



# A clinical prediction model for 10-year risk of self-reported osteoporosis diagnosis in pre- and perimenopausal women

Jacob Zeitlin<sup>1,2</sup> · Michael K. Parides<sup>3</sup> · Joseph M. Lane<sup>2,4</sup> · Linda A. Russell<sup>4</sup> · Kyle N. Kunze<sup>2</sup>

Received: 19 December 2022 / Accepted: 29 May 2023 / Published online: 5 June 2023  
© International Osteoporosis Foundation and Bone Health and Osteoporosis Foundation 2023

## Abstract

**Summary** A machine learning model using clinical, laboratory, and imaging data was developed to predict 10-year risk of menopause-related osteoporosis. The resulting predictions, which are sensitive and specific, highlight distinct clinical risk profiles that can be used to identify patients most likely to be diagnosed with osteoporosis.

**Purpose** The aim of this study was to incorporate demographic, metabolic, and imaging risk factors into a model for long-term prediction of self-reported osteoporosis diagnosis.

**Methods** This was a secondary analysis of 1685 patients from the longitudinal Study of Women's Health Across the Nation using data collected between 1996 and 2008. Participants were pre- or perimenopausal women between 42 and 52 years of age. A machine learning model was trained using 14 baseline risk factors—age, height, weight, body mass index, waist circumference, race, menopausal status, maternal osteoporosis history, maternal spine fracture history, serum estradiol level, serum dehydroepiandrosterone level, serum thyroid-stimulating hormone level, total spine bone mineral density, and total hip bone mineral density. The self-reported outcome was whether a doctor or other provider had told participants they have osteoporosis or treated them for osteoporosis.

**Results** At 10-year follow-up, a clinical osteoporosis diagnosis was reported by 113 (6.7%) women. Area under the receiver operating characteristic curve of the model was 0.83 (95% confidence interval, 0.73–0.91) and Brier score was 0.054 (95% confidence interval, 0.035–0.074). Total spine bone mineral density, total hip bone mineral density, and age had the largest contributions to predicted risk. Using two discrimination thresholds, stratification into low, medium, and high risk, respectively, was associated with likelihood ratios of 0.23, 3.2, and 6.8. At the lower threshold, sensitivity was 0.81, and specificity was 0.82.

**Conclusion** The model developed in this analysis integrates clinical data, serum biomarker levels, and bone mineral densities to predict 10-year risk of osteoporosis with good performance.

**Keywords** Bone mineral density · DXA · Screening · Aging · Menopause · Estrogen

## Introduction

Osteoporosis may not have clinical manifestations prior to the onset of a fragility fracture, which poses challenges for early prevention and diagnosis [1, 2]. This is of clinical importance because increased bone fragility and osteoporotic fractures may be associated with debilitating consequences such as spinal cord compression [3–5], avascular necrosis [6, 7], need for surgical intervention, and increased mortality rates [8–11]. Pain, disability, and loss of independence are additional noteworthy complications that occur in the setting of osteoporosis and fragility fractures. Women who are undergoing or have recently undergone menopause are at particularly elevated risk of developing osteoporosis

✉ Jacob Zeitlin  
jhz4001@med.cornell.edu

<sup>1</sup> Weill Cornell Medical College, New York, NY, USA

<sup>2</sup> Department of Orthopaedic Surgery, Hospital for Special Surgery, New York, NY, USA

<sup>3</sup> Department of Biostatistics and Bioinformatics, Hospital for Special Surgery, New York, NY, USA

<sup>4</sup> Metabolic Bone Health Center, Hospital for Special Surgery, New York, NY, USA

[12]. However, several other factors have been implicated in the development of osteoporosis, and a comprehensive understanding of the nuanced interactions of such factors on the magnitude of risk for osteoporosis development has not been well studied [13, 14].

Previous work has demonstrated that differences in serum concentrations of hormones such as estradiol [15, 16], dehydroepiandrosterone (DHEAS) [17–19], and thyroid-stimulating hormone (TSH) [20–22] are associated with important differences in risk of osteoporosis or osteoporotic fractures. Recent reports in the literature, however, have suggested that menopause and its associated changes in steroid hormone levels may not make as significant a contribution to the pathophysiology of the disease as was previously thought [23–25]. In fact, periodically updated guidelines from the International Society of Clinical Densitometry have included an increasing number of indications for bone density screening in recent years [26]. This suggests that the etiology of the disease is multifactorial, as risk may depend on individual patient phenotypes, and that the complete array of factors contributing to its pathogenesis is not fully known [27–30].

Existing prediction models pertaining to this patient demographic, such as the Fracture Risk Algorithm (FRAX) tool for predicting osteoporosis-related fracture risk in postmenopausal women and men aged 50 years and older [1], may serve as important clinical adjuncts in patients with osteoporosis. However, FRAX is not intended for premenopausal women or women younger than 50, two additional patient populations at risk of developing osteoporosis [31, 32]. In addition, the goal of FRAX is to predict fractures; however, a diagnosis of osteoporosis based on low BMD can guide clinical management even in the absence of a fracture. Moreover, patient awareness of their diagnosis likely plays a role in fracture risk and other clinical outcomes; however, there has been little investigation into factors influencing patient awareness. While several studies have demonstrated that machine learning models can outperform conventional clinical tools at predicting osteoporosis in postmenopausal women [33–36], the included populations have been limited to respondents of the Korea National Health and Nutrition Examination Surveys and a single medical center in northern Taiwan, thereby reducing the generalizability of their findings. Although fracture risk is an important metric when considering osteoporosis management, the likelihood of receiving a clinical diagnosis is a clinically relevant outcome that is also used to guide treatment decisions.

There is limited research on the prediction of long-term risk of osteoporosis in women younger than 50, premenopausal women, and perimenopausal women, especially in the USA. While several known risk factors for osteoporosis exist in clinical practice, a combined predictive model incorporating these factors to provide an individualized assessment of long-term osteoporosis risk does not exist in a Western

population. The extent to which these risk factors influence patient awareness of a diagnosis is also unclear. Treatment is generally recommended in postmenopausal women who have a bone density T-score of  $-2.5$  or less, a history of spine or hip fracture, or a FRAX score indicating increased fracture risk [37]. Therefore, a clinical model capable of predicting those outcomes at an earlier stage in the course of the disease may be of clinical benefit.

