

Research

JAMA Internal Medicine | Original Investigation

Race and Ethnicity and Fracture Prediction Among Younger Postmenopausal Women in the Women's Health Initiative Study

Carolyn J. Crandall, MD, MS, MACP; Joseph C. Larson, MS; John T. Schousboe, MD, PhD; JoAnn E. Manson, MD, DrPH; Nelson B. Watts, MD; John A. Robbins, MD; Peter Schnatz, DO; Rami Nassir, PhD; Aladdin H. Shadyab, PhD; Karen C. Johnson, MD, MPH; Jane A. Cauley, DrPH; Kristine E. Ensrud, MD, MPH

 Supplemental content

IMPORTANCE The best approach to identify younger postmenopausal women for osteoporosis screening is uncertain. The Fracture Risk Assessment Tool (FRAX), which includes self-identified racial and ethnic information, and the Osteoporosis Self-assessment Tool (OST), which does not, are risk assessment tools recommended by US Preventive Services Task Force guidelines to identify candidates for bone mineral density (BMD) testing in this age group.

OBJECTIVE To compare the ability of FRAX vs OST to discriminate between younger postmenopausal women who do and do not experience incident fracture during a 10-year follow-up in the 4 racial and ethnic groups specified by FRAX.

DESIGN, SETTING, AND PARTICIPANTS This cohort study of Women's Health Initiative participants included 67 169 women (baseline age range, 50-64 years) with 10 years of follow-up for major osteoporotic fracture (MOF; including hip, clinical spine, forearm, and shoulder fracture) at 40 US clinical centers. Data were collected from October 1993 to December 2008 and analyzed between May 11, 2022, and February 23, 2023.

MAIN OUTCOMES AND MEASURES Incident MOF and BMD (in a subset of 4607 women) were assessed. The area under the receiver operating characteristic curve (AUC) for FRAX (without BMD information) and OST was calculated within each racial and ethnic category.

RESULTS Among the 67 169 participants, mean (SD) age at baseline was 57.8 (4.1) years. A total of 1486 (2.2%) self-identified as Asian, 5927 (8.8%) as Black, 2545 (3.8%) as Hispanic, and 57 211 (85.2%) as White. During follow-up, 5594 women experienced MOF. For discrimination of MOF, AUC values for FRAX were 0.65 (95% CI, 0.58-0.71) for Asian, 0.55 (95% CI, 0.52-0.59) for Black, 0.61 (95% CI, 0.56-0.65) for Hispanic, and 0.59 (95% CI, 0.58-0.59) for White women. The AUC values for OST were 0.62 (95% CI, 0.56-0.69) for Asian, 0.53 (95% CI, 0.50-0.57) for Black, 0.58 (95% CI, 0.54-0.62) for Hispanic, and 0.55 (95% CI, 0.54-0.56) for White women. For discrimination of femoral neck osteoporosis, AUC values were excellent for OST (range, 0.79 [95% CI, 0.65-0.93] to 0.85 [95% CI, 0.74-0.96]), higher for OST than FRAX (range, 0.72 [95% CI, 0.68-0.75] to 0.74 [95% CI, 0.60-0.88]), and similar in each of the 4 racial and ethnic groups.

CONCLUSIONS AND RELEVANCE These findings suggest that within each racial and ethnic category, the US FRAX and OST have suboptimal performance in discrimination of MOF in younger postmenopausal women. In contrast, for identifying osteoporosis, OST was excellent. The US version of FRAX should not be routinely used to make screening decisions in younger postmenopausal women. Future investigations should improve existing tools or create new approaches to osteoporosis risk assessment for this age group.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Carolyn J. Crandall, MD, MS, MACP, Division of General Internal Medicine and Health Services Research, Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, 1100 Glendon Ave, Ste 850, Room 858, Los Angeles, CA 90024 (ccrandall@mednet.ucla.edu).

JAMA Intern Med. doi:10.1001/jamainternmed.2023.1253
Published online May 22, 2023.

Routine screening for osteoporosis is recommended for women 65 years or older.¹ However, for postmenopausal women younger than 65 years, the US Preventive Services Task Force (USPSTF) recommends the use of a formal osteoporosis risk assessment tool to select candidates for bone mineral density (BMD) testing. The Fracture Risk Assessment Tool (FRAX) is 1 of 5 risk assessment tools recommended by the USPSTF to identify candidates for screening among postmenopausal women younger than 65 years.¹ FRAX uses individual clinical risk factors to estimate 10-year risk of hip and major osteoporotic fracture (MOF), including clinical spine, forearm, hip, or shoulder fracture.²

The inclusion of race and ethnicity in clinical risk prediction algorithms is the topic of substantial controversy. In the US version, FRAX is race and ethnicity specific, and the 4 options for the user to select are “US (Caucasian),” “US (Black),” “US (Hispanic),” and “US (Asian).” To our knowledge, previous US studies have not examined the ability of the US version of FRAX to distinguish between younger postmenopausal women who do and do not experience fracture during the subsequent 10 years within each of the 4 racial and ethnic groups prespecified in FRAX. Similarly, a systematic review³ identified no published studies that met eligibility criteria and provided results of calibration for the US version of FRAX or other risk assessment instruments in the US population. The goal of this study was to examine the ability of FRAX (without BMD information) to distinguish between women aged 50 to 64 years who do and do not experience fracture within each of the 4 racial and ethnic groups (ie, the discrimination of FRAX). For comparison, we performed identical analyses with the Osteoporosis Self-assessment Tool (OST),⁴⁻⁶ which does not include race and ethnicity information and is another tool recommended by the USPSTF to identify postmenopausal women younger than 65 years who are candidates for osteoporosis screening.¹ Our secondary goal was to examine the ability of the US versions of FRAX and OST to identify postmenopausal women aged 50 to 64 years who have a BMD T score of -2.5 or less within each racial and ethnic category.

Methods

Women's Health Initiative Participants

Between 1993 and 1998, the Women's Health Initiative (WHI) recruited 161 808 postmenopausal women at 40 clinical centers in the US.⁷ Women were aged 50 to 79 years at baseline and were free of serious cardiac, pulmonary, renal, and hepatic conditions. The WHI Observational Study was designed to determine potential risk factors and natural course of important causes of morbidity and mortality in postmenopausal women. The WHI clinical trials tested use of menopausal hormone therapy (HT), calcium plus vitamin D supplementation, and low-fat eating patterns. Human participant review committees at each participating institution reviewed and approved the study. Each participant provided written informed consent. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Key Points

Question Does an osteoporosis risk assessment tool containing race and ethnicity information (Fracture Risk Assessment Tool [FRAX]) perform better than a tool without this information (Osteoporosis Self-assessment Tool [OST]) for osteoporosis risk assessment among younger postmenopausal women?

