2023

Machine-learning assessed abdominal aortic calcification is associated with long-term fall and fracture risk in community-dwelling older Australian women

Jack Dalla Via
Edith Cowan University

Abadi Gebre
Edith Cowan University

Cassandra Smith
Edith Cowan University

Zulqarnain Gilani
Edith Cowan University

David Suter
Edith Cowan University

See next page for additional authors

Authors

Jack Dalla Via, Abadi Gebre, Cassandra Smith, Zulqarnain Gilani, David Suter, Naeha Sharif, Pawel Szulc, John T. Schousboe, Douglas P. Kiel, Kun Zhu, William D. Leslie, Richard L. Prince, Joshua R. Lewis, and Marc Sim

Machine-Learning Assessed Abdominal Aortic Calcification is Associated with Long-Term Fall and Fracture Risk in Community-Dwelling Older Australian Women

Jack Dalla Via,[1] Abadi K. Gebre,[1,2] Cassandra Smith,[1,3] Zulqarnain Gilani,[1,4,5] David Suter,[1,4] Naeha Sharif,[1,4,5] Pawel Szulc,[6] John T. Schousboe,[7] Douglas P. Kiel,[8] Kun Zhu,[3,9] William D. Leslie,[10] Richard L. Prince,[1,3] Joshua R. Lewis,[1,3,11] and Marc Sim[1,3]

[1]Nutrition and Health Innovation Research Institute, School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia
[2]School of Pharmacy, College of Health Sciences, Mekelle University, Mekelle, Ethiopia
[3]Medical School, The University of Western Australia, Perth, Western Australia, Australia
[4]Centre for Artificial Intelligence and Machine Learning, School of Science, Edith Cowan University, Perth, Western Australia, Australia
[5]Department of Computer Science and Software Engineering, the University of Western Australia, Perth, Western Australia, Australia
[6]INSERM UMR 1033, University of Lyon, Lyon, France
[7]Park Nicollet Clinic and HealthPartners Institute, HealthPartners, Minneapolis, USA and Division of Health Policy and Management, University of Minnesota, Minneapolis, MN, USA
[8]Hinda and Arthur Marcus Institute for Aging Research, Hebrew Senior Life, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
[9]Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Perth, Western Australia, Australia
[10]Departments of Medicine and Radiology, University of Manitoba, Winnipeg, MB, Canada
[11]Centre for Kidney Research, Children's Hospital at Westmead School of Public Health, Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia

ABSTRACT

Abdominal aortic calcification (AAC), a recognized measure of advanced vascular disease, is associated with higher cardiovascular risk and poorer long-term prognosis. AAC can be assessed on dual-energy X-ray absorptiometry (DXA)-derived lateral spine images used for vertebral fracture assessment at the time of bone density screening using a validated 24-point scoring method (AAC-24). Previous studies have identified robust associations between AAC-24 score, incident falls, and fractures. However, a major limitation of manual AAC assessment is that it requires a trained expert. Hence, we have developed an automated machine-learning algorithm for assessing AAC-24 scores (ML-AAC24). In this prospective study, we evaluated the association between ML-AAC24 and long-term incident falls and fractures in 1023 community-dwelling older women (mean age, 75 $\pm$ 3 years) from the Perth Longitudinal Study of Ageing Women. Over 10 years of follow-up, 253 (24.7%) women experienced a clinical fracture identified via self-report every 4–6 months and verified by X-ray, and 169 (16.5%) women had a fracture hospitalization identified from linked hospital discharge data. Over 14.5 years, 393 (38.4%) women experienced an injurious fall requiring hospitalization identified from linked hospital discharge data. After adjusting for baseline fracture risk, women with moderate to extensive AAC (ML-AAC24 $\geq$ 2) had a greater risk of clinical fractures (hazard ratio [HR] 1.42; 95% confidence interval [CI], 1.10–1.85) and fall-related hospitalization (HR 1.35; 95% CI, 1.09–1.66), compared to those with low AAC (ML-AAC24 $\leq$ 1). Similar to manually assessed AAC-24, ML-AAC24 was not associated with fracture hospitalizations. The relative hazard estimates obtained using machine learning were similar to those using manually assessed AAC-24 scores. In conclusion, this novel automated method for assessing AAC, that can be easily and seamlessly captured at the time

of bone density testing, has robust associations with long-term incident clinical fractures and injurious falls. However, the performance of the ML-AAC24 algorithm needs to be verified in independent cohorts. © 2023 The Authors. *Journal of Bone and Mineral Research* published by Wiley Periodicals LLC on behalf of American Society for Bone and Mineral Research (ASBMR).

KEY WORDS: VASCULAR CALCIFICATION; ARTIFICIAL INTELLIGENCE; FRACTURE; INJURIOUS FALLS; AGING

Introduction

Dual-energy X-ray absorptiometry (DXA) machines measure bone mineral density (BMD) to assess osteoporosis and risk of fracture, with very low radiation. DXA machines are also increasingly being used to perform vertebral fracture assessment (VFA) to visually identify asymptomatic, clinically unrecognized vertebral fractures at the time of bone density screening.[1] This involves obtaining lateral images of the thoracolumbar spine. These lateral spine images can also be used to identify abdominal aortic calcification (AAC), a marker of advanced atherosclerotic vascular disease, most commonly assessed using a semiquantitative 24-point scale (AAC-24).[2,3] More extensive AAC is associated with an increased risk of cardiovascular disease (CVD) and related events, CVD and all-cause mortality,[4-6] as well as lower BMD, poorer muscle strength, and increased risk for injurious falls and fracture.[7-11]

A major limitation for the clinical assessment of AAC is that it needs to be manually scored by a trained expert, a time-consuming task that limits the possibility for AAC assessment to be scaled and incorporated into routine clinical assessment, especially as part of osteoporosis screening. To address this limitation, machine learning (ML) approaches have been investigated to automate the identification of AAC and AAC-24 scoring from lateral spine images.[12,13] This builds on work to automate measurement of aortic calcification from abdominal computed tomography (CT) scans.[14] We recently developed and tested a ML algorithm for the automated assessment of AAC (ML-AAC24) from DXA-derived lateral spine images acquired from various major manufacturers (and models) currently used clinically. These scores were validated against future major adverse cardiovascular events, cardiovascular, and all-cause mortality from Hologic and GE machines.[13]

Given that extensive AAC is associated with poorer musculoskeletal outcomes (e.g., falls[7] and fracture[8]), determining whether these relationships remain comparable after automation (i.e., ML-AAC24) is a critical step to enhance the clinical utility of lateral spine images at the time of BMD measurement. Therefore, this study examined whether ML-AAC24 is related to long-term falls and fracture risk in community-dwelling older women, representing a high-risk population.

