

Predicting Fracture Risk in Younger Postmenopausal Women: Comparison of the Garvan and FRAX Risk Calculators in the Women's Health Initiative Study

Carolyn J. Crandall, MD, MS¹, Joseph Larson, MS², Andrea LaCroix, PhD³, Jane A. Cauley, DrPH⁴, Meryl S. LeBoff, MD⁵, Wenjun Li, PhD⁶, Erin S. LeBlanc, MD, MPH⁷, Beatrice J. Edwards, MD, MPH⁸, JoAnn E. Manson, MD, DrPH⁹, and Kristine Ensrud, MD, MPH¹⁰

¹Division of General Internal Medicine and Health Services Research, Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, CA, USA; ²WHI Clinical Coordinating Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; ³Family and Preventive Medicine, University of California, San Diego, La Jolla, CA, USA; ⁴Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ⁵Endocrine, Diabetes and Hypertension Division, Brigham and Women's Hospital, Boston, MA, USA; ⁶Division of Biomedical Data Science, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA; ⁷Kaiser Permanente Center for Health Research NW, Portland, OR, USA; ⁸Department of Internal Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; ⁹Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital 3rd Floor, Harvard Medical School, Boston, MA, USA; ¹⁰Division of Epidemiology & Community Health, University of Minnesota Medical School, Minneapolis, MN, USA.

BACKGROUND: Guidelines recommend fracture risk assessment in postmenopausal women aged 50–64, but the optimal method is unknown.

OBJECTIVES: To compare discrimination and calibration of the Fracture Risk Assessment Tool (FRAX) and Garvan fracture risk calculator for predicting fractures in postmenopausal women aged 50–64 at baseline.

DESIGN: Prospective observational study.

PARTICIPANTS: Sixty-three thousand seven hundred twenty-three postmenopausal women aged 50–64 years participating in the Women's Health Initiative Observational Study and Clinical Trials.

MAIN MEASURES: Incident hip fractures and major osteoporotic fractures (MOF) during 10-year follow-up. Calculated FRAX- and Garvan-predicted hip fracture and MOF fracture probabilities.

KEY RESULTS: The observed 10-year hip fracture probability was 0.3% for women aged 50–54 years ($n = 14,768$), 0.6% for women aged 55–59 years ($n = 22,442$), and 1.1% for women aged 60–64 years ($n = 25,513$). At sensitivity thresholds $\geq 80\%$, specificity of both tools for detecting incident hip fracture during 10 years of follow-up was low: Garvan 30.6% (95% confidence interval [CI] 30.3–31.0%) and FRAX 43.1% (95% CI 42.7–43.5%). At maximal area under the receiver operating characteristic curve (AUC(c), 0.58 for Garvan, 0.65 for FRAX), sensitivity was 16.0% (95% CI 12.7–19.4%) for Garvan and 59.2% (95% CI 54.7–63.7%) for FRAX. At AUC(c) values, sensitivity was lower in African American and Hispanic women than among white women and lower in women aged 50–54 than those 60–64 years old. Observed hip fracture probabilities were similar to FRAX-predicted probabilities but greater than

Garvan-predicted probabilities. At AUC(c) values (0.56 for both tools), sensitivity for identifying MOF was also low (range 26.7–46.8%). At AUC(c) values (0.55 for both tools), sensitivity for identifying any clinical fracture ranged from 18.1 to 34.0%.

CONCLUSIONS: In postmenopausal women aged 50–64 years, the FRAX and Garvan fracture risk calculator discriminate poorly between women who do and do not experience fracture during 10-year follow-up. There is no useful threshold for either tool.

KEY WORDS: Garvan; fracture; osteoporosis; FRAX; fracture risk assessment.

J Gen Intern Med 34(2):235–42

DOI: 10.1007/s11606-018-4696-z

© Society of General Internal Medicine 2018

INTRODUCTION

One in two postmenopausal women will experience a fracture in their remaining lifetimes.¹ Because fractures can result in prolonged disability, loss of independence, and death, predicting the risk of fracture before it occurs is of great importance and in the scope of general internal medicine care.¹ The Fracture Risk Assessment Tool (FRAX), recommended by current clinical osteoporosis screening guidelines, is a Web-based fracture risk calculator that estimates the 10-year absolute probability of hip and major osteoporotic fracture (MOF) (hip, spine, wrist, and proximal humerus) in adults aged 40 through 90 years. FRAX includes the following risk factors: age, sex, weight, height, previous fracture (yes/no), parental hip fracture, current smoking, glucocorticoid use (prednisolone ≥ 5 mg/day for ≥ 3 months), rheumatoid arthritis, alcohol intake.² The FRAX tool can be used with or without bone mineral density (BMD) information.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11606-018-4696-z>) contains supplementary material, which is available to authorized users.