Findings In this cohort study of 67 169 women younger than 65 years, performance of FRAX and OST was suboptimal in discrimination of major osteoporotic fracture within each of 4 racial and ethnic categories specified by FRAX. Performance of OST (but not FRAX) was excellent in discriminating between women who did and did not have a bone density T score of -2.5 or less within each racial and ethnic category.

Meaning These findings suggest that the US FRAX is not useful for shared clinical decision-making regarding osteoporosis screening in younger postmenopausal women.

Of the 161 808 participants of the WHI observational study and WHI clinical trials, 90 769 were aged 50 to 64 years. We included data from 89 070 participants who self-reported as Asian, Black, Hispanic, or White. We then excluded data from participants who reported using osteoporosis medication at baseline (ie, bisphosphonate, calcitonin, selective estrogen receptor modulator [$n = 1091$]), participants for whom information regarding covariates was missing ($n = 378$), and participants who provided less than 10 years of follow-up time without having an MOF event ($n = 20 432$), resulting in an analytic sample size of 67 169 participants (Figure 1). Of these, 4607 participants had complete data for the BMD analysis. Data for this analysis were collected from October 1993 to December 2008, when follow-up was complete.

Assessment of Incident Fractures

Incident fractures were assessed from baseline to follow-up year 10 using annual questionnaires. The questionnaire asked, “Since the date on the front of this form, has a doctor told you for the first time that you have a new broken, crushed, or fractured bone?” The location of the fracture was also reported. Response choices included hip, upper leg (not hip), pelvis, knee (patella), lower leg or ankle, foot (not toe), tailbone (coccyx), spine or back (vertebra), lower arm or wrist, hand (not finger), elbow, and upper arm or shoulder. Fractures of the jaw, nose, face, skull, finger, toe, rib, and sternum were excluded. Self-reported hip fractures were confirmed using medical records. In accordance with the FRAX definitions,⁸ we defined MOF (primary outcome) as hip, clinical spine, forearm, or shoulder fracture.

Calculation of FRAX and OST Scores

The 10-year estimated absolute risk of MOF (primary outcome) was calculated for each participant at Sheffield University, Sheffield, UK, using FRAX without BMD (version 3.0),⁸ as previously described.⁹ FRAX calculations used weight and height measured at baseline. We calculated the OST score as $0.2 \times$ body weight in kilograms $-$ age in years; the score was truncated to yield an integer.⁴⁻⁶

BMD by Dual-Energy X-ray Absorptiometry

Women in the WHI BMD substudy that included participants from 3 of the 40 clinical centers (Tucson and Phoenix, Arizona; Pittsburgh, Pennsylvania; and Birmingham, Alabama) underwent BMD measurement at baseline using standardized protocols. Bone mineral density was measured using dual-energy x-ray absorptiometry (QDR2000 or 4500W machines [Hologic Inc]). Quality assurance procedures included cross-clinic calibration using hip and spine phantom scans, further evaluation of scans with specific problems, and a review of a random sample of all scans.¹⁰

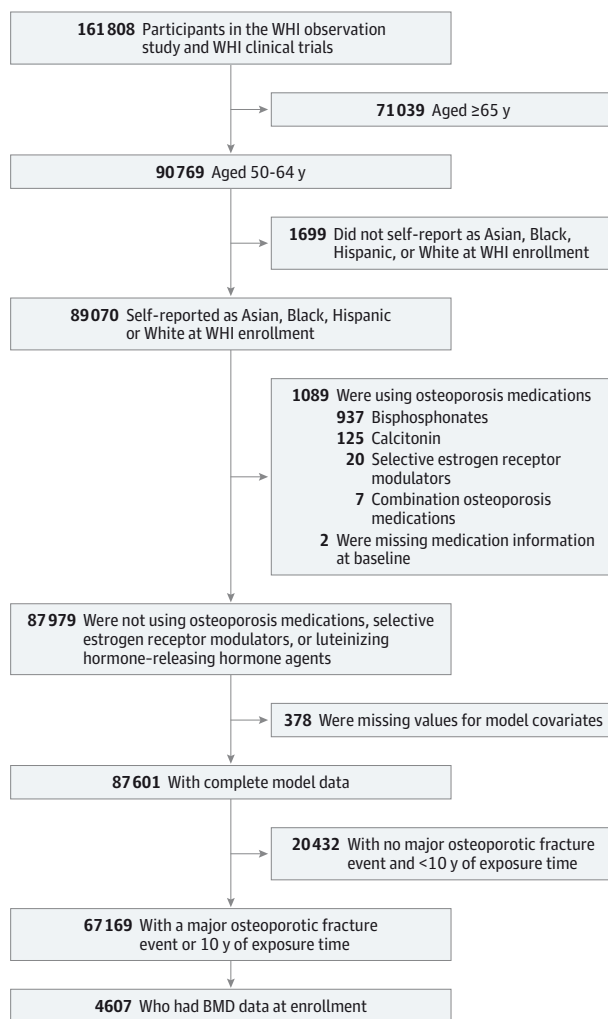
Other Covariates

Using self-assessment questionnaires at baseline, we collected information regarding participant age, race and ethnicity, smoking, alcohol intake, previous fracture before entry into WHI (≥ 55 years of age), parental hip fracture, frequency of falls in the past year, medical history (including previous diagnosis of rheumatoid arthritis, gastrointestinal tract malabsorption, liver disease, emphysema, or early menopause [< 45 years of age]), and medication use (including glucocorticoids and menopausal HT). Physical function was assessed using the RAND 36-item Short-Form Health Survey Physical Functioning construct (range, 0-100, with higher scores indicating a more favorable health state).¹¹⁻¹³

Statistical Analysis

Data were analyzed between May 11, 2022, and February 23, 2023. We used 2-sided, unpaired *t* tests (for continuous variables) and χ^2 tests (for categorical variables) to compare baseline characteristics for participants with vs without MOF events. We used logistic regression to calculate the area under the receiver operating characteristic curve (AUC) for FRAX and OST as continuous measures in the estimation of MOF risk (primary outcome). The AUC is a measure of discrimination—that is, the ability to differentiate between women who will and women who will not experience incident fracture during the 10 years of prospective follow-up. We stratified the AUC results by the 4 US FRAX racial and ethnic categories (Asian, Black, Hispanic, and White). An AUC value of 0.50 indicates no discrimination (ie, that a tool is no greater than chance in discrimination between participants who do and who do not experience MOF). An AUC value greater than 0.50 and less than 0.70 indicates poor discrimination; from 0.70 to less than 0.80, acceptable discrimination.¹⁴ An AUC value from 0.8 to less than 0.9 indicates excellent discrimination; 0.9 or greater, outstanding discrimination. In a sensitivity analysis, we repeated the AUC calculations using an alternative MOF outcome that included lower arm, upper arm, and hip, but excluding clinical spine fractures. Calibration, a comparison of the actual observed vs predicted risk of fracture, was calculated as number of observed fractures per number of predicted fractures. Calibration values greater than 1.00 indicate that the model underestimates the actual (observed) risk of fracture, whereas values less than 1.00 indicate that model overestimates the actual risk of fracture.¹⁵ We calculated calibration both overall and within each quintile of predicted risk.

