Novel methods to evaluate fracture risk models

M.G. Donaldson[1], **P. M. Cawthon**[1], **J.T. Schousboe**[2,3], **K.E. Ensrud**[3], **L.Y. Lui**[1], **J.A. Cauley**[4], **T.A. Hillier**[5], **B.C. Taylor**[3], **M.C. Hochberg**[6], **D.C. Bauer**[7], and **S.R. Cummings**[1] **for the Study of Osteoporotic Fractures (SOF)**

M.G. Donaldson: mdonaldson@sfcc-cpmc.net; P. M. Cawthon: pcawthon@sfcc-cpmc.net; J.T. Schousboe: john.schousboe@parknicollett.com; K.E. Ensrud: ensru001@umn.edu; L.Y. Lui: llui@sfcc-cpmc.net; J.A. Cauley: jcauley@edc.pitt.edu; T.A. Hillier: hunteresa@comcast.net; B.C. Taylor: brent.taylor2@va.gov; M.C. Hochberg: mhochber@medicine.umaryland.edu; D.C. Bauer: dbauer@psg.ucsf.edu; S.R. Cummings: scummings@sfcc-cpmc.net

[1]San Francisco Coordinating Center, San Francisco, CA, USA

[2]Park Nicollet Health Services, Minneapolis, MN, USA

[3]University of Minnesota and Veterans Affairs Medical Center, Minneapolis, MN, USA

[4]University of Pittsburgh, Pittsburgh, PA, USA

[5]Kaiser Permanente Center for Health Research, Portland, OR, USA

[6]University of Maryland, Baltimore, MD, USA

[7]University of California San Francisco, San Francisco, CA, USA

Abstract

Fracture prediction models help identify individuals at high risk who may benefit from treatment. Area Under the Curve (AUC) is used to compare prediction models. However, the AUC has limitations and may miss important differences between models. Novel reclassification methods quantify how accurately models classify patients who benefit from treatment and the proportion of patients above/below treatment thresholds. We applied two reclassification methods, using the NOF treatment thresholds, to compare two risk models: femoral neck BMD and age ("simple model") and FRAX ("FRAX model").

The Pepe method classifies based on case/non-case status and examines the proportion of each above and below thresholds. The Cook method examines fracture rates above and below thresholds. We applied these to the Study of Osteoporotic Fractures.

There were 6036 (1037 fractures) and 6232 (389 fractures) participants with complete data for major osteoporotic and hip fracture respectively. Both models for major osteoporotic fracture (0.68 vs. 0.69) and hip fracture (0.75 vs. 0.76) had similar AUCs. In contrast, using reclassification methods, each model classified a substantial number of women differently. Using the Pepe method, the FRAX model (vs. simple model), missed treating 70 (7%) cases of major osteoporotic fracture but avoided treating 285 (6%) non-cases. For hip fracture, the FRAX model missed treating 31 (8%) cases but avoided treating 1026 (18%) non-cases. The Cook method (both models, both fracture outcomes) had similar fracture rates above/below the treatment thresholds.

Compared with the AUC, new methods provide more detailed information about how models classify patients.

Corresponding author: Meghan G. Donaldson, PhD, Suite 5700, Lobby 5, 185 Berry Street, San Francisco, CA 94127, Fax. 415 514 8150, Phone. 604 506 0734.

Keywords

hip fracture; major osteoporotic fracture; FRAX; BMD; prediction

Introduction

Fracture prediction models are needed to identify individuals at high risk who may benefit from treatment. Currently, to compare prediction models, the area under the receiver operating characteristic (AUC) curve is used (e.g., old model vs. old model + new predictor). A predictor is considered to be beneficial if the AUC combining the old model and new predictor is greater than the AUC with the old model alone. The AUC is a summary measure and examines the performance of the model across a range of fracture probabilities. However, fracture probabilities at the extremes of the AUC may not be clinically relevant. Novel reclassification methods described by Pepe and colleagues (1–3) and Cook and colleagues (4,5), quantify how accurately individuals are classified who benefit from treatment and how accurately individuals are classified within each model as "high" and "low" risk (e.g., a threshold where treatment may or may not be initiated).