Given this important gap in the literature and its clinical implications, the purpose of the current study was to perform a predictive analysis integrating a comprehensive profile of risk factors in order to better understand osteoporosis risk. The authors hypothesized that (1) the integrated prediction model would demonstrate good to excellent risk discrimination and (2) long-term risk of osteoporosis in pre- and perimenopausal women would be predicted by clinical risk factors, serum biomarker levels, and bone mineral density (BMD) measurements.

## Methods

### Study design

This is a secondary analysis of 3302 patients using baseline and 10-year follow-up data from the longitudinal Study of Women's Health Across the Nation (SWAN) [38]. Between 1996 and 1997, participants joined SWAN through seven designated research centers throughout the USA—Ann Arbor, MI; Boston, MA; Chicago, IL; Alameda and Contra Costa County, CA; Los Angeles, CA; Jersey City, NJ; and Pittsburgh, PA. At the time of enrollment, women were pre- or perimenopausal with an intact uterus and at least one ovary, not taking hormones, and between 42 and 52 years of age. Premenopausal women were defined as those having no change in bleeding patterns; perimenopausal women were defined as those with a change in length of the bleeding or inter-bleeding interval with at least one occurrence of bleeding in the past 3 months. At each visit, data was recorded via physical measures, fasting morning blood draw, and both interview- and self-administered questionnaires.

BMD of the posterior-anterior lumbar spine and total hip was measured by dual x-ray absorptiometry (DXA) using either a Hologic QDR 2000 densitometer (Hologic Inc., Waltham, MA) or a Hologic QDR 4500A [39]. All centers employed a standard quality control program that included daily measurement of a Hologic anthropomorphic spine phantom at each site, cross-site calibration with a single anthropomorphic spine phantom, visual review of every scan image by a local site investigator experienced in bone densitometry, and random review of 5% of scans plus all problem scans by Synarc, Inc.

(Waltham, MA). Measurements of the local spine phantoms and the circulating spine phantom were analyzed by Synarc and used to adjust DXA measurements for minor temporal or geographic variations. Using standard reference values for the total hip, these measurements were then converted to T-scores [40]. Determination of the presence of an osteoporosis diagnosis was based on the participants' answers to the question, "Has a doctor, nurse practitioner or other health care provider told you that you had osteoporosis (brittle or thinning bones) or treated you for osteoporosis?"

Each longitudinal follow-up wave included participants from the initial wave. For this analysis, a total of 1,617 participants had incomplete baseline data or did not have 10-year follow-up and were excluded. Demographic data and values of risk factors for the remaining 1,685 participants are shown in Table 1. Additionally, this data is reported for the 1617 who were excluded from the study and compared with that of those who were included (Online Table 1 in the Supplementary Information).

## Statistical analysis

Of the 1685 women who met the inclusion criteria, 337 (20%) were randomly selected as a holdout set for model evaluation, while the other 1348 (80%) were used to train the predictive model. A generalized additive model with pairwise interactions (GA<sup>2</sup>M) was implemented in Python using the *Explainable Boosting Machine*, a visually interpretable machine learning model from the open-source *InterpretML* package [41, 42]. The primary outcome of interest was whether participants reported they had been told they had a diagnosis of osteoporosis by a doctor or other healthcare provider at 10-year follow-up. For comparison, we also ran a logistic regression.

Based on previous literature, we included 14 risk factors available in the SWAN dataset at baseline: age, height, weight, body mass index (BMI), waist circumference, race, menopausal status, maternal osteoporosis history, maternal spine fracture history, serum estradiol level, serum DHEAS level, serum TSH level, total spine BMD, and total hip

**Table 1** Values of each of the demographic characteristics and risk factors included in the model for those not diagnosed and those diagnosed with osteoporosis at 10-year follow-up. Data are expressed as mean  $\pm$  SD unless otherwise noted. *P*-values were calculated via Student's *t*-test, Pearson's chi-squared test, or Fisher's exact test with  $\alpha$  set to 0.05

	Not diagnosed ( <i>n</i> = 1572)	Diagnosed ( <i>n</i> = 113)	<i>P</i>
Age (years)	45.8 $\pm$ 2.7	46.9 $\pm$ 2.8	<0.0001
Height (cm)	162.4 $\pm$ 6.5	160.6 $\pm$ 6.5	0.0045
Weight (kg)	72.6 $\pm$ 19.1	65.0 $\pm$ 18.4	<0.0001
BMI (kg/m <sup>2</sup> )	27.4 $\pm$ 6.7	25.2 $\pm$ 6.9	0.0006
Waist circumference (cm)	84.5 $\pm$ 14.8	78.5 $\pm$ 14.3	<0.0001
Race ( <i>n</i> [%])			0.1688
Black	405 (25.8)	19 (16.7)	
Chinese	181 (11.5)	12 (10.6)	
Japanese	205 (13.0)	17 (15.0)	
White	781 (49.7)	65 (57.5)	
Menopausal status ( <i>n</i> [%])			0.0113
Premenopausal	895 (56.9)	49 (43.4)	
Perimenopausal	663 (42.2)	62 (54.9)	
Unknown	14 (0.9)	2 (1.8)	
Maternal osteoporosis history ( <i>n</i> [%])			0.0146
Yes	221 (14.1)	27 (23.9)	
No	1221 (77.7)	76 (67.3)	
Unknown	130 (8.3)	10 (8.8)	
Maternal spine fracture history ( <i>n</i> [%])			0.0732
Yes	40 (2.5)	7 (6.2)	
No	1453 (92.4)	100 (88.5)	
Unknown	79 (5.0)	6 (5.3)	
Estradiol (pg/mL)	77.2 $\pm$ 90.3	87.1 $\pm$ 91.6	0.2620
DHEAS (ug/dL)	135.7 $\pm$ 81.9	132.6 $\pm$ 80.4	0.6911
TSH (uIU/mL)	2.5 $\pm$ 4.3	2.3 $\pm$ 1.9	0.6034
Total spine BMD (g/cm <sup>2</sup> )	1.08 $\pm$ 0.13	0.95 $\pm$ 0.13	<0.0001
Total hip BMD (g/cm <sup>2</sup> )	0.96 $\pm$ 0.14	0.85 $\pm$ 0.14	<0.0001
T-score < -2.5 ( <i>n</i> [%])	7 (0.4)	2 (1.8)	0.1178

BMD. To maximize validity, the GA<sup>2</sup>M also considered six pairwise interaction effects among the 14 risk factors [41]. Summary statistics were reported for all variables and compared via Student's *t*-test, Pearson's chi-squared test, or Fisher's exact test with  $\alpha$  set to 0.05 (Table 1).