Figure 1. Study Flow Diagram



BMD indicates bone mineral density; WHI, Women's Health Initiative.

In secondary prespecified analyses, we repeated the analyses described above, substituting femoral neck BMD T score of -2.5 or less and femoral neck, total hip, or lumbar spine BMD T score of -2.5 or less as the outcome. We also conducted a secondary analysis excluding women who reported ever using HT at baseline or randomized to HT in the trials (analytic sample size, 21 313), and censoring data of women who initiated osteoporosis medication therapy any time during follow-up (analytic sample size, 60 135). Two-sided *P* $< .05$ indicated statistical significance.

Results

Characteristics of Participants at Baseline

Among the 67 169 participants, the mean (SD) age at baseline was 57.8 (4.1) years (Table 1). In terms of race and ethnicity, 1486 participants (2.2%) were Asian, 5927 (8.8%) were Black, 2545 (3.8%) were Hispanic, and 57 211 (85.2%) were White.

Table 1. Characteristics of Study Participants at Baseline Overall and by Major Osteoporotic Fracture Event

Baseline characteristic	Participant group ^a			P value ^b
	All (n = 67 169)	No (n = 61 575)	Yes (n = 5594)	
Age, mean (SD), y	57.8 (4.1)	57.8 (4.1)	58.6 (4.0)	<.001
Self-reported race and ethnicity				
Asian	1486 (2.2)	1407 (2.3)	79 (1.4)	<.001
Black	5927 (8.8)	5652 (9.2)	275 (4.9)	
Hispanic	2545 (3.8)	2341 (3.8)	204 (3.6)	
White	57 211 (85.2)	52 175 (84.7)	5036 (90.0)	
BMI				
Mean (SD)	28.0 (6.1)	28.0 (6.1)	28.2 (6.1)	.09
<25	23 759 (35.4)	21 858 (35.5)	1901 (34.0)	
25 to <30	22 666 (33.7)	20 732 (33.7)	1934 (34.6)	
≥30	20 396 (30.4)	18 669 (30.3)	1727 (30.9)	
Physical function score, mean (SD) ^c	85.3 (17.7)	85.6 (17.4)	81.9 (20.5)	<.001
Smoking				
Never	32 880 (49.0)	30 308 (49.2)	2572 (46.0)	<.001
Past	28 669 (42.7)	26 228 (42.6)	2441 (43.6)	
Current	4962 (7.4)	4456 (7.2)	506 (9.0)	
Alcohol intake				
Never	5790 (8.6)	5260 (8.5)	530 (9.5)	<.001
Past	11 055 (16.5)	10 033 (16.3)	1022 (18.3)	
Current (≥1 drink/mo)	49 941 (74.4)	45 940 (74.6)	4001 (71.5)	
Menopausal HT use ^d	37 476 (55.8)	34 792 (56.5)	2684 (48.0)	<.001
Daily oral glucocorticoid use ^e	193 (0.3)	157 (0.3)	36 (0.6)	<.001
Falls in the past year				
0	42 652 (63.5)	39 498 (64.1)	3154 (56.4)	<.001
1	12 762 (19.0)	11 543 (18.7)	1219 (21.8)	
2	5320 (7.9)	4755 (7.7)	565 (10.1)	
≥3	2821 (4.2)	2424 (3.9)	397 (7.1)	
History of fracture aged ≥55 y				
Yes	2808 (4.2)	2356 (3.8)	452 (8.1)	<.001
No	45 146 (67.2)	41 270 (67.0)	3876 (69.3)	
Not applicable (aged <55 y)	16 049 (23.9)	15 044 (24.4)	1005 (18.0)	
Parental history of hip fracture	7706 (11.5)	6900 (11.2)	806 (14.4)	<.001
Early menopause (aged <45 y)	13 457 (20.0)	12 287 (20.0)	1170 (20.9)	.009
History of rheumatoid arthritis	2478 (3.7)	2183 (3.5)	295 (5.3)	<.001
History of malabsorption ^f	193 (0.3)	169 (0.3)	24 (0.4)	.12
History of liver disease	1523 (2.3)	1361 (2.2)	162 (2.9)	.004
History of emphysema	1721 (2.6)	1503 (2.4)	218 (3.9)	<.001
FRAX-estimated 10-y risk of MOF				
Mean (SD), %	7.1 (4.1)	7.0 (4.0)	8.2 (4.9)	<.001
<8.4%	51 638 (76.9)	47 847 (77.7)	3791 (67.8)	
≥8.4%	15 531 (23.1)	13 728 (22.3)	1803 (32.2)	
OST score ^g				
Mean (SD)	3.0 (3.4)	3.0 (3.4)	3.0 (3.4)	.54
<2	26 201 (39.0)	23 991 (39.0)	2210 (39.5)	
≥2	40 968 (61.0)	37 584 (61.0)	3384 (60.5)	

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); FRAX, Fracture Risk Assessment Tool; HT, hormone therapy; MOF, major osteoporotic fracture; OST, Osteoporosis Screening Tool.

^a Unless otherwise indicated, data are expressed as No. (% of participants. Data were missing for body mass index (n = 348), physical function (n = 838), smoking (n = 658), alcohol intake (n = 383), falls in past year (n = 3614), history of fracture at 55 years or older (n = 3166), parental history of hip fracture (n = 2683), early menopause (n = 3651), rheumatoid arthritis (n = 2179), malabsorption (n = 2471), liver disease (n = 2), and emphysema (n = 5061).

^b Compares characteristic by MOF event status using t tests for continuous variables and χ^2 tests for categorical variables.

^c Calculated using the RAND 36-item Short-Form Health Survey Physical Functioning construct. Scores range from 0 to 100, with higher scores indicating a more favorable health state.

^d Incorporates both a participant's self-reported use at baseline as well as her intervention assignment in the Women's Health Initiative Hormone Therapy Trial. Women assigned to active hormone therapy intervention were characterized as "yes" for hormone therapy use, while women assigned to placebo were categorized as "no." Women not participating in the Hormone Therapy Trial were assigned their baseline self-report of hormone therapy use (yes, no).

^e Defined as at least 3 months of use of oral daily use of 5 mg or more of prednisone or equivalent.