Reclassification methods require specifying treatment thresholds *a priori*. These methods use *r×c* tables, the dimensions of which are determined by the number of risk groups of interest (e.g., high, medium and low). For each risk model of interest individuals are classified into these groups. The models are then cross-classified and the movement of individuals from low risk in the "old model" to higher risk groups using the "old model + new predictor", and vice versa is quantified, and vice versa. To date there are two reclassification methods. Pepe and colleagues suggest classifying individuals based on case and non-case status and examines the proportion of each above and below treatment thresholds. Similarly, reclassification is conditioned on case and non-case status. In contrast, Cook and colleagues suggest comparing the observed risk to the average predicted risk in the model comparisons of interest. This method does not condition on case and non-case status.

Ensrud and colleagues used AUC analyses to compare the predictive ability of the FRAX model with a simple model of femoral neck BMD and age (6). FRAX provides country specific estimates of individual 10-year absolute risk of hip fracture and major osteoporotic fracture (hip, spine, wrist and humerus) (7). They concluded that the FRAX model and simple model had similar predictive accuracy for major osteoporotic fracture (0.68 and 0.69 respectively) and hip fracture (0.75 and 0.76 respectively). However, it remains unknown how these models may classify women differently for treatment. Therefore, the aim of this analysis was to compare the AUC method to the new reclassification methods.

Methods

We used data from the Study of Osteoporotic Fractures (SOF) a prospective study of community-dwelling Caucasian women aged 65 years and older recruited from 4 communities in the United States: Baltimore, Maryland; Minneapolis, Minnesota; Portland, Oregon; and Monongahela Valley near Pittsburgh, Pennsylvania. Participants were recruited from population-based listings and mass mailings between 1986 and 1988 (8). SOF was initiated before widespread publicity about osteoporosis and women were not recruited on the basis of any risk factors for osteoporosis. All participants provided informed consent. This study was approved by the Institutional Review Board at each of the participating sites.

Demographic, anthropometric, lifestyle and medical history

Baseline examinations took place from 1986 to 1988 (n=9704). Women provided information regarding fracture history, smoking status, alcohol consumption, parental hip fracture history, rheumatoid arthritis and corticosteroid use.

Height was measured with a wall-mounted Harpenden stadiometer (Holtain Ltd., DyFed, United Kingdom). Weight was measured with a balance beam scale. Body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared.

Bone Mineral Density

Bone mineral density (BMD) was obtained between 1988 and 1990 (Visit 2) at the proximal femur by dual-energy x-ray absorptiometry (DXA) using QDR 1000 densitometers (Hologic, Bedford, Massachusetts). This was performed on 7959 (84%) of the 9451 surviving cohort (hip BMD by DXA was not available at baseline).

Ascertainment of fractures

After the baseline visit, participants were contacted every 4 months (by postcard or by telephone) to inquire about fractures during the 10-year follow-up period. More than 98% of these follow-up contacts were completed. Fractures were confirmed by review of radiographic reports. Incident fracture outcomes for this analysis include hip fracture and major osteoporotic fracture (hip, clinical spine, wrist and humerus).

FRAX 10-year absolute fracture risk with femoral neck BMD ("FRAX model"

The World Health Organization 10-year absolute risk of both hip fracture and major osteoporotic fracture (hip, clinical spine, forearm or shoulder) was calculated by the WHO Collaborating Centre for Metabolic Bone Disease. Calculation of absolute risk was done following the FRAX™ algorithm (7,9,10) including femoral neck BMD and is described in detail elsewhere. Briefly, the calculation of the 10-year probabilities is based on 9 risk factors (age, body mass index (BMI), parental history of hip fracture, patient history of previous fracture, presence of rheumatoid arthritis, smoking status, consumption of 3 or more alcoholic beverages per day, current use of glucocorticoids and secondary osteoporosis). The 10-year probabilities for both hip and major osteoporotic fracture can be calculated with or without FN BMD. We applied FRAX version 3.0 (11).