Importance scores for each risk factor in the *Explainable Boosting Machine* were calculated as the “logit,” or log-odds ratio (OR), averaged across all observations in the training dataset (Table 2) [41, 43]. Therefore, a greater importance score corresponds to a larger contribution in predicting risk of diagnosis. Age- and BMI-adjusted odds ratios calculated using logistic regression are also reported in Table 2. Performance of the model was assessed using area under the receiver operating characteristic curve (AUC), Brier score, sensitivity, specificity, accuracy, and kappa. Univariate and pairwise partial dependence plots were created to visualize how, if at all, each risk factor (and pairwise interaction of factors) was related to osteoporosis risk. AUC values between variations of the model were compared via paired *t*-test.

Two discrimination thresholds for predicting a diagnosis—one at which Youden's J index was maximized and one at which kappa was maximized—were calculated using an exhaustive search and then assessed. Specifically, a sequential search algorithm was implemented across thresholds between 0 and 1 to identify the threshold values at which

both of these predictive performance metrics reached a global maximum. These thresholds were used to separate women in the evaluation dataset into three categories (low, medium, and high) based on predicted risk. Confidence intervals (CIs) for AUC and Brier score were calculated using bootstrapping by resampling the validation dataset 200 times with replacement. A calibration curve [44] was also plotted using Python.

## Results

### Risk factor analysis

At 10-year follow-up, 113 (6.7%) of the 1685 women reported that they had been diagnosed with osteoporosis. Compared to those who were not diagnosed, women who reported diagnosis were, on average, older by 1.1 years ( $P < 0.0001$ ), shorter in height by 1.8 cm ( $P = 0.0045$ ), weighed 7.6 kg less ( $P < 0.0001$ ), had a lower BMI by 2.2 kg/m<sup>2</sup> ( $P = 0.0006$ ), were more likely to be perimenopausal ( $P = 0.0113$ ), and were more likely to have a mother who was diagnosed with osteoporosis ( $P = 0.0146$ ). In addition, those who were diagnosed had lower total spine BMD measurements by 0.13 g/cm<sup>3</sup> ( $P < 0.0001$ ) and lower total hip BMD measurements by 0.11 g/cm<sup>3</sup> ( $P < 0.0001$ ) at baseline

**Table 2** Importance scores and age- and BMI-adjusted odds ratios of each predictor as computed by GA<sup>2</sup>M and logistic regression, respectively

Predictor	Importance score	Adjusted OR (95% CI)
Total spine BMD	0.197160	0.42 (0.35–0.51) per 0.1 g/cm <sup>2</sup> increase
Total hip BMD	0.161848	0.43 (0.35–0.53) per 0.1 g/cm <sup>2</sup> increase
Age	0.120628	
Menopausal status	0.095324	1.55 (1.05–2.29) for perimenopausal
Weight	0.091292	0.96 (0.92–0.99) per kg increase
Serum DHEAS level	0.072141	1.00 (0.97–1.02) per 10 ug/dL increase
Waist circumference	0.069442	0.95 (0.91–0.99) per cm increase
Serum estradiol level	0.058510	1.01 (0.99–1.03) per 10 pg/mL increase
Total spine BMD x weight	0.052425	1.01 (1.01–1.02) per 0.1 g/cm <sup>2</sup> and kg increase
Serum TSH level	0.051661	0.98 (0.91–1.06) per uIU/mL increase
Height	0.051332	0.72 (0.53–0.96) per 10 cm increase
BMI	0.051255	
Maternal osteoporosis history	0.050803	1.70 (1.07–2.70)
Total spine BMD x waist circumference	0.050120	1.02 (1.01–1.03) per 0.1 g/cm <sup>2</sup> and cm increase
Race	0.048825	0.78 (0.46–1.33) for Black 0.74 (0.39–1.38) for Chinese 0.85 (0.49–1.48) for Japanese 1.42 (0.96–2.09) for White
Total spine BMD x total hip BMD	0.030806	1.19 (1.10–1.29) per 0.1 g/cm <sup>2</sup> and 0.1 g/cm <sup>2</sup> increase
Age x total hip BMD	0.030492	
Total spine BMD x menopausal status	0.030293	0.93 (0.64–1.34) per 0.1 g/cm <sup>2</sup> for perimenopausal
Age x total spine BMD	0.028619	
Maternal spine fracture history	0.010480	1.98 (0.85–4.60)

(Table 1). Importance scores for each of the 14 risk factors and six pairwise interaction effects, as well as age- and BMI-adjusted odds ratios calculated using logistic regression, are shown in Table 2.

For each of the six pairwise interactions that were selected by the GA<sup>2</sup>M, a partial dependence plot was created to visualize how the given combination of variables was incorporated into the model as a predictor of 10-year osteoporosis risk (Online Figure 1 in the Supplementary Information). Each plot represents the change in predicted risk as a function of a pairwise combination of risk factors. Total spine BMD was a member of five of the six pairs; total hip BMD and age were both members of two of the six pairs.

### Osteoporosis risk prediction and stratification

On the 337 observations that were initially held out for evaluation of model performance, the GA<sup>2</sup>M predicted 10-year risk of osteoporosis with an AUC of 0.83 (95% CI, 0.73–0.91) (Fig. 1). Brier score for the model was 0.054 (95% CI, 0.035–0.074).

