ORIGINAL RESEARCH

Electronic Health Records (EHRs) Can Identify Patients at High Risk of Fracture but Require Substantial Race Adjustments to Currently Available Fracture Risk Calculators

Rajesh K. Jain, MD[1], Mark Weiner, MD[2], Eric Polley, PhD[3], Amy Iwamaye, MD[4], Elbert Huang, MD[5], and Tamara Vokes, MD[1]

[1]Department of Medicine, Section of Endocrinology, Diabetes, and Metabolism, The University of Chicago, 5841 South Maryland Ave, MC 1027, Chicago, IL 60637, USA; [2]Weill Cornell Medicine, Clinical Population Health Sciences, New York, USA; [3]Department of Public Health Sciences, The University of Chicago, Chicago, USA; [4]Section of Endocrinology, Diabetes, and Metabolism, Lewis Katz School of Medicine at Temple University, Philadelphia, USA; [5]Department of Medicine and Department of Public Health Sciences, The University of Chicago, Chicago, USA

ABSTRACT

BACKGROUND: Osteoporotic fracture prediction calculators are poorly utilized in primary care, leading to underdiagnosis and undertreatment of those at risk for fracture. The use of these calculators could be improved if predictions were automated using the electronic health record (EHR). However, this approach is not well validated in multi-ethnic populations, and it is not clear if the adjustments for race or ethnicity made by calculators are appropriate.

OBJECTIVE: To investigate EHR-generated fracture predictions in a multi-ethnic population.

DESIGN: Retrospective cohort study using data from the EHR.

SETTING: An urban, academic medical center in Philadelphia, PA.

PARTICIPANTS: 12,758 White, 7,844 Black, and 3,587 Hispanic patients seeking routine care from 2010 to 2018 with mean 3.8 years follow-up.

INTERVENTIONS: None.

MEASUREMENTS: FRAX and QFracture, two of the most used fracture prediction tools, were studied. Risk for major osteoporotic fracture (MOF) and hip fracture were calculated using data from the EHR at baseline and compared to the number of fractures that occurred during follow-up.

RESULTS: MOF rates varied from 3.2 per 1000 patient-years in Black men to 7.6 in White women. FRAX and QFracture had similar discrimination for MOF prediction (area under the curve, AUC, 0.69 vs. 0.70, p=0.08) and for hip fracture prediction (AUC 0.77 vs 0.79, p=0.21) and were similar by race or ethnicity. FRAX had superior calibration than QFracture (calibration-in-the-large for FRAX 0.97 versus QFracture 2.02). The adjustment factors used in MOF prediction were generally accurate in Black women, but underestimated risk in Black men, Hispanic women, and Hispanic men.

LIMITATIONS: Single center design.

CONCLUSIONS: Fracture predictions using only EHR inputs can discriminate between high and low risk patients, even in Black and Hispanic patients, and could help primary care physicians identify patients who need screening or treatment. However, further refinements to the calculators may better adjust for race-ethnicity.

KEY WORDS: fracture; osteoporosis; electronic medical record; race; ethnicity; black; hispanic

Received December 19, 2022
Accepted July 21, 2023
Published online September 15, 2023

J Gen Intern Med 38(16):3451–9
DOI: 10.1007/s11606-023-08347-5

INTRODUCTION

Osteoporosis is a systemic skeletal disease characterized by reduced bone strength and pre-disposition to fracture [1]. Osteoporosis leads to two million fractures per year in the United States alone, resulting in pain and disability, and 54 million people in the United States have low bone density or osteoporosis [2]. Despite this morbidity, osteoporosis is usually not diagnosed or treated and, even after a hip fracture, only 3.3% of patients receive osteoporosis therapy [3]. Many patients never receive bone density screening or treatment, despite a high risk of fracture. Fracture risk is multi-faceted and includes risk factors including age, gender, prior fracture, co-morbid conditions, medication use, etc [4]. There are several fracture risk calculators available, such as the Fracture Risk Assessment Tool (FRAX) and QFracture. These calculators provide the risk of major fracture based on an individual's risk factors and have been well validated, even without bone density testing [5].

Unfortunately, osteoporosis and fracture risk are often not considered in busy primary care practices, leading to large numbers of osteoporotic fractures that could have been prevented with appropriate diagnosis and treatment [1]. The electronic health record (EHR) provides a unique opportunity to automate the identification of patients who are at risk of fractures within a primary care provider's regular workflow. However, there is relatively little data about the ability of the

EHR to identify patients at risk of fracture using currently available risk calculators. FRAX was derived from prospective, international cohorts using demographics, body mass index or bone density, and questionnaires, while QFracture was derived in the U.K., largely from EHR data [6, 7]. While FRAX has been previously validated using an EHR, this was done only in a large health fund in Israel [8]. Given that FRAX and QFracture were developed and validated predominantly outside of the U.S., it is important to examine their performance in a U.S. population, where there is greater racial and ethnic diversity.

Both FRAX and QFracture adjust the predicted risk by race or ethnicity. The use of race and ethnicity in medical calculators has come under scrutiny. Indeed, many of the differences seen in medicine that are attributed to race are related to historical and current social inequities and institutional racism. However, osteoporosis and fragility fractures stand out from many other areas of medicine. The rate of fracture varies widely throughout the world [9]. In the United States, while there are still disparities in post-fracture care, both Black and Asian patients fracture less than White patients, not more [10, 11]. Studies examining bone structure suggest significant differences in elements of bone microarchitecture by race or ethnicity that are poorly captured by bone mineral density or other clinical variables [12, 13]. However, there are very few studies validating fracture calculators in diverse U.S. multi-ethnic cohorts and none, to our knowledge, using the EHR. One of the largest of these studies, using data from the Women's Health Initiative (WHI), found that predictions by FRAX in African American and Hispanic women were no better than chance in those aged 50-64 [14].

