Full Length Article

Evaluation of patient tissue selection methods for deriving equivalent density calibration for femoral bone quantitative CT analyses

C. Winsor [a], X. Li [b,c,*], M. Qasim [b,c], C.R. Henak [a], P.J. Pickhardt [d], H. Ploeg [a,e], M. Viceconti [b,c,f,g]

[a] Mechanical Engineering, University of Wisconsin, USA
[b] Mechanical Engineering, University of Sheffield, UK
[c] INSIGNEO Institute for in silico Medicine, University of Sheffield, UK
[d] Radiology, University of Wisconsin, USA
[e] Mechanical and Materials Engineering, Queen's University, Canada
[f] Department of Industrial Engineering, Alma Mater Studiorum – University of Bologna, Italy
[g] Medical Technology Lab, IRCCS Rizzoli Orthopaedic Institute, Italy

ARTICLE INFO

ABSTRACT

Osteoporosis affects an increasing number of people every year and patient specific finite element analysis of the femur has been proposed to identify patients that could benefit from preventative treatment. The aim of this study was to demonstrate, verify, and validate an objective process for selecting tissues for use as the basis of phantomless calibration to enable patient specific finite element analysis derived hip fracture risk prediction. Retrospective reanalysis of patient computed tomography (CT) scans has the potential to yield insights into more accurate prediction of osteoporotic fracture. Bone mineral density (BMD) specific calibration scans are not typically captured during routine clinical practice. Tissue-based BMD calibration can therefore empower the retrospective study of patient CT scans captured during routine clinical practice. Together the method for selecting tissues as the basis for phantomless calibration coupled with the post-processing steps for deriving a calibration equation using the selected tissues provide an estimation of quantitative equivalent density results derived using calibration phantoms. Patient tissues from a retrospective cohort of 211 patients were evaluated. The best phantomless calibration resulted in a femoral strength (FS) [N] bias of 0.069 ± 0.07% over FS derived from inline calibration and a BMD [kg/cm^3] bias of 0.038 ± 0.037% over BMD derived from inline calibration. The phantomless calibration slope for the best method presented was within the range of patient specific calibration curves available for comparison and demonstrated a small bias of 0.028 ± 0.054 HU/(mg/cm^3), assuming the Mindways Model 3 BMD inline calibration phantom as the gold standard. The presented method of estimating a calibration equation from tissues showed promise for CT-based femoral fracture analyses of retrospective cohorts without readily available calibration data.

1. Introduction

Over 300,000 people experience an osteoporotic femoral fracture in the U.S. every year [1]. Despite available treatments, osteoporosis remains underdiagnosed [2], inspiring research towards a better understanding of osteoporotic fracture. In addition, the stratification accuracy of the prognostic standard of care (bone densitometry) is too low to reliably diagnose osteopenic patients, and to decide when to adopt second-line treatments such as Denosumab or Teriparatide [2,3]. This calls for more accurate prognostic methodologies. Various groups proposed quantitative computed tomography (QCT) based patient specific

finite element analyses (FEAs) for improved osteoporotic hip fracture risk assessment [4–6]. These FEAs have been shown to predict risk of hip fracture more accurately than areal bone mineral density (BMD) [4]. Retrospective reanalysis of patient computed tomography (CT) scans will further assist in the development of techniques to predict risk of osteoporotic fracture, potentially leading to improved prognostic accuracy. However, these models depend on the estimation of bone material properties, derived from CT X-ray attenuation. In phantom-based calibration this is achieved by placing an inline calibration phantom under the patient or by scanning offline a calibration phantom immediately after the patient, using the same CT scan settings. Phantom-based calibration is the gold standard in the development of patient specific FEAs.

Nomenclature

CT	computed tomography
QCT	quantitative computed tomography
ρ_{CT}	equivalent, CT or radiological density
ρ_{QCT}	quantitative equivalent density
BMD	bone mineral density
aBMD	areal bone mineral density
ρ_{app}	apparent density, wet
ρ_{ash}	ash density
ρ_{eff}	effective density
ρ_{HA}	effective calcium hydroxyapatite density
ρ_{K}	apparent density
ρ_{K2HPO4}	effective dipotassium phosphate density
ρ_{nom}	Nominal density
HA	calcium hydroxyapatite
HU	Hounsfield Units
FEA	finite element analysis
GCF	general calibration factor
FS	femoral strength
CTXA	computed tomography X-ray absorptiometry
SSE	sum of squared error
RMS	root mean square error

However, scanning the patient with an inline phantom is not a standard clinical practice, and delayed offline retrospective calibration is not always possible due to clinics regularly purchasing new CT scanners. Phantomless CT scan calibration can be derived from patient tissues and could therefore be a feasible alternative.

Before considering literature on existing phantomless methods, several variables should be identified and defined. There are several points in the process of capturing a CT scan that affect density assessment including: underlying theory and definitions, the chemical composition of the object being scanned, the acquisition settings and the reconstruction algorithms. Considering underlying theory, clinical CT images describe materials' X-ray attenuation in greyscale in terms of the Hounsfield Scale (in units HU),

$$CT\ Number = ((\mu_T - \mu_{water})/(\mu_{water} - \mu_{air}))*1000\ [HU] \tag{1}$$

Here μ, X-ray attenuation from the object, represents

$$\mu(E) = a_1 PE(E) + a_2 CS(E) = m_1\mu_1(E) + m_2\mu_2(E) \tag{2}$$

where E is the X-ray energy level, PE is the photoelectric basis function, CS is the Compton scattering effect basis function, and μ_1 and μ_2 are any two independent materials [7]. Compton scatter affects the definition of the Hounsfield scale such that X-ray attenuation measurements are roughly linearly proportional to density [8]. By definition this provides the basis for a linear estimate of the relationship between X-ray attenuation measurements and BMD [8]. Mathematically speaking, CT Numbers are non-unique and thus a plastic-composite mimicking BMD results in a similar measurement to scanning actual bone. By scanning a phantom of known chemical composition at a single energy, the variables can be simplified such that density can be calculated from X-ray attenuation measurements. After initial X-ray attenuation measurements have been captured, reconstruction algorithms generate a clearer image of a specific density range (i.e. soft tissue or bone). All of these variables impact the derivation of a conversion, between BMD and CT X-ray attenuation, that can be derived from CT X-ray attenuation measurements of a calibration phantom scanned in line with the patient [9,10]. Recently, some studies have begun to discuss how specific details of clinical CT scan protocols affect density estimates by examining repeatability [11], patient positioning [12], and reconstruction kernel [13–15].

Different inline calibration phantoms have appeared in previous studies [16–23]. These phantoms contain either calcium hydroxyapatite [21–23] $(Ca_5(PO_4)_3)$, abbreviated HA, or dipotassium phosphate [16–20] (K_2HPO_4). When these phantoms are CT scanned, HA or K_2HPO_4 equivalent density is generally ρ_{QCT} for an inline phantom or ρ_{CT} for an offline phantom. The material specific abbreviations are ρ_{HA} or ρ_{K2HPO4}, respectively [10]. Each phantom contains inserts with different known densities such as 0, 50, 100 and 200 mg/cm^3 of HA [21,23]. After scanning the phantom and segmenting the density references, both a calibration factor and a calibration equation can be calculated. The calibration equation for a HA phantom can be calculated using a linear regression with CT Number [HU] on the y-axis and known density [mg/cm^3] on the x-axis and then algebraically rearranging the equation to result in:

$$\rho_{HA} = (CT\ Number - b)/m, \tag{3}$$

where m [HU/(mg/cm^3)] and b [HU] are the slope and intercept, respectively, from the linear regression. When density reference phantoms are used, the derivation of the calibration equation naturally characterizes and accounts for CT number variations due to factors including manufacturer, model and protocol [24]. The use of standardized and stable references in modern density phantoms can provide a comparison for analyses across clinics. However, in the case of an inline phantom that is externally located under the patient, the phantom will be subjected to patient-moderated spectra variable with patient composition, size, and geometric position [25]. While scanning an offline phantom removes this variation, this calibration method does not capture differences such as those created by dosage-reducing variable current algorithms. Initially intended to create a standardized reference to characterize variations in CT number, differences in phantoms now introduce additional variances and limitations into the comparison of clinical assessment techniques. For example, K_2HPO_4 in place of HA was discussed by Cann et al. who argued K_2HPO_4 results in a slightly lower calibration slope than HA at equivalent densities [26,27], underestimating cortical bone density. They specifically pointed out that this difference is more pronounced at higher densities, visually demonstrated by Knowles et al. [10]. Phantomless calibration, by definition, removes the variations created by scanning a phantom, retains the potential to create a scan specific calibration equation and increases accuracy over an inline phantom by moving the density reference closer to the bone.