Femoral neck BMD and age 10-year absolute fracture risk model ("simple model"

We used a Weibull proportional hazards model to estimate the 10-year absolute risk of hip fracture and major osteoporotic fracture in SOF. We used age and femoral neck BMD as predictors of incident hip fracture and major osteoporotic fracture. Women were censored at 10-years to reflect the fracture prediction horizon in FRAX.

Treatment thresholds

To demonstrate the reclassification methods we used treatment thresholds of 20% for major osteoporotic fracture and 3% for hip fracture. We used these thresholds only to illustrate the methods. The NOF treatment guideline has adopted these thresholds. However, the guideline specifies that these thresholds should only be applied to women who have "low bone mass". These thresholds do not apply to women with normal femoral neck BMD, or a previous history of hip or vertebral fracture. We applied these thresholds to the entire SOF cohort in order to demonstrate how the new methods compare with the AUC method recently reported by Ensrud and colleagues (6).

Statistical Analysis

Women were excluded from the analysis if they were missing any of the risk factors required to calculate FRAX. We used reclassification methods described by Pepe and colleagues (1–3) and Cook and colleagues (4,5).

Net reclassification improvement (NRI) was calculated using the method outlined by Pencina and colleagues (12). The NRI is quantified as the sum of the differences in proportions of individuals moving up (from low risk to high risk) minus the proportion moving down (from high risk to low risk) for those with the outcome (cases), and the proportion of individuals moving down (from high risk to low risk) minus the proportion moving up (from low risk to high risk) for those without the outcome (non-cases) (13).

All analyses were completed using SAS (version 9.1) and STATA (version 9.0).

Results

Complete data to calculate FRAX was available for 6252 women (Table 1). The primary reason for not calculating FRAX on 3452 women was because 18.5% of the cohort was missing information regarding parental history of hip fracture (n=1797). Women who were missing risk factors for FRAX were on average older (72.3 vs. 71.3) and a larger proportion reported a previous history of fracture (42% vs. 34%). However, women in the analytic cohort and those missing risk factors had similar BMI (26.4) and similar femoral neck BMD (0.65 g/cm^2).

Although FRAX was calculated for 6252 women, only 6036 and 6232 women were included in the analyses for major osteoporotic fracture and hip fracture respectively. Women were excluded because they experienced a fracture of unknown type over the 10-year follow up period.

AUC

Ensrud and colleagues reported the results from the AUC analyses (6). The AUC for major osteoporotic fracture for the FRAX model (0.68) and for the simple model (0.69) were not significantly different from each other (p=0.51). The AUC for hip fracture for the FRAX model (0.75) and for the simple model (0.76) were not significantly different from each other (p=0.26).

Pepe Method

Major osteoporotic fracture—During 10 years of follow-up 1037 (17%) women experienced a major osteoporotic fracture (cases) and 4999 (83%) did not (non-cases) (Table 2). Applying the NOF treatment threshold of a 10-year fracture probability of 20% to the entire cohort, the simple model classified 36% of the cohort as "treat" and the FRAX model classified 30% of the cohort as "treat" (Table 2). Among cases, the simple model correctly classified 57% (596/1037) of cases or 9.9% (596/6036) of the entire cohort as "treat". The FRAX model however, classified 7% (70/1037) fewer cases or 1.1% (70/6036) fewer women in the entire cohort as "treat". Among non-cases, the simple model classified 31% (1567/4999) of non-cases or 26.0% (1567/6036) of the entire cohort as "treat". When using the FRAX model instead, 6% (285/4999) fewer non-cases or 4.7% (285/6036) of the entire cohort were classified as "treat" (Table 3).