It is clear, therefore, that further study of fracture risk calculators is warranted, especially in racial and ethnic minorities in the U.S. If fracture calculators could be automated into U.S. EHRs, it would be helpful in both reducing the osteoporosis treatment gap overall and reducing racial disparities in osteoporosis screening and treatment, which lags behind in Black women[15–17]. Therefore, we sought to first validate the ability of EHR-derived FRAX and QFracture to identify patients with high risk of fracture and to determine whether the current adjustments for race/ethnicity accurately capture the differences in fracture risk. We used the EHR data from a primary care population in an urban, tertiary care medical center with large portions of Black and Hispanic patients.

METHODS

We used routine clinical data from the EHR of Temple University Hospital in Philadelphia, PA (Epic Systems; Verona, WI) obtained from encounters between October 1, 2010 and December 1, 2018. Subjects at least 50 years of age were included if they averaged at least 1 visit per year with a primary care physician (PCP), defined as either Internal Medicine, Family Medicine, or Gerontology departments. Subjects were required to have at least 2 full years of follow-up. Of these two years, the first year was used for baseline data collection for fracture risk factors and subsequent time was used for observation of the outcome (i.e. fractures). Furthermore, because the target clinician population for risk calculators is the primary care setting, we required common health maintenance measures as a proxy for subjects receiving their primary care at the institution. This furthermore helped to signify that some of the visits were routine (i.e. not problem visits), that health maintenance could be addressed, and subjects were willing participants in routine health maintenance.

For women, our inclusion criteria were at least one measurement of LDL and at least one diagnosis code for both mammogram (ICD-10 Z12.31, ICD-9 Z76.12) and vaccination (ICD-9 V03-V06, ICD-10 Z23). For men, our inclusion criteria were at least one measurement of both PSA and LDL and at least one diagnosis code for vaccination (ICD-9 V03-V06, ICD-10 Z23). Of note, the screening tests could have occurred at any time and the actual screen or test (e.g. mammogram) did not have to be complete in order to be included in the study. Subjects with missing demographic data (age, race, or gender) or body mass index (BMI) data were excluded, given the use of these variables in fracture risk calculators. We also excluded subjects with a prescription for an osteoporosis medication at the time of entry into the study. If patient was later put on a medication for osteoporosis after entry, they were censored at the time of the prescription but earlier data was used. The study was approved by the Temple University Institutional Review Board (IRB).

Determination of fractures The presence and characteristics of fractures were determined by the presence of physician-billing codes (see ICD Coding algorithms in Supplemental Table 1). Major osteoporotic fractures (MOFs, a composite of fractures of the wrist, humerus, hip, and vertebrae) and hip fractures were analyzed separately. As the determination of vertebral fractures were through physician diagnosis codes, only clinical vertebral fractures, not morphometric vertebral fractures, were determined. Fractures temporally associated within 30 days of trauma codes were excluded, and we continued to follow these subjects to observe for non-traumatic fracture. To prevent double counting, subjects with prior fracture at any individual site (e.g. wrist or lumbar spine) were not counted as having an incident fracture at the same site. A subset of 140 random subjects with fracture codes were examined with 86% accuracy (121/140) for identifying MOF based on chart review of imaging studies and physician notes. The most common reasons for erroneous codes were incorrect site (e.g. hand fracture, instead of wrist fracture) or pain at the site with a normal imaging study.

Fracture calculators Supplemental Table 2 lists diagnosis codes and medications used to identify inputs for FRAX and QFracture for the main analysis. Some variables, such

as parental hip fracture history or nursing home resident, were poorly captured using the EHR. Diagnoses for secondary osteoporosis, which is not precisely defined, were generated from examples given on the FRAX website and a publication by the creators of FRAX[6]. QFracture was calculated using the full algorithm as published in 2012 and available online[18]. The calculator allows for a customized risk from 1-10 years, which allowed for direct calculation of each subject's risk based on the number of years of follow-up in the study. QFracture allows for "Other Race" which was used for U.S. Hispanics since they are not otherwise an option. Of note, the adjustment factors for "Other Race" in QFracture are numerically similar to the adjustment factors for U.S. Hispanic in FRAX.

FRAX Multi-entry Desktop was used to calculate the FRAX 10-year probabilities of major osteoporotic fracture and hip fracture. Because FRAX requires the race to be checked as U.S. Whites, Black, Hispanic, and Asian, we only included White, Black, and Hispanic patients in our analysis due to the small numbers of Asian patients. We further excluded patients who were multi-racial or of unknown race. We also separately calculated FRAX risk for Black and Hispanic subjects as if they were White to examine the adjustment factors used by FRAX. Each subjects' risk was adjusted for their years of follow-up by multiplying by the number of years of follow up divided by 10, as previously described and validated [8, 19]. For example, if the 10-year risk was calculated as 10%, a patient with 3 years of follow up would be adjusted to 3% risk.

Statistical analysis Subjects were compared across race or ethnicity on demographic variables, comorbidities, and predicted fracture risk using pairwise comparisons (t-tests for continuous variables or Chi-square for categorical variables). Incident fracture rates were standardized to 1,000 person-years and compared by race-ethnicity. Area under the Receiver Operating Curves (AUC) was used to compare between FRAX and QFracture's ability to distinguish between high and low risk subjects (i.e. discrimination) using a nonparametric approach [20]. Sensitivity, specificity, positive and negative predictive value were also calculated for those in the top 1%, top 10%, and top 20% of risk for each calculator. Calibration was determined by sorting subjects into deciles (for MOF) or quintiles (for hip fracture) of risk and comparing predicted vs. observed risk in each group. We designated quantile 1 as the group with the lowest predicted risk and increasing quantiles represent groups with escalating predicted risk. Adjustment factors for race or ethnicity were calculated by dividing the predicted risk of non-White individuals by the predicted risk for that individual as if they were marked as White. While there was slight variation in the adjustment factor, we used the mean adjustment factor by race, sex, and fracture site (MOF vs. hip fracture) and compared to the actual, observed fracture ratio.