Received January 22, 2018

Revised July 10, 2018

Accepted September 14, 2018

Published online October 17, 2018

For women aged ≥ 65 years, the United States Preventive Services Task Force (USPSTF) screening guideline `BMD testing. However, for postmenopausal women aged 50–64 years, screening is based on FRAX risk.¹ Specifically, the USPSTF recommends bone density testing for women aged 50–64 years who have a $\geq 9.3\%$ 10-year-predicted probability of MOF using FRAX (without BMD information).¹ The USPSTF chose a FRAX threshold of $\geq 9.3\%$ for women aged 50–64 years because it corresponds to the estimated 10-year probability of major osteoporotic fracture of a 65-year-old white female who has no other osteoporosis risk factors.¹ However, the USPSTF recommendation was not based on studies that evaluated the performance of FRAX for osteoporosis screening. In our prior work in the Women's Health Initiative (WHI) Study in women aged 50–64 years, the sensitivity of the USPSTF FRAX-based strategy for identifying incident MOF was low, ranging from 4.7% (women aged 50–54 years) to 37.3% (women aged 60–64 years).³

Given the low sensitivities for the USPSTF-recommended screening approach for FRAX in women aged 50–64 years, it is important to examine the performance of alternative screening tools. The Garvan fracture risk calculator (<http://garvan.org.au/promotions/bone-fracture-risk/calculator/>) includes sex, age, prior fracture since age 50 excluding major trauma (0, 1, 2, 3, or more), falls in last 12 months (0, 1, 2, 3, or more), and weight (kg) to provide the predicted absolute risk of hip fracture and any osteoporotic fracture over 5- and 10-year time horizons.^{4–6} Therefore, compared with the FRAX risk calculator, the Garvan risk calculator incorporates fewer risk factors, but includes falls, and counts the *number* of previous fractures. Unlike FRAX, the Garvan tool does not predict MOF risk. The Garvan risk calculator, developed in Australia, has never been externally validated in US women, but is commonly used in Australia and is recommended by the guidelines of the Royal Australian College of General Practitioners. Also, no published prospective studies have compared observed vs. predicted 10-year hip fracture risk for the FRAX and Garvan risk tools in postmenopausal women in the USA. A comparison of FRAX and Garvan hip fracture risk prediction is important to inform screening strategies.

Our primary objective was to evaluate and compare the discrimination and calibration of the FRAX and Garvan hip fracture risk prediction tools in postmenopausal women aged 50–64 years at baseline during 10 years of follow-up. Secondary objectives were to assess and compare the discrimination of the tools in the prediction of MOF and any clinical fracture.

METHODS

Design of the Women's Health Initiative Study

The Women's Health Initiative Observational Study (WHI-OS) and Clinical Trials (WHI-CT), conducted at 40 US clinical centers (1993–2005), enrolled postmenopausal women aged 50 to 79 years at baseline (1993–1998). The WHI-CT evaluated three interventions: a low-fat eating pattern, menopausal

hormone therapy, and calcium plus vitamin D supplementation.⁷ The study design was previously described.⁷ Inclusion criteria included freedom from serious medical conditions.

For the current study, we analyzed longitudinal data from baseline to year 10 of follow-up from all WHI-OS and WHI-CT participants who were aged 50–64 years at baseline ($n = 90,764$). We excluded data from participants who did not provide information regarding medication use at baseline, and from those who reported taking any medications known to influence osteoporosis at baseline: bisphosphonates, calcitonin, parathyroid hormone, selective estrogen receptor modulators, luteinizing hormone-releasing hormone agents, and somatostatin agents ($n = 1111$). We excluded data from participants who contributed incomplete information regarding osteoporosis risk factors ($n = 5714$, 6.4%) or less than 10 years of follow-up time before death or loss to follow-up ($n = 19,200$, 30.6%). Therefore, the analytic sample for this study consisted of 62,723 participants for hip fracture analyses, 63,621 for major osteoporotic fracture analyses, and 64,739 for clinical fracture analyses (Appendix Fig. 1). The analyses were performed using FRAX and Garvan calculators without BMD information.

Human subjects review committees at each participating institution approved the study. Each participant provided written informed consent.

Outcomes: Incident Fractures

Incident fractures were self-reported annually (WHI-OS) or semi-annually (WHI-CT) using questionnaires.

All hip fractures were confirmed by physician adjudicators using medical records. Other types of fractures were self-reported. In a medical record-based validation study, the average confirmation rate for all single-site self-reported fractures was 71%.⁸

MOF was defined as a first hip, lower arm/wrist, clinical spine, or upper arm fracture.^{2,9} We defined clinical fracture as fracture of the hip, vertebrae (clinical), wrist or forearm, hand, foot, humerus, scapula, clavicle, distal femur, proximal tibia, patella, pelvis, and sternum. Fractures of the fingers, toes, and skull were excluded.

Main Predictors: FRAX- and Garvan-Predicted Absolute Fracture Risk

The FRAX-predicted 10-year risks of hip fracture and of MOF were calculated by the World Health Organization Collaborating Centre for Metabolic Bone Disease (US FRAX version 3.0 without BMD), as described previously.³ For the Garvan risk tool, the 5-year- and 10-year-predicted risks of hip fracture and any fragility fracture (without BMD) were calculated for each participant using published formulas.⁴

Other Measurements

Information regarding age, race/ethnicity, medical history (previous fractures, rheumatoid arthritis, number of falls in

previous 12 months), medication use, parental hip fracture, smoking, alcohol intake, and use of supplemental calcium and vitamin D was assessed using baseline questionnaires. Weight and height were measured at baseline using standardized protocols.

Statistical Analysis

We determined the sensitivity, specificity, and AUC of the FRAX and Garvan models (without BMD information) in discriminating between participants who did and did not experience a hip fracture during (1) a 5-year time horizon for Garvan and (2) a 10-year time horizon for both FRAX and Garvan. AUCs plot the true positive rate (sensitivity) as a function of the false positive rate (100 minus specificity); they indicate how well a risk score can distinguish between diseased and non-diseased persons. AUC values range from 0.50 to 1.0; higher values indicate better discrimination. An AUC value of 0.50 indicates that the risk assessment tool is no better than chance in distinguishing between women who do and do not experience fracture during 10-year follow-up. AUC values were generated using logistic regression.

