



# Prediction of incident vertebral fracture using CT-based finite element analysis

B. T. Allaire<sup>1</sup> · D. Lu<sup>2</sup> · F. Johannesdottir<sup>1,3</sup> · D. Kopperdahl<sup>4</sup> · T. M. Keaveny<sup>5,6</sup> · M. Jarraya<sup>7,8</sup> · A. Guermazi<sup>8</sup> · M. A. Bredella<sup>9</sup> · E. J. Samelson<sup>10,11</sup> · D. P. Kiel<sup>10,11</sup> · D. E. Anderson<sup>1,3</sup> · S. Demissie<sup>2</sup> · M. L. Bouxsein<sup>1,3</sup>

Received: 21 December 2017 / Accepted: 19 September 2018 / Published online: 10 October 2018  
© International Osteoporosis Foundation and National Osteoporosis Foundation 2018

## Abstract

**Summary** Prior studies show vertebral strength from computed tomography-based finite element analysis may be associated with vertebral fracture risk. We found vertebral strength had a strong association with new vertebral fractures, suggesting that vertebral strength measures identify those at risk for vertebral fracture and may be a useful clinical tool.

**Introduction** We aimed to determine the association between vertebral strength by quantitative computed tomography (CT)-based finite element analysis (FEA) and incident vertebral fracture (VF). In addition, we examined sensitivity and specificity of previously proposed diagnostic thresholds for fragile bone strength and low BMD in predicting VF.

**Methods** In a case-control study, 26 incident VF cases (13 men, 13 women) and 62 age- and sex-matched controls aged 50 to 85 years were selected from the Framingham multi-detector computed tomography cohort. Vertebral compressive strength, integral vBMD, trabecular vBMD, CT-based BMC, and CT-based aBMD were measured from CT scans of the lumbar spine.

**Results** Lower vertebral strength at baseline was associated with an increased risk of new or worsening VF after adjusting for age, BMI, and prevalent VF status (odds ratio (OR) = 5.2 per 1 SD decrease, 95% CI 1.3–19.8). Area under receiver operating characteristic (ROC) curve comparisons revealed that vertebral strength better predicted incident VF than CT-based aBMD (AUC = 0.804 vs. 0.715,  $p = 0.05$ ) but was not better than integral vBMD (AUC = 0.815) or CT-based BMC (AUC = 0.794). Additionally, proposed fragile bone strength thresholds trended toward better sensitivity for identifying VF than that of aBMD-classified osteoporosis (0.46 vs. 0.23,  $p = 0.09$ ).

**Conclusion** This study shows an association between vertebral strength measures and incident vertebral fracture in men and women. Though limited by a small sample size, our findings also suggest that bone strength estimates by CT-based FEA provide equivalent or better ability to predict incident vertebral fracture compared to CT-based aBMD. Our study confirms that CT-based estimates of vertebral strength from FEA are useful for identifying patients who are at high risk for vertebral fracture.

**Keywords** Bone strength · Finite element analysis · Fracture prediction · Incident vertebral fracture

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00198-018-4716-1>) contains supplementary material, which is available to authorized users.

✉ M. L. Bouxsein  
mbouxsei@bidmc.harvard.edu

<sup>1</sup> Center for Advanced Orthopaedic Studies, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, RN 115, Boston, MA 02215, USA

<sup>2</sup> Boston University, Boston, MA, USA

<sup>3</sup> Department of Orthopedic Surgery, Harvard Medical School, Boston, MA, USA

<sup>4</sup> O.N. Diagnostics LLC, Berkeley, CA, USA

<sup>5</sup> Department of Mechanical Engineering, University of California, Berkeley, CA, USA

<sup>6</sup> Department of Bioengineering, University of California, Berkeley, CA, USA

<sup>7</sup> Department of Radiology, Mercy Catholic Medical Center, Darby, PA, USA

<sup>8</sup> Boston University School of Medicine, Boston, MA, USA

<sup>9</sup> Department of Radiology, Massachusetts General Hospital, Boston, MA, USA

<sup>10</sup> Institute for Aging Research, Hebrew SeniorLife, Boston, MA, USA

<sup>11</sup> Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA

## Introduction

Vertebral fractures (VFs) are the most common type of osteoporotic fracture [1]. Whereas dual-energy X-ray absorptiometry (DXA)-assessed areal bone mineral density (aBMD) is a standard tool for estimating fracture risk, close to half of individuals with vertebral fractures have normal aBMD by DXA [2, 3]. With the number of fractures continuing to rise due to aging of the population [4], more accurate identification of those at risk for vertebral fracture may promote efficient use of available interventions to reduce fractures. Moreover, with declining rates of osteoporosis screening, validation of techniques other than aBMD by DXA for fracture risk assessment may provide alternate methods to increase the number of individuals who are screened for osteoporosis [5, 6].

In the past decade, estimation of whole bone strength via finite element analysis (FEA) of quantitative computed tomography (CT) images has shown promise as an approach to assess fracture risk and treatment efficacy [7, 8]. Several studies show that FEA-derived estimates of vertebral strength are strongly associated with failure loads measured from *in vitro* mechanical testing of human cadaveric vertebrae ( $r^2 = 0.77$  to  $0.90$ ) [9–12]. In addition, FEA-derived vertebral strength measures are associated with prevalent vertebral fractures [13–16]. However, to date, only two studies have investigated FEA-derived vertebral strength for predicting incident vertebral fracture [12, 17]. One of these studies included both men and women but was limited to a rather homogeneous cohort (all subjects living in Iceland) [17], while the other was a multi-center US study but only examined men [12]. Thus, additional studies of incident vertebral fracture that include both women and men are needed to confirm the utility of the technique. Moreover, thresholds for low and fragile bone strength were recently proposed as guidelines for the clinical interpretation of FEA strength measurements [17]. Prior studies have suggested that FEA-based strength measurements appear to outperform aBMD for prediction of incident VF [12, 17]; however, no independent studies have tested the ability of proposed bone strength thresholds to identify those who eventually fracture, nor compared their sensitivity and specificity to those based on aBMD thresholds.

Thus, we aimed to test the hypothesis that lumbar vertebral strength derived from CT-based FEA estimates predicts incident vertebral fractures in a cohort of community-dwelling men and women. We also sought to compare the performance of proposed clinical guidelines and diagnostic thresholds for FEA strength with diagnostic thresholds for BMD. We hypothesized that vertebral strength thresholds would be more selective in identifying individuals who suffer VF than accepted clinical thresholds for both low areal (aBMD) and volumetric BMD (vBMD).