Although the addition of the FRAX variables to FN BMD and age (simple model) caused 20% (1205/6036) of the SOF cohort to change risk categories, there was little net reclassification (NRI= −1%). This was because the proportion of cases classified as low risk

by the FRAX model (incorrect risk category) was offset by the proportion of non-cases classified as low-risk by the FRAX model (correct risk category).

Hip fracture—During 10 years of follow-up 389 (6%) women experienced a hip fracture (cases) and 5843 (94%) did not (non-cases) (Table 2). Applying the NOF treatment threshold of a 10-year hip fracture probability of 3% to the entire cohort, the simple model classified 67% of the cohort as "treat" and the FRAX model classified 50% of the cohort as "treat". Among cases, the simple model correctly classified 92% (357/389) of cases or 5.8% (357/6154) of the entire cohort as "treat". The FRAX model however, classified 8% (31/389) fewer cases or 0.5% (31/6154) fewer women as "treat". Among non-cases, the simple model classified 66% (3835/5843) of non-cases or 62% (3835/6232) of women as "treat". When using the FRAX model instead, 17.7% (1032/6232) fewer non-cases or 16.6% (1032/6232) fewer women were classified as "treat" (Table 3).

The addition of the FRAX variables to FN BMD and age (simple model) caused 20% (1235/6232) of the SOF cohort to change risk categories, the net reclassification was only 9.8%. This was because the proportion of non-cases classified as low risk (correct risk category), was greater than the proportion of cases classified as low risk (incorrect risk category).

Cook Method

Major osteoporotic fracture—The overall rate of major osteoporotic fractures was 20% per 10 years of follow up. The FRAX model and the simple model were both well calibrated (the observed and expected rate of major osteoporotic fracture were similar). The rate of major osteoporotic fracture among those classified as "treat" (10-year risk >= 20%), was 37%/10-years for the FRAX model and 35% for the simple model. The rate of major osteoporotic fracture among those classified as "no treatment" (10-year risk <20%) was approximately 13%/10-years for both models (Table 4).

Among women classified as low risk with the simple model, 16% (425/3873) were reclassified as high risk using the FRAX model. The observed 10-year rate of major osteoporotic fracture among these women was 21.2%. This observed fracture rate is in agreement with the predicted fracture rate of the high risk category (20%). However, among women classified as high risk with the simple model, 36% (780/2163) were reclassified as low risk using the FRAX model. The observed 10-year rate of major osteoporotic fracture among these women was 22.6%. This observed fracture rate is higher that the predicted fracture rate of the low risk category (<20%). However, reclassified rates were very close to the treatment boundary of 20%.

Hip fracture—The overall rate of hip fracture was 6.8% per 10 years of follow up. The FRAX model and the simple model were both well calibrated (the observed and expected hip fracture rates were similar). The rate of hip fracture among those classified as "treat" (10-year risk >= 3%), was 11.7%/10-years for FRAX and 9.4% for BMD + age. The rate of hip fracture among those classified as "no treatment" (10-year risk <3%) was 2.1%/10-years for FRAX and 1.6%/10-years for BMD + age (Table 4).

Among women classified as low risk with the simple model, 4.2% (86/2040) were reclassified as high risk using the FRAX model. The observed 10-year rate of hip fracture among these women was 3.8%. This observed fracture rate is in agreement with the predicted fracture rate of the high risk category (3%). However, among women classified as high risk with the simple model, 27% (1149/4192) were reclassified as low risk using the FRAX model. The observed 10-year rate of major osteoporotic fracture among these women was 3.1%. This observed fracture rate is slightly higher that the predicted fracture rate of the

low risk category (<3%). However, as noted for major osteoporotic fractures, reclassified rates were very close to the treatment boundary of 3%.

Discussion

The FRAX model and the simple model had similar AUCs for major osteoporotic fracture and hip fracture (6). However, novel reclassification methods demonstrate that a large number of women move from high risk to low risk and vice versa using the FRAX model compared with the simple model. On the whole however, the net reclassification of women with the addition of the FRAX risk factors to femoral neck BMD and age (simple model) was balanced between: i) cases being classified as low risk (incorrect risk category) and non-cases being classified as low risk (correct risk category) and ii) the fracture rate among reclassified women were close to the threshold boundary.