We calculated the risk score thresholds that would correspond to sensitivities of $\geq 80\%$ for detection of (observed) incident hip fracture over 10 years, with associated specificities, positive predictive value (PPV), and AUC values. For each risk score, Youden's *J* statistic was calculated to identify the cut-point to maximize AUC values for identification of hip fracture (AUC(c)). We identified within age strata the Garvan and FRAX risk thresholds corresponding to AUC(c) values for identifying hip fracture. Because the FRAX and Garvan tools both provide an estimate of hip fracture probability, our primary analysis focused on prediction of hip fracture risk. In secondary analyses, we assessed discrimination for predicting the 10-year risk of MOF and 10-year risk of any clinical fracture. Because the Garvan risk tool does not provide an estimate of any clinical fracture or MOF, we used the Garvan-predicted probability of "any fragility fracture" when we assessed the discrimination of the Garvan risk tool for prediction of any observed clinical fracture or MOF. Because the FRAX risk tool does not provide an estimate of any clinical fracture, we used the FRAX-predicted probability of MOF when we assessed the discrimination of the FRAX risk tool for prediction of any clinical fracture. In additional secondary analyses designated a priori, we stratified our results according to the use of hormone therapy at baseline (yes vs. no) and race/ethnicity (white, African American, Hispanic, other).

In sensitivity analyses, we excluded data from participants who initiated osteoporosis medications any time during follow-up: bisphosphonates, calcitonin, parathyroid hormone, selective estrogen receptor modulators, luteinizing hormone-releasing hormone agents, and somatostatin agents. Finally, we repeated the main analyses after excluding participants ($n = 29,866$) who used estrogen therapy any time during study follow-up.

To assess calibration, we compared observed vs. predicted 10-year probability of hip fracture stratified by age category. Analyses were completed with SAS for Windows Version 9.4.

RESULTS

Baseline Characteristics

At baseline, the mean (standard deviation [SD]) participant age was 57.9 (4.1) years (Table 1).

Compared with the participants included in this study, participants excluded due to death or termination from the study after less than 10 years of follow-up time were older and less

Table 1 Selected Baseline Characteristics of the 64,739 Study Participants

	No. (%) of participants
Age (years)	
50–54	15,221 (23.5)
55–59	23,086 (35.7)
60–64	26,432 (40.8)
Body mass index (kg/m ²)	
<25	22,986 (35.5)
25–<30	21,698 (33.5)
≥ 30	19,703 (30.4)
Race/ethnicity	
Black	5719 (8.8)
Hispanic	2376 (3.7)
White	54,102 (83.6)
Other/unknown	2542 (3.9)
History of corticosteroid use	
Yes	402 (0.6)
No	64,337 (99.4)
Rheumatoid arthritis	
Yes	2612 (4.0)
No	61,725 (95.3)
Missing	402 (0.6)
Smoking	
Never smoke	31,634 (48.9)
Past smoker	27,720 (42.8)
Current smoker	4786 (7.4)
Missing	599 (0.9)
Number of fractures* ≥ 55	
0	46,600 (72.0)
1	2699 (4.2)
2	201 (0.3)
≥ 3	18 (0.0)
Participant < 55	15,221 (23.5)
Falls in past year	
None	43,994 (68.0)
1 time	12,644 (19.5)
2 times	5314 (8.2)
3 or more times	2787 (4.3)
Current menopausal hormone therapy	
Yes	35,969 (55.6)
No	28,725 (44.4)
Missing	45 (0.1)
≥ 3 alcoholic drinks per day	
Yes	762 (1.2)
No	63,822 (98.6)
Missing	155 (0.2)
Parental hip fracture	
Yes	7901 (12.2)
No	54,457 (84.1)
Missing	2381 (3.7%)

*Only one fracture is counted at each body site

likely to be white, and had higher body mass index (Appendix Table 1).

Fracture Incidence

The observed 10-year hip fracture probability was 0.3% for women aged 50–54 years ($n = 14,768$), 0.6% for women aged 55–59 years ($n = 22,442$), and 1.1% for women aged 60–64 years ($n = 25,513$). The observed 10-year MOF probability was 6.3% for women aged 50–54 years ($n = 14,947$), 8.0% for women aged 55–59 years ($n = 22,721$), and 9.9% for women aged 60–64 years ($n = 25,953$). The observed 10-year probability of any clinical fracture was 15.8% for women aged 50–54 years ($n = 15,221$), 17.2% for women aged 55–59 years ($n = 23,086$), and 19.0% for women aged 60–64 years ($n = 26,432$).

Discrimination and Calibration of Hip Fracture Probability by the FRAX and Garvan Models (Without BMD Information)

The overall AUC value for prediction of hip fracture was 0.68 (0.65–0.70) AUC for FRAX and 0.62 (95% CI 0.59–0.65) for Garvan.

At risk thresholds corresponding to 80% sensitivity, AUC(c) values for discrimination of hip fracture were low (0.64 [95% confidence interval (CI) 0.61–0.66] for FRAX; 0.57 [95% CI 0.55–0.60] for Garvan), and specificity was low (30.6% [95% CI 30.3–31.0%] for Garvan, 43.1% [95% CI 42.7–43.5%] for FRAX) (Table 2; Fig. 1a, b). In calibration

assessment, observed fracture probabilities were similar to FRAX-predicted probabilities but greater than Garvan-predicted probabilities in each age stratum (Fig. 2, Appendix Table 2).

In analyses excluding data from participants who used osteoporosis medication or estrogen therapy during the follow-up period, results were similar to those of the primary analysis (Appendix Table 3).