## Methods

### Participants

We used a case-control study design utilizing data from the Framingham Heart Study (FHS) Offspring and Third-Generation Multidetector CT (MDCT) Study [18]. Members of the Offspring [19] and Third-Generation Cohorts [20] included second-generation (and spouses) and third-generation offspring of the original FHS cohort, a community-based cohort in Massachusetts started in 1948 [21]. Members of the MDCT cohort underwent longitudinal CT scanning in 2002–2005 (baseline) and 2008–2011 (follow-up). Participants were eligible for the current study if they were at least 50 years old at the time of the baseline scan and had evaluable images from both the baseline and follow-up CT examinations ( $N = 1149$  individuals, 514 men, and 635 women). Vertebral fracture cases were defined as participants with a new or worsening vertebral fracture over a mean 6-year follow up time period. Controls included participants with no new or worsening vertebral fracture and were matched to cases by sex and age ( $\pm 5$  years).

We identified 26 participants who had at least 1 new ( $n = 23$ ) or worsening ( $n = 3$ ) vertebral fracture. Using available participants without new or worsening vertebral fracture, two to four controls were selected for each case depending on the number of available age- and sex-matched individuals. In total, 62 matching controls were found and were eligible for inclusion in the study.

### CT acquisition

As previously reported [18], FHS participants underwent volumetric CT scanning at baseline using a multidetector CT (LightSpeed Ultra, General Electric Medical Systems, Milwaukee, WI, USA) with tube voltage of 120 kVp, tube current of 320 to 400 mA (depending on body weight), gantry rotation of 500 ms, and slice thickness of 2.5 mm. After acquisition of frontal and lateral scout images, CT scan series were acquired to assess coronary and aortic calcification and included one CT scan of the chest (40 to 68 slices from the carina of the trachea to the diaphragm) and one of the abdominal region (150 mm superior to the L5/S1 junction) [18, 22]. Follow-up scanning was performed using a Discovery VCT 64-slice PET/CT scanner (General Electric Medical Systems, Milwaukee, WI, USA) with tube voltage of 120 kVp, tube current of 300 or 350 mA (depending on body weight), gantry rotation of 350 ms, and slice thickness of 2.5 mm. After acquisition of frontal and lateral scout images, CT series were acquired for the chest (from lung base to apices) and abdomen (150 mm superior to the top of the S1 vertebra). All participants were scanned concurrently with a hydroxyapatite phantom (Image Analysis, Inc., Lexington, KY, USA) for

calibrating the scan into units of equivalent-BMD of hydroxyapatite (in  $\text{mg}/\text{cm}^3$ ).

### Vertebral fracture identification

Vertebral fractures were evaluated from T4 to L4 by trained musculoskeletal radiologists (MJ, AG, MAB, HKG) using Genant's semiquantitative (SQ) method [23]. Baseline and follow-up CT images were viewed side-by-side with knowledge of CT chronology (but blinded to clinical data). Each vertebral body was scored according to a 4-point scale: no fracture (SQ0), mild (SQ1), moderate (SQ2), or severe (SQ3) fracture.

Prevalent vertebral fracture was defined as any vertebral body graded SQ1 or higher at baseline. Fracture cases included all individuals with new or worsening vertebral fractures. New fractures were defined as those vertebrae graded SQ0 (no fracture) at baseline and SQ1 or higher at follow-up, whereas worsening fractures were defined as those vertebrae graded SQ1 or SQ2 at baseline that increased at least one grade at follow-up. For each individual, severity of vertebral fracture was determined by the grade of the vertebral body with the highest score.

Initial readings were performed by a musculoskeletal radiologist (MJ) using the sagittal CT reformations as part of a larger study evaluating spinal degenerative features [24]. Since lateral radiographic images are most often used clinically to evaluate vertebral fracture, two independent readers (MAB, HKG) conducted a second level of review relying only on CT scout images, which extended from the sacrum to the upper thoracic spine. In cases of disagreement, scans were reviewed together by both readers to adjudicate the score.

### Vertebral strength and volumetric bone measures

Finite element analyses and volumetric bone measurements were performed on baseline CT scans by a trained analyst, blinded to case-control status, using VirtuOst software (O.N. Diagnostics, Berkeley, CA, USA), as previously described [9, 17, 25]. Analyses were performed on the L3 vertebra using the abdominal CT series. Briefly, bone was segmented and voxel intensity values were converted to BMD using a hydroxyapatite phantom. The bone volume was then resampled into isotropic voxels ( $1 \times 1 \times 1 \text{ mm}$ ), and each voxel was then converted into a hexahedral finite element and assigned material properties based on empirical relationships with BMD [26–28]. Displacement boundary conditions simulated uniform axial compression applied through a virtual layer of bone cement. Vertebral strength ( $N$ ) was defined as the compressive force at 2% deformation [17]. Trabecular and integral volumetric bone density were determined using the same software package, as well as vertebral body bone mineral content (BMC). Integral vBMD ( $\text{mg}/\text{cm}^3$ ) was defined as the average

density over the whole vertebral body, including endplates and excluding posterior elements. BMC was defined as the total mineral mass of the entire vertebral body, using the same contour as integral vBMD. Trabecular vBMD ( $\text{mg}/\text{cm}^3$ ) was defined as the average density of an ellipsoidal volume placed inside the trabecular compartment in the middle 10 mm of the vertebral volume.