To our knowledge, this is the first study to apply two new methods to compare two fracture risk models-- the FRAX model compared with a simple model of femoral neck BMD and age. Ideally, all cases would be classified as high risk and all non-cases would be classified as low risk. Using the Pepe method, which separates cases and non-cases, our results demonstrate that incorporating FRAX risk factors in addition to BMD and age for hip fracture prediction resulted in a net reclassification index (NRI) of −1% for major osteoporotic fracture and 10% for hip fracture. For major osteoporotic fracture the NRI was primarily driven by a balance between of cases moved to high risk and non-cases to low risk. For hip fracture, the NRI was largely driven by non-cases moving to low risk using the FRAX model.

Using the Cook method, which compares predicted and observed risk, the overall reclassification for major osteoporotic fracture was −6% and for hip fracture was approximately −17%. That is, in both cases, the addition of the FRAX risk factors compared with the simple model classified women less well. However, the results from the Cook method should be interpreted with caution as the observed fracture rates among those reclassified were very close to the treatment boundary.

For major osteoporotic fracture, both the simple model and the FRAX model only identified approximately 50% of the cases as "treatment" but, for hip fracture the simple model and the FRAX model correctly identified 92% and 84% of all cases respectively as "treatment". Although accurate identification of cases as "treatment" is important, it is also important to examine how well the models identify non-cases as "no treatment". The discrepancy in model performance between the two fracture outcomes could be due to two factors: strength of the association between risk factors and fracture outcome and, choice of treatment threshold. One potential reason that both models more accurately identified cases of hip fracture compared with major osteoporotic fracture is that BMD and age are most strongly associated with hip fracture than with fractures of the wrist, spine and humerus (these in combination with hip fractures constitute "major osteoporotic fractures") (14). The discrepancy between the two fracture outcomes is also evident from the AUC analysis— approximately 0.68 and 0.75 for major osteoporotic fracture and hip fracture respectively.

As we stated in our Methods section, we used the 20% 10-year risk of major osteoporotic fracture and 3% 10-year risk of hip fracture as a means of illustrating the two novel reclassification methods. Changes to these risk thresholds will have implications for the proportion of cases and non-cases classified as "treatment/no treatment". For example, the 3% threshold for hip fracture may be too low (e.g. increase cut point to 4/5/6%). While the 3% threshold does fairly well at identifying cases as "treatment" (simple model: 92% and FRAX model: 84%), the threshold performs moderately at identifying non-cases as "no

treatment" (simple model: 34% of non-cases identified as low risk; FRAX model: 52% of non cases identified as low risk). Thus, for hip fracture the 3% threshold is sensitive (identifies the majority of cases) but it is not particularly specific (does not identify the majority of non-cases as "no treatment"). For major osteoporotic fracture, the 20% threshold applied to both models does not identify cases as "treatment" particularly well (simple model: 57%; FRAX model: 51%) however it does better identify non-cases as "no treatment" (simple model: 69%; FRAX model: 74%). Thus, for major osteoporotic fracture the 20% threshold is moderately sensitive and moderately specific. As with any choice of threshold there is a trade-off between sensitivity and specificity and the costs, both financially in terms of BMD testing and personally in terms of incorrectly labeling individuals high risk who are in fact low risk and vice versa, need to be considered.

