION

Characteristics of the cohort of 6252 women (average age, 71.3 years) are shown in the Table. During an average (SD) follow-up ranging from 7.7 (3.1) years for clinical fracture to 8.7 (2.5) years for major osteoporotic fracture to 9.2 (1.8) years for hip fracture, 389 women (6.2%) experienced a hip fracture, 1037 (16.6%) experienced major osteoporotic fractures, and 1907 (30.5%) experienced clinical fractures. Compared with the 3452 women in the original SOF cohort excluded from this analysis because of missing data needed for calculation of FRAX model probabilities (primarily parental history of hip fracture), the 6252 women included in the analytical cohort were, on average, slightly younger (mean age, 71.4 vs 72.3 years; P=.37) and less likely to report poor to fair health status (15.0% vs 20.3%; P=.001) or previous history of fracture (34.5% vs 41.9%; P=.001). However, mean body mass index (26.4 vs 26.4; P=.99) and femoral neck BMD (0.65 vs 0.65; P=.37) were similar between the 2 groups.

FRAX MODEL WITH BMD VS MODEL WITH AGE AND BMD FOR PREDICTION OF FRACTURES

Across a range of sensitivities and specificities, the ROC curves for hip fracture (Figure 1A) were essentially superimposed for the FRAX model with femoral neck BMD and the simple model containing age and BMD. The figure displays ROC curves for hip fracture (A), major osteoporotic fracture (including hip, clinical vertebral, wrist, and humerus) (B), and clinical fracture (including nonvertebral and clinical vertebral) (C). The diagonal line indicates a reference area under the curve (AUC) of 0.50 (no better than chance alone). CI indicates confidence interval.
els containing age and BMD alone again performed similarly to FRAX models with BMD (Figure 1B and C).

There was no evidence that the proportion of women in each quartile of predicted risk who actually experienced a fracture outcome differed between the FRAX model with BMD and the model with age and BMD alone for the outcome of hip (P=.72), major osteoporotic (P=.78), or any clinical (P=.23) fracture (Figure 2A). Among the women classified in the highest quartile of predicted risk, similar proportions actually experienced a hip fracture (FRAX model with BMD, 14%; model with age and BMD, 15%), major osteoporotic fracture (29% and 30%, respectively), and any clinical fracture (46% and 45%, respectively) during the 10-year follow-up period. In addition, among the women classified in the lowest quartile of predicted risk, nearly identical proportions actually sustained a fracture outcome (hip fracture, 1% with either model; major osteoporotic fracture, 6% with the FRAX model with BMD and 7% with the model with age and BMD; and any clinical fracture, 19% with the FRAX model with BMD and 21% with the model with age and BMD).

**FRAX MODEL WITHOUT BMD VS MODEL WITH AGE AND FRACTURE HISTORY FOR PREDICTION OF FRACTURES**

The parsimonious model based on age and BMD alone was superior to the FRAX model without BMD, which had AUC statistics of 0.71 for hip fracture (P<.001), 0.64 for major osteoporotic fracture (P<.001), and 0.61 for any clinical fracture (P=.01). However, BMD measurement is not universally available in all practice settings. We therefore performed a secondary analysis in which we compared the predictive validity of FRAX models without BMD with that of simple models based on age and fracture history alone. In ROC analyses, the FRAX model without BMD performed similarly to a model containing age and fracture history for prediction of hip fracture (Figure 3A; AUC statistic, 0.71 for both models; P=.91). The AUC statistics were lower, but similar by either model, for major osteoporotic fractures (AUC, 0.64 for both models; P=.89; Figure 3B) and any clinical fracture (AUC, 0.61 for both models; P=.70; Figure 3C). The model based on age and fracture history was superior to a model containing age alone, which had AUC statistics of 0.69 for hip fracture (P=.01), 0.60 for major osteoporotic fracture (P<.001), and 0.56 for any clinical fracture (P<.001).