### CT-based areal bone mineral density

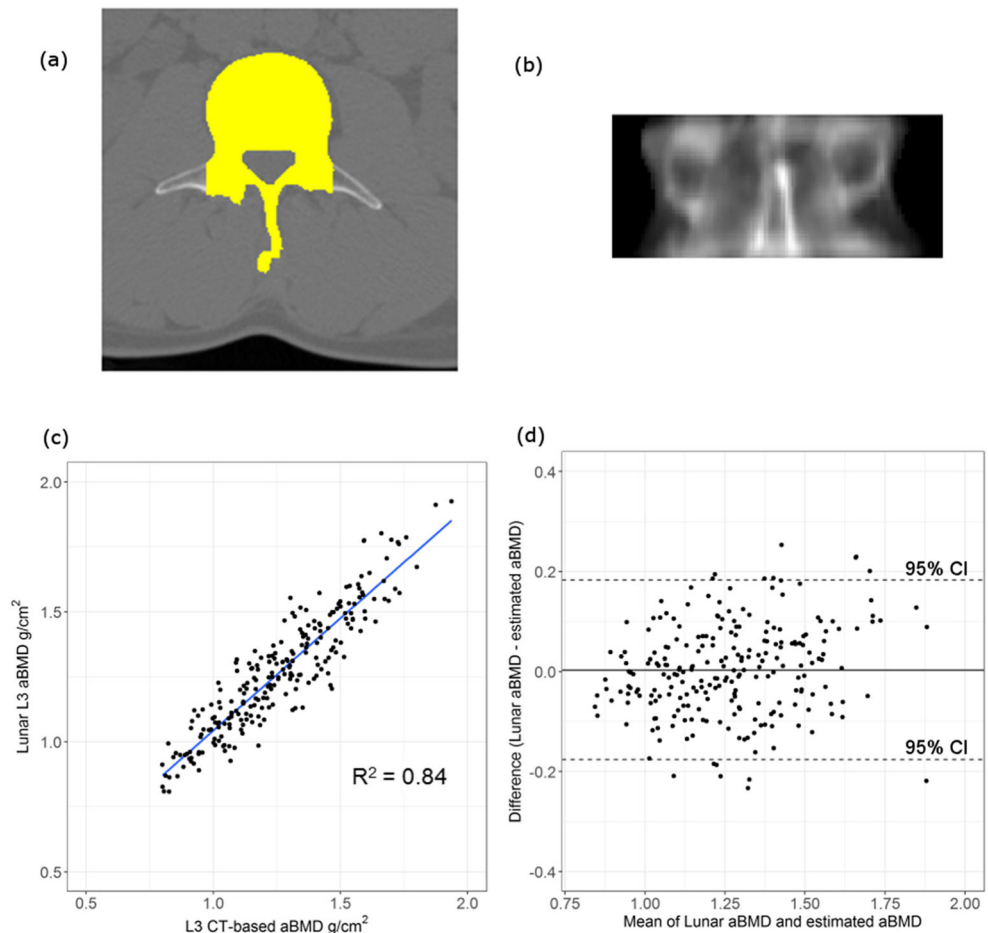
In the subjects eligible for this case-control study, DXA scans were not available. Scans were either not acquired or were performed many years prior to or after the CT exam. To address this, we measured a CT-based aBMD ( $\text{g}/\text{cm}^2$ ) of the L3 vertebrae via a semi-automated algorithm developed in our laboratory. In short, a custom program in Matlab (MathWorks, Natick, MA, USA) was created to contour the vertebral body, including posterior elements. Surfaces were identified in 3D and were then projected on a single plane. An anterior-posterior measurement of (PA) L3 aBMD was made using the average of BMD-calibrated attenuation values and known pixel size (Fig. 1a).

CT-based aBMD measurements were validated against direct DXA measurements in a separate cohort of 237 Framingham Heart Study participants (110 men and 127 women) who had PA lumbar spine aBMD measurements using a Lunar DPX-L (Lunar Corp, Madison, WI, USA) within 6 months of their corresponding CT scan. These subjects had no prevalent vertebral fractures in the lumbar spine, ranged in age from 44 to 81 years (mean 62.5 years) and had BMI between 17.5 and 36.9 (mean 27.3). The CT-based measurements of aBMD were strongly correlated with those from DXA ( $r^2 = 0.84$ ,  $p < 0.01$ , Fig. 1c). We used the resulting regression equation (Fig. 1c) to calibrate the values of the CT-based aBMD measurements to equal those from a Lunar DXA scanner. This was done to be able to evaluate the CT-based aBMD measurements on the same scale as clinically accepted Lunar aBMD osteoporosis thresholds for sensitivity analyses.

### Bone strength and bone density thresholds for sensitivity testing

Sex-specific thresholds for “fragile bone strength” for the proximal femur and lumbar spine have been previously reported [17]. In particular for the lumbar spine, fragile bone strength thresholds were 4500 N for women and 6500 N for men. These vertebral strength threshold values were validated for fracture prediction using the average strength of the L1 and L2 vertebral levels and are meant to correspond to osteoporosis thresholds of trabecular vBMD [17]. As stated earlier, CT scans available for this study did not cover L1 and L2. Our measurements were performed on L3. To account for this, we scaled our L3 strength measures to an equivalent L1/L2

**Fig. 1** **a** Contoured CT image including posterior elements. **b** CT-based aBMD image from projection of CT surfaces. **c** Linear regression between measured Lunar L3 aBMD and CT-based aBMD. **d** Bland-Altman plot looking at Lunar aBMD and estimated CT-based aBMD



average through simple multiplication by a scale factor of 0.89 so that we could directly compare our measurements against the L1/L2 thresholds established previously [17]. This scale factor was determined from a small independent cohort of 22 women and 3 men, age 50 to 81 years, with L1, L2, and L3 measured.

To compare the sensitivity of CT-based FEA and CT-based vBMD, we used the proposed diagnostic threshold for osteoporosis in trabecular vBMD ( $80 \text{ mg/cm}^3$ ) [29, 30]. For our CT-based L3 aBMD, we calculated T-scores using NHANES III reference data for the L3 vertebra, with 20–29-year-old white women as our reference group, converted to Lunar aBMD [30, 31]. Individuals with a T-score  $\leq 2.5$  were considered osteoporotic.

### Covariates

FHS participants underwent a standard clinical examination near the time of the baseline CT scan. The exam included a physician interview, physical exam, and standardized questionnaires. Age, sex, height, and weight were measured at this examination. Participants were considered as current smokers if they had smoked at least one cigarette per day over the past

year. Alcohol intake was determined by a self-reported questionnaire and was reported as ounces of alcohol per week.

### Statistical analyses

Baseline characteristics were compared between incident vertebral fracture cases and controls using a two-sample *T* test for continuous variables and a chi-square test of independence for categorical variables. We determined the associations between prediction variables and incident fracture case status using conditional logistic regression, computing odds ratios (ORs), and 95% confidence intervals, for 1 standard deviation (SD) change. Models were adjusted for age, BMI, and/or prevalent vertebral fracture status. An additional adjustment for CT-based aBMD was added to vertebral strength models. We observed the receiver operating characteristic (ROC) curves for CT-based aBMD, BMC, integral vBMD, and vertebral strength in models adjusted for age and BMI. Area under the curve (AUC) of the ROC curve was determined, and statistical comparison between AUCs was assessed. We also performed stratified analyses including only moderate/severe (SQ2+) fracture cases and their controls. Sensitivity, specificity, positive predictive value, and negative predictive value were

calculated using CT-based aBMD, trabecular vBMD, and vertebral strength thresholds [32]. Confidence intervals and *p*-values for significance were calculated for these parameters. Statistical analyses were performed using SAS 9.3 (SAS Institute, Inc., Cary, NC, USA) and R version 3.2.5 (R Foundation for Statistical Computing, Vienna, Austria).