To enable density assessment of patient scans where phantom-based calibration data were not captured, three approaches to phantomless calibration have been used in clinical research [28]: (1) using CT Numbers [HU] directly [29–32]; (2) using a calibration factor [21,23,33,34]; and, (3) substituting tissues as a calibration reference [15,16,18–23,26,31–38]. The first approach, using CT Numbers [HU] directly, is most accessible within current clinical practice limitations. Unfortunately, in order to be considered quantitative, the relevant BMD thresholds would have to be CT scanner and protocol specific. Trying to derive relevant FEAs based thresholds in terms of CT Numbers would pose challenges, such as requiring significant amounts of patient case studies. In the second approach, using a calibration factor, a general calibration factor (GCF) is calculated as the ratio of QCT derived BMD divided by CT Numbers [HU] and then rearranged to extrapolate phantomless BMD through multiplying CT Number [HU] by GCF [21]. While this approach is CT scanner and protocol specific, it is neither scan specific nor precise enough for FEAs. The third approach, substituting tissues as calibration references is scan specific, and has been applied in

FEAs of the femur [17,22,36,37,42]. This method is limited by the assumption that internal patient tissues have specific densities [42]. Previously, a variety of tissues served as the basis for deriving phantomless calibration: fat and muscle [19,20,25,35,39–41]; air and blood [17,36,37]; air and fat [17,36,37]; air, fat, and muscle [22]; and air, fat, blood, muscle, and cortical bone [22,42]. Many things are known to influence the ability of CT Numbers [HU] to measure tissues: hydration levels [20], patient pathologies [43], heterogeneous distributions of muscle and fat [20], and i.v. contrast [19,44]. Further, CT is unable to assess some pathologies known to affect CT Number, such as fatty atrophy of muscle [39].

While there is no standard method for determining which tissues to use as the basis for phantomless calibration, the literature provides some rationale for choosing specific tissues. Boden et al. showed fat and muscle offer reliable internal reference standards for measuring vertebral bone density with QCT using tabulated reference densities from White [25,45]. More recently, Michalski et al. used tabulated and standardized mass attenuation coefficients from the National Institute of Standards and Technology (NIST) [42,46]. Some researchers have attempted to determine their own ground truth values using a system of equations approach, finding: -69 mg/cm^3 for fat and 77 mg/cm^3 for muscle [40]; or -840 mg/cm^3 for air, -80 mg/cm^3 for fat, and 30 mg/cm^3 for muscle [22]. The limitation to deriving ground truth values, in lieu of using the standardized tables, is the unknown amount of pathological variation in the base cohort.

In the absence of phantom-based calibration data, computational researchers commonly estimate a linear relationship between a specific density and CT Number based on available literature. Two such densities include ash density, ash mass divided by bulk sample volume, and apparent density, wet mass without marrow divided by bulk sample volume [10]. Several studies are available in literature where bench researchers empirically derived linear relationships between either ash density or apparent density measurements of bone and CT number [47–54]. Ford et al. demonstrated a method for estimating a linear relationship between apparent density and CT Number for trabecular bone and cortical bone in mg/cm^3,

$$\rho_{app} = 1.106 * \text{CT Number} + 68.4 \qquad (4)$$

before using the relationship in a computational study [55]. Though not demonstrated in literature, another approach would be to estimate soft tissue density by estimating a theoretical calibration slope (CT$_{theoretical}$ = 1.025 HU/(mg/cm^3)) derived from theoretical air (1.205 mg/cm^3, -1024 HU) and theoretical water (1000 mg/cm^3, 0 HU). Both of these density estimation methods do not take into consideration CT scanner performance parameters or the anatomical area as phantom-based or tissue-based phantomless calibration estimates do.

The method for deriving Young's modulus, a measurement of material stiffness, from CT data for use in patient specific FEAs is sensitive to the relationship between a specific density and CT Number due to a power-law relationship [10]. CT data are converted to ρ_{ash} using Eq. (5) depending on the equivalent density, then ρ_{app} using Eq. (6), and finally Young's modulus using Eq. (7) [53].

$$\rho_{ash} = 0.8772 * \rho_{CT} + 0.07895 \qquad (5)$$

$$\rho_{app} = 0.598 * \rho_{ash} \qquad (6)$$

$$E = 6850 * \rho_{app}^{1.49} \qquad (7)$$

In addition to being specific to the phantom's reference material, these relationships are also specific to anatomic site, in this case the femur [56]. This suggests a need for a method flexible enough to consider anatomic site when selecting reference tissues for phantomless calibration.

The aim of this retrospective study was to demonstrate, verify and validate a method for selecting patient tissues from which to derive density for use in femur strength prediction. Using the selected tissue combinations, we present a method for using phantomless calibration to estimate bone material properties for later use in femoral fracture risk prediction. Using a 2^2-factorial design, we tested repeatability with and without theoretical data points, and with and without including multiple scans for each patient. For verification, we compared patient specific results against an offline custom CIRS BMD phantom, and an inline Mindways Model 3 BMD calibration phantom. For validation, we compared patient specific results against the inline Mindways Model 3 BMD phantom for the patients whose scans included the phantom.

2. Materials and methods

Patient scans were selected for a density-related sensitivity analysis from data gathered previously related to a cohort of 408 patients gathered at the University of Wisconsin-Madison hospital. Scans from this cohort were previously identified to examine femoral fracture in an age matched, case-control study. Full details of the previous study are available in Lee et al. from 2017 [57]. Retrospective CT scan analysis was Health Insurance Portability and Accountability Act compliant and approved by the UW-Madison Institutional Review Board (protocol number 2016-0168).

The pre-fracture cases analysed consisted of 43 patients, with 26 female patients (ages 50–93 years) and 16 male patients (ages 56–95 years). The average time to fracture after CT scan was 1 year, with the minimum occurring the same year and the maximum occurring within 4 years. The control cases analysed consisted of 168 patients, with 108 female patients (ages 50–90 years) and 60 male patients (ages 50–91 years).

2.1. Method of selecting patient scans for analysis

Scans analysed were limited to scans captured on a GE Lightspeed family CT scanner including: LightSpeed 16, LightSpeed Pro 16, LightSpeed Pro 32, LightSpeed Ultra, LightSpeed VCT, Discovery CT750HD, Optima 580, Optima 660 HD, and Revolution GSI. All scans analysed were captured at 120 kVp, and 1.25 mm slice thickness. The 258 scans analysed (Table 1) included images of 211 individual patients, both male and female (aged 50 to 95 years). Patients with surgical hardware were excluded from the study. Our goal in this selection was to cover a broad range of data for the phantomless calibration to be broadly applicable so we processed all data that met our inclusion criteria.

2.2. CT scanning protocol

Images were collected during routine abdominopelvic CT scans performed using 16- to 64-Multi-Detector CT scanners (LightSpeed Series, GE Healthcare). Hospital routine includes daily calibration scans on each machine to ensure the accuracy of the CT attenuation values. Standard scanning parameters for routine abdominopelvic CT scans are 120 kVp tube voltage, 1.25 mm slice thickness, 0.625 mm slice spacing, a medium or body type filter, a standard convolution kernel, and low doses of current, either static (50–100 mA) or modulated (noise index, 50; range 30–300 mA).

2.3. Inline quantitative equivalent density calibration using the Mindways model 3 BMD calibration phantom

Eight out of the 408 patient scans included an inline effective K_2HPO_4 density calibration phantom (Model 3 Phantom, Mindways Software, Inc., Austin, TX). Of those eight, three patients had existing surgical hardware and could not be analysed. Therefore, the analyses in this paper were limited to five patients. The calibration process for this phantom is described in detail by Mindways [58]. Manual calculation of the calibration slopes for the five patients scanned with the inline calibration phantom was conducted (Table 2). A power analysis for a two-

Table 1

The analysed cohort considered 211 distinct patients (bottom), with up to 258 scans (top). These scans were captured on nine different GE CT scanner models at the University of Wisconsin—Madison hospital.

GE CT scanner model	Number of scans	Number of patients with multiple scans	Pre-fracture (control)	Male (female)
LightSpeed 16	45	14	4 (41)	13 (32)
LightSpeed Pro 16	22	5	9 (13)	7 (15)
LightSpeed Pro 32	1	0	1 (0)	1 (0)
LightSpeed Ultra	139	49	3 (136)	56 (83)
LightSpeed VCT	35	10	26 (9)	12 (23)
Discovery CT750 HD	8	4	5 (3)	3 (5)
Optima 580	1	1	1 (0)	1 (0)
Optima 660HD	6	4	2 (4)	1 (5)
Revolution GSI	1	1	0 (1)	1 (0)
Totals	258	88	52 (206)	95 (163)
LightSpeed 16	39	–	2 (37)	12 (27)
LightSpeed Pro 16	19	–	7 (12)	7 (12)
LightSpeed Pro 32	1	–	1 (0)	1 (0)
LightSpeed Ultra	108	–	3 (105)	41 (67)
LightSpeed VCT	30	–	23 (7)	12 (18)
Discovery CT750 HD	8	–	5 (3)	3 (5)
Optima 660HD	6	–	2 (4)	1 (5)
Totals	211	–	43 (168)	77 (134)

Table 2

BMD [mg/cm^3] from equivalent density [mg/cm^3] equations for the five patients scanned with the inline K_2HPO_4 Mindways Model 3 BMD calibration Phantom.