Statistical analyses were done in Stata 17.1 (College Station, TX). We also used Microsoft Excel for Microsoft 365 Build 16.0 (Seattle, WA) to generate some of the figures. To ensure that performance was not tied to specific codes used to define FRAX and QFracture, we conducted sensitivity analyses by evaluating QFracture and FRAX's performance when requiring 2 diagnosis codes, instead of 1, for each criteria and when using different definitions of prior fracture.

RESULTS

Demographics

Table 1 shows the characteristics of the subjects by race/ethnicity. The White patients were older, had a higher proportion of men, and had significantly fewer visits overall. Overall, the average BMI was in the obese range. Years of follow-up varied from 1 to 7 years and averaged nearly 4 years. Hispanic patients had the lowest rate of COPD. The Charlson comorbidity score was significantly higher in Hispanics and Black subjects compared to Whites. Fracture rates, by race and sex, are shown in Figure 1. As expected, White women and men had the highest fracture rates (7.7 and 4.9 per 1,000 person-years, respectively), which was significantly higher than Black women and men (3.9 and 3.2 per 1,000 person-years, respectively). Hispanic women and men had intermediate fracture rates (6.3 and 3.9 per 1,000 person-years, respectively), which did not differ significantly from Whites. The hip fracture rates followed the same general trends as MOF, but due to the smaller numbers of events, only Black women had significantly lower hip fracture rates than White women.

Discrimination, Area Under the Curve, Sensitivity, and Specificity

Discrimination, or the ability to rank subjects according to risk, was not significantly different between QFracture and FRAX. For major osteoporotic fracture, the area under the curve (AUC) was 0.69 (95% CI 0.67-0.71) for FRAX and 0.70 (95% CI 0.68-0.73) for QFracture (p=0.08 for difference). For hip fracture the discrimination was overall better than MOF with the AUC of 0.77 (95% CI 0.72-0.83) for FRAX and 0.79 (0.74-0.84) for QFracture (p=0.21 for difference) [see Figure 2A and Supplemental Figure 1A]. Using the top 1%, 10%, or top 20% as cutoffs for both tools yielded similar sensitivities, specificities, and positive and negative predictive values (Supplemental Table 3).

Discrimination, as assessed by the AUC, did not significantly differ between races as shown in Figure 2 and Supplemental Figure 1. Discrimination did not significantly differ by age group.

Table 1 Demographics, Follow-Up Time and 10-Year Fracture Risks

	White (12,758)	Black (7,844)	Hispanic (3,587)
Age	64.5 ± 9.9	61.2 ± 9.3	60.2 ± 8.7
Male	53.1%	35.1%	44.9%
BMI	30.3 ± 6.4	32.3 ± 7.6	30.9 ± 6.4
Average visits per year	4.8 ± 4.1	7.2 ± 6.2	7.0 ± 5.2
Average years of follow-up	3.7 ± 1.8*	3.9 ± 1.8	3.7 ± 1.8*
Prior fracture (any)	5.0% *	4.0%	4.9%*
Chronic obstructive pulmonary disease	9.2%	10.3%	6.3%
Steroids, 3 months of use	0.5%*	0.9%	0.5%*
Rheumatoid arthritis	1.3%*	1.6%*	1.6%*
Secondary osteoporosis	18.6%	22.1%	20.2%
Charlson Comorbidity Score	0.6 ± 1.0	1.1 ± 1.6	0.9 ± 1.4
10-year QFracture Risk of MOF	4.7% ± 5.4%	2.2 % ± 2.5%	2.4% ± 2.6%
10-year FRAX Risk of MOF	7.5% ± 5.0%	3.0% ± 2.1%	3.5% ± 2.2%
10-year QFracture Risk of Hip Fracture	2.3% ± 4.7%	0.6% ± 2.0%	0.9% ± 2.0%
10-year FRAX Risk of Hip Fracture	1.9% ± 3.0%	0.6% ± 1.0%*	0.6% ± 1.1%*

Values marked with astericks (*) do not significantly differ

Calibration

Calibration, or the agreement between the predicted risk and observed outcomes, was superior for FRAX compared to QFracture for MOF (calibration-in-the-large, or observed to predicted ratio, for FRAX 0.97 vs. QFracture 2.02). For QFracture, prediction was lower than observed for all risk groups though this was less pronounced in the highest risk groups (Table 2). For FRAX, observed-to-predicted was close to 1 for most risk groups though there was some overestimation of risk in the highest risk group (See Table 2, Supplemental Figure 2). For hip fracture, there was overestimation of risk in the highest risk groups for both QFracture and FRAX (See Table 2, Supplemental Figure 3).

Supplemental Table 4 shows the performance of the calculators in different age groups. For MOF, there was underestimation of risk in the youngest age groups for QFracture, which substantially improved in the older age groups. FRAX also showed underestimation in the 50-59, slight overestimation in 60-79, and then nearly perfect calibration in age 80+. For hip fracture, there was slight underestimation in age 50-59 for both tools (albeit very low event rate), good calibration for QFracture in age 60-69, and overestimation in the other age groups for both tools.