## Results

Individuals with incident vertebral fracture had similar height, weight, smoking status, and alcohol intake as non-fracture controls but were more likely to have a prevalent vertebral fracture (Table 1). Integral vBMD, trabecular vBMD, CT-aBMD, BMC, and vertebral strength were significantly lower in the VF cases than controls ( $p < 0.05$ ). Deficits in vertebral bone density and strength in cases versus controls tended to be greater when including incident moderate and severe vertebral fractures only (Table 1).

Lower vertebral strength was significantly associated with incident VF and remained so after adjustment for age and BMI (OR = 3.8, 95% CI: 1.5–9.2) and further adjustment for prevalent VF (OR = 5.2, 95% CI 1.3–19.8, Supplemental Table 1). Of note, after additional adjustment for CT-aBMD, lower vertebral strength remained associated with incident VF (OR = 5.1, 95% CI 1.5–17). Lower integral vBMD, trabecular vBMD, and CT-based BMC were also associated with incident VF and remained so after adjusting for age and BMI (Table 2). The association between incident vertebral fracture and CT-based aBMD was weaker than the CT-derived volumetric measures (OR = 1.6, 95% CI 0.9–2.6,  $p = 0.1$ ). Associations between CT-aBMD and incident VF were weaker in all models, regardless of adjustment by covariates. AUC comparisons showed that vertebral strength was a significantly better predictor of incident VF than CT-based spine aBMD (AUC = 0.804 vs. 0.715,  $p = 0.05$ ). However, it not better than CT-based BMC (AUC = 0.804 vs. 0.794) or integral vBMD (AUC = 0.804 vs. 0.815). AUC comparisons revealed integral vBMD performed statistically better than CT-based aBMD ( $p = 0.03$ ) while BMC did not but was borderline ( $p = 0.07$ ).

When only considering moderate and severe incident fractures (SQ2+), all predictors were significantly associated with incident VF after adjusting for age and BMI. Trabecular vBMD ( $p = 0.053$ ) and CT-aBMD ( $p = 0.08$ ) were no longer significantly associated with incident VF after adjustment for prevalent VF status. Of note, odds ratios for moderate and severe vertebral fractures were generally higher than when considering all incident fractures. After adjustment for CT aBMD, lower vertebral strength remained associated with incident vertebral fracture (OR = 6.64, 95% CI: 1.2–37.8,  $p = 0.03$ ).

## Clinical thresholds and diagnostic performance

Classification of incident fracture cases with FEA-based fragile bone strength thresholds mostly exhibited similar diagnostic metrics to osteoporosis thresholds for CT-based aBMD and vertebral trabecular vBMD (Table 3). However, osteoporosis (t-score  $\leq -2.5$ ) from aBMD trended toward lower sensitivity than fragile bone strength (0.23 vs. 0.46,  $p = 0.09$ ) while maintaining similar values for specificity, positive, and negative predictive value. Additionally, osteoporosis classification from vertebral trabecular vBMD ( $\leq 80 \text{ mg/cm}^3$ ) tended to show lower diagnostic metrics than FEA-derived vertebral strength; however, no significant differences were found.

## Discussion

Our results indicate that vertebral strength estimates from finite element analysis of lumbar spine CT scans predict incident vertebral fracture risk in older men and women as well as or better than CT-based areal BMD. Furthermore, proposed fragile bone strength thresholds exhibit at least equivalent diagnostic performance compared to osteoporosis thresholds based on CT-based aBMD or vertebral trabecular vBMD.

Although finite element analysis of CT scans to estimate bone strength was first introduced more than two decades ago [33, 34], few prospective studies have tested the ability of FEA-estimated bone strength to predict vertebral fracture risk [12, 17, 35]. In the current study, we found odds ratios similar to previous studies. In particular, our unadjusted model showed an OR of 2.8 per 1SD decrease in vertebral strength, which was comparable to prior studies that reported odds ratios for vertebral fracture between 2.1 and 7.6 [12, 17]. These associations remained similar in models adjusting for additional covariates including age and BMI. Also, we found that vertebral strength was associated with incident VF after adjustment for prevalent VF, which itself is a very strong predictor of a new vertebral fracture [36]. Odds ratios seen in our study were generally lower than those reported by Wang et al. [12], who reported ORs for vertebral strength ranging from 7.2 to 9.6 across models adjusted for differing covariates. The heightened ORs are likely due to the use of clinical vertebral fractures in that study, which are likely more severe than those identified via radiographic review. Thus, it is probably more appropriate to compare our results from moderate and severe incident VF with the clinical fractures from that study. The age-adjusted OR for vertebral strength in our moderate and severe VF cases was 6.8, closer in magnitude to the odds ratio of 7.2 reported by Wang et al. [12].

Our study is the first to demonstrate the viability of CT-based FEA strength measures in the L3 vertebra for prediction of incident vertebral fracture. Prior studies have used different CT protocols and have examined different vertebrae. Namely,