Patient	CT scanner	Calibration slope [HU/(mg/cm^3)]	Calibration Y-intercept [HU]	Conversion Slope [(mg/cm^3)/HU]	Conversion Y-intercept [mg/cm^3]
1	LightSpeed VCT	1.03	5.59	0.97	−5.43
2	LightSpeed Pro 16	1.06	3.32	0.94	−3.12
3	LightSpeed Pro 16	1.05	13.54	0.95	−12.88
4	LightSpeed Pro 16	0.99	5.87	1.01	−5.93
5	Discovery CT750HD	1.06	−8.5	0.94	8.00
	Average values	1.04	3.96	0.96	−3.87

sample pooled *t*-test was conducted in MATLAB and the necessary sample size to meet 99% power ranged between 2 and 5 for the majority of the 40 phantomless slope combinations considered, with 3 outliers requiring a sample size of 8.

2.4. Offline equivalent density calibration using a custom BMD phantom

Retrospectively, we scanned offline a custom phantom with four HA density plugs at 100, 400, 1000 (part: 06217), and 1750 (part: 06221) mg/cm^3 (CIRS Inc., Norfolk, VA) submerged in water. Scan settings were 120 kVp, 1.25 mm slice thickness, 0.625 mm slice spacing, 100 mA, and a standard reconstruction kernel on the Discovery 750HD. HA plug densities were selected to be representative of human femoral bone [11]. Plugs were segmented by creating a virtual cylinder with a 10-

pixel diameter across 10 slices in the centre of the plug using Mimics v. 21 (Materialise, Leuven, Belgium). Linear regressions were calculated for CT Number (HU) as a function of known density, ρ_{HA}. Resulting equations were:

$$CT\ Number = 1.100 * \rho_{HA} + 26.29 \tag{8}$$

$$(CT\ Number – 26.29)/1.100 = (1.100 * \rho_{HA} + 26.29 – 26.29)/1.100 \tag{9}$$

$$\rho_{CT} = 0.9091 * CT\ Number – 23.90 \tag{10}$$

2.5. Identify most consistent reference densities across patients

We analysed phantomless calibration on 258 scans and considered five patients' nominal density references, including adipose tissue, aortic blood, skeletal muscle, urine, and air. Tissue segmentations were captured as virtual cylinders, with a diameter of 10 pixels and a depth of 10 slices, using Mimics v21.0 (Materialise, Leuven, Belgium). Due to the small size of the femoral artery, the virtual cylinder captured was reduced to a diameter of 8 pixels. For consistency, all virtual cylinders were created such that the centre of the virtual cylinder was around the same axial slices as the centre of the femoral head. An example of the virtual cylinder placement is shown in Fig. 1. Quality checks were conducted to ensure each virtual cylinder contained a volume of at least 100 voxels (ASTM E1935 2019). We were unable to segment urine in the patient's bladder for 167 out of the 258 scans due to empty bladders. Aortic blood was also difficult to segment due to their small sizes, resulting in measured values outside of 40 ± 20 HU for 46/258 left patient arteries and 47/258 right patient arteries. Table 3 shows the nominal density values assumed for the linear regression of HU and tissue density [46]. The 258 patients included in this study were segmented by a single operator. To assess the precision of results at the segmentation, BMD and FEAs levels, the five patients with inline phantoms were also segmented by three different operators. Inter- and intra-operator repeatability was calculated [59].

Each patient had up to nine potential data points that could be used for line fitting: theoretical air, segmented air, adipose tissue (right and left), aortic blood (right and left), skeletal muscle (right and left), and theoretical water. Any combination of at least two and up to nine data points could be used to derive a linear regression for the HU versus nominal density relationship, 502 possible combinations for each of the 258 scans. A custom MATLAB (v.2018b, The MathWorks, Inc., Natick, MA, US) script was developed to: (1) calculate all possible linear regressions, (2) discard all ill-conditioned calibration slope results, and (3) conduct a numerical analysis to sort density combination calibration slope results across patients. Ill-conditioned calibration slopes occurred when the algorithm fit a line with two values for the same tissue (i.e. right and left adipose). Sorting was accomplished by minimizing the sum of the squared error between the density calibration slope and a theoretical calibration slope,

$$\sum_1^n (m - 1.025)^2 \tag{11}$$

Recall from the introduction that the theoretical calibration slope [1.025 HU/(mg/cm^3)] is derived from theoretical air (1.205 mg/cm^3, −1024 HU) and theoretical water (1000 mg/cm^3, 0 HU). After discarding over-constrained combinations, the best 10 combinations and the worst combination were identified for further analysis.

2.6. Experimental design to test repeatability of tissue identification

Patient tissue segmentations were organized to form two groups: "Scans" included all scans eligible for processing for all patients, and "Patients" included only one scan for each patient. To form the Patients group, results from duplicate scans for patients were removed such that the results for CT scanners with fewer patient scans were kept except in the case of the Optima 580 and Revolution GSI, each of which only had
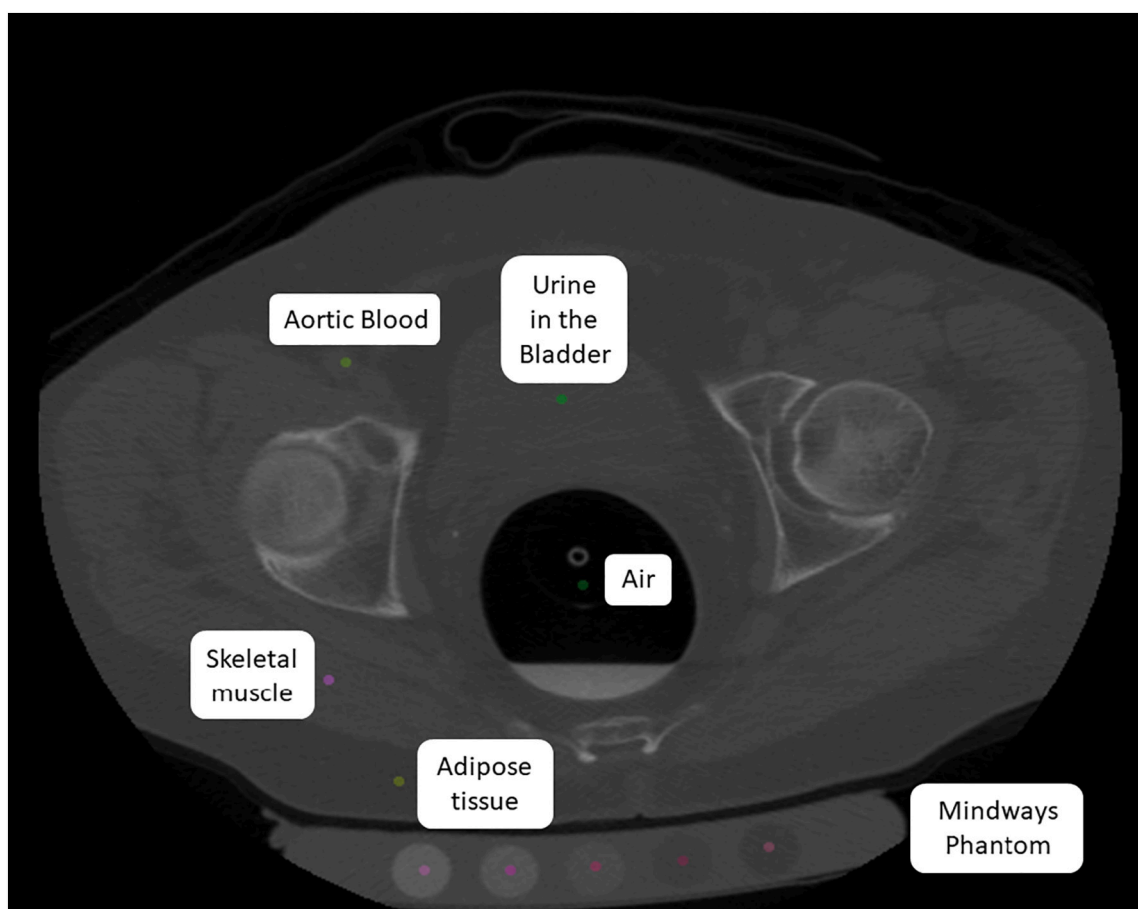
Fig. 1. Representative axial slice of a CT scan of an 85-year-old patient at the proximal femur level. The patient is lying supine. A Mindways Model 3 BMD Calibration Phantom is visible underneath the patient.