Methods

We have previously reported on the relationship between manually-assessed AAC with hospitalized fall[7] and fracture outcomes[8] in the study population as described in the following section. Development of the ML approach used in this study (ML-AAC24) has also been detailed.[13] The following section provides a brief overview.

Study population

The present study drew upon participants from the Perth Longitudinal Study of Ageing Women (PLSAW), which enrolled 1500 women aged 70 years or older in a 5-year, double-blind, randomized controlled trial (RCT) to determine the effect of calcium supplements on preventing osteoporotic fracture.[15] Following the trial, women were observed for a further 10 years. Women were excluded from the original RCT that forms the basis for our observational study analysis if they were taking any bone-active medication.[15] Lateral spine DXA images were obtained in 1024 women at baseline and/or 1 year of the RCT (1998/99). One participant's scan could not be read by our algorithm, leaving 1023 women for the current analysis. All 1023 women were included in fracture analyses and unadjusted and minimally-adjusted falls analysis, whereas 998 women were included in multivariable-adjusted falls analysis due to missing covariate data. Ethics approval for the 5-year RCT and a subsequent 10-year follow up was granted by the Human Research Ethics Committee at the University of Western Australia and the use of linked data was approved by Western Australian Department of Health (project number #2009/24). Written informed consent, including explicit authorization for future access to Western Australian Department of Health data, was obtained from all participants.

Baseline risk assessment

Body mass index (BMI, kg/m^2) was calculated from body weight measured using digital scales and height assessed using a wall-mounted stadiometer. Smoking status was classified as nonsmoker or ever smoked, which included individuals who had smoked at least one cigarette per day for 3 months or more, as well as current smokers. Participants' physical activity levels were assessed by asking them to report any sports, recreational activities, and regular physical activities they had participated in during the 3 months before their baseline visit. The energy expenditure in kJ/day for each participant was calculated, taking into account their body weight, using published energy costs for various activities.[16] The Socioeconomic Indexes for Areas developed by the Australian Bureau of Statistics was used to determine socioeconomic status by ranking residential postcodes based on their relative socioeconomic advantage/disadvantage. Participants were subsequently assigned to one of six groups, ranging from the most highly disadvantaged (top 10%) to the least disadvantaged (top 10%).[17] Medication use, including antihypertensives, statins, and oral hypoglycemic medications were confirmed by their primary care physician and classified according to the International Classification of Primary Care-Plus system (T89001–T90009). Prevalent diabetes at baseline was determined by insulin or oral hypoglycemic medication use. Prevalent atherosclerotic vascular disease (ASVD) was determined using primary discharge diagnoses from hospital data from 1980 to 1998 and retrieved from the Western Australian Data Linkage System, as described.[4] Atherosclerotic events were defined using primary diagnosis codes from the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM).[4] Previous falls were determined by asking participants if they had fallen in the 3 months prior to their baseline clinical visit.

Assessment of ACC

AAC was assessed from lateral spine images obtained at baseline or at 1-year (1998–1999). All lateral spine images were obtained from a Hologic 4500A bone densitometry machine, which captured digitally enhanced single-energy–derived lateral images of the thoracolumbar region. The ML-AAC24 scores used in the current study were obtained from our previous work focused on automated AAC-24 assessment for CVD outcomes, and the ML algorithm was only trained on expertly assessed AAC-24 scores.[13] A brief explanation of the training process is as follows[13]: a regression network was employed based on image features. This was based on the Kauppila AAC 24-point semiquantitative scoring method (AAC-24), which is the most widely used method to manually score AAC.[13,18] Briefly, the framework consists of three modules: image preprocessing, feature extraction, and regression. The pretrained EfficientNet-B3 was used for feature extraction due to its efficiency and high performance. The last fully connected convolutional layer of the EfficientNet-B3 model was replaced with a custom-designed regression network. Regression network was used to predict total AAC score on a continuous scale, ranging from 0 to 24, and was comprised of two dense layers with batch normalization and rectified linear unit activation function and a fully connected final layer with linear activation. This model was trained on the single-energy images from the Hologic 4500A machine.[13] The ML-AAC24 scores were categorized into three groups, based on previous work when assessed manually[4,8]: low (ML-AAC24 ≤ 1), moderate (ML-AAC24 ≥ 2 to <6), and extensive (ML-AAC24 ≥ 6). A 10-fold stratified cross-validation was performed to account for the imbalance in the different AAC categories in the training dataset. The average performance of the model was reported after 10 repetitions of training on nine folds, and testing on the remaining one as described.[13] For the current study, ML-AAC24 was categorized as low (ML-AAC24 ≤ 1) or moderate to extensive (ML-AAC24 ≥ 2), as in our previous work considering manually assessed AAC and fractures[8] in this cohort. As reported, there was substantial accuracy and agreement with the automatic ML-AAC24 scores and the manually assessed AAC-24 scores.[13] Specifically, Sharif et al.[13] reports that the Cohen's weighted kappa for manual AAC-24 versus ML-AAC24 assessed from lateral spine images obtained via the Hologic 4500A machine was 0.51, with an intraclass correlation coefficient of 0.76 (95% confidence interval [CI], 0.74–0.78). Performance of ML-AAC24 to classify individuals into the low versus moderate to extensive ML-AAC24 groups were also described[13]; accuracy 81.0% (79.8–82.0); sensitivity 82.9% (81.5–84.3); specificity 78.5% (76.7 to 80.2); positive predictive value 82.9% (81.7–84.0); and negative predictive value 78.5% (77.0–79.9). Of note, the agreement was substantial across different DXA makes and models (Hologic 4500A and Horizon, GE Lunar Prodigy and iDXA), suggesting that the algorithm trained on Hologic 4500A images is generalizable to other commonly used bone density machines.[13]

Prevalent and incident clinical fractures

Prevalent self-reported fractures were determined at baseline by asking the age and location of fracture sustained. Only fractures satisfying specific criteria were considered prevalent, including fractures that occurred after the age of 50 years and prior to participants baseline clinical visit due to minimal trauma (defined as falling from a height of 1 m or less), while also excluding fractures involving the face, skull, fingers, or toes. Prevalent vertebral fractures were determined from lateral single-energy images of the thoracolumbar spine, and scored using the Genant semiquantitative method.[19] A modification was that grade 1 fractures were considered fractures only if there was clear endplate depression or cortical discontinuity.[20] Any incident clinical fracture, including vertebral fractures, coming to clinical attention over 10 years were recorded in an adverse events diary, collected every 4 months during the first 5 years and every 6 months during the second 5 years. The diagnosis of all incident clinical vertebral and nonvertebral fractures was confirmed from radiographic records.