**Table 1** Baseline characteristics of cases and controls (mean  $\pm$  SD, or n, %)

	Any incident VF			Incident moderate/severe VF		
	Control (n = 62)	Case (n = 26)	<i>p</i> value	Control (n = 46)	Case (n = 20)	<i>p</i> value
Men	36 (58)	13 (50)	–	27 (59)	10 (50)	–
Women	26 (42)	13 (50)	–	19 (41)	10 (50)	–
Age (years)	66.3 (8.4)	68.1 (9.9)	0.381	66.3 (8.9)	68.2 (10.4)	0.451
Height (cm)	169.8 (9.8)	167.5 (9.2)	0.310	170.1 (9.9)	167.5 (9.5)	0.333
Weight (kg)	83.3 (16.9)	83.1 (18.6)	0.973	83.3 (18.2)	83.4 (20.0)	0.974
BMI	28.9 (6.0)	29.7 (7.0)	0.612	28.8 (6.4)	29.8 (7.6)	0.585
Current smoker (yes/no)	4 (6.5)	5 (19.2)	0.117	2 (4.3)	3 (15)	0.159
Total alcohol intake (oz/wk)	2.4 (3.3)	2.0 (3.1)	0.584	2.4 (3.5)	1.75 (3.4)	0.511
Prevalent VF, any (yes/no)	3 (4.8)	14 (53.9)	<0.001	3 (6.5)	10 (50)	<0.001
Prevalent VF, moderate/severe (yes/no)	1 (1.6)	3 (11.5)	0.076	1 (2.2)	2 (10)	0.216
Vertebral strength, CT-FEA (N)	8706 (2914)	6552 (2391)	0.001	8843 (3178)	6168 (2076)	<0.001
L3 Trabecular vBMD (mg/cm <sup>3</sup> )	107.3 (35.3)	86.0 (30.9)	0.009	108.8 (38.8)	84.9 (29.6)	0.017
L3 Integral vBMD (mg/cm <sup>3</sup> )	200.2 (38.3)	165.9 (33.8)	<0.001	200.3 (42.1)	160.9 (30.6)	<0.001
L3 CT-based aBMD (g/cm <sup>2</sup> )	1.08 (0.21)	0.97 (0.23)	0.032	1.08 (0.21)	0.92 (0.19)	0.006
L3 CT-based BMC (g)	9.16 (2.57)	7.31 (1.96)	<0.001	9.28 (2.73)	6.95 (1.74)	<0.001

prior studies in the spine utilized L1 and L2 imaging data. Additionally, slice thickness for the CT images in our study were higher (2.5 mm) than previous studies (1 mm) [12, 17]. Taken together, this demonstrates the efficacy of CT-based FEA for prediction of incident vertebral fracture risk across different lumbar spine levels and different scanning protocols.

We examined the efficacy of proposed CT-based FEA thresholds of vertebral strength [17] in a new and independent

cohort of men and women. Our data suggest that these thresholds provide equivalent or better sensitivity identifying those at risk for VF compared to BMD thresholds of osteoporosis, both in aBMD t-score ( $\leq -2.5$ ) and trabecular vBMD (80 mg/cm<sup>3</sup>). Specifically, we found that 23% of our fracture cases were osteoporotic by CT-based aBMD. This closely matches recent literature [3] from NHANES 2013–2014 participants that found that 26% of fracture cases aged 50 and older were osteoporotic by CT-based aBMD. In comparison, 46% of our fracture cases met the criteria for fragile bone strength by FEA without much loss in specificity. Due to the limited number of fracture cases in our study, further studies are needed to verify the promising diagnostic capability of these thresholds as an alternative clinical tool to DXA for fracture risk assessment.

Although we found that FE vertebral strength was significantly more associated with incident vertebral fracture than our CT-based aBMD measure, vertebral strength did not predict fracture better than integral vBMD or CT-based BMC measures in our cohort. This is not all that surprising given the high correlation between these parameters (Supplemental Table 2) and small sample size. Still, FE vertebral strength estimates may be beneficial. First, FE produces an estimated strength value that can be used to calculate a factor-of-risk when loading can also be estimated or measured. Second, FE can take into account structure and material properties in a way that other summary measures do not. Additionally, advances in modeling techniques, simulated loading conditions, and boundary conditions could further improve strength approximations and ultimately fracture prediction.

**Table 2** Risk of incident mild (SQ1+) and moderate (SQ2+) vertebral fracture per 1 SD decrease in bone measures at L3 vertebral level, expressed as odds ratio (OR) with 95% confidence intervals

	Unadjusted	Adjusted for age and BMI
Any incident vertebral fracture (26 VF cases, 62 controls)		
<i>Integral vBMD</i>	<b>2.8 (1.4–5.5)</b>	<b>3.6 (1.5–8.6)</b>
<i>Trabecular vBMD</i>	<b>2.1 (1.1–4.1)</b>	<b>2.4 (1.1–5.3)</b>
<i>CT-based aBMD</i>	1.6 (0.9–2.6)	1.8 (1.0–3.3)
<i>CT-based BMC</i>	<b>3.3 (1.4–7.9)</b>	<b>3.8 (1.6–9.2)</b>
<i>FEA vertebral strength</i>	<b>2.8 (1.3–6.2)</b>	<b>3.8 (1.5–9.2)</b>
Incident moderate or severe vertebral fracture (20 VF cases, 46 controls)		
<i>Integral vBMD</i>	<b>3.6 (1.4–9.3)</b>	<b>5.7 (1.6–20.1)</b>
<i>Trabecular vBMD</i>	<b>2.5 (1.1–5.9)</b>	<b>2.9 (1.1–7.9)</b>
<i>CT-based aBMD</i>	<b>2.2 (1.1–4.5)</b>	<b>3.7 (1.4–10.0)</b>
<i>CT-Based BMC</i>	<b>6.2 (1.5–25.6)</b>	<b>7.4 (1.7–31.8)</b>
<i>FEA vertebral strength</i>	<b>4.9 (1.5–16.0)</b>	<b>8.4 (1.9–36.6)</b>

Bold indicates significance ( $p < 0.05$ ). Model did not converge to proper solution

**Table 3** Sensitivity and specificity for identification of incident vertebral fracture according to L3 vertebral strength, CT-based aBMD, and trabecular vBMD thresholds (95% confidence intervals in parentheses)

	Vertebral strength (from CT-FEA) [17]	CT-based aBMD (from CT)	Trabecular vBMD (from CT) [17, 29]
Threshold/cut point	Women $\leq$ 4500N Men $\leq$ 6500N	t-score $\leq$ -2.5	vBMD $\leq$ 80 mg/cm <sup>3</sup>
Sensitivity	0.46 (0.27–0.67)	0.23 (0.09–0.44)	0.27 (0.12–0.48)
Specificity	0.87 (0.76–0.94)	0.94 (0.84–0.98)	0.81 (0.69–0.90)
Pos. predictive value	0.60 (0.36–0.81)	0.60 (0.26–0.88)	0.37 (0.16–0.62)
Neg. predictive value	0.79 (0.68–0.88)	0.74 (0.63–0.84)	0.72 (0.60–0.83)

Our study also demonstrates that it may be of interest to conduct studies aimed at determining interventional thresholds of integral vBMD or CT-based BMC given that they yield similar ORs to vertebral strength, particularly in the spine. Thresholds for these measures may provide equivalent or better sensitivity than DXA aBMD thresholds but future studies would be needed to develop the thresholds and confirm this.