There was no evidence that the proportion of women in each quartile of predicted risk who actually experienced a fracture outcome differed between the FRAX model without BMD and the model with age and fracture history alone for the outcome of hip (P=.26), major osteoporotic (P=.69), or any clinical (P=.16) fracture (Figure 2B). Among the women classified in the highest quartile of predicted risk, similar proportions actually experienced a hip fracture (13% with either model), major osteoporotic fracture (27% with either model), and any clinical fracture (44% with the FRAX model without BMD and 43% with the model with age and fracture history) during the 10-year follow-up period. In addition, among the women classified in the lowest quartile of predicted risk, essentially identical proportions actually sustained a fracture outcome (hip fracture, 2% with either model; major osteoporotic fracture, 9% with the FRAX model without BMD and 10% with the model with age and fracture history; any clinical fracture, 23% with the FRAX model without BMD and 24% with the model with age and fracture history).

**COMMENT**

In this cohort of older community-dwelling women, a simple model based on age and BMD alone predicted 10-year risk of hip, major osteoporotic, and any clinical fracture as well as more complex FRAX models with BMD. Similarly, a parsimonious model based on age and fracture history alone predicted 10-year risk of these 3 fracture outcomes as well as more complex FRAX models without BMD. These findings suggest that use of the FRAX risk assessment tool does not enhance fracture prediction in older women beyond that provided by simple models based on age and BMD or age and fracture history alone.

Our results are in general agreement with published findings regarding the predictive validity of the FRAX risk assessment tool. In the FRAX development and validation study, findings concerning hip fracture prediction indicated that a model combining BMD with FRAX clinical risk factors vs a model with BMD alone resulted in a small increment (5% to 19% increase, depending on age group) in the gradient of risk per 1-SD change in risk score. However, the performance of the model with BMD alone (assessed with ROC analysis) was nearly identical to that for the model with BMD plus FRAX clinical risk factors, suggesting that hip fracture prediction did not substantially improve by adding FRAX clinical risk factors to the model with BMD alone. There was also a small increment (11% to 21% increase, depending on age group) in the gradient of risk for prediction of other osteoporotic fractures (defined variably depending on the cohort, and excluding hip fractures) when FRAX clinical risk factors were used in combination with BMD; however, a comparison of the performance of the model containing BMD plus clinical risk factors with that of the model with BMD alone was not reported. Findings from this study and the FRAX development and validation study both suggest that models for predicting hip fracture perform better than those for predicting other fracture outcomes. Neither this study nor the FRAX development and validation study examined the ability of models to predict radiographic vertebral fracture, the most common manifestation of osteoporosis. However, a study using data collected in the Fracture Intervention Trial reported that the FRAX model with BMD and a simple model based on age and BMD alone were similarly accurate in terms of discrimination of new radiographic vertebral fracture.

Although gradients of risk and areas under ROC curves are widely used methods for model comparison, the clinical usefulness of a risk prediction model may also be judged by the extent to which the risk calculated from the model reflects the fraction of individuals in the
population who actually sustained the outcome of interest. To address this issue, we calculated the proportion of women in each quartile (and decile) of predicted risk who actually experienced a fracture outcome by means of FRAX and simple models and compared these proportions between models. These results also suggest that use of the FRAX models does not enhance fracture prediction beyond that provided by more parsimonious models be-

Figure 2. Proportion of women classified in each quartile of predicted risk who actually experienced a fracture outcome. The figure displays proportions according to the FRAX model (a web-based risk assessment tool) with bone mineral density (BMD) and the model with age and BMD (A), and the FRAX model without BMD and the model with age and previous fracture (B).
cause there was no evidence that use of the FRAX models improved the classification of high- (and low-) risk categories such that a higher (and lower) proportion of women who actually experienced a fracture outcome were identified. Similar findings have been reported in the study comparing FRAX and simple models for the prediction of new radiographic vertebral fracture.26