Using the same predictors and outcome measure, a logistic regression model had an AUC of 0.62 (95% CI, 0.49–0.76) and Brier score of 0.057 (95% CI, 0.039–0.079). This AUC was significantly less than that of the GA<sup>2</sup>M ( $P = 0.0115$ ). Furthermore, to assess the degree to which baseline BMD and age contributed to predictive performance alone, two additional models were developed and tested. In the first, BMD at both the hip and the spine, as well as age, were excluded from the predictors. This model had an AUC of 0.70 (95% CI, 0.60–0.80) and a Brier score of 0.057 (95% CI, 0.038–0.079). The AUC was not significantly less than that of the full GA<sup>2</sup>M ( $P = 0.082$ ). In the second, hip and spine BMD and age were the only predictors included. This model had an AUC of 0.82 (95% CI, 0.71–0.90) and a Brier

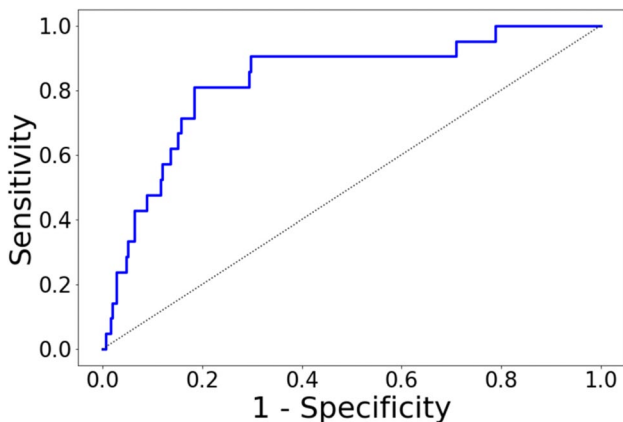


Fig. 1 Receiver operating characteristic curve

score of 0.054 (95% CI, 0.036–0.074). The AUC was not significantly less than that of the full GA<sup>2</sup>M ( $P = 0.822$ ).

Two discrimination thresholds for predicted probability of diagnosis ( $p_{dx}$ ) were calculated: the threshold at which Youden’s  $J$  index was maximized, which was determined to be  $p_{dx} = 0.063$ , and the threshold at which kappa was maximized, which was determined to be  $p_{dx} = 0.140$ . Using  $p_{dx} = 0.063$ , sensitivity was 0.81, specificity 0.82, accuracy 82%, and kappa 0.28. Using  $p_{dx} = 0.140$ , sensitivity was 0.43, specificity 0.94, accuracy 91%, and kappa 0.31.

These two thresholds established three risk-stratified categories of women according to predicted 10-year risk of osteoporosis (Fig. 2). In the lowest-risk category were those with  $p_{dx} \leq 0.063$  ( $n = 262$ ), 4 of whom (1.5%) were diagnosed with osteoporosis at 10-year follow-up (likelihood ratio [LR] of 0.23). In the medium-risk category were those with  $0.063 < p_{dx} \leq 0.140$  ( $n = 46$ ), 8 of whom (17.4%) were ultimately diagnosed (LR of 3.2). In the high-risk category were those with  $p_{dx} > 0.140$  ( $n = 29$ ), 9 of whom (31.0%) were ultimately diagnosed (LR of 6.8). Across the three risk categories, observed rates of osteoporosis diagnosis were determined to be significantly different via Pearson’s chi-squared test ( $P < 0.0001$ ).

### Calibration of prediction model

A calibration curve was plotted to visualize the relationship between predicted risk of diagnosis at baseline and actual diagnosis rates at follow-up in each of the ten risk-stratified deciles of the evaluation dataset (Fig. 3).  $R^2$  for the curve was 0.81 based on a least-squares regression, which yielded a line of best fit with a slope of 1.32 and y-intercept of  $-0.02$ . In four of the ten deciles, the GA<sup>2</sup>M overestimated

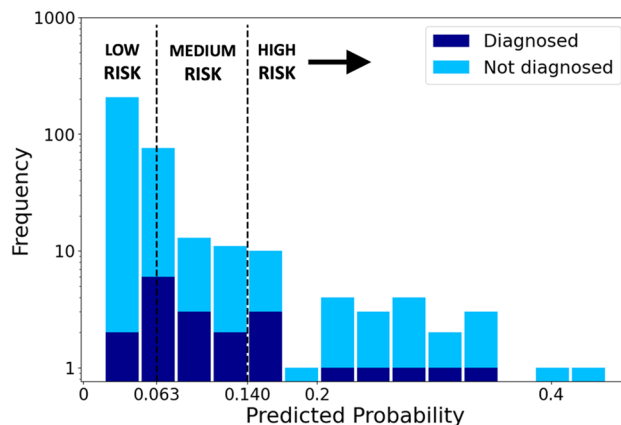
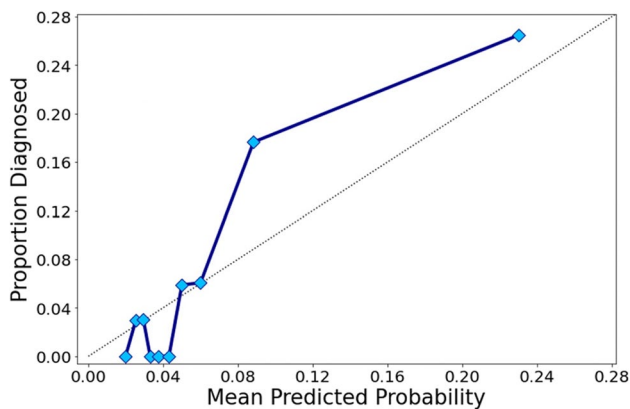


Fig. 2 Histogram showing the distribution of 10-year osteoporosis risk predictions made by the model at baseline, categorized by follow-up outcome. Frequencies are represented using a logarithmic scale on the y-axis. Dotted lines indicate stratifications of low-, medium-, and high-risk individuals





**Fig. 3** Calibration curve. Each plotted point represents 10 percent of the evaluation dataset ( $n = 337$ )

risk of osteoporosis. In the other six, the GA<sup>2</sup>M underestimated risk. Importantly, of the ten deciles, only one—the ninth—had a mean  $p_{dx}$  that was more than one standard error away from the observed rate.