Table 3
Nominal density values from NIST [46].

Reference material	Nominal density [mg/cm^3]
Theoretical air (not segmented)	1.205
Adipose tissue	950
Aortic blood	1060
Theoretical water (not segmented)	1000
Skeletal muscle	1050

one patient scan. A 2^2 factorial designed experiment was conducted by running the MATLAB script used to identify the most consistent reference densities across patient populations two levels for each group including and excluding values for theoretical air and water in the combinatorial analysis.

2.7. Finite element model BMD and femur strength

Five finite element models were developed for each patient to investigate the impact of different calibration equations on BMD and femoral strength (FS) calculations model I: patient specific inline K_2HPO_4 calibration; model II: the average of the patient specific inline K_2HPO_4 calibrations; model III: the offline HA calibration; model IV: phantomless calibration derived from air, aortic blood, and skeletal muscle (AABSM); and, model V: phantomless calibration derived from air and adipose (AA). One femur was segmented for each patient: four were segmented in Mimics v19.0 or 21 (Materialise, Leuven, Belgium) and one was segmented in ITK-Snap (ITK-Snap 3.6.0, University of Pennsylvania). Each geometry was discretized into ten-node tetrahedral

elements using ICEM CFD (ICEM CFD 16.2, Ansys Inc., PA, USA) with a maximum edge length of 3 mm based on a previous mesh convergence study [60]. Note that each patient had the same mesh for all models.

Elastic moduli were mapped onto the meshed bone using the equations described in the introduction and Bonemat (V3.2, Istituto Ortopedico Rizzoli, Bologna, Italy). BMD was calculated for each model as the summation across groups of the density in each material group multiplied by the number of elements with that material group. Femur strength was calculated using a sideways fall loading scenario with fixed boundary constraints at the estimated knee centre and a simulated planar bearing at the lateral coordinate on the trochanter [4,61]. A concentrated point load, 1000 N, was applied to the centre of the femoral head in thirty-three different force directions from $-30°$ to $30°$ (posteriorly to anteriorly directed) in the transverse plane and $0°$ to $30°$ (x-axis to medially directed) in the frontal plane [61]. FEAs strain results were post-analysed using a maximum principal strain failure criterion with limiting values at 0.73% for tensile and 1.04% for compressive strains as previously defined by Bayraktar et al. [62]. FS was defined as the minimum force (N) at failure across all 33 side-fall loading conditions. All FEAs were conducted in ANSYS 16.2 (Ansys Inc., PA, USA).

2.8. Statistical analysis

The mean and standard deviation were calculated for patient tissue segmentation measurement results in HU for both the "Scans" and "Patients" groups. Once patient specific density calibration slopes were calculated, statistical measurements were mean, standard deviation, and 95% confidence interval. Bland-Altman analyses were conducted for the five patients with inline Mindways Model 3 BMD phantoms included in

their scans. The hypotheses that no statistically significant difference exists between calibration methods were tested using a Students' t-test ($\alpha = 0.01$) for the calibration slopes, the calibration intercepts, BMD, and FS. Normality was tested using the Shapiro-Wilk test in IBM SPSS Statistics for Windows, version 26 (IBM Corp., Armonk, N.Y., USA); however, all other statistics were calculated in MATLAB 2018b.

3. Results

Phantomless calibration was valid when compared against inline phantom calibration for FS, BMD, calibration equation (Figs. 2, 3, 4). The algorithm produced calibration equation results consistent with those from inline phantom calibration (Fig. 5). Intra- and inter-operator repeatability found the method highly repeatable for FS, BMD, and calibration equations (Table 4). Of the tissues segmented, adipose was the most repeatable and the bladder was the least repeatable (Table 4).

The AABSM combination produced the best slope result for 3 of the 4 categories in the 2^2 factorial designed experiment. The 4th category, excluding multiple scans per patient and theoretical air and water, found the AA combination produced the best slope. The first category, including theoretical air and theoretical water for all scans ($n = 258$), found AABSM scan specific slope values [HU/(mg/cm^3)] of mean $\pm$ std. dev (lower–upper) = 1.021 ± 0.006 (1.008–1.034) and found measured air and theoretical water produced the worst combination with slope values of 1.379 ± 6.185 (−10.99–13.75). The second category, including theoretical air and theoretical water for 1 scan per patient (n = 211), found AABSM scan specific slope values of mean $\pm$ std. dev (lower–upper) = 1.021 ± 0.006 (1.009–1.034) for the best combination and found measured air and theoretical water produced the worst combination with slope values of 1.468 ± 6.856 (−12.24–15.18). The third category, excluding theoretical air and theoretical water for all scans ($n = 258$), found AABSM scan specific slope values of mean $\pm$ std. dev (lower–upper) = 1.017 ± 0.010 (0.998–1.037) for the best combination and found aortic blood and skeletal muscle produced the worst result with values of 0.893 ± 2.151 (−3.458–5.195). The final category, excluding theoretical air and water for 1 scan per patient ($n = 211$), found AA scan specific slope values of mean $\pm$ std. dev (lower–upper) = 0.975 ± 0.010 (0.956–0.994) and found aortic blood and skeletal muscle produced the worst result with values of 0.839 ± 2.149 (−3.458–5.137).

For FS results, the AABSM calibration resulted in a 6.9% bias over scan specific inline calibration, a 7.3% bias over averaged inline calibration, and a 22% bias over offline calibration; and the AA calibration resulted in a 9.9% bias over scan specific inline calibration, a 10% bias over averaged inline calibration, and a 25% bias over offline calibration (Fig. 2). For BMD results, the AABSM calibration resulted in a 3.9% bias over scan specific inline calibration, a 3.7% bias over averaged inline calibration, and a 17% bias over offline calibration; and the AA calibration resulted in a 6.1% bias over scan specific inline calibration, a 6.0% bias over averaged inline calibration, and a 19% bias over offline calibration (Fig. 3). When considering the calibration slopes directly, the AABSM and AA combinations resulted in biases of 2.6% and 6.3% over scan specific inline calibration, respectively (Fig. 4). For the calibration
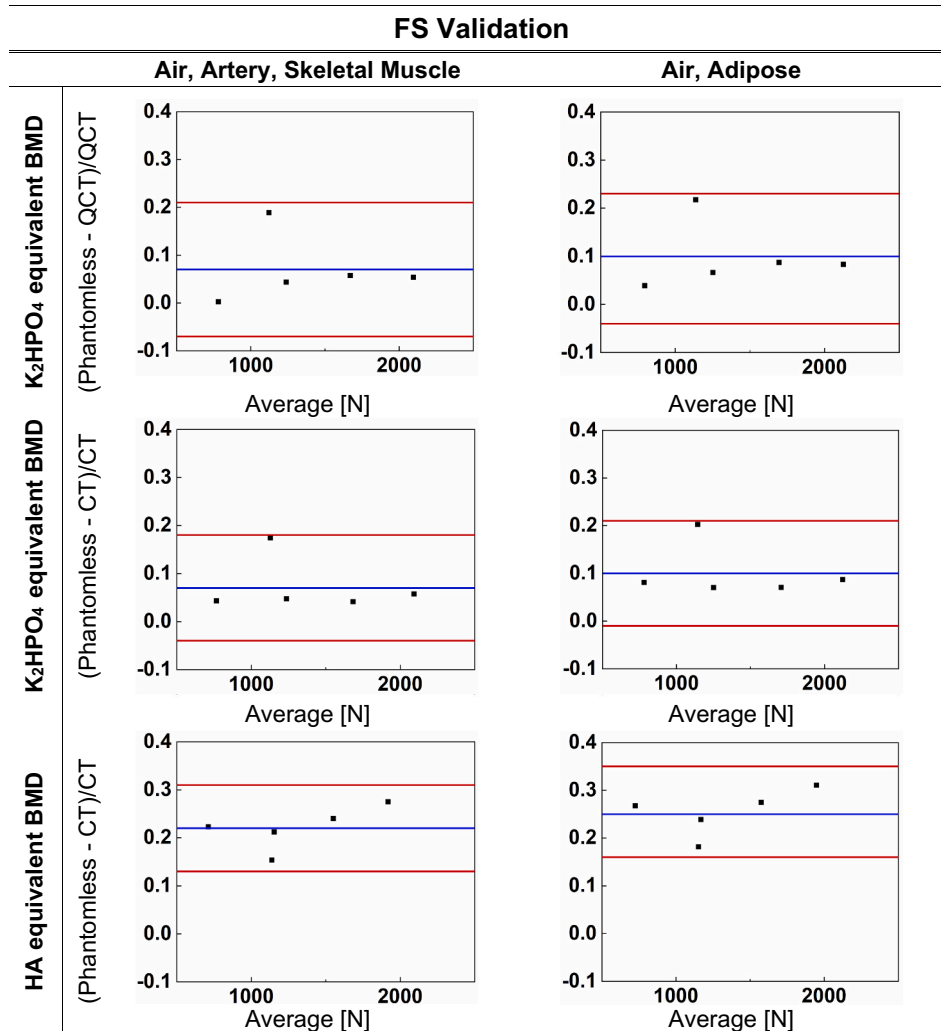