The FRAX risk assessment tool represents a major advance in the field of osteoporosis for several reasons. The tool is based on data collected from cohorts in the United States, Europe, Australia, and Asia and is applicable to both the developed and the developing world. Modeling techniques incorporated into the FRAX tool take into account country-specific fracture and death rates. Its aim to move forward risk assessment from a strategy based on BMD alone to an approach based on the absolute risk of fracture is appealing because absolute risk classification systems overcome several of the drawbacks posed by relative risk classification systems and may be more intuitive to both clinicians and patients.27 However, despite these merits, findings from this study in older white women suggest that use of the FRAX tool does not lead to substantial improvements in fracture risk prediction. Cost-effective analyses25 supporting use of the FRAX tool to select men and women 50 years or older with osteopenia for pharmacologic therapy rely on a critical, but controversial, assumption28-31 that drug treatment is effective in reducing the risk of all clinical fractures regardless of BMD status. Moreover, application of the new NOF guidelines may result in recommending pharmacologic treatment to a very large proportion of women 65 years or older.29 Thus, randomized trials evaluating the effectiveness of treatment approaches based on absolute risk of fracture, including the approach recommended by the NOF (incorporating selected risk cutoff points based on FRAX tool probabilities), should be conducted before these approaches are widely implemented into clinical decision making regarding whether to initiate drug therapy.

This study has a number of strengths, including the comprehensive set of measurements and duration and completeness of follow-up. However, the study also has several limitations. Our findings are based on a single cohort of older white women. Although we cross-validated our findings regarding the performance of models, results concerning the predictive validity of FRAX and simple models require confirmation in other cohorts. In particular, these findings may not be generalizable to younger women or men. Other than rheumatoid arthritis, data on 6 specific conditions associated with secondary osteoporosis that compose an additional component in the FRAX risk assessment tool were not collected in SOF. However, these medical conditions are uncommon in healthy older women, and their association with increased fracture risk is in large part due to lower BMD among those with disease. For this reason, checking the “secondary osteoporosis” box in FRAX does not alter the risk score once BMD is entered into the algorithm.32 For a given patient, FRAX and simple models will not necessarily agree on the classification of risk status despite nearly identical ROC curves. Other methods are available to assess and interpret the utility of a new model, such as measures of reclassification33,34 that determine how often an alternative model successfully reclassifies individuals from one risk class to another and thus alters individual treatment decisions. However, the use of these measures requires the existence of predefined treatment thresholds at which treatments would be altered, as well as the availability of effective treatments at different risk levels. In the case of osteoporosis, these requirements are not met at present, espe-
cally the availability of treatments at different risk levels. Although model weights in the parsimonious models were derived from the SOF population and the potential for overfitting exists, findings from this study are consistent with those reported in the FRAX development and validation study. Because the equations and algorithms that generate the fracture risk probabilities have not been published for the FRAX tool, it was not feasible to directly quantify the change in the AUC statistic that accompanies adding each component of the FRAX model. Finally, both the FRAX models and the simple models are limited in their ability to predict fracture, especially nonhip fractures. Thus, risk prediction is challenging, even when robust risk factors like BMD and age are available.

We conclude that a simple model based on age and BMD alone predicted 10-year risk of hip, major osteoporotic, and any clinical fracture as well as more complex FRAX models with BMD. Similarly, a parsimonious model based on age and fracture history alone predicted 10-year risk of these fractures as well as more complex FRAX models without BMD. These results suggest that use of the FRAX risk assessment tool does not improve fracture prediction in older women beyond that provided by simple models based on age and BMD or age and previous fracture alone.

Accepted for Publication: August 9, 2009.