^f Defined as self-report of special diet prescribed for malabsorption, celiac sprue, ulcerative colitis, or Crohn disease.

^g Calculated as 0.2 × body weight in kilograms – age in years.

Compared with the 61 575 participants who did not experience an MOF event, the 5594 participants who experienced an MOF event during follow-up were less likely to be Asian (79 [1.4%] vs 1407 [2.3%]) or Black (275 [4.9%] vs 5652 [9.2%]) and

more likely to experience at least 3 falls in the 12 months prior to study baseline (397 [7.1%] vs 2424 [3.9%]) and report a history of parental hip fracture (806 [14.4%] vs 6900 [11.2%]) ($P < .001$ for all). The mean (SD) body mass index (BMI; calcu-

lated as weight in kilograms divided by height in meters squared) was higher in Black (31.4 [6.8]) and Hispanic (29.0 [5.6]) women compared with Asian (25.3 [4.8]) and White (27.7 [5.9]) women (eTable 1 in Supplement 1). Black women were more likely to be current smokers (702 [11.8%]) compared with Asian (73 [4.9%]), Hispanic (176 [6.9%]), and White (4011 [7.0%]) women; Black women were also less likely to report menopausal HT use (2324 [39.2%]) compared with Asian (880 [59.2%]), Hispanic (1283 [50.4%]), and White (32 989 [57.7%]) women. Compared with Hispanic and White women, Asian women were more likely (1083 [72.9%]) and Black women were less likely (1133 [19.1%]) to have an OST score less than 2 (an OST threshold associated with higher risk of having BMD T score of ≤ -2.5).

FRAX Without BMD and OST: Discrimination of MOF by Race

The AUC values for MOF overall were low: 0.59 (95% CI, 0.59-0.60) for FRAX and 0.55 (95% CI, 0.54-0.56) for OST (eTable 2 in Supplement 1). For discrimination of MOF, AUC values for FRAX were 0.65 (95% CI, 0.58-0.71) for Asian, 0.55 (95% CI, 0.52-0.59) for Black, 0.61 (95% CI, 0.56-0.65) for Hispanic, and 0.59 (95% CI, 0.58-0.59) for White women. The AUC values for OST were 0.62 (95% CI, 0.56-0.69) for Asian, 0.53 (95% CI, 0.50-0.57) for Black, 0.58 (95% CI, 0.54-0.62) for Hispanic, and 0.55 (95% CI, 0.54-0.56) for White women. The AUC curves for MOF were lower (closer to the reference line that corresponded to AUC of 0.50) for OST than those for FRAX. However, *P* values for comparisons of AUCs for FRAX vs OST within each racial and ethnic group were not statistically significant, except that among White participants, among whom the AUC was significantly higher for FRAX (0.59 [95% CI, 0.58-0.59]) than for OST (0.55 [95% CI, 0.54-0.56]; *P* < .001) (Figure 2). The sensitivity analysis excluding clinical spine fractures showed similar AUC values (eTable 3 in Supplement 1).

Calibration of FRAX and OST by Race and Ethnicity for Estimation of MOF

In the overall sample, the calibration (ie, the ratio of observed to predicted MOF) ranged from 0.95 (in the lowest quantile of predicted risk) to 1.06 (in the fourth quantile of predicted risk) for FRAX and from 0.94 (in the third quantile of predicted risk) to 1.08 (in the lowest quantile of predicted risk) for OST (eTable 4 in Supplement 1 and Figure 3). The calibration slopes were similar among Asian, Hispanic, and White women for OST (range, 0.87-0.92) and for FRAX (range, 1.00-1.12), but the calibration slope was higher in Black women for OST (1.31) and FRAX (1.26), due to less agreement between observed and estimated fracture risk in the upper quintile of predicted risk. Specifically, FRAX and OST underestimated fractures among high-risk Black women.

FRAX Without BMD and OST: Discrimination of Femoral Neck BMD T Score of ≤ -2.5 or Less by Race and Ethnicity

Overall, the AUC values for identifying a femoral neck BMD T score of ≤ -2.5 or less were 0.72 (95% CI, 0.68-0.75) for FRAX and 0.83 (95% CI, 0.80-0.85) for OST (Table 2). Only 1 Asian woman had a femoral neck BMD T score of ≤ -2.5 or less, so we

report results only for Black, Hispanic, and White women. The AUC values for FRAX were similar in the racial and ethnic groups: 0.74 (95% CI, 0.61-0.87) for Black women, 0.74 (95% CI, 0.60-0.88) for Hispanic women, and 0.72 (95% CI, 0.68-0.75) for White women. In each of the racial and ethnic groups, AUC values were higher for OST than for FRAX, ranging from 0.79 (95% CI, 0.65-0.93) to 0.85 (95% CI, 0.74-0.96) for OST (for OST vs FRAX, *P* < .001 for White women and *P* > .05 for Black and Hispanic women). A similar pattern was apparent for having a BMD T score of ≤ -2.5 or less at any of 3 BMD sites (femoral neck, total hip, lumbar spine).