Fall and fracture-related hospitalization

Fall and fracture-related hospitalizations were collected from the Western Australia Hospital Morbidity Data Collection (HMDC) using the Western Australian Data Linkage System, which provides a complete validated record of every participant's primary diagnosis at hospital discharge using coded data from all hospitals in Western Australia. The HMDC records of all women were obtained from their baseline clinical visit in 1998 and over the next 14.5 years for hospitalized falls and 10 years for fracture-related hospitalization to enable comparisons to previous work in this cohort.[7,8] Diagnosis codes were defined using the ICD-9-CM codes 1998 to 1999[21] mapped to the ICD-10 Australian Modification (ICD-10-AM) for 1999 to 2013.[22] Hip and fracture-related hospitalizations were identified using the following ICD-10 codes: S02, S12, S22, S32, S42, S52, S62, S72, S82, S92, M80, T02, T08, T10, T12, and T14.2. Fractures of the face (S02.2–S02.6), fingers (S62.5–S62.7), and toes (S92.4–S92.5), as well as fractures caused by motor vehicle injuries (External Cause of Injury codes V00–V99) were excluded. Major osteoporotic fractures included those of the hip, spine, humerus, and wrist, as defined by the Fracture Risk Assessment Tool (FRAX). Fall-related hospitalizations were identified based on the international classification of external causes of injury codes and ICD-coded discharge data for all public and private inpatient admissions. Falls from standing height or less that were not caused by an external force, were identified using the following ICD-10 codes: W01, W05, W06, WO7, W08, W10, W18, and W19. The HMDS captures coded diagnosis data pertaining to all public and private inpatient contacts in Western Australia, it allows ascertainment of both fall and fracture-related hospitalizations independently of a patient report with the associated problems such as loss to follow-up.

Bone measurements

Quantitative ultrasound (QUS) of the left calcaneal was obtained in duplicate using a Lunar Achilles Ultrasound machine (Lunar Corp., Madison, WI, USA) at baseline (1998) in 989 women in this study. The average measurement of broadband ultrasound attenuation, speed of sound, and stiffness index were recorded as detailed.[23] Total hip and femoral neck BMD of subjects was measured by DXA (Hologic Acclaim 4500A fan-beam densitometer; Hologic Corp, Waltham, MA, USA). The coefficient of variation at the total hip was 1.2% in our laboratory.[8]

Estimated fracture risk

Fracture risk prediction using the FRAX website (https://frax.shef.ac.uk/FRAX/) was conducted using FRAX Australia. The estimated 10-year major osteoporotic fracture risk was calculated using baseline data, with and without BMD. The dataset comprised

the following clinical parameters: age, sex, weight (in kg), height (cm), fracture history (including previous fracture and parental history of hip fracture), current smoking status, glucocorticoid use, history of rheumatoid arthritis or secondary osteoporosis, and alcohol consumption (three or more units per day). Data were inputted with and without femoral neck BMD (g/cm^2).

Muscle strength and function assessment

Grip strength was assessed using an isometric Jamar hand dynamometer, with the peak value recorded from three attempts using the participants' dominant hand. The timed-up and-go test

(TUG), involves timing the participant's ability to rise from a chair, walk 3 m, turn, and return to sit on the chair. Participants were allowed to practice the test once before being timed.

Statistical analysis

Differences in baseline characteristics between the low and moderate to extensive ML-AAC24 categories were compared using independent t tests, the Mann–Whitney U test, or chi-square test where appropriate. Kaplan-Meier survival analysis was used to determine the univariate association of ML-AAC24 categories with 14.5-year fall and 10-year fracture-related

Table 1. Baseline Characteristics and Fall and Fracture Related Variables Stratified by Machine Learning Assessed Abdominal Aortic Calcification

Demographics	All participants	Low AAC (ML-AAC24 ≤ 1)	Moderate to extensive AAC (ML-AAC24 ≥ 2)
Number	1023	454	569
Age, years	75.0 ± 2.6	**74.7 ± 2.6**	**75.2 ± 2.6**
Body mass index (BMI)[a], kg/m^2	27.1 ± 4.4	**27.6 ± 4.6**	**26.7 ± 4.2**
Randomization			
Placebo, n (%)	511 (50.0)	234 (51.5)	277 (48.7)
Calcium, n (%)	484 (47.3)	208 (45.8)	276 (48.5)
Calcium plus vitamin D, n (%)	28 (2.7)	12 (2.6)	16 (2.8)
Smoker ever[a], n (%)	367 (36.0)	**136 (30.0)**	**231 (40.7)**
Physical activity, kJ/day	493 (186–868)	497 (191–867)	488 (181–871)
Prevalent atherosclerotic vascular disease, n (%)	112 (10.9)	**34 (7.5)**	**78 (13.7)**
Prevalent diabetes mellitus, n (%)	57 (5.6)	22 (4.8)	35 (6.2)
Statins, n (%)	194 (19.0)	**68 (15.0)**	**126 (22.1)**
Blood pressure lowering medication, n (%)	434 (42.4)	186 (41.0)	248 (43.6)
Socioeconomic status[b]			
Top 10% most highly disadvantaged, n (%)	43 (4.2)	15 (3.3)	28 (5.0)
Highly disadvantaged, n (%)	115 (11.3)	56 (12.4)	59 (10.5)
Moderate-highly disadvantaged, n (%)	168 (16.6)	86 (19.1)	82 (14.5)
Low-moderately disadvantaged, n (%)	151 (14.9)	71 (15.8)	80 (14.2)
Low disadvantaged, n (%)	221 (21.8)	97 (21.6)	124 (22.0)
Top 10% least disadvantaged, n (%)	316 (31.2)	125 (27.8)	191 (33.9)
Bone			
Hip			
Total hip BMD, g/cm^2	0.81 ± 0.12	**0.83 ± 0.12**	**0.80 ± 0.12**
Femoral neck BMD, g/cm^2	0.69 ± 0.10	**0.70 ± 0.10**	**0.69 ± 0.10**
MOF FRAX (Aust) without FN BMD, %	11.1 ± 4.7	**10.4 ± 4.4**	**11.7 ± 4.9**
MOF FRAX (Aust) with FN BMD, %	8.0 ± 3.8	**7.5 ± 3.3**	**8.4 ± 4.1**
Calcaneal[c]			
Broadband ultrasound attenuation, db/Mhz	100.6 ± 7.9	**101.5 ± 7.9**	**99.9 ± 7.8**
Speed of sound, m/sec	1513.4 ± 25.3	**1515.2 ± 25.3**	**1512.0 ± 25.2**
Stiffness index, %	70.8 ± 11.3	**71.9 ± 11.3**	**70.0 ± 11.2**
Fracture			
Prevalent fracture from age 50 years			
Any fracture, n (%)	263 (25.7)	108 (23.8)	155 (27.2)
Prevalent vertebral fracture from VFA[d]			
None, n (%)	929 (90.9)	418 (92.1)	511 (90.0)
One, n (%)	75 (7.3)	31 (6.8)	44 (7.7)
Two or more, n (%)	18 (1.8)	5 (1.1)	13 (2.3)
Grade of vertebral fracture[d]			
None, n (%)	929 (90.9)	418 (92.1)	511 (89.8)
Grade one only, n (%)	35 (3.4)	19 (4.2)	16 (2.8)
Grade two or three, n (%)	58 (5.7)	17 (3.7)	41 (7.2)