Adjustment Factors for Race or Ethnicity

Both risk calculators apply an adjustment factor for race or ethnicity to the score as calculated for Whites (rather than provide a separate calculator for each race). For example, if a White subject's fracture risk is calculated as 10% and the

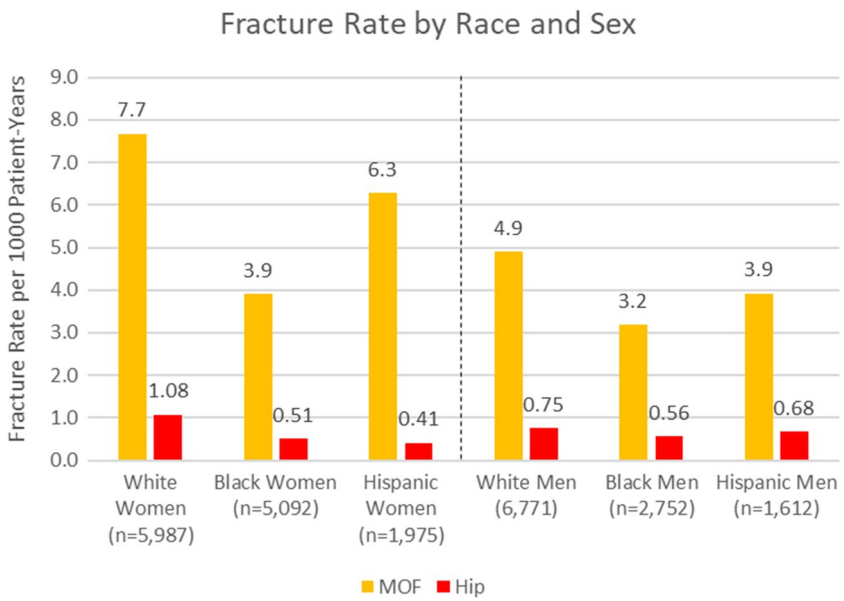

Figure 1 Fracture rates by race/ethnicity and sex.

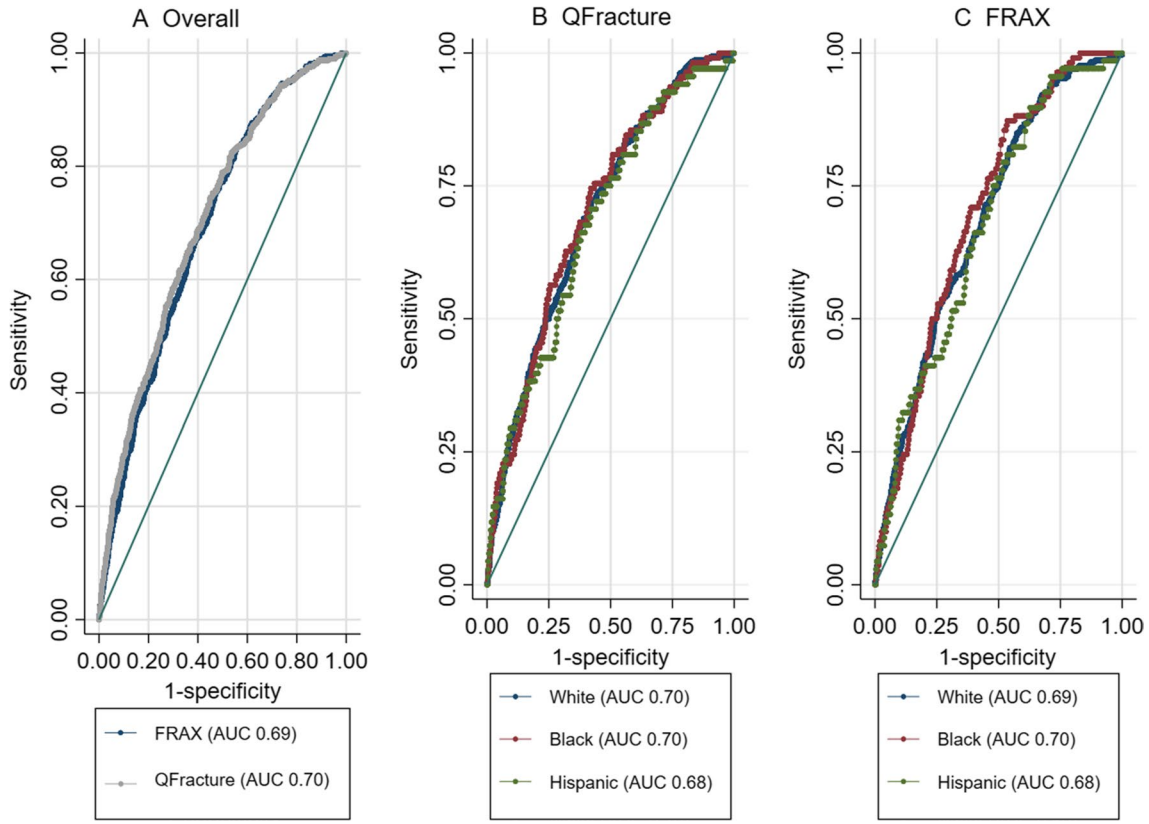

Figure 2 Receiver Operating Curves (ROCs) for major osteoporotic fracture A) in the overall cohort, B) By Race for QFracture, and C) By Race for FRAX.

adjustment factor for that calculator was 0.5, a Black patient with equivalent risk factors would be calculated at 5%. We examined these adjustment factors for race or ethnicity in FRAX and QFracture to the actual, observed fracture rates (i.e. fracture rate in Black or Hispanic divided by fracture rate in White; Fig. 3). For both calculators, MOF adjustment factors for Black women were accurate (observed: 0.51 versus adjustment factors: 0.45 and 0.48 for FRAX and QFracture, respectively); however, adjustment factors underestimated risk in Black men, Hispanic women, and Hispanic