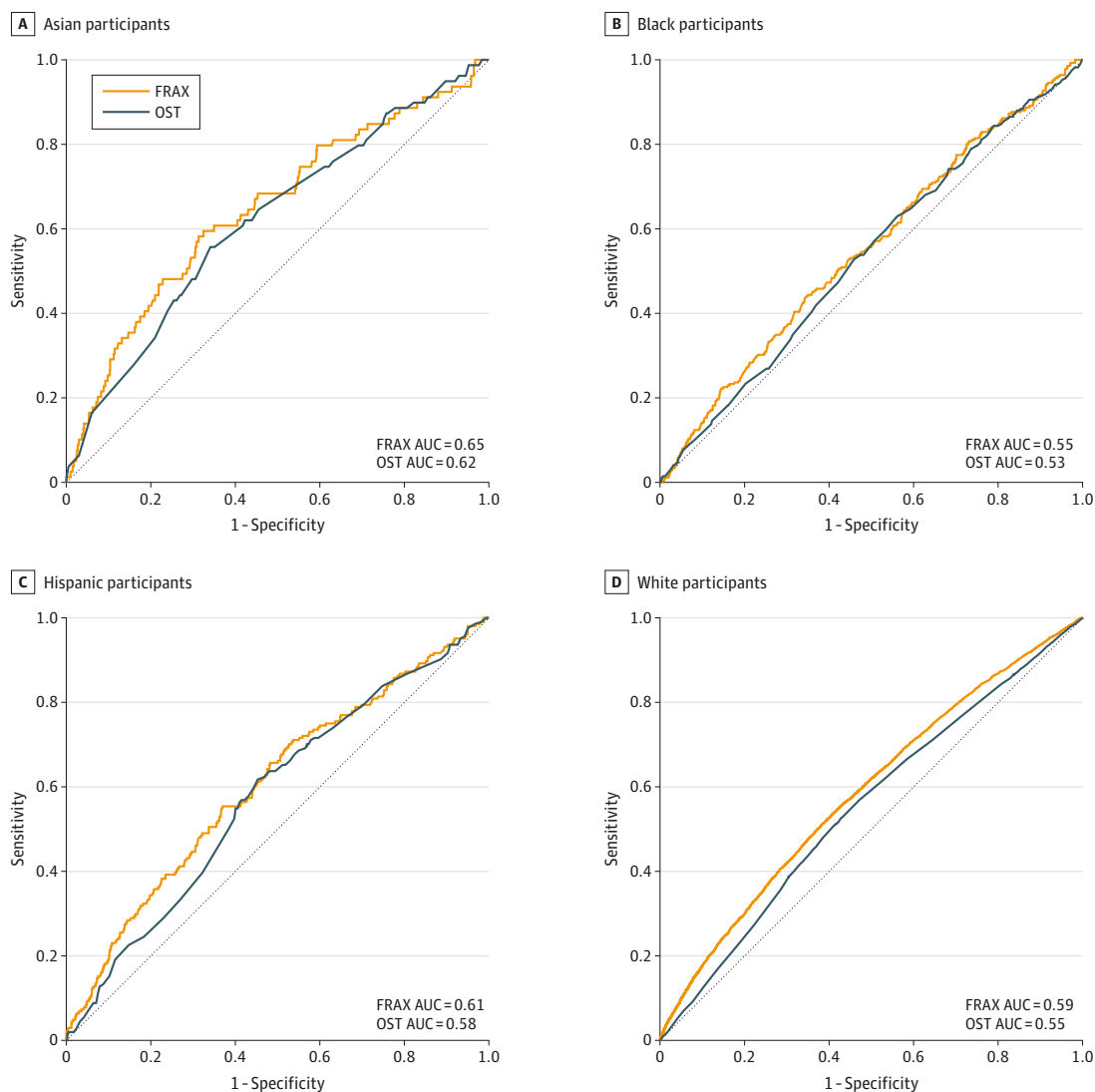
Secondary Analysis: Results Among Women not Using HT or Osteoporosis Medication

We performed a sensitivity analysis limited to women reporting no current or previous HT use at baseline and who were not assigned to the active HT arm in the HT trials, and another sensitivity analysis censoring data of women who initiated osteoporosis medication at any time during study follow-up. Results of these sensitivity analyses were very similar to those of the primary analyses (eTables 5 and 6 in Supplement 1).

Discussion

The inclusion of race and ethnicity in clinical risk prediction algorithms is the focus of increasing attention.^{16,17} The 2011 International Society for Clinical Densitometry Official Positions stated that “separate FRAX models are available for US Asians, Black[s], and Hispanics because hip and MOF rates are lower [in] these ethnic groups than in US Whites.”¹⁸ (The current statement does not address this specific topic.¹⁹) In this study of 67 169 postmenopausal women aged 50 to 64 years at baseline, when examined within each of the racial and ethnic groups specified by the FRAX risk prediction tool (Asian, Black, Hispanic, and White women), FRAX and OST each performed poorly in discriminating between women who did and did not experience MOF (AUC values ranging from 0.55 to 0.65). In contrast to our results regarding MOF, the ability of FRAX and OST to discriminate between women who do and do not have a femoral neck BMD T score in the treatment range (≤ -2.5) was higher, with AUC values ranging from 0.72 to 0.74 for FRAX and 0.79 to 0.85 for OST, and similar among Black, Hispanic, and White women. Low prevalence of a femoral neck BMD T score of ≤ -2.5 or less among Asian women precluded reliable assessment of AUC values for a BMD T score of ≤ -2.5 or less in Asian participants. FRAX and OST were well calibrated for MOF among Asian, Hispanic, and White women, but not Black women; they underestimated the actual observed risk of MOF among high-risk Black women. Results were similar in sensitivity analyses excluding data from women who initiated osteoporosis medication therapy during follow-up. We cannot compare the current results regarding discrimination of FRAX and OST for MOF with those of previously published studies because, to our knowledge, studies that examined FRAX in US study cohorts did not evaluate discrimination of FRAX for MOF according to race and ethnicity.²⁰⁻²⁶

Figure 2. Receiver Operating Characteristic Curves for Major Osteoporotic Fractures During 10 Years of Follow-up Among Women Aged 50 to 64 Years at Baseline by Race and Ethnicity



For all comparisons of the Fracture Risk Assessment Tool (FRAX) vs the Osteoporosis Self-assessment Tool (OST), $P < .001$ among all participants, $P = .59$ among Asian participants, $P = .50$ among Black participants, $P = .35$ among

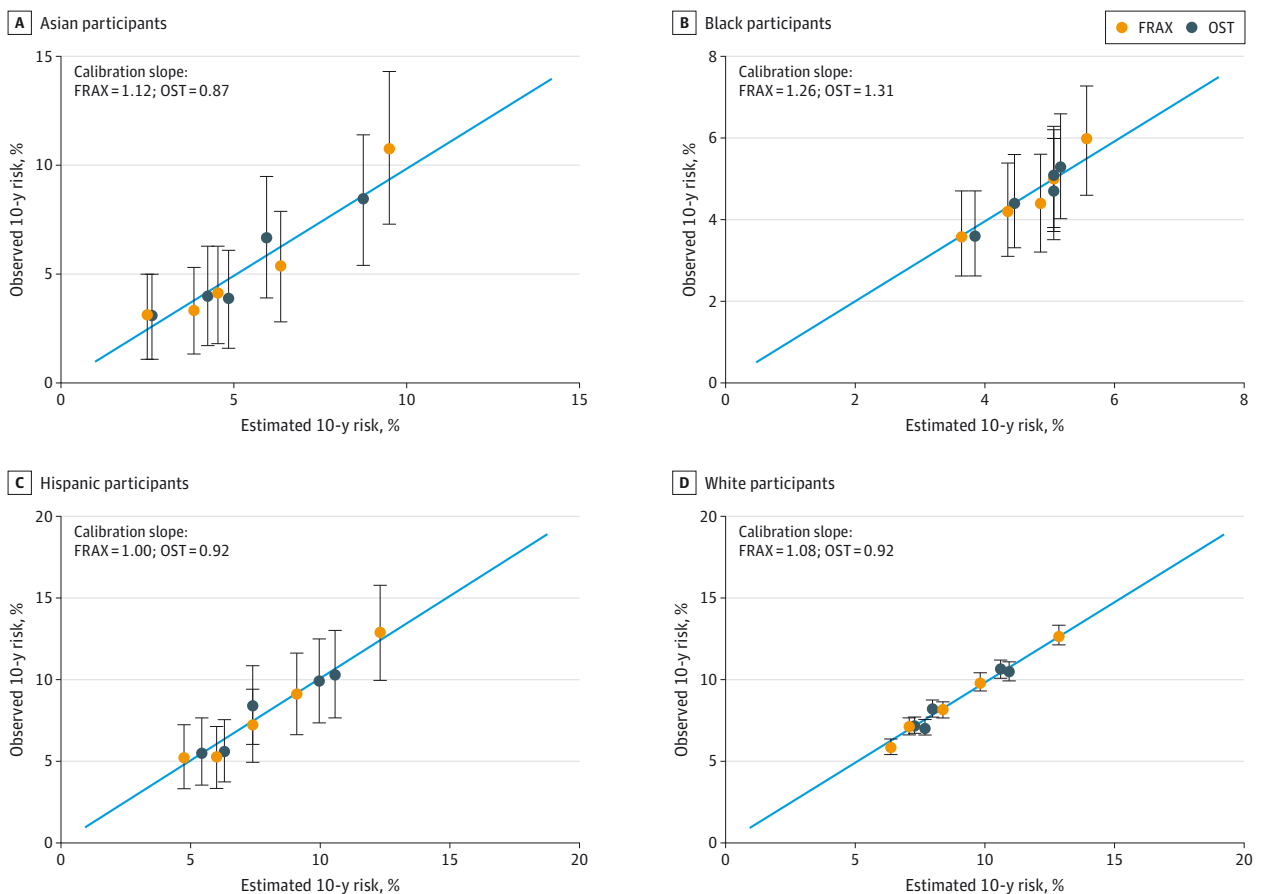
Hispanic participants, and $P < .001$ among White participants (calculated from a χ^2 statistic on 1 *df* testing of the difference between each paired group). AUC indicates area under the receiver operating characteristic curve.