A decision curve is shown in Online Figure 2 in the Supplementary Information.

## Discussion

The main finding of the current study was that a novel clinical prediction model integrating clinical risk factors, serum biomarker levels, and BMD measurements demonstrated good performance across several metrics for predicting the 10-year likelihood of having been told of a diagnosis of osteoporosis. Of the 14 risk factors included in this analysis, total spine BMD and total hip BMD were the two features with the highest predictive power in estimating 10-year risk of osteoporosis. However, several other factors were found to be of predictive importance, including those pertaining to demographics, physical measurements, and metabolic profiles. Finally, we implicated several discrimination thresholds that reliability classified patients into low-, medium- and high-risk categories for osteoporosis development. With further validation of this model, it may be possible for clinicians to diagnose osteoporosis earlier in the course of the disease and potentially reduce adverse events associated with osteoporosis by identifying those at greatest long-term risk and subsequently implementing prophylactic strategies.

As independent risk factors, total spine BMD (importance score = 0.197) and total hip BMD (importance score = 0.162) were the two top-ranked factors in predicting 10-year risk of osteoporosis (Table 2). These importance scores, expressed in logits, had corresponding age- and BMI-adjusted ORs of 0.42 (95% CI, 0.35–0.51) per 0.1 g/cm<sup>2</sup> increase and 0.43 (95% CI, 0.35–0.53) per 0.1 g/cm<sup>2</sup>

increase, respectively. This finding is in accordance with previously published analyses of longitudinal data from SWAN, which have demonstrated that during this 10-year observation period, BMD decreased by between 8.5 and 11.2% at both the lumbar spine and femoral neck [45–48]. The machine learning-based analysis we present in this paper extends the existing literature by quantifying the increase in risk of a clinical diagnosis of osteoporosis associated with bone loss in this population. Notably, while we found that spine BMD demonstrated a higher importance score than did hip BMD, only T-score from the femoral neck is included as input to the FRAX model, possibly limiting its predictive ability. Furthermore, while knowledge of a low baseline BMD would hold the potential to influence diagnosis awareness in this study population, this is not a likely source of bias in the current study, as the raw BMD measurements required cross-calibration and quality control for validation, and only 9 (0.5%) of the included participants had baseline T-scores in the osteoporotic range. This distinction is important to consider, as low baseline BMD may be a significant independent risk factor for osteoporosis and/or fractures irrespective of the rate of bone loss.

After BMD, the third- and fourth-ranked risk factors in our model were age (importance score = 0.121) and menopausal status (importance score = 0.095), corresponding to ORs of 1.32 and 1.25, respectively (Table 2). This finding corroborates previously published work suggesting that osteoporosis is associated with aging independently of its association with menopause [11, 23–26, 30, 49]. Additionally, age exhibited pairwise interaction effects with both spine BMD and hip BMD: therefore, in women with BMD measurements indicative of inferior bone quality, 10-year risk of osteoporosis was higher at a younger age and earlier menopausal status (Online Figures 1D and 1F in the Supplementary Information). Notably, while a model with BMD and age alone does not perform significantly worse than the model with all 14 predictors (BMD and age included), a separate model composed exclusively of the other 11 factors demonstrated good predictive performance. This result suggests that some combination of these 11 predictors might be clinically meaningful in predicting long-term diagnosis, and their marginal predictive value when incorporated into a model based on BMD and age alone would likely become more apparent in a larger cohort. However, further work is needed to evaluate the clinical importance of their contribution to risk assessment. Even in the absence of such work, these findings are nevertheless pertinent for clinical consideration, given the lack of indications for routine BMD screening in this age group. In addition, compared to DXA scans, it is less expensive and less invasive to obtain the remainder of these patient factors, as they require only taking a complete history, performing a physical examination, and collecting bloodwork.

This modeling technique highlighted other pairwise interaction effects as well. The interaction between spine BMD and weight had the highest importance score among all pairwise interactions at 0.052, corresponding to an OR of 1.13 (Online Figure 1A in the Supplementary Information). Overall, the combination of low spine BMD and low bodyweight at baseline conferred a particularly elevated 10-year risk of osteoporosis. However, this relationship was nonlinear with respect to bodyweight: for those at the lower extreme of spine BMD, a weight at the upper extreme was associated with a greater 10-year risk than an intermediate weight. This suggests that those at the extremes of bodyweight (i.e., both underweight and overweight) may be at elevated risk of osteoporosis. These findings might be explained by the effects of the increased release of adipokines in patients with obesity on bone quality [50, 51] or lifestyle factors (e.g., diet and exercise) that concomitantly contribute to the pathogenesis of osteoporosis [52]. Given that BMD is both one of the inputs and a key component of the outcome variable, there does exist some degree of inherent circularity within the model; however, we believe that this model still yields several important insights into osteoporosis risk.

While the current standard of care is to perform DXA screening in women aged 65 or older without risk factors for osteoporosis, the findings in the current study suggest several patient risk profiles in which it may be beneficial to initiate screening at an earlier age. Of the 113 women in this dataset who were ultimately diagnosed with osteoporosis, the mean age at baseline, when the DXA scans were captured, was 46.9 years. Even at that age, which is 18 years younger than guidelines call for screening to take place, the women exhibited significantly decreased BMD in both the spine (0.95 versus 1.08 g/cm<sup>2</sup>,  $P < 0.0001$ ) and the hip (0.85 versus 0.96 g/cm<sup>2</sup>,  $P < 0.0001$ ) compared to those who were not ultimately diagnosed. While some of those women had risk factors that would otherwise justify early screening, for those who did not, adherence to the current guidelines would have precluded a physician from taking advantage of data that could be used to diagnosis osteoporosis at an earlier age. By the age of 65, regardless of symptoms and risk factors, it may be too late to intervene effectively.