Table 2 Calibration for QFracture and FRAX MOF and Hip Fracture Prediction by Risk Quantile

		QFracture				FRAX			
	Quantile	Number of people	Fracture Rate (%)	Mean Predicted Risk (%)	Observed to Predicted Ratio	Number of people	Fracture Rate (%)	Mean Predicted Risk (%)	Observed to Predicted Ratio
MOF	1	2419	0.25	0.07	3.5	2421	0.21	0.26	0.8
	2	2419	0.58	0.15	3.9	2418	0.62	0.51	1.2
	3	2419	0.83	0.23	3.6	2430	0.74	0.74	1.0
	4	2419	1.32	0.32	4.1	2411	1.37	0.98	1.4
	5	2419	1.28	0.43	3.0	2416	1.49	1.27	1.2
	6	2419	1.94	0.58	3.3	2421	1.94	1.61	1.2
	7	2419	2.03	0.80	2.5	2418	2.48	2.03	1.2
	8	2419	2.89	1.13	2.6	2419	2.56	2.58	1.0
	9	2419	2.77	1.75	1.6	2417	3.23	3.47	0.9
	10	2418	5.54	4.17	1.3	2418	4.80	6.56	0.7
Hip Fracture	1	4838	0.02	0.009	2.3	4896	0.04	0.02	1.8
	2	4838	0.06	0.03	1.8	4942	0.08	0.07	1.2
	3	4838	0.13	0.09	1.7	4694	0.13	0.15	0.9
	4	4838	0.33	0.22	1.5	4840	0.33	0.35	0.9
	5	4837	0.75	1.34	0.6	4817	0.79	1.67	0.5

MOF, Major osteoporotic fracture

Quantile 1 is the group with the lowest predicted fracture risk, and increasing quantiles represent increasing predicted fracture risk.

men (observed: 0.65-0.82 versus adjustment factors: 0.42-0.64). For hip fracture, adjustment factor for Black women for FRAX was accurate (observed: 0.47 versus adjustment factor: 0.45), but there was underestimation of hip fracture in Black and Hispanic men (observed: 0.75-0.90 versus adjustment factors: 0.42-0.56). For QFracture, adjustment factor for Black women substantially underestimated risk (observed: 0.47 versus adjustment factor: 0.10), though was relatively accurate for the other groups.

Sensitivity Analysis

We re-calculated QFracture and FRAX while requiring 2 diagnosis codes for each criteria, instead of 1, and there was no significant differences in area under the curve for either MOF or hip fracture prediction (all AUCs within 0.01 of the original analysis). We also examined FRAX while limiting prior fractures to prior non-traumatic, prior MOFs, or with exclusion of hand and foot fractures--again, there was no significant difference in performance. We also calculated QFracture and FRAX in subjects using a 2 year lookback to determine the calculator criteria, instead of just 1 year (in n=21,998 subjects with at least 2 years of lookback and 1 year of follow-up data), and performance was nearly identical (AUC for MOF: 0.68 for FRAX vs. 0.69 QFracture; for hip fracture, 0.78 for FRAX, 0.80 for QFracture, p=0.14 and 0.34, respectively).

DISCUSSION

This is the first study to demonstrate that FRAX and QFracture derived from the EHR of a large urban medical center can accurately discriminate between high and low fracture risk subjects receiving primary care in the U.S. Furthermore, our study includes Black and Hispanic subjects and men which have usually been under-represented in osteoporosis risk investigations. This is a crucial finding that supports routine fracture risk assessment for primary care physicians without any manual effort. These findings also support the use of the EHR in population health management approaches to facilitate care of patients at high risk of fracture. These strategies could greatly increase targeted osteoporosis screening and treatment and improve patient outcomes. However, we found that the adjustments for race built in FRAX and QFracture did not provide accurate prediction in our population especially in Hispanic subjects and men. This suggests that integration of fracture calculators into the EHR would require local and race-specific calibration.

The discrimination for the two fracture calculators was reasonable and largely in-line with prior studies [5]. Since we used the EHR to capture all inputs for both calculators, this approach could allow for automation of fracture risk prediction. A National Institute of Health Pathways to Prevention Workshop wrote that "Inadequate time is most likely the biggest contributing factor to the lack of attention to osteoporosis among primary care physicians," [1]. It is uncertain how many primary care physicians routinely use fracture risk calculators given the time it requires, and our study validates the use of EHR-generated fracture prediction. Since QFracture is freely available, it may be logistically easier to include in an EHR system, though it requires more significant re-calibration than FRAX to account for underestimation of risk.

Automated fracture risk calculation could improve osteoporosis screening in several ways. For example, fracture risk calculations could be directly integrated into the EHR—allowing primary care physicians to be alerted at the point of care for patients at high risk of fracture. However, the exact approach would need to be tested, as many EHR alerts are simply ignored due to "alert fatigue," [21]. There may be alternative effective approaches—for example, one study demonstrated that physicians responded to dashboards that compare their statin prescribing rates to their peers, and a similar system for bone density screening rates could be effective [22]. Healthcare system approaches, such as targeted outreach to patients at high risk, may also be appropriate.