Fig. 2. FS results derived using phantomless calibration displayed the least bias when compared against results derived using the average of the patient and scan-specific K$_2$HPO$_4$ calibration as shown by Bland-Altman analyses. Overall results using phantomless calibration were more consistent with results from the K$_2$HPO$_4$ phantom than the HA phantom. The blue lines are the means and the red lines are the 95% confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
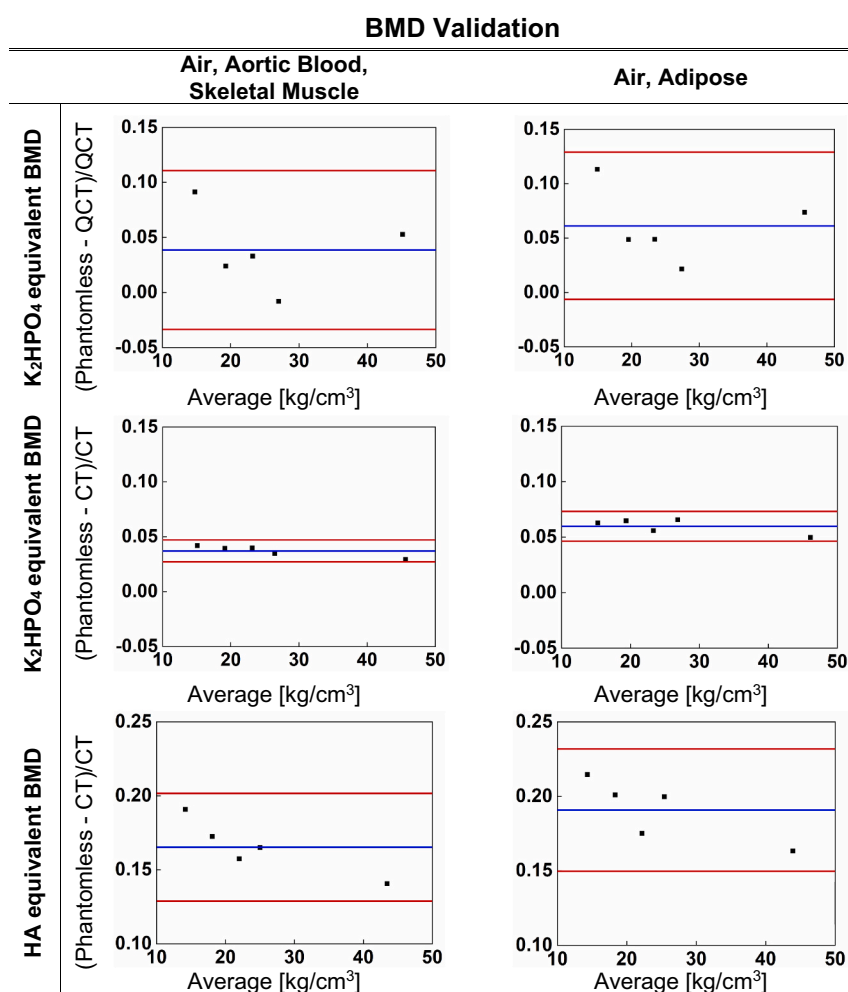
BMD Validation

Fig. 3. BMD results derived using phantomless calibration displayed the least bias when compared against results derived using the average of the patient and scan specific K_2HPO_4 calibration as shown by Bland-Altman analyses. Overall results using phantomless calibration were more consistent with results from the K_2HPO_4 phantom than the HA phantom. The blue lines are the means and the red lines are the 95% confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

intercepts, the AABSM and AA combinations resulted in biases of 110% and 110% over scan specific inline calibration, respectively (Fig. 4). When comparing scan specific results for all 211 patient scans against the scan specific inline calibration, the ten best AABSM slope combinations all resulted in the majority of patients falling within the range demonstrated by the inline calibration (Fig. 5). The three best AA slope combinations did not fall within the range demonstrated by the inline calibration; however, the inter-quartile range for the next seven best did fall within the range demonstrated by the inline calibration (Fig. 5). All intercepts for the ten best combinations for both AABSM and AA fell within the range demonstrated by the inline calibration (Fig. 5). Biases for the best ten tissue combination results for all four categories, compared with the inline calibration slope, were found to be less than or equal to 0.068 ± 0.064 HU/(mg/cm^3) for the five patients with inline calibration available. The resulting 40 calibration slopes and the scan specific inline calibration slopes were found to be normally distributed using a Shapiro-Wilk test.

Differences in FS between calibration methods were only statistically significant for AABSM versus the average of the inline calibrations ($p < 0.01$). Differences in BMD between calibration methods were not statistically significant for either phantomless calibration combination (AABSM and AA) and the inline phantom calibration ($p = 0.03, 0.10$). However, differences in BMD between calibration methods were statistically significant for both phantomless calibration combinations (AABSM and AA) versus the average of the inline calibrations ($p = 0.003, 0.002$) and the offline phantom ($p = 0.004, 0.003$). Differences in calibration equation followed the same trend. For the slopes, differences were not statistically significant between either phantomless calibration

combination (AABSM and AA) and the inline phantom ($p = 0.04, 0.17$). Conversely, differences were statistically significant between both phantomless calibration combinations (AABSM and AA) versus the average of the inline calibrations ($p < 0.001, 0.001$) and the offline phantom ($p < 0.001, 0.001$). For the intercepts, differences were not statistically significant between either phantomless calibration combination (AABSM and AA) and the inline phantom ($p = 0.08, 0.26$). Continuing with the trend, differences were statistically significant between both phantomless calibration combinations (AABSM and AA) versus the average of the inline calibrations ($p < 0.001, 0.001$) and the offline phantom ($p < 0.001, 0.001$).

Both average intra-operator and inter-operator repeatability were better for AABSM than for AA when analysing FS, BMD, or calibration equation (Table 4). Segmentation CT Number [HU] results found similar means and standard deviations for tissues compared between the "all scans" and "one scan per patient" categories, respectively: adipose -99.12 ± 9.44 and -98.98 ± 9.62; aortic blood 52.42 ± 17.28 and 51.83 ± 17.41; and muscle 43.58 ± 13.42 and 44.01 ± 14.05.

4. Discussion

The main aim of this study was to demonstrate, verify and validate a method for selecting basis patient tissues for deriving an equivalent density equation in femoral bone QCT analyses. As an example, this method identified AABSM as the best combination of tissues for phantomless calibration. This method was shown to be valid for FS, BMD, and calibration equation results. The validity of phantomless calibration for FEAs of the femur is consistent with other studies [17,37,42]. To verify
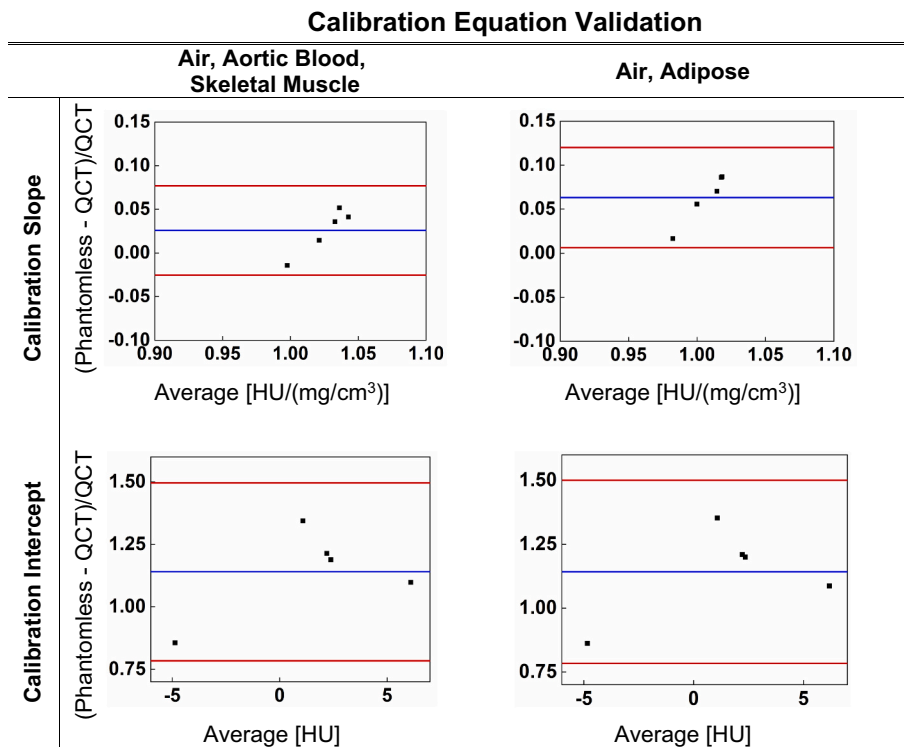
Calibration Equation Validation