Using the two discrimination thresholds calculated to maximize Youden's  $J$  index and kappa, respectively  $p_{dx} = 0.063$  and  $p_{dx} = 0.140$ , this analysis established a systematic approach to classify women as low risk, medium risk, or high risk based on estimated probability of osteoporosis diagnosis within 10 years (Fig. 3). On the evaluation dataset, 1.5% of the low-risk, 17.4% of the medium-risk, and 31.0% of the high-risk women were ultimately diagnosed at follow-up. While stratification into the high-risk category yielded a sensitivity of only

0.43, it was highly specific, yielding a specificity of 0.94. Even at the more aggressive  $p_{dx}$  of 0.063, where sensitivity increased to 0.81, specificity only fell to 0.82. This suggests that women in the high- and medium-risk groups might be candidates for prophylactic strategies against osteoporosis. These strategies might include conservative management, such as educational programs or protective equipment; medical management, such as early administration of bisphosphonates or monoclonal antibodies; or even novel surgical techniques.

We believe the relatively small size of the high-risk group is acceptable, because in a clinical setting this highest-risk designation should be adequately specific in order to minimize unwarranted treatment of women in whom the development of osteoporosis is not certain or almost certain, as treatment has its own downsides. This group may be relatively smaller due to the relative paucity of women with the extensive combinations of risk factors seen in the high-risk group. Given the numerous unmeasured and immeasurable factors that influence risk, it may be difficult for this model to identify members of this high-risk group with high specificity, especially given the limited amount of available training data.

At an even finer level of granularity, a calibration curve divided the evaluation data into ten deciles based on predicted osteoporosis risk (Fig. 3). This curve demonstrated a strong correlation between predicted risk and observed risk, with  $R^2 = 0.81$ , an indicator of robust model validity. In the future, a regression model might be used to predict follow-up BMD as a continuous outcome, rather than the classification model presented here for predicting a binary outcome.

This analysis has limitations. First, for the purpose of this study, time to diagnosis was not considered in this model; however, this is an important factor in predicting fracture risk. Further work is required to validate this model on external datasets, as discussed above. However, the trained model has excellent internal validity, demonstrated by its performance on an evaluation dataset to which it was naïve during training. This is indicated by several performance metrics including AUC, Brier score, sensitivity, specificity, accuracy, kappa, and calibration curve analysis. Second, the model also designates as its outcome variable a self-reported osteoporosis diagnosis, which may fail to capture a population of patients with poor awareness of their condition. Further work is warranted to investigate underlying risk factors for and causes of lack of patient awareness of a diagnosis in the presence of low BMD. Despite its subjectivity, we contend that a strength of this self-reported questionnaire response is its consideration of the multifactorial nature of osteoporosis, which is a complex pathological process. Moreover, given stark differences in biological processes including aging and metabolism, it is reasonable to believe that in populations characterized by different demographic features, it would

be necessary to develop a separate model with a different set of risk factors. Third, because of the longitudinal design of SWAN, which lacks a randomization protocol, it is not possible to establish causative relationships between risk factors and predicted 10-year probability of osteoporosis. However, most of the predictors included in this model are unmodifiable and knowing how each relates to risk of osteoporosis would nevertheless be quite influential in motivating a clinical approach. Fourth, limiting the analysis to the 1685 women with 10-year follow-up does not allow for the consideration of validity for those without 10-year follow-up, who may exhibit different risk profiles. Fifth, this dataset does not include BMD for the femoral neck in isolation, which is incorporated into the FRAX model and commonly utilized when performing BMD assessments. In the absence of complete data, it is impossible for the model to account for this fact. Finally, the exclusion of approximately half of the original participants based on data availability may have biased results, as those who died might have had lower BMD given that low BMD is a predictor of mortality and loss to follow-up is unlikely to be completely random. However, no differences were observed in hip or spine BMD, nor in prevalence of having a baseline T-score below  $-2.5$ , between those ultimately included and those excluded (Online Table 1 in the Supplementary Information).

The model developed in this analysis integrates clinical data, serum biomarker levels, and bone mineral densities to predict 10-year risk of osteoporosis with good performance. This statistical modeling technique highlighted the importance of baseline BMD measurements, especially those of the spine, as long-term predictors of osteoporosis risk, and had superior discriminative ability to that of a logistic regression model. However, the current data demonstrates that BMD alone does not entirely determine risk, as numerous other factors including demographics and metabolic profiles contribute to predicted probabilities of diagnosis, both independently and as parts of pairwise interaction effects. Given the powerful discriminative ability of this model, physicians may reliably stratify pre- and perimenopausal women into categories based on predicted osteoporosis risk and determine potential target populations for the investigation of preventative strategies during earlier stages of the disease. Contextualized alongside the clinically important assessment of fracture risk, these risk assessments add novel and meaningful information regarding a distinct clinical endpoint. While further research and validation will be important to assess performance and generalizability outside of this cohort, this clinical prediction model effectively integrates multiple categories of risk factors, highlighting its potential to facilitate earlier identification for more robust prevention and possible interventions in the management of osteoporosis.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11657-023-01292-0>.

## Declarations

**Competing interests** JZ: None. MP: None. JL: CollPlant, Inc.: stock or stock options; Kuros: paid consultant; Lenoss Medical, Inc.: paid consultant; Merck: research support; Mesentech: paid consultant; Norvartis: research support; On Foundation: paid consultant; Radius Health, Inc.: paid consultant; Terumo BCT: paid consultant. LR: Arthritis Foundation: board or committee member. KK: Arthroscopy: editorial or governing board