(Continues)

Table 1. Continued

Demographics	All participants	Low AAC (ML-AAC24 ≤ 1)	Moderate to extensive AAC (ML-AAC24 ≥ 2)
Location of vertebral fracture from VFA[d,e]			
Thoracic VF, n (%)	72 (7.0)	28 (6.2)	44 (7.7)
Lumbar VF, n (%)	24 (2.3)	9 (2.0)	15 (2.6)
Physical function			
Grip strength[f], kg	20.7 ± 4.6	20.7 ± 4.7	20.8 ± 4.5
Timed-up-and-go[g], sec	9.2 (8.0–10.8)	9.2 (8.1–10.7)	9.2 (8.0–10.8)
Falls			
Prevalent falls[h], n (%)	108 (10.7)	43 (9.6)	65 (11.6)

Note: Data expressed as mean ± SD, median (interquartile range), or number and (%). Bolded values represent significant differences (*p* value <0.05) between AAC categories using *t* test, chi-square test, or Mann–Whitney *U* test where appropriate.

Abbreviations: AAC = abdominal aortic calcification; BMC = bone mineral content; BMD = bone mineral density; FN = femoral neck; FRAX = fracture risk assessment tool; ML = machine learning; MOF = major osteoporotic fracture; VF = vertebral fracture; VFA = vertebral fracture assessment.

[a]*n* = 1020.
[b]*n* = 1014.
[c]*n* = 989.
[d]*n* = 1022.
[e]Some participants had both thoracic and lumbar spine fractures.
[f]*n* = 1016.
[g]*n* = 1021.
[h]*n* = 1008.

hospitalization, as well as incident self-reported clinical fractures. Cox proportional hazards regression models were used to investigate the association between ML-AAC24 categories with fall and fracture outcomes. To more closely align with our previous work,[7,8] the covariates included in the models differed for the falls and fracture analysis. For fractures, the model was adjusted for treatment (placebo/calcium) and FRAX with and without femoral neck BMD. For falls, the minimally-adjusted model included age, treatment, and BMI. The multivariable-adjusted model included the minimally-adjusted model plus prevalent ASVD, smoked ever, prevalent diabetes mellitus, statin use, blood pressure lowering medication use, socioeconomic status, physical activity, and self-reported prevalent falls. Treatment and FRAX with BMD fracture models as well as the multivariable-adjusted falls model were repeated using manually assessed AAC, to compare the association of ML-AAC24 and manual AAC with fracture and fall outcomes. No violations of the Cox proportional hazards assumptions were detected. All analyses were performed using IBM SPSS (Version 29; IBM Corp., Armonk, NY, USA) and STATA version 9.2 statistical software programs (Stata Corporation, College Station, TX, USA).

Additional analyses

To further explore the relationship between ML-AAC24 and falls and fracture outcomes, we analyzed fracture outcomes using a multivariable-adjusted model based on the covariates included for falls. However, the multivariable model differed slightly to that used for falls due to the inclusion of FRAX in the model, which includes age, BMI, smoking status, and prevalent fractures. Therefore, this revised multivariable fracture model included treatment code, FRAX with femoral neck BMD, prevalent ASVD, prevalent diabetes mellitus, statin use, blood pressure–lowering medication use, socioeconomic status, physical activity, and prevalent falls. We also analyzed the relationship of AAC with fall and fracture outcomes using two alternative categorizations of ML-AAC24 used previously.[4,7] Specifically, we compared AAC presence (ML-AAC24 ≥ 1) and absence (ML-AAC24 = 0), as well

as across three groups of AAC extent (low [ML-AAC24 ≤ 1], moderate [ML-AAC24 ≥ 2 to <6], and extensive [ML-AAC24 ≥ 6]).

Results

A total of 1023 women were included in the current study, with a mean ± standard deviation (SD) age of 75.0 ± 2.6 years at baseline. Baseline characteristics of the total sample, as well as for those with low and moderate to extensive ML-AAC24, are presented in Table 1. Compared with women with low ML-AAC24, women with moderate to extensive ML-AAC24 were slightly older, had lower BMI, were more commonly current or former smokers, and were more likely to have prevalent ASVD and be taking lipid-lowering medication. Compared with women with low ML-AAC24, women with moderate to extensive ML-AAC24 had lower hip BMD, lower ultrasound attenuation and stiffness index, and higher estimated major osteoporotic fracture risk. However, the proportion of women with any prevalent osteoporotic fracture since the age of 50 years, as well as the number, grade, or location of prevalent vertebral fractures did not differ (Table 1). Grip strength, TUG, and the proportion of women with a self-reported prevalent fall in the 3 months prior to baseline did not differ between women with low or moderate to extensive ML-AAC24 (Table 1).