Fig. 4. Phantomless calibration slopes derived from air, aortic blood and skeletal muscle segmentations displayed less bias than those derived from air and adipose when compared with patient and scan specific K_2HPO_4 calibration as shown by Bland-Altman analyses. While both sets of phantomless calibration intercepts displayed similar and large bias, all averages were within the performance expectations for a GE CT scanner (0 ± 7 HU) [24]. The blue lines are the mean and the red lines are the 95% confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
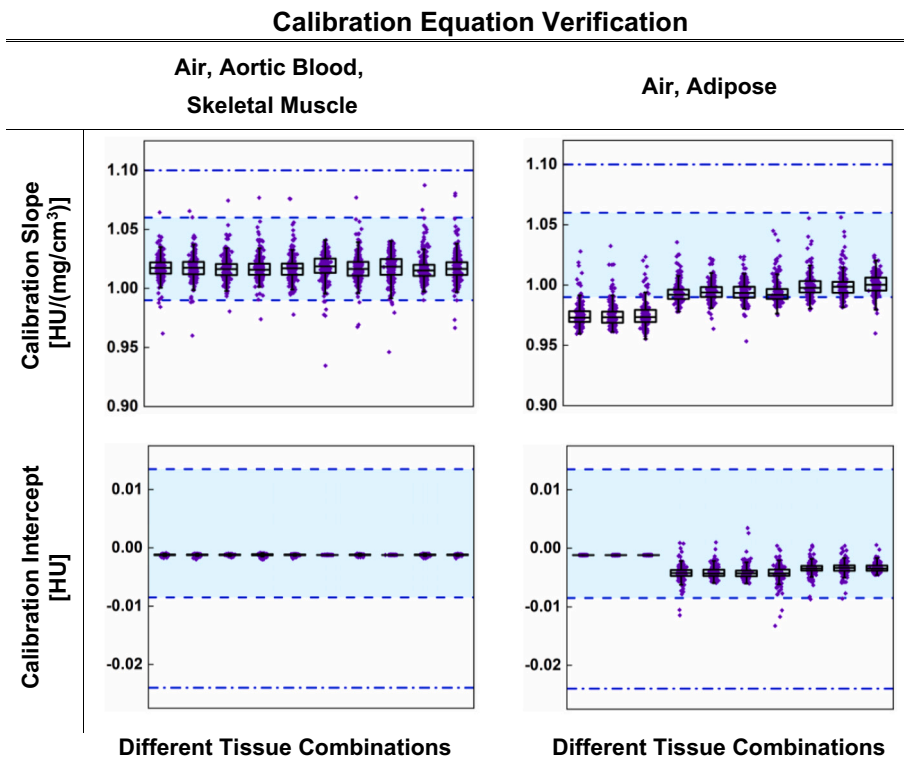
Calibration Equation Verification

Fig. 5. These plots compare the ten best combinations of tissues in terms of calibration slopes and intercepts. Boxplots are overlaid on scatter plots of the patient specific calibration slopes and intercepts (purple points). For the slopes plots, the three blue lines include: the dashed lines for the minimum (0.99 HU/(mg/cm^3)) and maximum (1.06 HU/(mg/cm^3)) slopes across patients from the K_2HPO_4 calibration phantom, and the dash-dot line is for the calibration slope for the custom phantom scanned offline in water (1.10 HU/(mg/cm^3)). All slopes are in HU/(mg/cm^3). For the intercept plots, the three blue lines include: the dashed lines for the minimum (-0.0085 HU) and maximum (0.0135 HU) patient specific results for the K_2HPO_4 phantom, and the dash-dot line is calibration intercept for the custom phantom scanned in water (-0.0239 HU). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

this method, results for 258 scans were shown to be within the range of those from the inline calibration of five scans. This method shows promise for use in the retrospective analysis of patient cohorts without available calibration information and can be applied opportunistically to any CT scan.

This study differs from previous studies in several ways including different CT scanners, CT scan protocols, tissues used as the basis for phantomless calibration, assumed tissue densities, methods of segmentation, and FEA pipelines. Focusing in on which tissues are used as the basis for phantomless calibration, this study's selection of the AABSM combination of tissues is different from prior combinations in literature for FEAs of the femur, including: fat and muscle [39]; air and fat

Table 4

Intra- and Inter-operator reanalysis precision error (root-mean-square) for FS, BMD, calibration equation, and tissue segmentations at the femur for $n = 5$. Coefficients of variation (CV_{RMS}, in %) and standard deviations (SD_{RMS}, in absolute units) are presented.

Measurement	Intra-operator		Inter-operator	
	CV_{RMS} (%)	SD_{RMS}	CV_{RMS} (%)	SD_{RMS}
Femoral strength [N]				
AABSM	0.28	9.67	0.42	13.27
AA	1.27	63.85	4.14	224.44
Inline	0.37	11.68		
Bone mineral density [kg/cm³]				
AABSM	0.22	0.13	0.39	0.20
AA	0.65	0.48	1.52	1.28
Inline	0.57	0.57		
Slope [HU/(mg/cm³)]				
AABSM	0.22	0.00	0.41	0.01
AA	0.41	0.01	1.01	0.02
Inline	0.60	0.01		
Intercept [HU]				
AABSM	2.70	0.08	4.56	0.14
AA	6.20	7.45	17.03	20.12
Inline	9.31	2.22		
Segment [HU]				
Adipose	1.68	3.77	1.78	3.98
Air	0.35	7.44	0.93	20.10
Aortic blood	10.09	10.67	13.93	15.99
Skeletal muscle	10.25	7.19	11.24	8.12
Bladder	50.82	3.67	68.11	12.21

[17,36,37]; air, fat, and muscle [22]; and air, fat, blood, muscle, and cortical bone [22,42]. The variety of different combinations shows the need for a universally accessible objective method, such as that presented in this study, for identifying the best tissues for use as the basis for phantomless calibration within the existing constraints of CT scanners and CT scan protocols for the application specific anatomic site. Algorithms for decision making, such as that presented in the current study, can be more robust than correlation approaches such as those presented by Eggermont et al. [22].

Despite the differences in FEA pipelines, the bias introduced by phantomless calibration is comparable across studies, with all other variables held constant within the respective studies. This study's calculated FS mean absolute difference, 90 N (6.9%), was similar to recent studies such as Lee et al. 30 N (0.8%) [17], and Michalski et al. −40 N (17%) [39]. The calculated BMD biases 0.92 kg/cm³ (0.04%) were larger than a recent study on a more developed method presented by Lee et al., 2 mg/cm³ (0.9%) [17]. Note that differences observed in FS measurements were expected to be greater than differences observed in BMD measurements for two reasons. First, differences that appear small when examining preliminary results (i.e. segmentation, calibration equation, BMD) are amplified by the power-law component of the density-elastic modulus relationship (Eq. (7)) making FEAs sensitive to changes in the calibration equation. Second, the side-ways fall load case is more sensitive to changes in mechanical properties of materials due to the stress gradient from bending in the combined-loading. Both the results of this study and the results from literature show greater differences in FS biases than BMD biases. From a clinical perspective, this drives the reasonable assumption that variables known to effect CT Number [HU] or BMD measurement would have an amplified effect on FS.

Recent studies have proposed the use of QCT derived FEAs for improved osteoporotic hip fracture risk prediction [4–6] and the use of phantomless calibration in this context [17,42]. Limited studies have been conducted to identify and quantify the impact of relevant factors. Michalski et al., who conducted part of their analysis on ten full body cadavers, iteratively correlated ROI specific CT Numbers across energy levels setting the example of taking these factors into account during the development of their phantomless calibration method [42]. Several authors have noted the improvements in phantomless calibration results due to the decreased distance between the patient and the reference [17,20,25]. The current study controlled for some factors known to create variations in CT Number [HU] by limiting data analysed to scans captured on GE LightSpeed CT Scanners with 120 kVp, variable mA, slice thickness of 1.25 mm, slice increments of 0.625 mm, and a standard reconstruction kernel. Lee et al. used similar inclusion criteria, identifying 120 kVp and a standard reconstruction kernel as the most important imaging technique factors and their decision to analyse a single protocol as a limitation [17]. Although attempting to work with a standardized protocol, Eggermont et al. found that a small number of their patients were scanned with a different reconstruction kernel allowing them to make relevant observations (1) changing reconstruction kernel had no significant effect on phantom-based or air-fat-muscle calibration, and (2) changing reconstruction kernel resulted in significantly higher failure loads when using their non-patient specific calibration [22]. Michalski et al. observed that by using consistent imaging acquisition and a single imaging protocol there were fewer confounding variables when measuring methodological precision [42]. Beyond the limitation of only considering one clinical protocol, this study was also limited to pre-fracture cases that went on to experience femoral fragility fracture.