Calibration values quantify how well FRAX estimates of 10-year fracture probability match the actual observed 10-year cumulative probability of that fracture outcome. A systematic review performed for the USPSTF^{3,27} identified no published studies that provided results of calibration for the US version of FRAX in the US population. However, we note that a study of the Manitoba Bone Mineral Density Program registry data identified significant ethnic differences in performance (ie, calibration) of the Canadian FRAX tool, with fracture probability overestimated among Asian and Black women.²⁸ Further studies regarding calibration of risk assessment methods across race and ethnicity are needed.

Our results have clinical implications. First, clinicians should be aware that among younger postmenopausal women,

neither FRAX nor OST distinguishes between those who do and do not subsequently experience MOF. Second, the poor discrimination of FRAX observed in this age group suggests that it is difficult to identify which women will experience future MOF. In contrast, our results suggest that it is reasonable for clinicians to use OST to identify young postmenopausal women who are potential candidates for osteoporosis drug therapy (those with a BMD T score of ≤ -2.5). In contrast to the MOF results, the AUC values for BMD T scores of -2.5 or less were higher. The OST values (AUC range, 0.79-0.85) had excellent ability to distinguish between women who do and do not have a femoral neck BMD T score of -2.5 or less among Black, Hispanic, and White women. Our results have implications for shared decision-making with patients because young post-

Figure 3. Observed and Expected Major Osteoporotic Fracture of Estimated Fracture Risk Among Women Aged 50 to 64 Years at Baseline by Race and Ethnicity



Dots represent the intersection of the predicted event and observed event for each quintile; whiskers represent the 95% CIs for the observed event percentage. FRAX indicates Fracture Risk Assessment Tool; OST, Osteoporosis Self-assessment Tool.

menopausal women should be informed about the suboptimal performance of race- and ethnicity-specific US fracture risk calculators for fracture risk estimation.

The current results suggest that the USPSTF guidelines regarding osteoporosis screening in this age group should be reassessed. The finding that OST has excellent discrimination (without incorporating race and ethnicity information) and was superior to FRAX for identifying younger postmenopausal women with a BMD T score of -2.5 or less in the 4 racial and ethnic groups examined will inform future clinical guidelines. The USPSTF guidelines mention FRAX (which includes Asian, Black, Hispanic, and White race and ethnicity categories) and the Simple Calculated Osteoporosis Risk Estimation Score (ISCORE), which includes an item regarding non-Black race) as 2 of the 5 risk assessment tools. Consequently, the USPSTF's approach is essentially recommending race- and ethnicity-specific tools as an appropriate approach for selecting candidates for osteoporosis screening among younger postmenopausal women.¹ The remaining 3 tools mentioned by the USPSTF (OST, Osteoporosis Index of Risk [OSIRIS], and Osteoporosis Risk Assessment Instrument [ORAI]) do not include race as a risk factor. OST does not contain race and eth-

nicity information and is simpler to use (simpler calculation), but nonetheless both FRAX and OST have poor ability to distinguish between women who do and do not experience MOF within each of the 4 racial and ethnic groups. Our results suggest that even if MOF rates are lower among younger postmenopausal Asian, Black, and Hispanic women compared with rates in White women, the separate FRAX models for US Asian, Black, and Hispanic women perform poorly for discriminating between younger postmenopausal women who do and do not experience incident MOF.

Strengths and Limitations

Strengths of our study include the prospective 10-year follow-up in a large racially and ethnically diverse group of younger postmenopausal women. A potential limitation of our study is that data regarding nonhip fractures were based on self-report. The validity of information regarding self-reported fractures is good in the WHI; the validity for the self-reported MOF category at the exact anatomical site was 80% and was higher for hip fractures (78%), forearm fractures (81%), and humerus fractures (82%) than for clinical spine fractures (51%).²⁹ Also, despite including 1486 participants who self-

Table 2. AUC Values for FRAX and OST Estimated Risk of BMD T Score of -2.5 or Less in the BMD Subset (N = 4607)^a

Outcome	Race and ethnicity	No. of participants	No. of events	AUC (95% CI) ^b	
				FRAX	OST
Femoral neck T score ≤ -2.5	All	4607	235	0.72 (0.68-0.75)	0.83 (0.80-0.85)
	Black	628	14	0.74 (0.61-0.87)	0.85 (0.74-0.96)
	Hispanic	320	14	0.74 (0.60-0.88)	0.79 (0.65-0.93)
	White	3642	206	0.72 (0.68-0.75)	0.82 (0.80-0.85)
Any BMD site T score ≤ -2.5 ^c	All	4607	653	0.64 (0.62-0.66)	0.74 (0.72-0.76)
	Black	628	130	0.68 (0.63-0.73)	0.76 (0.71-0.81)
	Hispanic	320	37	0.68 (0.59-0.76)	0.76 (0.68-0.84)
	White	3642	484	0.68 (0.65-0.70)	0.75 (0.73-0.78)

Abbreviations: AUC, area under the receiver operating characteristic curve; BMD, bone mineral density; FRAX, Fracture Risk Assessment Tool; OST, Osteoporosis Screening Tool.