Author Affiliations: Department of Medicine (Drs Ensrud, Taylor, Schousboe, and Fink) and Division of Epidemiology and Community Health, School of Public Health (Drs Ensrud, Taylor, and Fink), University of Minnesota, Minneapolis; Center for Chronic Disease Outcomes Research, Section of General Internal Medicine, Minneapolis Veterans Affairs Medical Center, Minneapolis (Drs Ensrud, Taylor, and Fink); California Pacific Medical Center Research Institute, San Francisco (Ms Lui and Drs Donaldson, Browner, and Cummings); Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania (Dr Cauley); and Center for Health Research, Kaiser Permanente Northwest/Hawaii, Portland, Oregon (Dr Hillier).

Correspondence: Kristine E. Ensrud, MD, MPH, Department of Medicine, Minneapolis Veterans Affairs Medical Center, 1 Veterans Dr, Section of General Internal Medicine (111-0), Minneapolis, MN 55417 (ensru001@umn.edu).

Author Contributions: Ms Lui performed the statistical analyses and is independent of any commercial funder. She had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analyses. Study concept and design: Ensrud, Cauley, Browner, and Cummings. Acquisition of data: Ensrud, Cauley, Hillier, Browner, and Cummings. Analysis and interpretation of data: Ensrud, Lui, Taylor, Schousboe, Donaldson, Fink, Cauley, and Hillier. Drafting of the manuscript: Ensrud. Critical revision of the manuscript for important intellectual content: Lui, Taylor, Schousboe, Donaldson, Fink, Cauley, Hillier, Browner, and Cummings. Statistical analysis: Lui, Taylor, and Browner. Obtained funding: Ensrud, Cauley, Hillier, Browner, and Cummings. Administrative, technical, and material support: Ensrud and Cauley. Study supervision: Ensrud.


Financial Disclosure: None reported.

Funding/Support: The Study of Osteoporotic Fractures (SOF) is supported by National Institutes of Health funding. The National Institute on Aging provides support under the following grant numbers: AG05407, AR35582, AG05394, AR35584, AR35583, AG005407, AG027576, AG005394, and AG027574.

Role of the Sponsor: The funding agencies had no direct role in the conduct of the study; the collection, management, analyses, and interpretation of the data; or the preparation or approval of the manuscript.

Additional Contributions: Kyle A. Moen provided assistance with the manuscript and with the preparation and formatting of the table and figures.

REFERENCES

Should There Be a Fracas Over FRAX and Other Fracture Prediction Tools?

Clinical practice guidelines regarding when to initiate osteoporosis treatment have evolved from the use of bone mineral density (BMD) thresholds to a more complex consideration of the patient’s 10-year absolute fracture risk.1 There are sound reasons for this shift: most fractures occur in patients with BMD T scores above −2.5, and other clinical risk factors, especially age, affect fracture risk.2 Assuming equal relative treatment efficacy, prescribing therapy for a patient with a higher 10-year fracture risk will result in a greater absolute fracture reduction and be more cost-effective for society.

The World Health Organization’s FRAX tool was developed to help clinicians calculate the 10-year fracture risk for individual patients. In the United Kingdom, fracture risk thresholds identify those who should be referred for BMD testing and those in whom treatment decisions can be made without using BMD testing.2 In the United States, where universal screening of older women is recommended, National Osteoporosis Foundation guidelines suggest using the 10-year fracture risk to identify patients with osteopenia whose high fracture risk warrants treatment.

In this issue of Archives, Ensrud et al used the Study of Osteoporotic Fractures cohort to ask whether the FRAX model, with its 9 risk factors and femoral neck BMD, can predict fracture better than simpler models using, for example, only age and BMD. They found that the simple models performed as well as FRAX when they were evaluated using receiver operating characteristic curves or according to the proportion of patients in each quartile of risk who later experience a fracture. Similarly, age and history of fracture performed similarly to FRAX models without BMD. All of the clinical prediction models performed best for hip fractures (area under the curve, 0.74–0.75) but had only modest discriminative ability for major osteoporotic fracture and clinical fracture (area under the curve, 0.63–0.68). Only women were used in Ensrud and colleagues’ analysis, but other