In stratified analyses, at AUC(c) thresholds, the sensitivities of the FRAX and Garvan risk tools for identifying incident hip fracture were much lower for women aged 50–54 (Garvan 0.0% [95% CI 0.0–0.0%], FRAX 10.4% [95% CI 1.5–19.4%]) than those for women aged 60–64 (Garvan 24.2% [95% CI 19.1–29.3%], FRAX 80.5% [95% CI 75.8–85.2%]) (Table 3). Results were very similar among women who were current users of hormone therapy at baseline and those who were not. However, in race/ethnicity-stratified analyses, the sensitivities of FRAX and Garvan models for identifying incident hip fractures were markedly lower in African American and Hispanic women than those among white women. At AUC(c) values, the sensitivity in African American women was 0.0% (95% CI 0.0–0.0%) for FRAX and 7.7% (95% CI 0.0–24.5%) for Garvan. The sensitivity in Hispanic women was 12.5% (95% CI 0.0–42.1%) for FRAX and 0.0% (95% CI 0.0–0.0%) for Garvan. The sensitivity in white women was 62.2% (95% CI 57.7–66.8%) for FRAX and 16.2% (95% CI 12.8–19.7%) for Garvan.

For both the FRAX and Garvan tool, AUC values for prediction of hip fracture were low across racial/ethnic groups,

Table 2 Sensitivity, Specificity, Positive Predictive Value (PPV), and Area Under the Receiver Operating Characteristic Curve (AUC) of the Garvan Tool and FRAX for Identifying Incident Hip Fracture Over 10 Years of Follow-up Under Various Construct Cut-points (Sensitivity Range of 80–99%)

Outcome	Tool					
Hip fracture 468 events	Garvan	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	AUC (95% CI)*	
	≥ 0.097	80.6 (77.0–84.2)	30.6 (30.3–31.0)	0.9 (0.8–1.0)	0.57 (0.55–0.60)	
	≥ 0.079	85.3 (82.0–88.5)	23.1 (22.7–23.4)	0.8 (0.7–0.9)	0.56 (0.54–0.59)	
	≥ 0.065	90.2 (87.5–92.9)	17.4 (17.1–17.7)	0.8 (0.7–0.9)	0.56 (0.54–0.58)	
	≥ 0.036	95.1 (93.1–97.1)	7.3 (7.1–7.5)	0.8 (0.7–0.8)	0.53 (0.51–0.56)	
	≥ 0.012	99.1 (98.3–100.0)	1.4 (1.3–1.5)	0.8 (0.7–0.8)	0.53 (0.51–0.55)	
	FRAX	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	AUC (95% CI)	
	≥ 0.430	80.6 (77.0–84.2)	43.1 (42.7–43.5)	1.1 (0.9–1.2)	0.64 (0.61–0.66)	
	≥ 0.350	85.0 (81.8–88.3)	33.2 (32.8–33.6)	0.9 (0.9–1.0)	0.61 (0.59–0.63)	
	≥ 0.280	90.0 (87.2–92.7)	23.7 (23.3–24.0)	0.9 (0.8–1.0)	0.59 (0.57–0.61)	
	≥ 0.210	95.1 (93.1–97.1)	14.0 (13.7–14.2)	0.8 (0.7–0.9)	0.57 (0.55–0.59)	
	≥ 0.090	99.1 (98.3–100.0)	1.9 (1.8–2.0)	0.8 (0.7–0.8)	0.53 (0.51–0.55)	
	MOF 5332 events	Garvan	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	AUC (95% CI)
		≥ 7.371	80.0 (79.0–81.1)	27.1 (26.7–27.4)	9.1 (8.9–9.4)	0.55 (0.54–0.56)
≥ 6.863		85.1 (84.1–86.0)	19.8 (19.5–20.1)	8.8 (8.6–9.1)	0.54 (0.53–0.55)	
≥ 6.365		90.0 (89.2–90.8)	13.1 (12.9–13.4)	8.7 (8.4–8.9)	0.53 (0.53–0.54)	
≥ 5.686		95.0 (94.4–95.6)	6.6 (6.4–6.8)	8.5 (8.3–8.7)	0.53 (0.52–0.53)	
≥ 4.405		99.0 (98.8–99.3)	1.3 (1.2–1.4)	8.4 (8.2–8.6)	0.52 (0.52–0.53)	
FRAX		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	AUC (95% CI)	
≥ 4.800		80.2 (79.1–81.3)	28.6 (28.2–28.9)	9.3 (9.0–9.6)	0.56 (0.55–0.57)	
≥ 4.330		85.2 (84.2–86.1)	22.8 (22.4–23.1)	9.2 (8.9–9.4)	0.56 (0.55–0.56)	
≥ 3.750		90.2 (89.4–91.0)	16.1 (15.8–16.4)	8.9 (8.7–9.2)	0.55 (0.54–0.56)	
≥ 2.980		95.0 (94.5–95.6)	9.0 (8.8–9.3)	8.7 (8.5–9.0)	0.54 (0.53–0.55)	
≥ 1.790		99.0 (98.8–99.3)	1.8 (1.6–1.9)	8.4 (8.2–8.7)	0.53 (0.52–0.53)	

FRAX denotes the Fracture Risk Assessment Tool. Sensitivity thresholds were chosen to reflect thresholds of clinical relevance (i.e., sensitivity ≥ 80%). Cut-points selected to maximize area under the curve via Youden's J statistic

*Logistic models to calculate AUC are additionally adjusted for WHI hormone therapy and calcium/vitamin D interventions