## References

1. Cosman F, de Beur SJ, LeBoff MS et al (2014) Clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int* 25:2359–2381. <https://doi.org/10.1007/S00198-014-2794-2>
2. Siris ES, Adler R, Bilezikian J et al (2014) The clinical diagnosis of osteoporosis: a position statement from the National Bone Health Alliance Working Group. *Osteoporos Int* 25:1439–1443. <https://doi.org/10.1007/S00198-014-2655-Z>
3. Alexandru D, So W (2012) Evaluation and management of vertebral compression fractures. *Perm J* 16:46. <https://doi.org/10.7812/TPP/12-037>
4. Mccarthy J, Davis A (2016) Diagnosis and management of vertebral compression fractures. *Am Fam Physician* 94(1):44–50
5. Varacallo MA, Fox EJ (2014) Osteoporosis and its complications. *Med Clin North Am* 98:817–831. <https://doi.org/10.1016/J.MCNA.2014.03.007>
6. Dedrick DK, Mackenzie JR, Burney RE (1986) Complications of femoral neck fracture in young adults. *J Trauma* 26:932–937. <https://doi.org/10.1097/00005373-198610000-00013>
7. Nauth A, Creek AT, Zellar A et al (2017) Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. *The Lancet* 389:1519–1527. [https://doi.org/10.1016/S0140-6736\(17\)30066-1](https://doi.org/10.1016/S0140-6736(17)30066-1)
8. Center JR, Nguyen TV, Schneider D et al (1999) Mortality after all major types of osteoporotic fracture in men and women: an observational study. *The Lancet* 353:878–882. [https://doi.org/10.1016/S0140-6736\(98\)09075-8](https://doi.org/10.1016/S0140-6736(98)09075-8)
9. Fernandez MA, Achten J, Parsons N et al (2022) Cemented or uncemented hemiarthroplasty for intracapsular hip fracture. *N Engl J Med* 386:521–530. [https://doi.org/10.1056/NEJMOA2108337/SUPPL\\_FILE/NEJMOA2108337\\_DATA-SHARING.PDF](https://doi.org/10.1056/NEJMOA2108337/SUPPL_FILE/NEJMOA2108337_DATA-SHARING.PDF)
10. Guzon-Illescas O, Perez Fernandez E, Crespí Villarias N et al (2019) Mortality after osteoporotic hip fracture: incidence, trends, and associated factors. *J Orthop Surg Res* 14:1–9. <https://doi.org/10.1186/S13018-019-1226-6/FIGURES/3>
11. Lane JM, Russell L, Khan SN (2000) Osteoporosis. *Clin Orthop Relat Res* 372:139–150. <https://doi.org/10.1097/00003086-20003000-00016>
12. Karlamangla AS, Burnett-Bowie SAM, Crandall CJ (2018) Bone health during the menopause transition and beyond. *Obstet Gynecol Clin North Am* 45:695–708. <https://doi.org/10.1016/J.OGC.2018.07.012>
13. Weaver CM, Gordon CM, Janz KF et al (2016) The National Osteoporosis Foundation's position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. *Osteoporos Int* 27:1281–1386. <https://doi.org/10.1007/s00198-015-3440-3>