^a All models are adjusted for current hormone therapy use (yes or no) and the Calcium plus Vitamin D Trial intervention assignment (active, placebo, or not randomized). Because only 1 Asian participant had a femoral neck BMD T score of -2.5 or less, we do not display results for Asian participants.

^b Femoral neck AUC comparisons by race and ethnicity are as follows. For FRAX, $P > .99$ for Black vs Hispanic women, $P = .77$ for Black vs White women, and $P = .79$ for Hispanic vs White women. For OST, $P = .50$ for Black vs Hispanic women, $P = .60$ for Black vs White women, and $P = .67$ for Hispanic vs White women. For comparison of FRAX AUC vs OST AUC at the femoral neck within

each racial and ethnic group, $P < .001$ for all participants, $P = .22$ for Black women, $P = .61$ for Hispanic women, and $P < .001$ for White women. Any BMD site AUC comparisons by race and ethnicity are as follows. For FRAX, $P = .96$ for Black vs Hispanic women, $P = .86$ for Black vs White women, and $P = .96$ for Hispanic vs White women. For OST, $P = .96$ for Hispanic vs White women, $P = .74$ for Black vs White women, and $P = .88$ for Hispanic vs White women. For comparisons of FRAX AUC vs OST AUC at any BMD site within each racial and ethnic group, $P < .001$ for all participants, $P = .03$ for Black women, $P = .19$ for Hispanic women, and $P < .001$ for White women. P values were calculated using a χ^2 statistic on 1 *df* testing the difference between each paired group.

^c Includes femoral neck, total hip, or lumbar spine.

identified as Asian, there were few incident fracture events among Asian women, corresponding with the known lower fracture incidence in Asian women compared with Black, Hispanic, and White women. Similarly, we had few Asian women with a BMD T score of -2.5 or less. Therefore, our results regarding calibration among Asian women may require confirmation in future studies. Measurements of BMD were available at 3 of the 40 clinical centers. We note that FRAX was originally designed to estimate fracture risk, while OST was designed to detect BMD-defined osteoporosis, but we compared these tools because they are recommended by USPSTF guidelines for screening decisions in this age group, making the comparison highly clinically relevant.

Conclusions

This cohort study found that within each of the racial and ethnic categories specified by the FRAX risk assessment tool, the US version of FRAX and OST for risk assessment both had suboptimal performance with poor to fair discrimination for MOF among younger postmenopausal women. The US version of FRAX should not be routinely used to make screening decisions in younger postmenopausal women. Future investigations should improve existing tools or create new approaches to osteoporosis risk assessment for this age group.

ARTICLE INFORMATION

Accepted for Publication: March 7, 2023.

Published Online: May 22, 2023.

doi:10.1001/jamainternmed.2023.1253

Author Affiliations: Division of General Internal Medicine and Health Services Research, Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles (Crandall); Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington (Larson); HealthPartners Institute, Park Nicollet Clinic and University of Minnesota, Minneapolis (Schousboe); Division of Preventive Medicine, Department of Medicine, Harvard Medical School, Boston, Massachusetts (Manson); Mercy Health Osteoporosis and Bone Services, Cincinnati, Ohio (Watts); Center for Healthcare Policy and Research, Department of Medicine, University of California, Davis, Medical Center, Sacramento (Robbins); Department of Obstetrics and Gynecology, Reading Hospital/Tower Health and Drexel University, Philadelphia, Pennsylvania (Schnatz); Department of Internal Medicine, Reading Hospital/Tower Health and Drexel

University, Philadelphia, Pennsylvania (Schnatz); Department of Biochemistry and Molecular Medicine, University of California, Davis (Nassir); Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla (Shadyab); Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis (Johnson); Department of Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (Cauley); Division of Epidemiology & Community Health, University of Minnesota Medical School, Minneapolis (Ensrud).

Author Contributions: Dr Larson had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Crandall, Manson, Robbins, Schnatz, Johnson, Cauley.

Acquisition, analysis, or interpretation of data: Crandall, Larson, Schousboe, Manson, Watts, Robbins, Nassir, Shadyab, Johnson, Cauley, Ensrud.

Drafting of the manuscript: Crandall.

Critical revision of the manuscript for important intellectual content: Crandall, Larson, Schousboe,

Manson, Watts, Robbins, Schnatz, Nassir, Shadyab, Johnson, Cauley, Ensrud.

Statistical analysis: Larson, Schousboe.

Obtained funding: Robbins, Johnson, Cauley.

Administrative, technical, or material support: Manson, Johnson.

Supervision: Manson, Schnatz, Nassir, Johnson.

Conflict of Interest Disclosures: None reported.

Funding/Support: The Women's Health Initiative (WHI) program is supported by contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, and 75N92021D00005 from the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services.

Role of the Funder/Sponsor: Representatives of the National Heart, Lung, and Blood Institute had a role in the design and conduct of this study; collection, management, and analysis of the data; approval of the manuscript. They had no role in the interpretation of the data, preparation of the manuscript, decision to submit the manuscript for publication, or review of the manuscript.

Data Sharing Statement: See Supplement 2.

Additional Contributions: We thank the WHI study participants, investigators, and staff for their exceptional dedication.

Additional Information: The short list of WHI investigators includes the following: Program Office (National Heart, Lung, and Blood Institute, Bethesda, Maryland): Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller. Clinical Coordinating Center (Fred Hutchinson Cancer Research Center, Seattle, Washington): Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg. Investigators and academic centers: Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, JoAnn E. Manson; MedStar Health Research Institute, Howard University, Washington, DC, Barbara V. Howard; Stanford Prevention Research Center, Stanford, California, Marcia L. Stefanick; The Ohio State University, Columbus, Rebecca Jackson; University of Arizona, Tucson and Phoenix, Cynthia A. Thomson; University at Buffalo, Buffalo, New York, Jean Wactawski-Wende; University of Florida, Gainesville and Jacksonville, Marian Limacher; University of Iowa, Iowa City and Davenport, Jennifer Robinson; University of Pittsburgh, Pittsburgh, Pennsylvania, Lewis Kuller; Wake Forest University School of Medicine, Winston-Salem, North Carolina, Sally Shumaker; University of Nevada, Reno, Robert Brunner; and WHI Memory Study at Wake Forest University School of Medicine, Winston-Salem, North Carolina, Mark Espeland. For a list of all the investigators who have contributed to WHI science, visit <https://s3-us-west-2.amazonaws.com/www-whi-org/wp-content/uploads/WHI-Investigator-Long-List.pdf>.