Of note, to date, CT-based FEA (including the current study) has relied on a hydroxyapatite phantom for calibration of CT-based X-ray attenuation values (Hounsfield units) to equivalent BMD values. While this limits the widespread use of CT scans acquired for other purposes in the assessment of fracture risk, recent studies have demonstrated utility of CT scans acquired for inflammatory bowel disease [37] and colorectal cancer screening [38] for assessment of bone density and strength. Further, methods are now published for phantomless calibration of CT scans [39], and a recent study showed that FEA of routine clinical CT scans is feasible for assessment of hip bone density t-scores and femoral strength, and that femoral strength values derived from these CT scans via FEA are as effective as DXA-hip BMD in predicting hip fracture risk [40]. Additionally, a recent cost-effectiveness study [41] suggests it may be cost-effective to use a new hip CT for assessing osteoporosis. Given the results of our study, it may be worth examining use of routine abdominal CT scans for osteoporosis assessment, and further, whether the use of a dedicated spine CT for osteoporosis assessment would be cost-effective.

Our study has a number of limitations. The number of individuals with incident vertebral fracture was small, and thus limited our ability to detect significant differences in the associations between different bone measurements and incident VF. The small number of fractures is likely due to the younger age of the Framingham Heart Study cohort relative to other cohorts where VFs have been examined. The mean age for our cases and controls were 68 and 66 years respectively, with ages ranging from 50 to 85 years. In comparison, previous studies included only those aged 65 or older with a mean of around 75 years. This age difference is also reflected in vertebral trabecular BMD differences between our study and the Icelandic population-based cohort (AGES). Mean

vertebral trabecular density values in the AGES cohort were 64 and 79 mg/cm<sup>3</sup> for cases and controls, respectively. These values are notably lower than the 89 and 110 mg/cm<sup>3</sup> mean values for vBMD in the cases and controls in the current study. Regardless, we have shown vertebral strength predicts vertebral fracture risk even in a younger and, perhaps, healthier cohort. This complements previous studies by offering a wider age range and highlighting the robustness of CT-based FEA strength estimates in their ability to predict fracture risk.

One limitation of this study is use of vertebral strength and CT-based aBMD at a single vertebral level (L3), whereas spine DXA for osteoporosis diagnosis averages over a number of lumbar vertebrae to get a more robust measure of overall aBMD [30]. However, given the high correlation between vertebral strength measures at adjacent vertebrae, the prediction of vertebral fracture is likely to be similar if FE strength or CT-based aBMD were averaged over multiple lumbar vertebral levels.

Another potential limitation is the use of lumbar strength and BMD measurements to predict vertebral fractures anywhere in the spine. We previously reported that single-level lumbar vertebral strength measures are associated with prevalent fracture anywhere in the spine and that thoracic vertebral strength measures were not better at predicting prevalent thoracic fractures [42]. That observation, coupled with the strong associations seen between L3 vertebral strength and incident fracture in our study, provides evidence for the robustness of single-level lumbar FE strength measures for prediction of VF anywhere in the spine. However, future studies examining the ability of site-specific measurements to predict site-specific incident VF could be of interest—though they may not be of significant clinical relevance given the increased radiation exposure associated with more extensive spine imaging.

Related, we determined sensitivity of fragile strength thresholds based on scaled L3 data. Scale factors between L1/L2 and L3 were determined in a small cohort of mostly women. Thus, we acknowledge the potential error incurred by using this scale factor to determine threshold sensitivity. Additional validation of scale factors across the spine, or specific fragile strength thresholds at each vertebral level, would

help the broader applicability of spine CT-FEA as a diagnostic tool. A further limitation is that these vertebral strength thresholds are specific to the specific modeling approach used here (VirtuOst, O.N. Diagnostics). A different modeling approach likely could be used in conjunction with these thresholds, given that approach was externally validated against cadaver experiments conducted under the same conditions. However, future studies could look at comparing approaches or establishing thresholds specific to other modeling techniques.

An additional shortcoming of our study was the lack of spine aBMD measurements by DXA. DXA is the clinical standard for the diagnosis of osteoporosis and low BMD [31]. To overcome this limitation and compare with previous studies [12], we developed a robust approach to estimate aBMD from CT images. Though we demonstrated our CT-based measure was highly correlated with aBMD measures by DXA ( $r = 0.92$ ) in a separate cohort, some variation between CT-based aBMD and measured DXA aBMD is still unexplained that may be related to inherent differences in scanning modalities. So, it is important to note that even though calibrated CT-based aBMD does a good job of approximating DXA aBMD, we are making an assumption that they are equivalent measures in order to use the aBMD osteoporosis threshold (i.e., T-score  $< -2.5$ ).

In conclusion, our study has demonstrated the utility of CT-based vertebral strength estimates for prediction of vertebral fracture risk in men and women from a community-based cohort. Though we were limited by a small sample size, our findings suggest equivalent or better performance of diagnostic thresholds for vertebral strength by CT-based FEA compared to those of CT-based aBMD and trabecular vBMD in identifying those who suffered a vertebral fracture. Considering all available reports so far, CT-based FEA-strength estimates consistently identify patients at high risk for vertebral fracture and therefore may be an important diagnostic tool.

**Acknowledgements** The contents are solely the responsibility of the authors and do not necessarily represent the views of the NIH. We acknowledge Dr. Harry K. Genant's contribution in reviewing vertebral fracture assessment.

**Funding information** This work was supported by grants from the National Institutes of Health (R01 AR053986, R00 AG042458, R01 AG041658, R01 AR041398), and by the National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study (NIH/NHLBI Contract N01-HC-25195).

## Compliance with ethical standards

**Conflicts of interest** Brett T. Allaire, Darlene Lu, Fjola Johannesdottir, Mohamed Jarraya, Ali Guermazi, Miriam A. Bredella, Elizabeth J. Samelson, Douglas P. Kiel, Dennis E. Anderson, Serkalem Demissie, and Mary L. Bouxsein declare that they have no conflicts of interest. Dr. Keaveny has been a consultant for Amgen, AgNovos Healthcare, and O.N. Diagnostics and has equity in O.N. Diagnostics. Drs. Keaveny and Kopperdahl have financial interests in O.N. Diagnostics and both they and the company may benefit from the results of this work.