14. Leslie WD, Lix LM (2014) Comparison between various fracture risk assessment tools. *Osteoporos Int* 25:1–21. <https://doi.org/10.1007/S00198-013-2409-3/FIGURES/8>
15. Cauley JA (2015) Estrogen and bone health in men and women. *Steroids* 99:11–15. <https://doi.org/10.1016/J.STEROIDS.2014.12.010>
16. Larsson SC, Melhus H, Michaëlsson K (2018) Circulating serum 25-hydroxyvitamin d levels and bone mineral density: Mendelian randomization study. *J Bone Miner Res* 33:840–844. <https://doi.org/10.1002/JBMR.3389>
17. Ohlsson C, Nethander M, Kindmark A et al (2017) Low serum DHEAS predicts increased fracture risk in older men: the MrOS Sweden study. *J Bone Miner Res* 32:1607–1614. <https://doi.org/10.1002/JBMR.3123>
18. Quester J, Nethander M, Eriksson A, Ohlsson C (2022) Endogenous DHEAS is causally linked with lumbar spine bone mineral density and forearm fractures in women. *J Clin Endocrinol Metab* 107:e2080–e2086. <https://doi.org/10.1210/CLINEM/DGAB915>
19. Yokomoto-Umakoshi M, Umakoshi H, Iwahashi N et al (2021) Protective role of DHEAS in age-related changes in bone mass and fracture risk. *J Clin Endocrinol Metab* 106:e4580–e4592. <https://doi.org/10.1210/CLINEM/DGAB459>
20. Flynn RW, Bonellie SR, Jung RT et al (2010) Serum thyroid-stimulating hormone concentration and morbidity from cardiovascular disease and fractures in patients on long-term thyroxine therapy. *J Clin Endocrinol Metab* 95:186–193. <https://doi.org/10.1210/JC.2009-1625>
21. Mazziotti G, Porcelli T, Patelli I et al (2010) Serum TSH values and risk of vertebral fractures in euthyroid post-menopausal women with low bone mineral density. *Bone* 46:747–751. <https://doi.org/10.1016/J.BONE.2009.10.031>
22. Yang L, Wang H, Guo J et al (2020) Low normal TSH Levels and thyroid autoimmunity are associated with an increased risk of osteoporosis in euthyroid postmenopausal women. *Endocr Metab Immune Disord Drug Targets* 21:859–865. <https://doi.org/10.2174/1871530320666200810144506>
23. Khosla S, Melton LJ, Riggs BL (2011) The unitary model for estrogen deficiency and the pathogenesis of osteoporosis: is a revision needed? *J Bone Miner Res* 26:441–451. <https://doi.org/10.1002/JBMR.262>
24. Manolagas SC (2010) From estrogen-centric to aging and oxidative stress: a revised perspective of the pathogenesis of osteoporosis. *Endocr Rev* 31:266–300. <https://doi.org/10.1210/ER.2009-0024>
25. Recker R, Lappe J, Davies K, Heaney R (2000) Characterization of perimenopausal bone loss: a prospective study. *J Bone Miner Res* 15:1965–1973. <https://doi.org/10.1359/JBMR.2000.15.10.1965>
26. Shuhart CR, Yeap SS, Anderson PA et al (2019) Executive summary of the 2019 ISCD position development conference on monitoring treatment, DXA cross-calibration and least significant change, spinal cord injury, peri-prosthetic and orthopedic bone health, Transgender Medicine, and Pediatrics. *J Clin Densitom* 22:453–471. <https://doi.org/10.1016/J.JOCD.2019.07.001>
27. Clynes MA, Gregson CL, Bruyère O et al (2021) Osteosarcopenia: where osteoporosis and sarcopenia collide. *Rheumatology* 60:529–537. <https://doi.org/10.1093/RHEUMATOLOGY/KEAA755>
28. Diemar SS, Sejling AS, Eiken P et al (2019) An explorative literature review of the multifactorial causes of osteoporosis in epilepsy. *Epilepsy & Behav* 100:106511. <https://doi.org/10.1016/J.YEBEH.2019.106511>
29. Locantore P, del Gatto V, Gelli S et al (2020) The interplay between immune system and microbiota in osteoporosis. *Mediators Inflamm* 2020:3686749. <https://doi.org/10.1155/2020/3686749>
30. Saul D, Kosinsky RL (2021) Epigenetics of aging and aging-associated diseases. *Int J Mol Sci* 22(22):401. <https://doi.org/10.3390/IJMS22010401>
31. Callegari ET, Garland SM, Gorelik A, Wark JD (2017) Determinants of bone mineral density in young Australian women; results from the safe-D study. *Osteoporos Int* 28:2619–2631. <https://doi.org/10.1007/S00198-017-4100-6/TABLES/5>
32. Rozenberg S, Bruyère O, Bergmann P et al (2020) How to manage osteoporosis before the age of 50. *Maturitas* 138:14–25. <https://doi.org/10.1016/J.MATURITAS.2020.05.004>
33. Kim SK, Yoo TK, Oh E, Kim DW (2013) Osteoporosis risk prediction using machine learning and conventional methods. *Annu Int Conf IEEE Eng Med Biol Soc* 2013:188–191. <https://doi.org/10.1109/EMBC.2013.6609469>
34. Shim JG, Kim DW, Ryu KH et al (2020) Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. *Arch Osteoporos* 15:1–9. <https://doi.org/10.1007/S11657-020-00802-8/FIGURES/2>
35. Yang WYO, Lai CC, Tsou MT, Hwang LC (2021) Development of machine learning models for prediction of osteoporosis from clinical health examination data. *Int J Environ Res Public Health* 18:7635. <https://doi.org/10.3390/IJERPH18147635/S1>
36. Yoo TK, Kim SK, Kim DW et al (2013) Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. *Yonsei Med J* 54:1321–1330. <https://doi.org/10.3349/YMJ.2013.54.6.1321>
37. Black DM, Rosen CJ (2016) Clinical practice. Postmenopausal Osteoporosis. *N Engl J Med* 374:254–262. [https://doi.org/10.1056/NEJMCP1513724/SUPPL\\_FILE/NEJMCP1513724\\_DISCL\\_OSURES.PDF](https://doi.org/10.1056/NEJMCP1513724/SUPPL_FILE/NEJMCP1513724_DISCL_OSURES.PDF)
38. Santoro N, Sutton-Tyrrell K (2011) The SWAN song: study of women's health across the nation's recurring themes. *Obstet Gynecol Clin North Am* 38:417–423. <https://doi.org/10.1016/J.OGC.2011.05.001>
39. Finkelstein JS, Brockwell SE, Mehta V et al (2008) Bone mineral density changes during the menopause transition in a multiethnic cohort of women. *J Clin Endocrinol Metab* 93:861–868. <https://doi.org/10.1210/JC.2007-1876>
40. Looker AC, Johnston CC, Wahner HW et al (1995) Prevalence of low femoral bone density in older U.S. women from NHANES III. *J Bone Miner Res* 10:796–802. <https://doi.org/10.1002/JBMR.5650100517>
41. Lou Y, Caruana R, Gehrke J, Hooker G (2013) Accurate intelligible models with pairwise interactions. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, NY, USA, pp 623–631
42. Nori H, Jenkins S, Koch P, Caruana R (2019) InterpretML: a unified framework for machine learning interpretability. *ArXiv* 1909.09223. <https://doi.org/10.48550/arxiv.1909.09223>
43. McCullagh P, Nelder JA (2019) An outline of generalized linear models. *Generalized Linear Models* 2:21–47. <https://doi.org/10.1201/9780203753736>
44. Steyerberg EW (2009) Clinical prediction models: a practical approach to development, validation, and updating 1:497. <https://doi.org/10.1007/978-0-387-77244-8>
45. Greendale GA, Sowers M, Han W et al (2012) Bone mineral density loss in relation to the final menstrual period in a multiethnic cohort: Results from the Study of Women's Health Across the Nation (SWAN). *J Bone Miner Res* 27:111–118. <https://doi.org/10.1002/JBMR.534>
46. Greendale GA, Tseng CH, Han W et al (2015) Dietary isoflavones and bone mineral density during midlife and the menopausal transition: cross-sectional and longitudinal results from the Study of Women's

- Health Across the Nation Phytoestrogen Study. *Menopause* 22:279–288. <https://doi.org/10.1097/GME.0000000000000305>
47. Crandall CJ, Tseng CH, Karlamangla AS et al (2013) Serum sex steroid levels and longitudinal changes in bone density in relation to the final menstrual period. *J Clin Endocrinol Metab* 98:E654–E663. <https://doi.org/10.1210/JC.2012-3651>
  48. Sowers MR, Finkelstein JS, Ettinger B et al (2003) The association of endogenous hormone concentrations and bone mineral density measures in pre- and perimenopausal women of four ethnic groups: SWAN. *Osteoporos Int* 14:44–52. <https://doi.org/10.1007/S00198-002-1307-X/METRICS>
  49. Riggs BL, Khosla S, Melton LJ (1998) A unitary model for involutional osteoporosis: estrogen deficiency causes both type I and type II osteoporosis in postmenopausal women and contributes to bone loss in aging men. *J Bone Miner Res* 13:763–773. <https://doi.org/10.1359/JBMR.1998.13.5.763>
  50. Conde J, Scotece M, Gomez RG et al (2011) Adipokines and osteoarthritis: novel molecules involved in the pathogenesis and progression of disease. *Arthritis* 2011:203901. <https://doi.org/10.1155/2011/203901>
  51. Hou J, He C, He W et al (2020) Obesity and bone health: a complex link. *Front Cell Dev Biol* 8:1620. <https://doi.org/10.3389/FCELL.2020.600181/BIBTEX>
  52. Christianson MS, Shen W (2013) Osteoporosis prevention and management: nonpharmacologic and lifestyle options. *Clin Obstet Gynecol* 56:703–710. <https://doi.org/10.1097/GRF.0B013E3182A9D15A>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.