The current study's segmentation method may be less repeatable than the segmentation methods presented in other studies. Where this study conducted manual segmentation using the mean CT Number [HU] over the digital volume, other studies used higher fidelity segmentation methods. Examples relevant to multiple studies include: Lee et al. who have automated their segmentation using gradient-profile algorithms independent of absolute attenuation [17,36,37], or the popular histogram and peak fitting approach [20,22,25,41,42]. Boden et al. designed the histogram and peak fitting approach specifically to overcome the challenge of reliably locating a conventional ROI to calculate the mean CT Number [HU] of the digital volume [25]. This implies that methods using this approach would naturally account for the heterogeneity included in patient tissues and improving the precision of phantomless calibration. The differences in segmentation methods are a major reason why this method was less repeatable than those presented previously in literature (Table 5). This comparison shows using a higher fidelity segmentation method may improve the repeatability of the current study's phantomless calibration method.

This study showed phantomless calibration results were close to results derived from the Mindways Model 3 BMD inline phantom which relies on K_2HPO_4 as a reference material. Further, the phantomless calibration derived results were not significantly different from the inline calibration derived results and were significantly different from both the averaged inline calibration and the offline calibration. Both the inline phantom, which ranges from −53.4 to 375.8 of equivalent K_2HPO_4, and this phantomless calibration technique, require extrapolation in order to define in vivo BMD [58]. The potential for extrapolation errors has been raised as a concern in several studies [20,22,41]. In their phantomless study, Lee et al. demonstrated their method to

Table 5

Comparison of precision errors between the current study and literature.

Measurement	Intra-operator		Inter-operator	
	CV_{RMS} (%)	SD_{RMS}	CV_{RMS} (%)	SD_{RMS}
Femoral strength				
Lee [17] [N]			0.4	20
Michalski [42] [N]			6.0	84
Current study [N]	0.28	9.67	0.42	13
Bone mineral density				
Lee [17] [mg/cm³]			<0.3	≤1
Michalski [42] [mg/cm³]	4.3	12	5.3	11
Current study [kg/cm³]	0.22	0.13	0.39	0.20

calibrate CT scans was equivalent to traditional phantom-based calibration [17]. If assumptions are made about the density of bone and included when deriving phantomless calibration, the results become less accurate as showed by the correlation analysis in the pilot study written by Eggermont et al. [22].

There were several limitations to this study. CT scans of the proximal femur region include a limited choice of tissues to segment: adipose tissue, skeletal muscle, aortic blood, and in some cases the bladder is empty. In addition to population variance across patients, tissues also depend on a variety of patient specific variables such as: hydration level [20], patient pathologies [43], heterogenous distributions of muscle and fat [20], i.v. contrast [19,44], exercise habits, and body mass index. The cohort studied here did not include patient details about exercise habits, body mass index or comorbidities. Future studies should consider a more detailed examination of factors known to cause variance across patients and a larger sample size to further develop the phantomless calibration methodology. In this study, GE LightSpeed family CT scanners were used to demonstrate the calibration process. CT scanners from other manufacturers were not analysed due to lack of available data. Future work should consider a multi-centre study comparing the same model of CT scanner across different hospitals and consider CT scanners from other manufacturers. Also of note was the small sample size of available calibration curves for comparison.

This study did not examine several potential confounding variables. When reassigning pre-fracture/control pairings, researchers were not blind to CTXA, a method for measuring areal BMD from CT data mathematically equivalent to dual-energy X-ray absorptiometry, density measurements. Stratification accuracy between pre-fracture and control cases when using phantomless calibration was not examined. Additional confounding factors may have been present such as: other diseases, routine exercise habits, differences in body-mass index/height/weight, comorbidities, or different pathologies. These were not considered due to lack of readily available cohort information. Several of these variables could be considered in a prospective study or in a reanalysis of retrospective data prospectively gathered. A more systemic method of randomly assigning controls to pre-fracture cases could be developed and implemented to mitigate the potential alignment of CTXA density measurements between pre-fracture and control cases. Future studies could be designed to fully test stratification accuracy between pre-fracture and control cases when using phantomless calibration.

Overall, results derived from the phantomless calibration slopes were a valid substitute for those derived from the inline calibration. When considering FS, the phantomless calibration resulted in a small 7% increase over inline calibration. For BMD, the phantomless calibration resulted in a small 4% increase over inline calibration. The phantomless calibration slopes were consistently comparable with the range demonstrated by the patient specific Mindways Model 3 BMD phantom calibration slopes, with our best method displaying a small bias of 0.028 $\pm$ 0.054 HU/(mg/cm^3). The study shows the proposed method for phantomless calibration is valid for FEA studies of retrospective cohorts lacking calibration data. This method can be applied opportunistically to CT scans captured for analyses other than hip fracture. Further examination of the error introduced when the proposed method for phantomless calibration is applied in patient specific FEA derived FS should be conducted.

The segmentation measurements, BMD measurements and femoral strength results created during this study are available by contacting the corresponding author.

CRediT authorship contribution statement

C. Winsor: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Funding acquisition. **X. Li:** Conceptualization, Methodology, Software, Resources, Writing - review & editing, Supervision, Funding acquisition. **M. Qasim:** Methodology, Software, Formal analysis, Writing - review & editing. **C.**

R. Henak: Writing - review & editing, Supervision. **P.J. Pickhardt:** Investigation, Resources, Writing - review & editing. **H. Ploeg:** Conceptualization, Writing - review & editing, Funding acquisition. **M. Viceconti:** Conceptualization, Methodology, Formal analysis, Resources, Writing - review & editing, Supervision, Funding acquisition.

Declaration of competing interest

Acknowledgements

References

[1] R. Burge, B. Dawson-Hughes, D.H. Solomon, J.B. Wong, A. King, A. Tosteson, Incidence and economic burden of osteoporosis-related fractures in the United States, 2005–2025, J. Bone Miner. Res. 22 (3) (2007) 465–475, https://doi.org/10.1359/jbmr.061113.

[2] E.S. Siris, Y.T. Chen, T.A. Abbott, et al., Bone mineral density thresholds for pharmacological intervention to prevent fractures, Arch. Intern. Med. 164 (10) (2004) 1108–1112, https://doi.org/10.1001/archinte.164.10.1108.

[3] M. Viceconti, M. Qasim, P. Bhattacharya, X. Li, Are CT-based finite element model predictions of femoral bone strengthening clinically useful? Curr Osteoporos Rep. (2018) 1–8, https://doi.org/10.1007/s11914-018-0438-8.

[4] M. Qasim, G. Farinella, J. Zhang, et al., Patient-specific finite element estimated femur strength as a predictor of the risk of hip fracture: the effect of methodological determinants, Osteoporos. Int. 27 (9) (2016) 2815–2822, https://doi.org/10.1007/s00198-016-3597-4.

[5] T.M. Keaveny, B.L. Clarke, F. Cosman, et al., Biomechanical Computed Tomography analysis (BCT) for clinical assessment of osteoporosis, Osteoporos. Int. (2020) 1025–1048, https://doi.org/10.1007/s00198-020-05384-2.

[6] Katz Y, Dahan G, Sosna J, Shelef I, Cherniavsky E, Yosibash Z. *Scanner Influence on the Mechanical Response of QCT-Based Finite Element Analysis of Long Bones.* Vol 86.; 2019. doi:https://doi.org/10.1016/j.jbiomech.2019.01.049.

[7] T.P. Szczykutowicz, Dual-Energy and Spectral Imaging, Elsevier B.V., 2014, https://doi.org/10.1016/b978-0-444-53632-7.00207-0.

[8] Kamalian S, Lev MH, Gupta R. *Computed Tomography Imaging and Angiography – Principles.* Vol 135. 1st ed. Elsevier B.V.; 2016. doi:https://doi.org/10.10 16/B978-0-444-53485-9.00001-5.

[9] K. Engelke, C. Libanati, T. Fuerst, P. Zysset, H.K. Genant, Advanced CT based in vivo methods for the assessment of bone density, structure, and strength, Curr Osteoporos Rep. 11 (3) (2013) 246–255, https://doi.org/10.1007/s11914-013-0147-2.

[10] N.K. Knowles, J.M. Reeves, L.M. Ferreira, Quantitative Computed Tomography (QCT) derived bone mineral density (BMD) in finite element studies: a review of the literature, J Exp Orthop. 3 (1) (2016) 36, https://doi.org/10.1186/s40634-016-0072-2.