Our study is one of the largest validations of FRAX and QFracture in U.S. minority populations. FRAX was derived

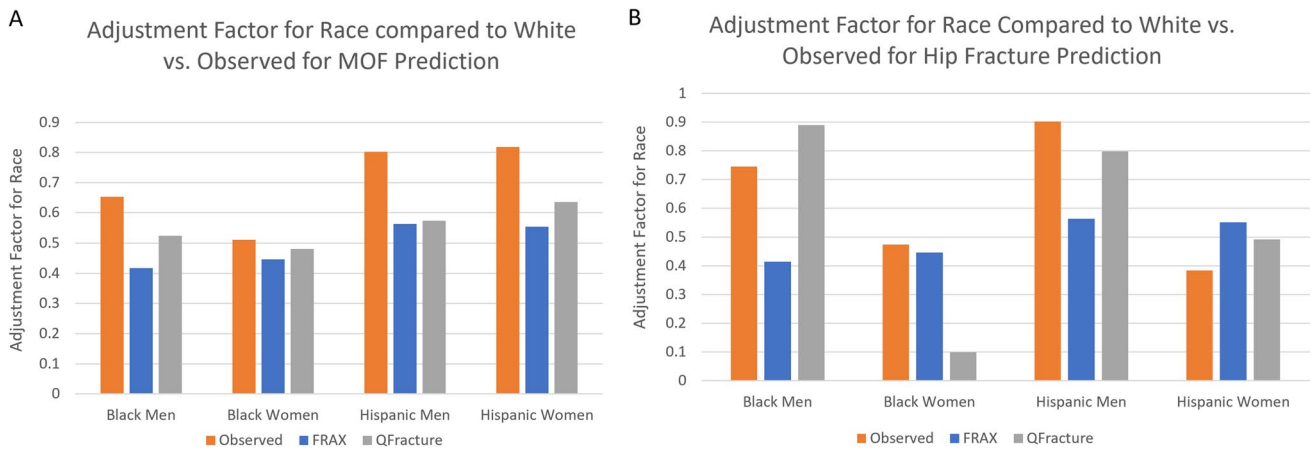

Figure 3 Adjustment Factors for Race vs. Observed for A) Major Osteoporotic Fracture and B) Hip Fracture.

from prospective cohort studies across North America, Europe, and Asia, but there was only 1 U.S. site (Rochester, MN) and few U.S. minorities were part of the original derivation [23]. FRAX adjusts its estimates in Black and Hispanic patients based on fracture rates in the U.S., though what data are used for the adjustments is not clear. The inclusion of race or ethnicity in medical algorithms is controversial, and our study addresses the need to evaluate fracture algorithm performance in multiethnic populations [24]. In prior studies of the Women's Health Initiative (WHI), FRAX performed poorly in Black and Hispanic women, especially in younger, premenopausal women [14, 25, 26]. In our study, FRAX and QFracture had good discrimination in Black and Hispanic patients and was comparable to that of White patients. The study population of WHI is healthier than our clinical population, which might explain the differences between our results and WHI. In particular, the WHI excluded patients with substantial comorbidities or contraindications to estrogen use, such as breast cancer, acute myocardial infarction, stroke, severe hypertension, dementia, and alcoholism [27].

We did find underestimation of risk in younger patients aged 50-59 though discrimination did not differ by age group. Validation studies of fracture calculators in this age group have had conflicting findings. In the WHI, discrimination of FRAX and Garvan were poor for younger premenopausal women, but the calibration for FRAX appeared to be good [14]. In other U.S. and European cohorts, fracture calculators have shown both good and poor calibration [8, 28–31]. This age group may be challenging for fracture calculators to predict due to low fracture rates at the population level. Studies of clinical or "real world" populations, such as ours, may be enriched with younger patients with many risk factors for fracture, including those that are not accounted for by risk calculators.

Our study demonstrated that the adjustment factors for Hispanic patients generally underestimated risk. Of note, QFracture, which was derived in the United Kingdom, does not have a specific designation for Hispanics so "Other Race" was used. In the United States, observed fracture rates in Hispanic subjects have not been consistent across studies. In the WHI, Hispanic women fractured at nearly half the rate of White women (which is similar to FRAX's adjustment), while in the National Osteoporosis Risk Assessment (NORA) study, Hispanic women fractured at the same rate as White women [10, 11]. It is not clear if the difference in results was related to differences in the study population, the use of estrogen in WHI, intra-ethnicity variation, or another reason. In our study, like NORA, fractures rates in Hispanic women were similar to White women. Further work is needed to accurately define the fracture rates of Hispanic people in the United States and better understand the reasons for differences between studies.

We also found that the fracture rates in minority men were higher than predicted. There is far less data about fragility fractures in minority men than women. The large, prospective studies that examined fracture risk in racial-ethnic minorities (WHI and NORA) were done exclusively in women. While there have been major studies that have improved our knowledge base about fracture risk in men in general, such as the MrOS studies, even these studies include few Black or Hispanic men (244 Black men, 127 Hispanic men) [32]. Given the lack of data, it is not clear if adjustments in FRAX are extrapolated from differences in rates in women. In contrast, QFracture's derivation did include over 25,000 Black subjects, of which about 50% of their overall population was male. It is of note then that the adjustments in Black men made by QFracture for MOF and hip fracture were closer to the observed rates than FRAX.

Underestimation of fracture risk in Black and Hispanic patients could also be related to differences in health status of the minority populations, which may not be adequately captured by the fracture calculators. The Black and Hispanic subjects had different characteristics, including age, gender, and higher comorbidity burdens, as measured by the Charlson Comorbidity Score. However, the fracture calculators already contain age, gender, and co-morbid conditions and should account for their effects on fracture risk. Therefore, our analyses of calculator performance by race-ethnicity should not have substantially been impacted by the differences in characteristics. For example, the adjustment factors for Black women for FRAX performed very well despite this difference. It may be that adjustment factors for Black women are particularly suitable given there are more prospective fracture studies in Black women than Hispanic subjects or Black men, such as the Study of Osteoporotic fractures [33]. Our results also suggest that some of the risk factors in the Charlson Comorbidity Score not already in FRAX or QFracture could improve fracture prediction. Other variables in the EHR, including healthcare utilization or lab values, may also improve predictive ability, and an EHR-optimized fracture tool may be worthy of investigation.