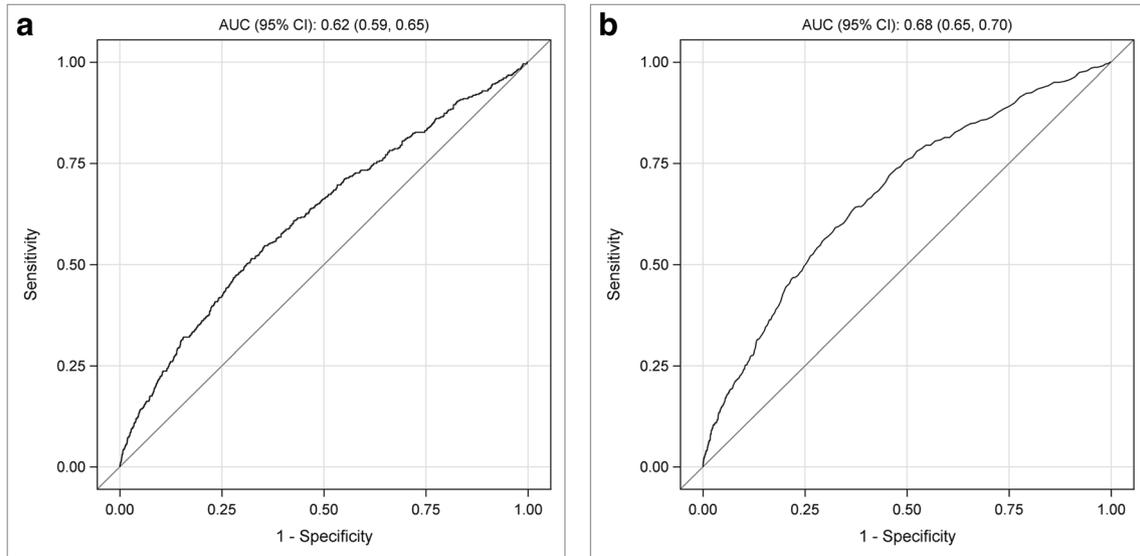


Figure 1 Area under the receiver operating characteristic (ROC) curves for the prediction of incident 10-year hip fracture by the Garvan tool and Fracture Risk Assessment Tool (FRAX). a ROC curve for identifying hip fracture by the Garvan tool. b ROC curve for identifying hip fracture by the Fracture Risk Assessment Tool (FRAX).

albeit slightly higher among white women (Appendix Table 4). AUC values ranged from 0.53 to 0.58 in African American and Hispanic participants, indicating that the tools performed no better than chance in discriminating between African American and Hispanic participants who did and did not experience hip fractures during the follow-up period.

At AUC(c) values, the sensitivity of the 5-year Garvan hip fracture prediction model was lower than that of the FRAX (10-year) hip prediction model, and lower among 50- to 54-year-old women (0% [95% CI 0.0–100%] for Garvan 5 years, 10.4% [95% CI 1.5–19.4%] for FRAX) than among 60- to 64-year-old women (12.6% [95% CI 6.1–19.1%] for Garvan 5 years, 80.5% [95% CI 75.8–85.2%] for FRAX) (Appendix Table 5).

Discrimination of 10-Year Probability of MOF and Any Clinical Fracture for the FRAX and Garvan Models (Without BMD Information)

The overall AUC values for MOF were 0.58 (95% CI 0.57–0.59) for FRAX and 0.57 (95% CI 0.57–0.58) for Garvan; overall AUC values for clinical fracture were 0.55 (95% CI 0.54–0.56) for FRAX and 0.55 (95% CI 0.54–0.55) for Garvan.

Neither FRAX nor Garvan performed well in discriminating between women who did and did not experience MOF (Appendix Fig. 2a, b) or clinical fractures (Appendix Fig. 3a, b). AUC(c) values were identical, 0.56 in each age stratum, and corresponding sensitivity values were low for the Garvan tool and FRAX in all strata (range 26.7–46.8%) (Table 4). Findings were similar regarding prediction of any clinical fracture (Appendix Table 6).

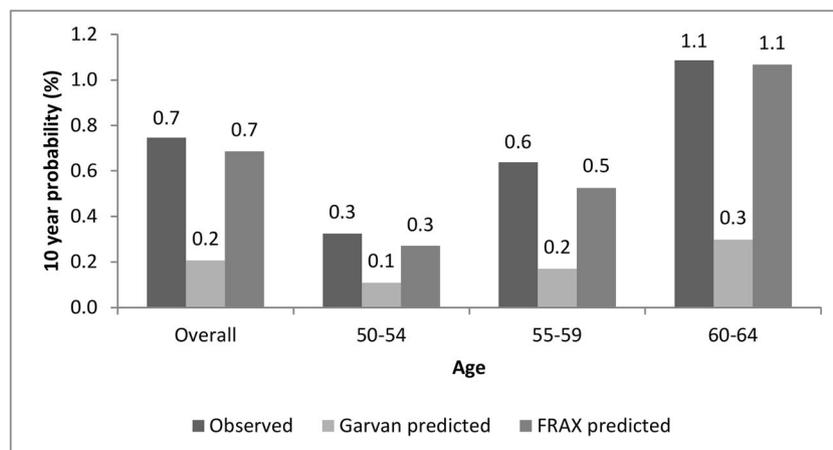


Figure 2 Observed versus predicted 10-year probability of hip fracture overall and by age group according to the Garvan tool and Fracture Risk Assessment Tool (FRAX).