## References

- Melton LJ, Lane AW, Cooper C et al (1993) Prevalence and incidence of vertebral deformities. *Osteoporos Int* 3:113–119. <https://doi.org/10.1007/BF01623271>
- Schuit SCE, Van Der Klift M, Weel AEAM et al (2004) Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam study. *Bone* 34:195–202. <https://doi.org/10.1016/j.bone.2003.10.001>
- Cosman F, Kregge JH, Looker AC, Schousboe JT, Fan B, Sarafrazi Isfahani N, Shepherd JA, Krohn KD, Steiger P, Wilson KE, Genant HK (2017) Spine fracture prevalence in a nationally representative sample of US women and men aged  $\geq 40$  years: results from the National Health and Nutrition Examination Survey (NHANES) 2013–2014. *Osteoporos Int* 28:1857–1866. <https://doi.org/10.1007/s00198-017-3948-9>
- Schneider EL, Guralnik JM (1990) The aging of America. *JAMA* 263:2335–2340. <https://doi.org/10.1001/jama.1990.03440170057036>
- Gillespie CW, Morin PE (2017) Trends and disparities in osteoporosis screening among women in the United States, 2008–2014. *Am J Med* 130:306–316. <https://doi.org/10.1016/j.amjmed.2016.10.018>
- Binkley N, Blank RD, Leslie WD, Lewiecki EM, Eisman JA, Bilezikian JP (2017) Osteoporosis in crisis: it's time to focus on fracture. *J Bone Miner Res* 32:1391–1394. <https://doi.org/10.1002/jbmr.3182>
- Orwoll ES, Marshall LM, Nielson CM, Cummings SR, Lapidus J, Cauley JA, Ensrud K, Lane N, Hoffmann PR, Kopperdahl DL, Keaveny TM, for the Osteoporotic Fractures in Men (MrOS) Study Group (2009) Finite element analysis of the proximal femur and hip fracture risk in older men. *J Bone Miner Res* 24:475–483. <https://doi.org/10.1359/JBMR.081201>
- Ward J, Wood C, Rouch K, Pienkowski D, Malluche HH (2016) Stiffness and strength of bone in osteoporotic patients treated with varying durations of oral bisphosphonates. *Osteoporos Int* 27:2681–2688. <https://doi.org/10.1007/s00198-016-3661-0>
- Crawford RP, Cann CE, Keaveny TM (2003) Finite element models predict in vitro vertebral body compressive strength better than quantitative computed tomography. *Bone* 33:744–750. [https://doi.org/10.1016/S8756-3282\(03\)00210-2](https://doi.org/10.1016/S8756-3282(03)00210-2)
- Imai K, Ohnishi I, Bessho M, Nakamura K (2006) Nonlinear finite element model predicts vertebral bone strength and fracture site. *Spine (Phila Pa 1976)* 31:1789–1794. <https://doi.org/10.1097/01.brs.0000225993.57349.df>
- Zysset PK, Dall'Ara E, Varga P, Pahr DH (2013) Finite element analysis for prediction of bone strength. *Bonekey Rep* 2:1–9. <https://doi.org/10.1038/bonekey.2013.120>
- Wang X, Sanyal A, Cawthon PM, Palermo L, Jekir M, Christensen J, Ensrud KE, Cummings SR, Orwoll E, Black DM, for the Osteoporotic Fractures in Men (MrOS) Research Group, Keaveny TM (2012) Prediction of new clinical vertebral fractures in elderly men using finite element analysis of CT scans. *J Bone Miner Res* 27:808–816. <https://doi.org/10.1002/jbmr.1539>
- Graeff C, Marin F, Petto H, Kayser O, Reisinger A, Peña J, Zysset P, Glüer CC (2013) High resolution quantitative computed tomography-based assessment of trabecular microstructure and strength estimates by finite-element analysis of the spine, but not DXA, reflects vertebral fracture status in men with glucocorticoid-induced osteoporosis. *Bone* 52:568–577. <https://doi.org/10.1016/j.bone.2012.10.036>
- Melton LJ, Riggs BL, Keaveny TM et al (2010) Relation of vertebral deformities to bone density, structure, and strength. *J Bone Miner Res* 25:1922–1930. <https://doi.org/10.1002/jbmr.150>