ML-AAC24 and 10-year clinical fractures

Over 10 years of follow-up, 24.7% (253/1023) of women experienced a clinical fracture (8124 person-years; mean ± SD follow-up 7.9 ± 3.0 years), and 16.5% (169/1023) of women had a fracture hospitalization (9111 person-years; mean ± SD follow-up 8.9 ± 2.2 years). Compared with women with low ML-AAC24, women with moderate to extensive ML-AAC24 had an increased risk of any clinical fracture, including vertebral and nonvertebral fractures (Table 2, Fig. 1). Hospitalization risk due to any fracture or hip fracture did not differ between women with low or moderate to extensive ML-AAC24. These results were consistent across all models of adjustment.

ML-AAC24 and 14.5-year fall-related hospitalizations

Over 14.5 years of follow-up (11,548 person-years; mean ± SD follow-up 11.3 ± 3.9 years), 38.4% (393/1023) of women experienced a fall-related hospitalization. Compared with women with low ML-AAC24, women with moderate to extensive ML-AAC24 had between 35% and 37% increased hazard for a fall-related hospitalization in the multivariable-adjusted model (Table 2, Fig. 1).

ML-AAC24 compared with manually assessed AAC24

Hazard ratios (HRs) for ML-AAC24 and 10-year incident fractures were consistent when AAC groupings based on manually assessed AAC24 was used in the current fracture analysis (Fig. 2). Similarly, HRs for ML-AAC24 and 14.5-year fall-related hospitalizations were consistent when manually-assessed AAC24 was used in the multivariable-adjusted model (Fig. 2).

Additional analyses

When additional covariates were included in the analysis, to align with the primary multivariable-adjusted analysis for falls, HRs for moderate to extensive compared with low ML-AAC24 for all fracture outcomes were nearly identical (Table S1). When AAC was categorized as presence and absence (Table S2), or into three categories of extent (Table S3), results were largely consistent with that of the primary analysis. However, for presence versus absence of ML-AAC24, the relationship with clinical vertebral fractures was not significant, despite the HRs being similar to the primary analysis. Across three categories of AAC extent, extensive but not moderate AAC was associated with an increased risk of clinical vertebral fracture, compared with low AAC. Despite, extensive AAC being associated with increased risk of clinical nonvertebral fracture and hip fracture hospitalization, neither association persisted once adjusted for FRAX with femoral neck BMD.

Table 2. Hazard Ratios (HRs) for Fracture Outcomes Over 10 Years and Fall-Related Hospitalization Risk Over 14.5 Years By Machine Learning Assessed Categories of Abdominal Aortic Calcification

	Number of events (%)	Unadjusted		Adjusted for treatment and FRAX without FN BMD		Adjusted for treatment and FRAX with FN BMD	
		HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p
Clinical fracture[a]							
Low AAC	90/454 (19.8)	**1 (reference)**		**1 (reference)**		**1 (reference)**	
Moderate to extensive AAC	163/569 (28.6)	**1.54 (1.19–1.99)**	**0.001**	**1.48 (1.14–1.91)**	**0.003**	**1.42 (1.10–1.85)**	**0.008**
Clinical vertebral fracture							
Low AAC	20/454 (4.4)	**1 (reference)**		**1 (reference)**		**1 (reference)**	
Moderate to extensive AAC	48/569 (8.4)	**1.94 (1.15–3.28)**	**0.013**	**1.82 (1.08–3.07)**	**0.026**	**1.74 (1.03–2.95)**	**0.039**
Clinical nonvertebral fracture							
Low AAC	74/454 (16.3)	**1 (reference)**		**1 (reference)**		**1 (reference)**	
Moderate to extensive AAC	136/569 (23.9)	**1.55 (1.17–2.06)**	**0.002**	**1.48 (1.11–1.96)**	**0.007**	**1.43 (1.08–1.90)**	**0.014**
Fracture hospitalization							
Low AAC	64/454 (14.1)	1 (reference)		1 (reference)		1 (reference)	
Moderate to extensive AAC	105/569 (18.5)	1.34 (0.98–1.82)	0.068	1.23 (0.90–1.69)	0.192	1.17 (0.85–1.60)	0.341
Hip fracture hospitalization							
Low AAC	23/454 (5.1)	1 (reference)		1 (reference)		1 (reference)	
Moderate to extensive AAC	38/569 (6.7)	1.33 (0.80–2.24)	0.275	1.28 (0.76–2.16)	0.349	1.19 (0.70–2.00)	0.526

	Number of events (%)	Unadjusted		Minimally adjusted[b]		Multivariable adjusted[c]	
		HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
Injurious fall-related hospitalization							
Low AAC	150/454 (33.0)	**1 (reference)**		**1 (reference)**		**1 (reference)**	
Moderate to extensive AAC	243/569 (42.7)	**1.44 (1.17–1.76)**	**<0.001**	**1.37 (1.12–1.68)**	**0.003**	**1.35 (1.09–1.66)**	**0.006**

Note: Bolded values represent significant differences (*p* < 0.05). HRs (95% CI) analyzed using Cox-proportional hazard models. AAC categorized as low (ML-AAC24 ≤ 1) and moderate to extensive (ML-AAC24 ≥ 2). n = 1023 for all fracture models, and for unadjusted and minimally adjusted falls models, n = 998 for multivariable adjusted falls model.

Abbreviations: AAC = abdominal aortic calcification; BMD = bone mineral density; FN = femoral neck; FRAX = fracture risk assessment tool; HR = hazard ratio; ML = machine learning.

[a]Clinical vertebral and nonvertebral fracture numbers do not equal total clinical fracture numbers as women may have suffered more than one fracture type.

[b]Minimally adjusted = age, treatment code and BMI.