Table 3 Sensitivity, Specificity, and Area Under the Receiver Operating Characteristic Curve (AUC) of the Garvan Tool and Fracture Risk Assessment Tool (FRAX) for Identifying Incident Hip Fracture After 10 Years, Stratified by Age Group, at Thresholds Maximizing the Area Under the Receiver Operating Characteristic Curve (AUC)

	Sensitivity (95% CI)	Specificity (95% CI)	AUC
All participants (n = 63,621)			
Garvan \geq 0.462	16.0 (12.7–19.4)	93.5 (93.3–93.7)	0.58 (0.55–0.60)
FRAX \geq 0.706	59.2 (54.7–63.7)	67.6 (67.2–67.9)	0.65 (0.62–0.67)
Age			
50–54 (n = 14,768)			
Garvan \geq 0.462	0.0 (0.0–0.0)	99.7 (99.7–99.8)	0.54 (0.49–0.60)
FRAX \geq 0.706	10.4 (1.5–19.4)	98.1 (97.9–98.3)	0.58 (0.52–0.65)
55–59 (n = 22,442)			
Garvan \geq 0.462	5.6 (1.8–9.4)	96.9 (96.7–97.1)	0.56 (0.52–0.60)
FRAX \geq 0.706	34.3 (26.4–42.1)	81.2 (80.7–81.7)	0.61 (0.56–0.65)
60–64 (n = 25,513)			
Garvan \geq 0.462	24.2 (19.1–29.3)	86.9 (86.5–87.3)	0.60 (0.57–0.63)
FRAX \geq 0.706	80.5 (75.8–85.2)	37.7 (37.1–38.3)	0.62 (0.59–0.65)
HT use [†]			
No (n = 27,617)			
Garvan \geq 0.462	13.2 (8.8–17.7)	93.4 (93.1–93.7)	0.55 (0.52–0.58)
FRAX \geq 0.706	57.3 (50.8–63.8)	67.7 (67.1–68.2)	0.63 (0.60–0.67)
Yes (n = 35,065)			
Garvan \geq 0.462	18.7 (13.7–23.6)	93.6 (93.3–93.8)	0.60 (0.57–0.63)
FRAX \geq 0.706	61.0 (54.8–67.2)	67.5 (67.0–68.0)	0.66 (0.63–0.70)
Years of HT use			
0 years (n = 12,250)			
Garvan \geq 0.462	16.5 (8.7–24.3)	92.7 (92.3–93.2)	0.57 (0.53–0.62)
FRAX \geq 0.706	57.1 (46.8–67.5)	64.0 (63.2–64.9)	0.63 (0.57–0.68)
5+ years (n = 18,762)			
Garvan \geq 0.462	21.8 (13.0–30.7)	93.1 (92.8–93.5)	0.61 (0.57–0.66)
FRAX \geq 0.706	64.4 (54.1–74.6)	65.8 (65.2–66.5)	0.67 (0.62–0.72)
Race/ethnicity			
White (n = 52,536)			
Garvan \geq 0.462	16.2 (12.8–19.7)	93.2 (93.0–93.5)	0.57 (0.55–0.60)
FRAX \geq 0.706	62.2 (57.7–66.8)	62.9 (62.5–63.3)	0.64 (0.61–0.66)
African American (n = 5475)			
Garvan \geq 0.462	7.7 (0.0–24.5)	97.5 (97.0–97.9)	0.61 (0.48–0.74)
FRAX \geq 0.706	0.0 (0.0–0.0)	96.4 (95.9–96.9)	0.63 (0.51–0.75)
Hispanic (n = 2262)			
Garvan \geq 0.462	12.5 (0.0–42.1)	94.5 (93.5–95.4)	0.71 (0.58–0.83)
FRAX \geq 0.706	0.0 (0.0–0.0)	92.8 (91.7–93.8)	0.71 (0.61–0.81)
Other/unknown* (n = 2450)			
Garvan \geq 0.462	20.0 (0.0–50.2)	89.4 (88.2–90.6)	0.67 (0.56–0.78)
FRAX \geq 0.706	50.0 (12.3–87.7)	78.8 (77.1–80.4)	0.74 (0.62–0.86)

Youden's J statistic was calculated to identify the cut-point to maximize AUC values for identification of hip fracture (AUC(c))

*Other/unknown category includes American Indian, Asian, and unknown race/ethnicity participants

[†]Menopausal hormone therapy (HT) status over the interval has varying definitions depending on whether the participant was enrolled in the Women's Health Initiative Observational Study (OS) or the HT Clinical Trials (CT): OS, Non-HT CT Participants: Status determined by baseline current HT use (yes/no)

HT participants: status determined by HT intervention arm (active/placebo)

OS participants: participants who contributed data for at least 6 of the 8 possible years of self-report covered by WHI were divided into groups. Participants who reported \geq 50% of study period using HT were included in the 5+ years HT use category. If the participant reported 0% of her time on HT, she was put into the no HT use group

HT participants: active arm participants who were on intervention for 5+ years were put into the 5+ years HT use category. Placebo arm participants who were on placebo for 5+ years and then reported not starting hormones on their post-intervention form were put into the no HT use group

Non-HT, CT participants: not included in analysis

For both the FRAX and Garvan tool, AUC values for prediction of MOF were low and similar across racial/ethnic groups (range 0.53–0.61) (Appendix Table 7).