15. Imai K, Ohnishi I, Matsumoto T, Yamamoto S, Nakamura K (2009) Assessment of vertebral fracture risk and therapeutic effects of alendronate in postmenopausal women using a quantitative computed tomography-based nonlinear finite element method. *Osteoporos Int* 20:801–810. <https://doi.org/10.1007/s00198-008-0750-8>
16. Melton LJ, Riggs BL, Keaveny TM et al (2007) Structural determinants of vertebral fracture risk. *J Bone Miner Res* 22:1885–1892. <https://doi.org/10.1359/jbmr.070728>
17. Kopperdahl DL, Aspelund T, Hoffmann PF, Sigurdsson S, Siggeirsdottir K, Harris TB, Gudnason V, Keaveny TM (2014) Assessment of incident spine and hip fractures in women and men using finite element analysis of CT scans. *J Bone Miner Res* 29:570–580. <https://doi.org/10.1002/jbmr.2069>
18. Hoffmann U, Massaro JM, Fox CS, Manders E, O'Donnell CJ (2008) Defining normal distributions of coronary artery calcium in women and men from the Framingham heart study. *Am J Cardiol* 102:1136–1141. <https://doi.org/10.1016/j.amjcard.2008.06.038>. Defining
19. Feinleib M, Kannel WB, Garrison RJ et al (1975) The Framingham offspring study. Design and preliminary data. *Prev Med (Baltim)* 4: 518–525. [https://doi.org/10.1016/0091-7435\(75\)90035-3](https://doi.org/10.1016/0091-7435(75)90035-3)
20. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D (2007) The third generation cohort of the National Heart, Lung, and Blood Institute's Framingham heart study: design, recruitment, and initial examination. *Am J Epidemiol* 165:1328–1335. <https://doi.org/10.1093/aje/kwm021>
21. Dawber TR, Meadors GF, Moore FE (1951) Epidemiological approaches to heart disease: the Framingham study. *Am J Public Health* 41:279–281. <https://doi.org/10.2105/AJPH.41.3.279>
22. Chan JJ, Cupples LA, Kiel DP, O'Donnell CJ, Hoffmann U, Samelson EJ (2015) QCT volumetric bone mineral density and vascular and valvular calcification: the Framingham study. *J Bone Miner Res* 30:1767–1774. <https://doi.org/10.1002/jbmr.2530>
23. Genant HK, Wu CY, van Kuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* 8:1137–1148. <https://doi.org/10.1002/jbmr.5650080915>
24. Yau MS, Demissie S, Zhou Y, Anderson DE, Lorbergs AL, Kiel DP, Allaire BT, Yang L, Cupples LA, Travison TG, Boussein ML, Karasik D, Samelson EJ (2016) Heritability of thoracic spine curvature and genetic correlations with other spine traits: the Framingham study. *J Bone Miner Res* 31:2077–2084. <https://doi.org/10.1002/jbmr.2925>
25. Keaveny TM, Donley DW, Hoffmann PF, Mitlak BH, Glass EV, San Martin JA (2007) Effects of teriparatide and alendronate on vertebral strength as assessed by finite element modeling of QCT scans in women with osteoporosis. *J Bone Miner Res* 22:149–157. <https://doi.org/10.1359/jbmr.061011>
26. Morgan EF, Keaveny TM (2001) Dependence of yield strain of human trabecular bone on anatomic site. *J Biomech* 34:569–577. [https://doi.org/10.1016/S0021-9290\(01\)00011-2](https://doi.org/10.1016/S0021-9290(01)00011-2)
27. Morgan EF, Bayraktar HH, Keaveny TM (2003) Trabecular bone modulus-density relationships depend on anatomic site. *J Biomech* 36:897–904. [https://doi.org/10.1016/S0021-9290\(03\)00071-X](https://doi.org/10.1016/S0021-9290(03)00071-X)
28. Kopperdahl DL, Morgan EF, Keaveny TM (2002) Quantitative computed tomography estimates of the mechanical properties of human vertebral trabecular bone. *J Orthop Res* 20:801–805
29. Engelke K, Adams JE, Armbrecht G, Augat P, Bogado CE, Boussein ML, Felsenberg D, Ito M, Prevrhal S, Hans DB, Lewiecki EM (2008) Clinical use of quantitative computed tomography and peripheral quantitative computed tomography in the management of osteoporosis in adults: the 2007 ISCD official positions. *J Clin Densitom* 11:123–162. <https://doi.org/10.1016/j.jocd.2007.12.010>
30. Schousboe JT, Shepherd JA, Bilezikian JP, Baim S (2013) Executive summary of the 2013 International Society for Clinical Densitometry Position Development Conference on bone densitometry. *J Clin Densitom* 16:455–466. <https://doi.org/10.1016/j.jocd.2013.08.004>
31. Kanis JA, McCloskey EV, Johansson H, Oden A, Melton LJ III, Khaltsev N (2008) A reference standard for the description of osteoporosis. *Bone* 42:467–475
32. Altman DG, Bland JM (1994) Diagnostic tests. 1: sensitivity and specificity. *BMJ* 308:1552. <https://doi.org/10.1136/bmj.308.6943.1552>
33. Keyak JH, Rossi SA, Jones KA, Skinner HB (1997) Prediction of femoral fracture load using automated finite element modelling. *J Biomech* 31:125–133. [https://doi.org/10.1016/S0021-9290\(97\)00123-1](https://doi.org/10.1016/S0021-9290(97)00123-1)
34. van Rietbergen B, Weinans H, Huiskes R, Odgaard A (1995) A new method to determine trabecular bone elastic properties and loading using micromechanical finite-elements models. *J Biomech* 28:69–81
35. Keyak JH, Sigurdsson S, Karlsdottir G, Oskarsdottir D, Sigmarsdottir A, Zhao S, Kornak J, Harris TB, Sigurdsson G, Jonsson BY, Siggeirsdottir K, Eiriksdottir G, Gudnason V, Lang TF (2011) Male-female differences in the association between incident hip fracture and proximal femur strength: a finite element analysis study. *Bone* 48:1239–1245. <https://doi.org/10.1016/j.bone.2011.03.682>. Male-female
36. Lindsay R, Silverman SL, Cooper C, Hanley DA, Barton I, Broy SB, Licata A, Benhamou L, Geusens P, Flowers K, Stracke H, Seeman E (2001) Risk of new vertebral fracture in the year following a fracture. *JAMA* 285:320–323. <https://doi.org/10.1001/jama.285.3.320>
37. Weber NK, Fidler JL, Keaveny TM, Clarke BL, Khosla S, Fletcher JG, Lee DC, Pardi DS, Loftus EV, Kane SV, Barlow JM, Murthy NS, Becker BD, Bruining DH (2014) Validation of a CT-derived method for osteoporosis screening in IBD patients undergoing contrast-enhanced CT enterography. *Am J Gastroenterol* 109:401–408. <https://doi.org/10.1038/ajg.2013.478>
38. Fidler JL, Murthy NS, Khosla S, Clarke BL, Bruining DH, Kopperdahl DL, Lee DC, Keaveny TM (2016) Comprehensive assessment of osteoporosis and bone fragility with CT colonography. *Radiology* 278:172–180. <https://doi.org/10.1148/radiol.2015141984>
39. Lee DC, Hoffmann PF, Kopperdahl DL, Keaveny TM (2017) Phantomless calibration of CT scans for measurement of BMD and bone strength—inter-operator reanalysis precision. *Bone* 103: 325–333. <https://doi.org/10.1016/j.bone.2017.07.029>
40. Adams A, Fischer H, Kopperdahl D et al (2017) The fracture, osteoporosis, and CT utilization study (FOCUS)—utilizing pre-existing CT to assess risk of hip fracture in a large real-world clinical setting. *J Bone Miner Res* 32(Suppl):1
41. Agten CA, Ramme AJ, Kang S, Honig S, Chang G (2017) Cost-effectiveness of virtual bone strength testing in osteoporosis screening programs for postmenopausal women in the United States. *Radiology* 161259:506–517. <https://doi.org/10.1148/radiol.2017161259>
42. Anderson DE, Demissie S, Allaire BT, Bruno AG, Kopperdahl DL, Keaveny TM, Kiel DP, Boussein ML (2014) The associations between QCT-based vertebral bone measurements and prevalent vertebral fractures depend on the spinal locations of both bone measurement and fracture. *Osteoporos Int* 25:559–566. <https://doi.org/10.1007/s00198-013-2452-0>