[c]Multivariable adjusted = minimally adjusted model plus prevalent atherosclerotic vascular disease, smoked ever, prevalent diabetes mellitus, statin use, blood pressure lowering medication use, socioeconomic status, physical activity, and self-reported prevalent falls.
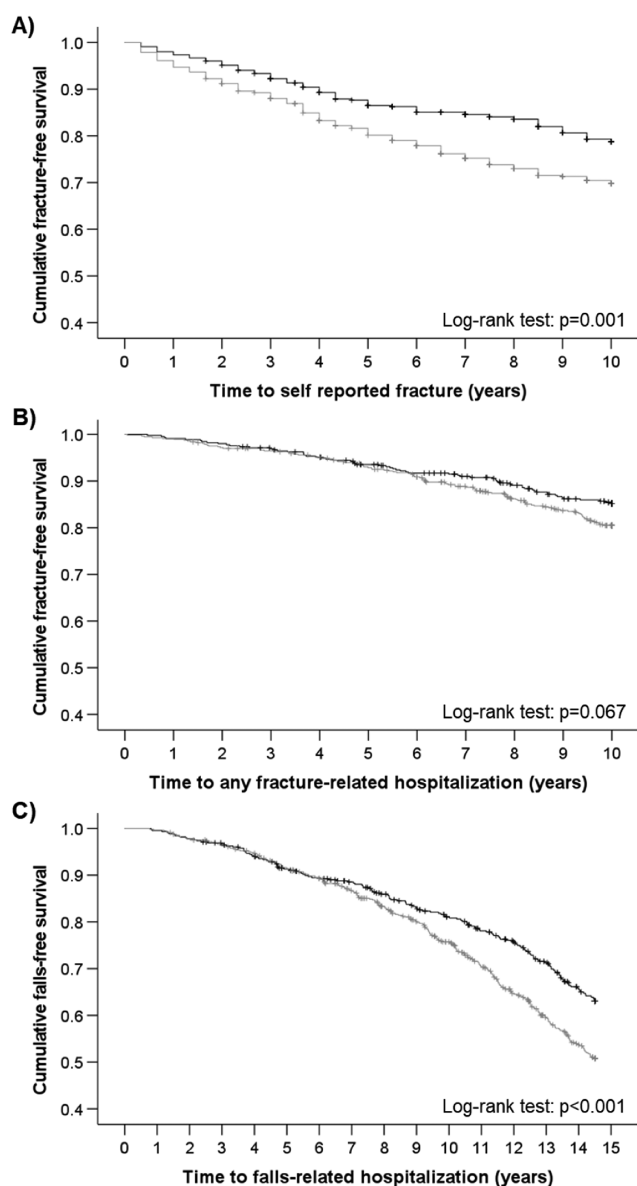
A)

B)

C)

Fig. 1. Kaplan–Meier survival curve for abdominal aortic calcification (AAC) score categories for (*A*) 10-year incident self-reported fractures, (*B*) 10-year incident fracture-related hospitalizations, and (*C*) 14.5-year incident fall-related hospitalizations. Low AAC (ML-AAC24 ≤ 1) and moderate to extensive AAC (ML-AAC24 ≥ 2) categories are represented by black and gray lines, respectively.

Discussion

This study reports that older women with moderate to extensive AAC, assessed using an automated ML algorithm, had an increased risk of incident fall-related hospitalizations and clinical fractures, compared with women with low ML-AAC24. Of note, ML-AAC24 was not associated with incident fracture-related hospitalizations. These findings remained consistent after adjusting for various covariates. Women with moderate to extensive ML-AAC24 also had lower BMD compared to women with low ML-AAC24. However, neither TUG performance nor prevalent fractures or
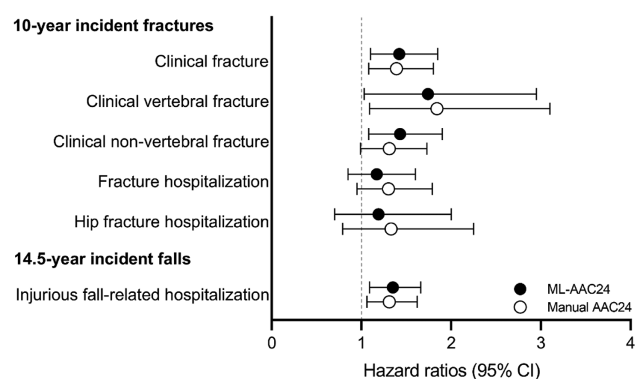

Fig. 2. Hazard ratios (HRs) for fracture outcomes over 10 years and fall-related hospitalization risk over 14.5 years using machine learning AAC24 (ML-AAC24) and manually assessed AAC-24. HRs are for moderate to extensive AAC (AAC-24 ≥ 2) compared with low AAC (AAC-24 ≤ 1) as the reference category. The fracture models ($n = 1023$) are adjusted for treatment code and FRAX with femoral neck bone mineral density (BMD). The falls model ($n = 998$) is adjusted for age, treatment code, body mass index (BMI), prevalent atherosclerotic vascular disease, smoked ever, prevalent diabetes mellitus, statin use, blood pressure–lowering medication use, socioeconomic status, physical activity, and self-reported prevalent falls.

falls differed between ML-AAC24 groups. These findings support the use of ML-AAC24 to routinely assess AAC on lateral spine images in clinical practice, to provide important prognostic information regarding clinical musculoskeletal outcomes.

The potential mechanisms underlying the observed relationships between AAC and falls and fracture risk have been discussed.[7,8,11] Briefly, this may relate to impaired blood flow to the skeleton distal to the abdominal aorta negatively influencing BMD, which may contribute to increasing fracture risk. Vascular disease may also contribute to increased falls propensity, and subsequent fracture, via various mechanisms. For example, vascular disease may lead to increased prescription of certain cardiovascular medications that may increase falls risk due to side effects such as orthostatic hypotension, confusion, or dehydration.[24] Falls may also result from transient ischemic events or syncope, or cardiac abnormalities such as arrhythmias and reduced cardiac output, which may be more common in individuals with vascular disease.[24] There are also shared genetic and environmental risk factors for muscle, bone, and vascular health.[7,8,11] However, the potential mechanisms require further investigation. Although moderate to extensive ML-AAC24 was associated with incident self-reported clinical fractures, it was not associated with incident fracture-related hospitalizations. This was similar to our previous analysis[8] where manually assessed AAC was not associated with fracture hospitalizations after accounting for BMD. In our additional analyses, extensive but not moderate ML-AAC24 was associated with an increased risk of hip fracture hospitalization; however, this was attenuated when FRAX with BMD was included in the model. These results may be in part due to reduced statistical power because there were fewer fracture hospitalizations, and particularly hip fracture hospitalizations, compared with clinical fractures. The HRs for ML-AAC24 and incident fracture and falls risk in this study were largely consistent with those observed using manually assessed