DISCUSSION

In this cohort of US postmenopausal women aged 50–64 years, we compared the Garvan and FRAX fracture risk assessment tools. The FRAX risk tool was only modestly better than chance, and the Garvan risk tool was no better than chance, in discriminating women with and without incident

hip fractures during 10 years of prospective follow-up. For prediction of MOF and any clinical fracture, performance of the two tools was similar and poor. At AUC(c) values, sensitivity of both tools was extremely low (ranging 0–13%) for African American and Hispanic women and women aged 50–54 years. At thresholds yielding sensitivity 80% for identification of observed hip fractures, PPV was very low (< 2%) for both tools. Observed fracture probabilities were similar to FRAX-predicted probabilities but greater than Garvan-predicted probabilities, indicating that the FRAX is well calibrated, but the Garvan tool markedly underestimates actual 10-year fracture probabilities.

Table 4 Sensitivity, Specificity, and Area Under the Receiver Operating Characteristic Curve (AUC) of the Garvan Tool and Fracture Risk Assessment Tool (FRAX) for Identifying Incident Major Osteoporotic Fracture After 10 Years, Stratified by Baseline Age, at Thresholds Maximizing the Area Under the Receiver Operating Characteristic Curve (AUC)

	Sensitivity (95% CI)	Specificity (95% CI)	AUC
50–54 years (<i>n</i> = 14,947)			
Garvan \geq 7.2	33.2 (30.2–36.2)	74.7 (74.0–75.4)	0.56 (0.54–0.58)
FRAX \geq 5.10	26.7 (23.9–29.5)	79.3 (78.7–80.0)	0.55 (0.53–0.57)
55–59 years (<i>n</i> = 22,721)			
Garvan \geq 8.95	46.8 (44.5–49.1)	63.1 (62.4–63.7)	0.56 (0.54–0.57)
FRAX \geq 7.04	33.6 (31.4–35.7)	75.4 (74.8–76.0)	0.56 (0.54–0.57)
60–64 years (<i>n</i> = 25,953)			
Garvan \geq 13.58	27.1 (25.4–28.9)	81.6 (81.1–82.1)	0.56 (0.55–0.57)
FRAX \geq 9.27	37.5 (35.6–39.4)	72.2 (71.6–72.8)	0.56 (0.55–0.57)

Youden's J statistic was calculated to identify the cut-point to maximize AUC values for identification of major osteoporotic fracture (AUC(c))

To our knowledge, this is the first study to compare actual vs. predicted 10-year hip fracture risk for the FRAX and Garvan risk tools in postmenopausal women aged 50–64 years in the USA. A few studies were performed outside the USA and/or in older women. First, the GLOW study of women from 10 countries including the USA had a 2-year duration of follow-up; two thirds of the participants were aged 65 and older, i.e., older than the present study of women aged 50–64 years.¹⁰ In the GLOW study, AUC values for hip fracture were 0.65 for FRAX and 0.61 for the Garvan risk calculator.¹⁰ Second, the study by Bolland and colleagues in New Zealand, which was limited to women with normal BMD for their age (mean 74 years), found similar AUC values for the FRAX and Garvan risk tools, ranging between 0.60 and 0.70 for hip fractures during a mean 8.8-year follow-up.¹¹ The previous studies did not report AUC values separately in women aged 50–64, precluding a direct comparison with our current study. Finally, a recent Israeli study with an average follow-up of 4.7 years found underestimation of hip fracture risk by FRAX and Garvan and superior calibration of FRAX compared with Garvan, as in our study, but higher AUC values (0.82 for FRAX, 0.78 for Garvan) compared with our study.¹² The differences in AUC values between the studies may be due to inclusion of both men and women, a wider age range, and/or shorter follow-up duration in the Israeli study compared with our study.¹²

Our results are clinically relevant to general internists because of the several different existing guidelines regarding screening in postmenopausal women aged 50–64. Although the FRAX is recommended by several guidelines for screening decisions in postmenopausal women in this age group,^{1, 9} general internists should be aware that neither the FRAX nor Garvan risk calculator performs well in identifying postmenopausal women aged 50–64 years who will experience hip fractures, MOF, or any clinical fracture during 10 years of follow-up. Thus, our results do not support the use of either FRAX or the Garvan hip fracture risk prediction tool in postmenopausal women aged 50–64 years. Moreover, the ability of the tools to distinguish between women who do and do not experience fracture was particularly low among African American and Hispanic women, and no better than

chance. The role of the clinical use of these tools among minority women should be examined in future studies. Given the unsatisfactory performance of the tools for fracture risk prediction and low 10-year observed probability of disabling fracture (i.e., hip fracture), our results suggest that emphasis on lifestyle strategies rather than drug treatment for fracture risk reduction is appropriate for most women in this age group.

Limitations of this study include that participants who were excluded (due to providing less than 10 years of study follow-up) were older. However, there is no indication that this would lead to bias favoring one fracture prediction method over another. Our analyses evaluating tool performance within racial/ethnic groups require confirmation in other studies because of the relatively low incidence of fractures in African American and Hispanic participants. Strengths of our study include the large sample size, the 10-year duration of follow-up, medical record-based confirmation of hip fractures, and detailed information regarding fracture risk factors.

In conclusion, the FRAX and Garvan fracture risk calculator discriminated poorly between postmenopausal women aged 50–64 years who do and do not experience hip fracture during 10-year follow-up; discrimination was similarly poor for MOF and clinical fractures. For women of color, the discriminative ability of both tools was very poor, and no greater than chance. There is not a useful threshold for either tool for prediction of these fracture outcomes. Calibration was acceptable for the FRAX, but the Garvan tool underestimated observed fracture probabilities. Fracture prediction in this age group requires assessment of risk factors not included in currently available strategies. These results will help to inform future osteoporosis screening guidelines.