REFERENCES

- Curry SJ, Krist AH, Owens DK, et al; US Preventive Services Task Force. Screening for osteoporosis to prevent fractures: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2018;319(24):2521-2531. doi:10.1001/jama.2018.7498
- World Health Organization Collaborating Centre for Metabolic Bone Diseases. FRAX WHO Fracture Risk Assessment Tool. Web version 4.2. University of Sheffield. September 26, 2020. Accessed June 24, 2021. <https://www.sheffield.ac.uk/FRAX/index.aspx>
- Viswanathan M, Reddy S, Berkman N, et al. Screening to prevent osteoporotic fractures: an evidence review for the US Preventive Services Task Force. Agency for Healthcare Research and Quality (US); 2018. Report No. 15-05226-EF-1. Accessed April 12, 2023. <https://pubmed.ncbi.nlm.nih.gov/30325616/>
- Cadarette SM, McIsaac WJ, Hawker GA, et al. The validity of decision rules for selecting women with primary osteoporosis for bone mineral density testing. *Osteoporos Int*. 2004;15(5):361-366. doi:10.1007/s00198-003-1552-7
- Geusens P, Hochberg MC, van der Voort DJ, et al. Performance of risk indices for identifying low bone density in postmenopausal women. *Mayo Clin Proc*. 2002;77(7):629-637. doi:10.4065/77.7.629
- Gourlay ML, Miller WC, Richey F, Garrett JM, Hanson LC, Reginster JY. Performance of osteoporosis risk assessment tools in postmenopausal women aged 45-64 years. *Osteoporos Int*. 2005;16(8):921-927. doi:10.1007/s00198-004-1775-2
- The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998;19(1):61-109. doi:10.1016/S0197-2456(97)00078-0
- 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. doi:10.1007/s00198-007-0543-5
- Crandall CJ, Larson J, Gourlay ML, et al. Osteoporosis screening in postmenopausal women 50 to 64 years old: comparison of US Preventive Services Task Force strategy and two traditional strategies in the Women's Health Initiative. *J Bone Miner Res*. 2014;29(7):1661-1666. doi:10.1002/jbmr.2174
- LaCroix AZ, Cauley JA, Pettinger M, et al. Statin use, clinical fracture, and bone density in postmenopausal women: results from the Women's Health Initiative observational study. *Ann Intern Med*. 2003;139(2):97-104. doi:10.7326/0003-4819-139-2-200307150-00009
- Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. *Med Care*. 1992;30(6):473-483. doi:10.1097/00005650-199206000-00002
- Andresen EM, Bowley N, Rothenberg BM, Panzer R, Katz P. Test-retest performance of a mailed version of the Medical Outcomes Study 36-Item Short-Form Health Survey among older adults. *Med Care*. 1996;34(12):1165-1170. doi:10.1097/00005650-199612000-00001
- Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ*. 1993;2(3):217-227. doi:10.1002/hec.4730020305
- Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. Wiley; 2013. doi:10.1002/9781118548387
- Schousboe JT, Langsetmo L, Taylor BC, Ensrud KE. Fracture risk prediction modeling and statistics: what should clinical researchers, journal reviewers, and clinicians know? *J Clin Densitom*. 2017;20(3):280-290. doi:10.1016/j.jocd.2017.06.012
- Reid HW, Selvan B, Batch BC, Lee RH. The break in FRAX: equity concerns in estimating fracture risk in racial and ethnic minorities. *J Am Geriatr Soc*. 2021;69(9):2692-2695. doi:10.1111/jgs.17316
- Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9):874-882. doi:10.1056/NEJMs2004740
- Cauley JA, El-Hajj Fuleihan G, Arabi A, et al; FRAX® Position Conference Members. Official positions for FRAX® clinical regarding international differences from Joint Official Positions Development Conference of the International Society for Clinical Densitometry and International Osteoporosis Foundation on FRAX®. *J Clin Densitom*. 2011;14(3):240-262. doi:10.1016/j.jocd.2011.05.015
- The International Society for Clinical Densitometry. 2019 ISCD official positions: adult. Updated 2019. Accessed March 1, 2023. <https://iscd.org/learn/official-positions/adult-positions/>
- Adler RA, Hastings FW, Petkov VI. Treatment thresholds for osteoporosis in men on androgen deprivation therapy: T-score versus FRAX. *Osteoporos Int*. 2010;21(4):647-653. doi:10.1007/s00198-009-0984-0
- Bansal S, Pecina JL, Merry SP, et al. US Preventative Services Task Force FRAX threshold has a low sensitivity to detect osteoporosis in women ages 50-64 years. *Osteoporos Int*. 2015;26(4):1429-1433. doi:10.1007/s00198-015-3026-0
- Cass AR, Shepherd AJ, Asirof R, Mahajan M, Nizami M. Comparison of the Male Osteoporosis Risk Estimation Score (MORES) with FRAX in identifying men at risk for osteoporosis. *Ann Fam Med*. 2016;14(4):365-369. doi:10.1370/afm.1945
- Ensrud KE, Lui LY, Taylor BC, et al; Study of Osteoporotic Fractures Research Group. A comparison of prediction models for fractures in older women: is more better? *Arch Intern Med*. 2009;169(22):2087-2094. doi:10.1001/archinternmed.2009.404
- Kälvesten J, Lui LY, Brismar T, Cummings S. Digital X-ray radiogrammetry in the study of osteoporotic fractures: comparison to dual energy X-ray absorptiometry and FRAX. *Bone*. 2016;86:30-35. doi:10.1016/j.bone.2016.02.011
- Pressman AR, Lo JC, Chandra M, Ettinger B. Methods for assessing fracture risk prediction models: experience with FRAX in a large integrated health care delivery system. *J Clin Densitom*. 2011;14(4):407-415. doi:10.1016/j.jocd.2011.06.006
- Wu Q, Xiao X, Xu Y. Performance of FRAX in predicting fractures in US postmenopausal women with varied race and genetic profiles. *J Clin Med*. 2020;9(1):285. doi:10.3390/jcm9010285
- Viswanathan M, Reddy S, Berkman N, et al. Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2018;319(24):2532-2551. doi:10.1001/jama.2018.6537
- Leslie WD, Morin SN, Lix LM, et al. Fracture prediction from FRAX for Canadian ethnic groups: a registry-based cohort study. *Osteoporos Int*. 2021;32(1):113-122. doi:10.1007/s00198-020-05594-8
- Chen Z, Kooperberg C, Pettinger MB, et al. Validity of self-report for fractures among a multiethnic cohort of postmenopausal women: results from the Women's Health Initiative observational study and clinical trials. *Menopause*. 2004;11(3):264-274. doi:10.1097/O1.GME.0000094210.15096.FD