AAC in the analysis. Furthermore, the observed HRs for moderate to extensive ML-AAC24 and incident fracture outcomes are consistent with those reported in a recent meta-analysis of 86 studies and 61,553 participants.[11] In 28 studies (n = 33,748 participants) reporting fracture risk with any/advanced AAC compared with no/low AAC, the higher AAC group had a pooled relative risk of 1.51 (95% CI, 1.25–1.82) for any incident fracture.[11] In support of the current results showing poorer bone health with moderate to extensive ML-AAC24, a meta-analysis of 30 studies showed that those with any/advanced AAC had lower BMD at the total hip, femoral neck, and lumbar spine.[11] Overall, falls and fracture risk estimates using ML-AAC24 appear consistent with manually assessed AAC, which aligns with the good levels of agreement reported between the ML-AAC24 and AAC scored by trained imaging specialists.[13]

We and others have shown that the presence and extent of manually assessed AAC is associated with impaired musculoskeletal health,[7,8,11] increased risk for all-cause mortality and cardiac events,[6] as well as late-life dementia hospitalizations and deaths.[25] Recently, we reported that women with higher ML-AAC24 had increased risk of all-cause mortality and CVD-related mortality.[13] The present study builds on this evidence, demonstrating that higher ML-AAC24 is also associated with increased incident falls and clinical fracture risk. Statin use has been proposed to increase vascular calcification,[26,27] and there were a higher proportion of statin users in the moderate to extensive ML-AAC24 group of the current study. However, when statin use was included as a covariate in multivariable adjusted falls and fracture models, the point estimates were nearly identical for all outcomes, suggesting the observed associations were independent of statin use. Importantly, the predictive value of ML-AAC24 does not appear to decrease with increasing follow-up duration, suggesting it is a valuable measure of long-term fall and fracture risk. However, the predictive value appears to be lower for imminent falls and fractures. In both the previous[13] analysis and present analyses on falls and fractures, the observed associations with ML-AAC24 were consistent with that of manually assessed AAC. Together, these results highlight that ML-AAC24 provides important prognostic information about cardiovascular and musculoskeletal health, which is similar to that obtained when AAC is assessed manually by an experienced imaging specialist.

Clinical guidelines increasingly recommend performing lateral spine images at the time of DXA testing to identify asymptomatic vertebral fractures.[1] Further, even if a lateral spine image is obtained, AAC is not routinely assessed in clinical practice, largely due to practical considerations such as the time taken for manual scoring and the lack of expert assessors globally who are appropriately trained to score AAC. Therefore, methods to automate AAC assessment are needed to address these key barriers and support more widespread and routine use of AAC clinically at the time of osteoporosis screening. For example, although a lateral spine DXA scan takes 2–5 min to reposition and scan (depending on make and model), it takes even the most experienced reader approximately 5–6 min to obtain an AAC score from the image. In comparison, our recently developed ML algorithm takes less than a minute to predict the AAC score for hundreds of images. The ultimate goal of this work is for an automated AAC assessment to be incorporated into the software of common DXA machines. Although Hologic DXA scanners include software that helps to more easily visualize AAC from lateral spine images, it does not automatically detect or quantify AAC, so still requires considerable time and expertise to obtain an AAC-24 score. An automated algorithm would allow AAC to be instantaneously assessed and reported on in clinical practice for any lateral spine image captured. Importantly, our algorithm was shown to have good levels of agreement with AAC that was manually assessed by trained experts.[13] This is consistent with the reported agreement between automatic and manual scoring of aortic calcification from abdominal CT scans.[14] The current study builds on this work demonstrating that ML-AAC24 provides prognostic information about falls and fracture risk in older women. Given bone density scans involve a very low radiation dose and are routinely used for osteoporosis screening, performed most commonly in older women, the ability of this ML approach to provide AAC readings at the time of bone density testing is of clear clinical value, especially in the context of opportunistic health screening for other health outcomes including all-cause mortality and cardiovascular events,[6] as well as late-life dementia hospitalizations and deaths.[25]

Strengths of the present study include the large number of older Australian women included, who are representative of the population typically undergoing bone densitometry for osteoporosis screening, the long-term prospective follow-up (10 and 14.5 years for fracture and falls outcomes, respectively), the capture of clinical fracture events through verified self-report (by General Practitioner), as well as the capture of fall- and fracture-related hospitalizations from the Western Australian Data Linkage System independent of self-report. There are also several limitations that should be considered. First, as acknowledged in the original study,[13] in the absence of an unseen hold-out separate dataset, the machine learning algorithm to assess AAC-24 scores was trained on the same set of images that were used in this study. However, to mitigate this limitation, a 10-fold cross-validation approach was employed to ensure that the algorithm was evaluated on a data partition (fold) that it had not seen during training, as described.[13] This process was repeated 10 times to obtain the test AAC24 scores across the entire dataset, ensuring that the algorithm's performance was assessed on unseen data with no prior knowledge of the test images. An important next step of this work will be to verify the performance of the ML-AAC24 algorithm in independent cohorts. Additionally, it is important to note that the algorithm was trained only on AAC-24 scores and has not been optimized/trained to predict long-term fall- or fracture-related risks/hospitalizations. Second, this was an observational study so causality cannot be established. Third, we cannot exclude the possibility of bias being introduced, particularly as the lateral spine scans were only completed in approximately 70% of the larger cohort. Finally, the cohort consisted predominantly of community-dwelling older white women with a mean age of 75 years; therefore, these findings may not be generalizable to younger women, men, older women of other ethnicities, or to any clinical populations.

In conclusion, we showed that higher ML-assessed AAC was associated with poorer bone health and an increased risk of incident clinical fractures and fall-related hospitalizations. Compared to manually assessed AAC, ML-AAC24 provides similar risk estimates for incident falls and clinical fractures. This work supports the assessment of ML-AAC24 on lateral spine images in clinical practice to identify individuals at greatest need of interventions to mitigate declines in musculoskeletal health to prevent falls and fracture, independent of traditional fall and fracture risk factors.

Author Contributions

Jack Dalla Via: Writing – original draft; writing – review and editing; formal analysis; conceptualization; methodology. **Abadi K Gebre:** Writing – review and editing; formal analysis; conceptualization. **Cassandra Smith:** Writing – review and editing; methodology. **Zulqarnain Gilani:** Writing – review and editing; software; conceptualization. **David Suter:** Writing – review and editing; software; conceptualization. **Naeha Sharif:** Writing – review and editing; software; conceptualization. **Pawel Szulc:** Writing – review and editing; conceptualization; methodology. **John T Schousboe:** Writing – review and editing; investigation; conceptualization; methodology. **Douglas P Kiel:** Writing – review and editing; conceptualization; methodology. **Kun Zhu:** Writing – review and editing; methodology; investigation; data curation. **William D Leslie:** Writing – review and editing; conceptualization; methodology. **Richard L Prince:** Writing – review and editing; funding acquisition; conceptualization; project administration; methodology; supervision; investigation; data curation; resources. **Joshua R Lewis:** Writing – review and editing; conceptualization; supervision; data curation; methodology; investigation; funding acquisition. **Marc Sim:** Writing – review and editing; writing – original draft; conceptualization; formal analysis; supervision; methodology; data curation; funding acquisition.