Acknowledgements:

Short List of WHI Investigators
 Program Office: (National Heart, Lung, and Blood Institute, Bethesda, MA) Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller.
 Clinical Coordinating Center: Clinical Coordinating Center: (Fred Hutchinson Cancer Research Center, Seattle, WA) Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg.
 Investigators and Academic Centers: (Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn E. Manson; (MedStar Health Research Institute/Howard University, Washington, DC)

Barbara V. Howard; (Stanford Prevention Research Center, Stanford, CA) Marcia L. Stefanick; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Arizona, Tucson/Phoenix, AZ) Cynthia A. Thomson; (University at Buffalo, Buffalo, NY) Jean Wactawski-Wende; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Iowa, Iowa City/Davenport, IA) Jennifer Robinson; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker; (University of Nevada, Reno, NV) Robert Brunner; (University of Minnesota, Minneapolis, MN) Karen L. Margolis. Women's Health Initiative Memory Study; (Wake Forest University School of Medicine, Winston-Salem, NC) Mark Espeland.

For a list of all the investigators who have contributed to WHI science, please visit: <https://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf>

We thank the WHI study participants and investigators for their tremendous dedication and commitment to the study.

Authorship Roles: CJC was responsible for the conception of the study. All authors participated in the analysis and interpretation of data and critical revisions of the manuscript for important intellectual content. JL was responsible for performing the statistical analyses. AL, JAC, MSL, and JEM were responsible for acquisition of data.

Corresponding Author: Carolyn J. Crandall, MD, MS; Division of General Internal Medicine and Health Services Research, Department of Medicine David Geffen School of Medicine at University of California, Los Angeles, CA, USA (e-mail: ccrandall@mednet.ucla.edu).

Funding The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C.

Compliance with Ethical Standards: Human subjects review committees at each participating institution approved the study. Each participant provided written informed consent.

Conflict of Interest: The authors declare that they do not have a conflict of interest.

Data Analysis: JL had full access to all the data in the study and takes responsibility for the integrity of the data analysis.

Data Sharing: Women's Health Initiative Study data are available via the BioLINCC website of the National Heart, Lung, and Blood Institute at <https://biolincc.nhlbi.nih.gov/home/>

Role of the Funding Source: This study was partially funded through contracts with the WHI Coordinating Center. The WHI Study

was funded by the National Institutes of Health. The National Institutes of Health designated representatives who participated in the design and monitoring of the WHI. The researchers are independent from the funders.

REFERENCES

1. Screening for Osteoporosis: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med.* 2011;154(5):356–64. doi:<https://doi.org/10.1059/0003-4819-154-5-201103010-00307>
2. Centre for Metabolic Bone Diseases UoS, UK. FRAX® Fracture Risk Assessment Tool. <https://www.shef.ac.uk/FRAX/>. Accessed 8/16/2018.
3. Crandall CJ, Larson JC, Watts NB, Gourlay ML, Donaldson MG, LaCroix A, et al. Comparison of Fracture Risk Prediction by the US Preventive Services Task Force Strategy and Two Alternative Strategies in Women 50–64 Years Old in the Women's Health Initiative. *J Clin Endocrinol Metab* 2014;99(12):4514–22. doi:<https://doi.org/10.1210/jc.2014-2332>
4. Ahmed LA, Nguyen ND, Bjornerem A, Joakimsen RM, Jorgensen L, Stormer J, et al. External Validation of the Garvan Nomograms for Predicting Absolute Fracture Risk: the Tromso Study. *PLoS One* 2014;9(9):e107695. doi:<https://doi.org/10.1371/journal.pone.0107695>
5. Nguyen ND, Frost SA, Center JR, Eisman JA, Nguyen TV. Development of a Nomogram for Individualizing Hip Fracture Risk in Men and Women. *Osteoporos Int* 2007;18(8):1109–17. doi:<https://doi.org/10.1007/s00198-007-0362-8>
6. Nguyen ND, Frost SA, Center JR, Eisman JA, Nguyen TV. Development of Prognostic Nomograms for Individualizing 5-Year and 10-Year Fracture Risks. *Osteoporos Int* 2008;19(10):1431–44. doi:<https://doi.org/10.1007/s00198-008-0588-0>
7. Jacobsen BK, Nilssen S, Heuch I, Kvale G. Reproductive factors and fatal hip fractures. A Norwegian prospective study of 63,000 women. *J Epidemiol Community Health* 1998;52(10):645–50.
8. Chen Z, Kooperberg C, Pettinger MB, Bassford T, Cauley JA, LaCroix AZ, 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):264–74.
9. National Osteoporosis Foundation. Clinician's Guide to Prevention and Treatment of Osteoporosis. Washington, DC: National Osteoporosis Foundation 2014.
10. Sambrook PN, Flahive J, Hooven FH, Boonen S, Chapurlat R, Lindsay R, et al. Predicting Fractures in an International Cohort Using Risk Factor Algorithms Without BMD. *J Bone Miner Res* 2011;26(11):2770–7. doi:<https://doi.org/10.1002/jbmr.503>
11. Bolland MJ, Siu AT, Mason BH, Horne AM, Ames RW, Grey AB, et al. Evaluation of the FRAX and Garvan Fracture Risk Calculators in Older Women. *J Bone Miner Res* 2011;26(2):420–7. doi:<https://doi.org/10.1002/jbmr.215>
12. Dagan N, Cohen-Stavi C, Leventer-Roberts M, Balicer RD. External Validation and Comparison of Three Prediction Tools for Risk of Osteoporotic Fractures Using Data from Population Based Electronic Health Records: Retrospective Cohort Study. *BMJ.* 2017;356:i6755. doi:<https://doi.org/10.1136/bmj.i6755>