A previous analysis by Ensrud and colleagues used AUC to compare the FRAX model and the simple model (6). In this analysis, area under the ROC curve, or c statistic, for FRAX and FN BMD and age models were compared for hip fracture and major osteoporotic fracture outcomes. The authors reported that the AUC for FRAX and the FN BMD and age model were not statistically different from each other for both hip fracture (AUCs of 0.75 and 0.76 respectively) and major osteoporotic fracture (AUCs of 0.68 and 0.69 respectively) outcomes. The authors concluded that the addition of other risk factors included in FRAX to femoral neck BMD and age did not improve fracture prediction. However, the c statistic, a measure of discrimination, may not be optimal when comparing models that predict future risk above or below specific thresholds (4) . The AUC compares models across a range of probabilities and does not classify patients based on treatment thresholds. The c statistic is a function of the ROC curve and is the probability that when a case and non-case present as a pair the probability that the case will have a higher predicted probability of the outcome compared with the non-case, aggregated over all probabilities from 0% to 100%. However, patients do not present in pairs and clinicians are often not interested in the probability of a positive test give the individual has the outcome of interest. Reclassification methods however provide clinically meaningful outcomes by providing the proportion of individuals who change categories of risk around a specific cutpoint. Despite similar AUCs for the FRAX model and simple model (6), reclassification methods demonstrate that substantial proportions of women change classification. Furthermore, the method proposed by Cook and colleagues makes use of longitudinal data where follow-up time across participants is not uniform.

One of the limitations in interpreting net reclassification index is that currently reclassification of cases from high risk to low risk is weighted equally to non-cases reclassified from low risk to high risk. Thus, using a summary measure to describe reclassification, such as the NRI (12), should be interpreted with caution. It remains important to consider the costs associated with labeling individuals as high risk when they are in fact low risk. Conversely, it is also important to consider the costs associated with missing treatment benefits among those reclassified inappropriately to low risk. Different patients may attribute different levels of importance to being labeled high risk when they are in fact at low risk compared to being labeled low risk when in fact they are at high risk.

However, reclassification methods do require the a priori specification of treatment thresholds, and the magnitude and direction of reclassification will change with different risk thresholds. In this study we used as an example published treatment thresholds (15) that are incorporated into existing US national treatment guidelines. These treatment thresholds, as detailed in the methods section, do not apply to all of the women who participated in SOF. We applied these treatment thresholds to illustrate the implementation of these novel reclassification methods.

In the cardiovascular literature (16) and breast cancer literature (17) the risk threshold denoting high and low risk is not dichotomous. For example, in the cardiovascular literature 5-year risk is divided into four risk categories. Having several categories of risk allows for greater spread of individuals across risk. In addition, the performance of the model with new or additional risk factors may be easier to evaluate as a better model pushes a greater proportion of individuals to the extremes and has fewer individuals in the intermediate risk categories or grey zone.

Our analysis has several strengths. We used data from the Study of Osteoporotic fractures, a large prospective cohort that enrolled women who were not selected on the basis of risk factors for osteoporosis. Furthermore, we were able to use incident fractures over a 10-year follow up period and therefore we did not have to extrapolate this from the data. However, there are limitations. We only studied older and not younger women. A recent study by Dawson-Hughes and colleagues suggests that the addition of FRAX to the new NOF treatment guidelines would recommend fewer younger women for treatment compared to the old NOF guidelines (18). Therefore, FRAX may reclassify younger women more appropriately. Finally, our results may not be generalizable to other ethnicities or to men.

Novel methods for comparing risk models provide more information about how individual patients would be reclassified as treatment or not treatment by risk model. In our example, the two models classify a substantial number of women differently. Whether these classifications are beneficial or not depends on the value placed on the reclassification: is it better to avoid treating patients who do not have fractures/low risk or miss treating patients who do have fractures/high risk. This will depend on the value placed on these tradeoffs. Reclassification methods present an opportunity to further examine this issue.