There are limitations to our study. First, this was a single center study, and our results may not be generalizable to other medical centers, especially those in non-academic or rural settings. Second, we did not use bone mineral density for FRAX prediction, which is known to improve FRAX's performance [5]. Only a small minority of subjects in the study had BMD testing available. However, this is also a strength of our study: if integrated into an EHR, automated fracture risk calculation could be used to identify those who need BMD testing, rather than requiring BMD testing already be done. Third, we used the EHR to identify fractures and all risk factors. Some risk factors were approximated or poorly captured. However, while this could be seen as a 'limitation', the fact that the calculators still performed reasonably well is evidence that it was likely not a substantial limitation. Also, because all fractures were obtained from the EHR, some fractures may have been missed, and this may have been

more of an issue for fractures that affect ambulation, such as a hip fracture, and impede the ability to seek outpatient care. A possible final limitation is that we chose a primary care population based on visits and health screenings with long-term follow up. We chose this population both because they would be the best candidates for automated fracture screening and because they reliably followed up, which allowed us to accurately capture fractures that occurred in follow up. It is possible these criteria screened for healthier patients (i.e. "healthy user bias") or reflect implicit racial bias related to healthcare access disparities [34, 35]. However, the institution serves a diverse population, and minorities still comprised of nearly half the study population. Further, the criteria may screen out patients who did not undergo vaccine administration but would have accepted osteoporosis screening or treatment. The PSA screening criteria in men may also have excluded otherwise appropriate subjects since PSA testing recommendations have changed over the years, and testing may have been withheld even in patients otherwise receiving health maintenance [36, 37].

Our study has notable strengths. To our knowledge, it is the first study to demonstrate that the EHR can be used to identify patients who are at risk of fracture and need a more aggressive diagnostic and therapeutic approach. In addition, this is one of the largest evaluations of FRAX and QFracture in U.S. Black and Hispanic subjects, including relatively large numbers of men. We demonstrated that these tools are valid using EHR data, which supports the calculation of automated fracture risk calculation in EHR systems. We conducted several sensitivity analyses that demonstrated the results were not related to specific diagnosis codes or the lookback time used in derivation of QFracture or FRAX. We also used the genuine algorithm for FRAX rather than estimating FRAX risk as has been done previously [8].

Overall, our study validates the use of EHR-generated fracture predictions in the United States for the first time and adds to the evidence base for the use of race or ethnicity in fracture risk calculators. Our study demonstrates that EHR inputs allow automated fracture risk prediction that provides good discrimination over several years of follow-up without patient or provider effort. The calculators performed well in a clinical population with high rates of comorbid disease and with substantial racial-ethnic variation in fracture rates. While there were problems with calibration and these calculators could better account for race or ethnicity, overall, our study demonstrates that the inclusion of race improved fracture prediction and could help target those in need of osteoporosis screening or treatment.

Supplementary Information *The online version contains supplementary material available at https://doi.org/10.1007/s11606-023-08347-5.*

Corresponding Author: *Rajesh K. Jain, MD; Department of Medicine, Section of Endocrinology, Diabetes, and Metabolism, The University of Chicago, 5841 South Maryland Ave, MC 1027, Chicago, IL, 60637, USA (e-mail: rjain2@bsd.uchicago.edu).*

Funding *Amgen Foundation.*

Declarations

Disclosures *RJ, TV are investigators for Radius Health. MW, EP, AI, and EH have no disclosures.*

REFERENCES

1. **Siu A, Allore H, Brown D, Charles ST, Lohman M.** National Institutes of Health Pathways to Prevention Workshop: Research Gaps for Long-Term Drug Therapies for Osteoporotic Fracture Prevention. Ann Intern Med. 2019;171:51–57.
2. Bone Health & Osteoporosis Foundation [Internet]. [cited 2023 Apr 18]. General Facts. Available from: https://www.bonehealthandosteoporosis.org/preventing-fractures/general-facts/. Accessed 18 Apr 2023
3. **Desai RJ, Mahesri M, Abdia Y, Barberio J, Tong A, Zhang D, et al.** Association of osteoporosis medication use after hip fracture with prevention of subsequent nonvertebral fractures: an instrumental variable analysis. JAMA Netw Open. 2018;1(3):e180826–e180826.
4. **Chandran M.** Fracture risk assessment in clinical practice: why do it? What to do it with? J Clin Densitom. 2017;20(3):274–9.
5. **Beaudoin C, Moore L, Gagné M, Bessette L, Ste-Marie LG, Brown JP, et al.** Performance of predictive tools to identify individuals at risk of non-traumatic fracture: a systematic review, meta-analysis, and meta-regression. Osteoporos Int. 2019;30:721–740.
6. **Kanis JA, Johnell O, Odén A, Johansson H, McCloskey E.** FRAX™ and the assessment of fracture probability in men and women from the UK. Osteoporos Int. 2008;19(4):385–97.
7. **Hippisley-Cox J, Coupland C.** Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. Bmj. 2012;344:e3427.
8. **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.
9. **Kanis JA, Odén A, McCloskey EV, Johansson H, Wahl DA, Cooper C, et al.** A systematic review of hip fracture incidence and probability of fracture worldwide. Osteoporos Int. 2012;23(9):2239–56.
10. **Cauley JA, Wu L, Wampler NS, Barnhart JM, Allison M, Chen Z, et al.** Clinical Risk Factors for Fractures in Multi-Ethnic Women: The Women's Health Initiative. J Bone Miner Res. 2007;22(11):1816–26.
11. **Barrett-Connor E, Siris ES, Wehren LE, Miller PD, Abbott TA, Berger ML, et al.** Osteoporosis and fracture risk in women of different ethnic groups. J Bone Miner Res. 2005;20(2):185–94.
12. **Putman MS, Yu EW, Lee H, Neer RM, Schindler E, Taylor AP, et al.** Differences in skeletal microarchitecture and strength in African-American and white women. J Bone Miner Res. 2013;28(10):2177–85.
13. **Wang XF, Wang Q, Ghasem-Zadeh A, Evans A, McLeod C, Iuliano-Burns S, et al.** Differences in macro-and microarchitecture of the appendicular skeleton in young Chinese and white women. J Bone Miner Res. 2009;24(12):1946–52.
14. **Crandall C, Larson J, LaCroix A, Cauley J, LeBoff M, Li W, et al.** Predicting Fracture Risk in Younger Postmenopausal Women: Comparison of the Garvan and FRAX Risk Calculators in the Women's Health Initiative Study. J Gen Intern Med. 2019;34(2):235–42.
15. **Gillespie CW, Morin PE.** Trends and disparities in osteoporosis screening among women in the United States, 2008-2014. Am J Med. 2017;130(3):306–16.
16. **Curtis JR, McClure LA, Delzell E, Howard VJ, Orwoll E, Saag KG, et al.** Population-based fracture risk assessment and osteoporosis treatment disparities by race and gender. J Gen Intern Med. 2009;24(8):956–62.
17. **Fuggle NR, Curtis B, Clynes M, Zhang J, Ward K, Javaid MK, et al.** The treatment gap: the missed opportunities for osteoporosis therapy. Bone. 2021;144:115833.
18. opensource - Revision 5: /qfracture/c [Internet]. [cited 2022 May 3]. Available from: https://svn.clinrisk.co.uk/opensource/qfracture/c/. Accessed 3 May 2022