Acknowledgments

Disclosures

The authors declare no conflicts of interest.

Peer Review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/jbmr.4921.

Data Availability Statement

Data is available from the authors upon reasonable request inline with governing ethical considerations.

References

1. Schousboe JT, Vokes T, Broy SB, et al. Vertebral fracture assessment: the 2007 ISCD official positions. J Clin Densitom. 2008;11(1):92–108.
2. Schousboe JT, Lewis JR, Kiel DP. Abdominal aortic calcification on dual-energy X-ray absorptiometry: methods of assessment and clinical significance. Bone. 2017;104:91–100.
3. Schousboe JT, Wilson KE, Hangartner TN. Detection of aortic calcification during vertebral fracture assessment (VFA) compared to digital radiography. PloS One. 2007;2(8):e715.
4. Lewis JR, Schousboe JT, Lim WH, et al. Long-term atherosclerotic vascular disease risk and prognosis in elderly women with abdominal aortic calcification on lateral spine images captured during bone density testing: a prospective study. J Bone Miner Res. 2018;33(6):1001–1010.
5. Schousboe JT, Taylor BC, Kiel DP, Ensrud KE, Wilson KE, McCloskey EV. Abdominal aortic calcification detected on lateral spine images from a bone densitometer predicts incident myocardial infarction or stroke in older women. J Bone Miner Res. 2008;23(3):409–416.
6. Leow K, Szulc P, Schousboe JT, et al. Prognostic value of abdominal aortic calcification: a systematic review and meta-analysis of observational studies. J Am Heart Assoc. 2021;10(2):e017205.
7. Gebre AK, Sim M, Rodríguez AJ, et al. Abdominal aortic calcification is associated with a higher risk of injurious fall-related hospitalizations in older Australian women. Atherosclerosis. 2021;328:153–159.
8. Lewis JR, Eggermont CJ, Schousboe JT, et al. Association between abdominal aortic calcification, bone mineral density, and fracture in older women. J Bone Miner Res. 2019;34(11):2052–2060.
9. Rodríguez AJ, Lewis JR, Scott DS, et al. Aortic calcification is associated with five-year decline in handgrip strength in older women. Calcif Tissue Int. 2018;103:589–598.
10. Szulc P, Blackwell T, Schousboe JT, et al. High hip fracture risk in men with severe aortic calcification: MrOS study. J Bone Miner Res. 2014;29(4):968–975.
11. Gebre AK, Lewis JR, Leow K, et al. Abdominal aortic calcification, bone mineral density and fractures: a systematic review and meta-analysis of observational studies. J Gerontol A Biol Sci Med Sci. 2023;78(7):1147–1154.
12. Reid S, Schousboe JT, Kimelman D, Monchka BA, Jozani MJ, Leslie WD. Machine learning for automated abdominal aortic calcification scoring of DXA vertebral fracture assessment images: a pilot study. Bone. 2021;148:115943.
13. Sharif N, Gilani SZ, Suter D, et al. Machine learning for abdominal aortic calcification assessment from bone density machine-derived lateral spine images. eBioMedicine. 2023;94:104676.
14. Graffy PM, Liu J, O'Connor S, Summers RM, Pickhardt PJ. Automated segmentation and quantification of aortic calcification at abdominal CT: application of a deep learning-based algorithm to a longitudinal screening cohort. Abdom Radiol. 2019;44:2921–2928.
15. Prince RL, Devine A, Dhaliwal SS, Dick IM. Effects of calcium supplementation on clinical fracture and bone structure: results of a 5-year, double-blind, placebo-controlled trial in elderly women. Arch Intern Med. 2006;166(8):869–875.
16. Bruce DG, Devine A, Prince RL. Recreational physical activity levels in healthy older women: the importance of fear of falling. J Am Geriatr Soc. 2002;50(1):84–89.
17. Australian Bureau of Statistics. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia. Canberra, Australia: Australian Bureau of Statistics; 1991.
18. Kauppila LI, Polak JF, Cupples LA, Hannan MT, Kiel DP, Wilson PW. New indices to classify location, severity and progression of calcific lesions in the abdominal aorta: a 25-year follow-up study. Atherosclerosis. 1997;132(2):245–250.

19. Genant HK, Wu CY, Van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res. 1993; 8(9):1137–1148.

20. Cawthon PM, Haslam J, Fullman R, et al. Methods and reliability of radiographic vertebral fracture detection in older men: the osteoporotic fractures in men study. Bone. 2014;67: 152–155.

21. World Health Organization. Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death: Based on the Recommendations of the Ninth Revision Conference, 1975, and Adopted by the Twenty-Ninth World Health Assembly. Geneva, Switzerland: World Health Organization; 1977.

22. World Health Organization. International Statistical Classification of Diseases and Related Health Problems: Alphabetical Index. Geneva, Switzerland: World Health Organization; 2004.

23. Gebre AK, Prince R, Schousboe J, et al. Calcaneal quantitative ultrasound is associated with all-cause and cardiovascular disease mortality independent of hip bone mineral density. Osteoporos Int. 2022;33(7):1557–1567.

24. Denfeld QE, Turrise S, MacLaughlin EJ, et al. Preventing and managing falls in adults with cardiovascular disease: a scientific statement from the American heart association. Circ Cardiovasc Qual Outcomes. 2022;15(6):e000108.

25. Porter T, Sim M, Prince RL, et al. Abdominal aortic calcification on lateral spine images captured during bone density testing and late-life dementia risk in older women: a prospective cohort study. Lancet Reg Health West Pac. 2022;26:100502.

26. Puri R, Nicholls SJ, Shao M, et al. Impact of statins on serial coronary calcification during atheroma progression and regression. J Am Coll Cardiol. 2015;65(13):1273–1282.

27. Xian JZ, Lu M, Fong F, et al. Statin effects on vascular calcification: microarchitectural changes in aortic calcium deposits in aged hyperlipidemic mice. Arterioscler Thromb Vasc Biol. 2021;41(4): e185–e192.