Acknowledgments

References

1. Pepe MS, Feng Z, Gu JW. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med. 2008; 27(2):173–181. [PubMed: 17671958]

2. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. Ann Intern Med. 2008; 149(10):751–760. [PubMed: 19017593]

3. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol. 2008; 167(3):362–368. [PubMed: 17982157]

4. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007; 115(7):928–935. [PubMed: 17309939]

5. Cook NR. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med. 2008; 27(2):191–195. [PubMed: 17671959]

6. Ensrud KE, Lui LY, Taylor BC, Schousboe JT, Donaldson MG, Fink HA, Cauley JA, Hillier TA, Browner WS, Cummings SR. A comparison of prediction models for fractures in older women: is more better? Arch Intern Med. 2009; 169(22):2087–2094. [PubMed: 20008691]

7. World Health Organisation. FRAX WHO Fracture Risk Assessment Tool. 2008.

8. Cummings S, Nevitt M, Browner W, Stone K, Fox K, Ensrud K, Carley J, Black D, Vogt T. Risk factors for hip fracture in white women: The Study of Osteoporotic Fractures reserach group. New England Journal of Medicine. 1995; 332(12):767–773. [PubMed: 7862179]

9. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. Osteoporos Int. 2008; 19(4):385–397. [PubMed: 18292978]

10. Kanis JA, Oden A, Johnell O, Johansson H, De Laet C, Brown J, Burckhardt P, Cooper C, Christiansen C, Cummings S, Eisman JA, Fujiwara S, Gluer C, Goltzman D, Hans D, Krieg MA, La Croix A, McCloskey E, Mellstrom D, Melton LJ 3rd, Pols H, Reeve J, Sanders K, Schott AM, Silman A, Torgerson D, van Staa T, Watts NB, Yoshimura N. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. Osteoporos Int. 2007; 18(8):1033–1046. [PubMed: 17323110]

11. Ettinger B, Black DM, Dawson-Hughes B, Pressman AR, Melton LJ 3rd. Updated fracture incidence rates for the US version of FRAX. Osteoporos Int. 2010; 21(1):25–33. [PubMed: 19705048]

12. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008; 27(2):157–172. discussion 207-12. [PubMed: 17569110]

13. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 21(1):128–138. [PubMed: 20010215]

14. Stone KL, Seeley DG, Lui LY, Cauley JA, Ensrud K, Browner WS, Nevitt MC, Cummings SR. BMD at multiple sites and risk of fracture of multiple types: long-term results from the Study of Osteoporotic Fractures. J Bone Miner Res. 2003; 18(11):1947–1954. [PubMed: 14606506]

15. Tosteson AN, Melton LJ 3rd, Dawson-Hughes B, Baim S, Favus MJ, Khosla S, Lindsay RL. Cost-effective osteoporosis treatment thresholds: the United States perspective. Osteoporos Int. 2008; 19(4):437–447. [PubMed: 18292976]

16. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA. 2007; 297(6):611–619. [PubMed: 17299196]

17. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med. 2008; 148(5):337–347. [PubMed: 18316752]

18. Dawson-Hughes B, Looker AC, Tosteson AN, Johansson H, Kanis JA, Melton LJ 3rd. The potential impact of new National Osteoporosis Foundation guidance on treatment patterns. Osteoporos Int. 2010; 21(1):41–52. [PubMed: 19705046]

Table 1

Baseline Characteristics of 6252 participants

Characteristic	Finding
Age, mean (SD), years	71.3 (5.1)
BMI, kg/m^2,mean (SD)	26.4 (4.6)
Previous fracture, no. (%)	2155 (34.5)
Oral glucocorticoid therapy, No. (%)	741 (11.9)
Rheumatoid arthritis, No. (%)	429 (6.9)
Parental history of hip fracture, No. (%)	925 (14.8)
Current smoker, No. (%)	583 (9.3)
Alcohol intake, 3 drinks per day, No. (%)	184 (2.9)
Femoral neck BMD, mean (SD), g/cm2	0.65 (0.11)
Women with hip fracture, No. (%)	389 (6.2%)
Women with major osteoporotic fracture, No. (%)	1037 (16.6%)
FRAX with FN BMD, hip fracture, mean (SD), 10-year risk	5.2 (7.0)
FRAX with FN BMD, major osteoporotic fracture, mean (SD), 10-year risk	17.5 (9.9)
FN BMD + Age, hip fracture, mean (SD), 10-year risk	7.1 (7.8)
FN BMD + Age, major osteoporotic fracture, mean (SD), 10-year risk	18.4 (10.6)

Table 2

Proportion of major osteoporotic fracture cases and non-cases classified as "high" (20%) and "low" (<20%) risk and proportion of hip fracture cases and non-cases classified as "high" (3%) and "low" (<3%) risk for two models of fracture prediction using thresholds from the National Osteoporosis Treatment Guideline.