19. **Leslie WD, Majumdar SR, Morin SN, Lix LM, Johansson H, Oden A, et al.** FRAX for fracture prediction shorter and longer than 10 years: the Manitoba BMD registry. Osteoporos Int. 2017;28(9):2557–64.
20. **DeLong ER, DeLong DM, Clarke-Pearson DL.** Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;837–45.
21. **Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R.** Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak. 2017;17(1):1–9.
22. **Patel MS, Kurtzman GW, Kannan S, Small DS, Morris A, Honeywell S, et al.** Effect of an automated patient dashboard using active choice and peer comparison performance feedback to physicians on statin prescribing: the PRESCRIBE cluster randomized clinical trial. JAMA Netw Open. 2018;1(3):e180818–e180818.
23. **Kanis JA, Odén A, Johnell O, Johansson H, De Laet C, Brown J, et al.** 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):1033–46.
24. **Vyas DA, Eisenstein LG, Jones DS.** Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. Vol. 383, New England Journal of Medicine. Mass Medical Soc; 2020:383;874–882.
25. **Cauley J, LaCroix AZ, Wu C, Lewis B, Wactawski-Wende J, Masaki K, et al.** FRAX: does fracture prediction differ by race/ethnicity. J Bone Miner Res 2010;25 (Suppl 1):S320. Available at https://asbmr.onlinelibrary.wiley.com/doi/epdf/10.1002/jbmr.5650251304.
26. **Crandall CJ, Larson J, Cauley JA, Schousboe JT, LaCroix AZ, Robbins JA, et al.** Do Additional Clinical Risk Factors Improve the Performance of Fracture Risk Assessment Tool (FRAX) Among Postmenopausal Women? Findings From the Women's Health Initiative Observational Study and Clinical Trials. JBMR Plus. 2019;3(12):e10239.
27. **Anderson G, Cummings S, Freedman LS, Furberg C, Henderson M, Johnson SR, et al.** Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials. 1998;19(1):61–109.
28. **Collins GS, Mallett S, Altman DG.** Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores. Bmj. 2011;342:d3651.
29. **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):407–15.
30. **Azagra R, Zwart M, Encabo G, Aguyé A, Martin-Sánchez JC, Puchol-Ruiz N, et al.** Rationale of the Spanish FRAX model in decision-making for predicting osteoporotic fractures: an update of FRIDEX cohort of Spanish women. BMC Musculoskelet Disord. 2016;17(1):1–10.
31. **Cordomí CT, Del Río LM, Di Gregorio S, Casas L, Estrada MD, Kotzeva A, et al**. Validation of the FRAX predictive model for major osteoporotic fracture in a historical cohort of Spanish women. J Clin Densitom. 2013;16(2):231–7.
32. **Orwoll E, Blank JB, Barrett-Connor E, Cauley J, Cummings S, Ensrud K, et al.** Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study—a large observational study of the determinants of fracture in older men. Contemp Clin Trials. 2005;26(5):569–85.
33. **Cauley JA, Lui LY, Ensrud KE, Zmuda JM, Stone KL, Hochberg MC, et al.** Bone mineral density and the risk of incident nonspinal fractures in black and white women. Jama. 2005;293(17):2102–8.
34. **Lees KA, Wortley PM, Coughlin SS.** Comparison of racial/ethnic disparities in adult immunization and cancer screening. Am J Prev Med. 2005;29(5):404–11.
35. **Ahmed AT, Welch BT, Brinjikji W, Farah WH, Henrichsen TL, Murad MH, et al.** Racial disparities in screening mammography in the United States: a systematic review and meta-analysis. J Am Coll Radiol. 2017;14(2):157–65.
36 **Moyer VA**. Force* UPST creening for prostate cancer: US Preventive Services Task Force recommendation statement. Ann Intern Med. 2012;157(2):120–34.
37. **Grossman DC, Curry SJ, Owens DK, Bibbins-Domingo K, Caughey AB, Davidson KW, et al.** Screening for prostate cancer: US Preventive Services Task Force recommendation statement. Jama. 2018;319(18):1901–13.