	Major Osteoporotic Fracture		Hip Fracture	
	FN BMD + Age	FRAX with FN BMD	FN BMD + Age	FRAX with FN BMD
% (n) of total classified "high risk" [*]	36% (2163/6036)	30% (1808/6036)	67% (4192/6232)	50% (3129/6232)
Cases				
% (n) classified "high risk"	57%(596/1037)	51%(526/1037)	92% (357/389)	84% (326/389)
% (n) classified "low risk" [†]	43%(441/1037)	49%(511/1037)	8% (32/389)	16% (63/389)
Non-cases				
% (n) classified "high risk"	31%(1567/4999)	26%(1282/4999)	66% (3835/5843)	48% (2803/5843)
% (n) classified "low risk"	69%(3432/4999)	74%(3717/4999)	34% (2008/5843)	52% (3040/5843)

[*] low risk is <20% 10-year risk for major osteoporotic fracture and <3% 10-year risk for hip fracture

[†] high risk is 20% 10-year risk for major osteoporotic fracture and 3% 10-year risk for hip fracture

Table 3

Reclassification using the Pepe method for major osteoporotic fracture and hip fracture using the National Osteoporosis Foundation threshold for treatment

	Major osteoporotic fracture: 10-year risk FRAX with FN BMD					
	Cases (n=1037)			Non-cases (n=4999)		
10-year risk from model with FN BMD and age	<20%	20%	Total	<20%	20%	Total
<20%	362	79	441	3086	346	3432
20%	149	447	596	631	936	1567
Total	511	526	1037	3717	1282	4999

	Hip fracture: 10-year risk FRAX with FN BMD					
	Cases (n=389)			Non-cases (n=5765)		
10-year risk from model with FN BMD and age	<3%	3%	Total	<3%	3%	Total
<3%	29	3	32	1925	83	2008
3%	34	323	357	1115	2720	3835
Total	63	326	389	3040	2803	5843

Net Reclassification Index (NRI) major osteoporotic fracture= (79/1037-149/1037)- (346/4999-631/4999)= −1.0%

NRI hip fracture= (3/389-34/389)-(83/5843-1115/5843)=9.7%

Table 4

Reclassification using Cook method of major osteoporotic fractures and hip fracture using NOF threshold for treatment

10-year risk from model with FN BMD and age	Major osteoporotic fracture: 10-year risk from FRAX with FN BMD		
	<20%	20%	Total
<20%			
Women, n	3448	425	3873
Events, n	362	79	441
Observed 10-year fracture rate, %	11.6	21.2	12.6
20%			
Women, n	780	1383	2163
Events, n	149	447	596
Observed 10-year fracture rate, %	22.6	42.1	34.6
Total			
Women, n	4228	1808	6036
Events, n	511	526	1037
Observed 10-year fracture rate, %	13.5	36.7	19.8

10-year risk from model with FN BMD and age	Hip fracture: 10-year risk from FRAX with FN BMD		
	<3%	3%	Total
<3%			
Women, n	1954	86	2040
Events, n	29	3	32
Observed 10-year fracture rate, %	1.5	3.8	1.6
3%			
Women, n	1149	3043	4192
Events, n	34	323	357
Observed 10-year fracture rate, %	3.1	12.0	9.4
Total			
Women, n	3103	3129	6232

Major osteoporotic fracture: 10-year risk from FRAX with FN BMD

10-year risk from model with FN BMD and age	<20%	20%	Total
Events, n	63	326	389
Observed 10-year fracture rate, %	2.1	11.7	6.8