[11] Crookshank M, Ploeg HL, Ellis R, MacIntyre NJ. Repeatable calibration of Hounsfield units to mineral density and effect of scanning medium. *Adv Biomech Appl.* 2013;1(1):15–22. doi:10.12989/aba.2013.1.1.015.

[12] T.P. Szczykutowicz, A. DuPlissis, P.J. Pickhardt, Variation in CT number and image noise uniformity according to patient positioning in MDCT, Am. J. Roentgenol. 208 (5) (2017) 1064–1072, https://doi.org/10.2214/AJR.16.17215.

[13] A.S. Michalski, W.B. Edwards, S.K. Boyd, The influence of reconstruction kernel on bone mineral and strength estimates using quantitative computed tomography and finite element analysis, J. Clin. Densitom. 22 (2) (2019) 219–228, https://doi.org/10.1016/j.jocd.2017.09.001.

[14] H. Giambini, D. Dragomir-Daescu, A. Nassr, M.J. Yaszemski, C. Zhao, Quantitative computed tomography protocols affect material mapping and quantitative computed tomography-based finite-element analysis predicted stiffness, J. Biomech. Eng. 138 (9) (2016) 1–7, https://doi.org/10.1115/1.4034172.

[15] H. Giambini, D. Dragomir-Daescu, P.M. Huddleston, J.J. Camp, K.-N. An, A. Nassr, The effect of quantitative computed tomography acquisition protocols on bone mineral density estimation, J. Biomech. Eng. 137 (11) (2015) 114502, https://doi.org/10.1115/1.4031572.

[16] Y.H. Lee, J.J. Kim, I.G. Jang, Patient-specific phantomless estimation of bone mineral density and its effects on finite element analysis results: a feasibility study, Comput Math Methods Med. 2019 (2019) 1–10, https://doi.org/10.1155/2019/4102410.

[17] D.C. Lee, P.F. Hoffmann, D.L. Kopperdahl, T.M. Keaveny, Phantomless calibration of CT scans for measurement of BMD and bone strength—inter-operator reanalysis precision, Bone. 103 (2017) 325–333, https://doi.org/10.1016/j.bone.2017.07.029.

[18] S. Lee, Gong H. Hyeon, Hyun J. Young, et al., Estimation of stature from femur length measured using computed tomography after the analysis of three-dimensional characteristics of femur bone in Korean cadavers, Int. J. Legal Med. 131 (5) (2017) 1355–1362, https://doi.org/10.1007/s00414-017-1556-z.

[19] J. Kaesmacher, H. Liebl, T. Baum, J.S. Kirschke, Bone mineral density estimations from routine multidetector computed tomography: a comparative study of contrast and calibration effects, J. Comput. Assist. Tomogr. 41 (2) (2017) 217–223, https://doi.org/10.1097/RCT.0000000000000518.

[20] D.K. Mueller, A. Kutscherenko, H. Bartel, A. Vlassenbroek, P. Ourednicek, J. Erckenbrecht, Phantom-less QCT BMD system as screening tool for osteoporosis without additional radiation, Eur. J. Radiol. 79 (3) (2011) 375–381, https://doi.org/10.1016/j.ejrad.2010.02.008.

[21] M.J. Budoff, J.M. Malpeso, I. Zeb, et al., Measurement of phantomless thoracic bone mineral density on coronary artery calcium CT scans acquired with various CT scanner models, Radiology. 267 (3) (2013) 830–836, https://doi.org/10.1148/radiol.13111987.

[22] F. Eggermont, N. Verdonschot, Y. van der Linden, E. Tanck, Calibration with or without phantom for fracture risk prediction in cancer patients with femoral bone metastases using CT-based finite element models, PLoS One 14 (7) (2019) 1–13, https://doi.org/10.1371/journal.pone.0220564.

[23] S.S. Mao, Y. Luo, H. Fischer, M.J. Buodff, D. Li, Routine coronary calcium scan can precisely measure vertebral bone density without a quantitative calibration phantom, J. Comput. Assist. Tomogr. 40 (1) (2016) 126–130, https://doi.org/10.1097/RCT.0000000000000330.

[24] R.J. Cropp, P. Seslija, D. Tso, Y. Thakur, Scanner and kVp dependence of measured CT numbers in the ACR CT phantom, J Appl Clin Med Phys. 14 (6) (2013) 338–349, https://doi.org/10.1120/jacmp.v14i6.4417.

[25] S.D. Boden, D.J. Goodenough, C.D. Stockham, E. Jacobs, T. Dina, R.M. Allman, Precise measurement of vertebral bone density using computed tomography without the use of an external reference phantom, J. Digit. Imaging 2 (1) (1989) 31–38, https://doi.org/10.1007/BF03168013.

[26] Cann CE, Adams JE, Brown JK, Brett AD. CTXA hip - An extension of classical DXA measurements using Quantitative CT. *PLoS One*. 2014;9(3). doi:https://doi.org/10.1371/journal.pone.0091904.

[27] M.M. Goodsitt, Conversion relations for quantitative CT bone mineral densities measured with solid and liquid calibration standards, Bone Miner. 19 (2) (1992) 145–158, https://doi.org/10.1016/0169-6009(92)90922-Z.

[28] S.S. Mao, D. Li, Y. Luo, Y.S. Syed, M.J. Budoff, Application of quantitative computed tomography for assessment of trabecular bone mineral density, microarchitecture and mechanical property, Clin. Imaging 40 (2) (2016) 330–338, https://doi.org/10.1016/j.clinimag.2015.09.016.

[29] P.J. Pickhardt, B.D. Pooler, T. Lauder, A.M. del Rio, R.J. Bruce, N. Binkley, Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications, Ann. Intern. Med. 158 (8) (2013) 588–595, https://doi.org/10.7326/0003-4819-158-8-201304160-00003.

[30] P.J. Pickhardt, L.J. Lee, A. Muñoz Del Rio, et al., Simultaneous screening for osteoporosis at CT colonography: bone mineral density assessment using MDCT attenuation techniques compared with the DXA reference standard, J. Bone Miner. Res. 26 (9) (2011) 2194–2203, https://doi.org/10.1002/jbmr.428.

[31] D.L. Christensen, K.E. Nappo, J.A. Wolfe, et al., Ten-year fracture risk predicted by proximal femur Housfield units, Osteoporos. Int. 31 (11) (2020) 2123–2130, https://doi.org/10.1007/s00198-020-05477-y.

[32] D.L. Christensen, K.E. Nappo, J.A. Wolfe, et al., Proximal femur hounsfield units on CT colonoscopy correlate with dual-energy X-ray absorptiometry, Clin. Orthop. Relat. Res. 477 (4) (2019) 850–860, https://doi.org/10.1097/CORR.0000000000000480.

[33] A.H. Habashy, X. Yan, J.K. Brown, X. Xiong, S.C. Kaste, Estimation of bone mineral density in children from diagnostic CT images: a comparison of methods with and without an internal calibration standard, Bone. 48 (5) (2011) 1087–1094, https://doi.org/10.1016/j.bone.2010.12.012.

[34] R.M. Summers, N. Baecher, J. Yao, et al., Feasibility of simultaneous CT colonography and fully-automated bone mineral densitometry in a single examination, J. Comput. Assist. Tomogr. 35 (2) (2011) 212–216, https://doi.org/10.1097/RCT.0b013e3182032537.

[35] J. Therkildsen, J. Thygesen, S. Winther, et al., Vertebral bone mineral density measured by quantitative computed tomography with and without a calibration phantom: a comparison between 2 different software solutions, J. Clin. Densitom. 21 (3) (2018) 367–374, https://doi.org/10.1016/j.jocd.2017.12.003.

[36] B.J. Schwaiger, L. Facchetti, A.S. Gersing, et al., Vertebral and femoral bone mineral density and bone strength in prostate cancer patients assessed in phantomless PET/CT examinations, Bone. 101 (2017) 62–69, https://doi.org/10.1016/j.bone.2017.04.008.

[37] J.L. Fidler, N.S. Murthy, S. Khosla, et al., Comprehensive assessment of osteoporosis and bone fragility with CT colonography, Radiology. 278 (1) (2016) 172–180, https://doi.org/10.1148/radiol.2015141984.

[38] M. Saffarzadeh, R.C. Hightower, J.W. Talton, A.N. Miller, J.D. Stitzel, A.A. Weaver, Multicenter analysis of CIREN occupant lumbar bone mineral density and correlation with age and fracture incidence, Traffic Inj Prev. 17 (2016) 34–41, https://doi.org/10.1080/15389588.2016.1203429.

[39] M.F. Boomsma, I. Slouwerhof, J.A. van Dalen, et al., Use of internal references for assessing CT density measurements of the pelvis as replacement for use of an external phantom, Skelet. Radiol. 44 (11) (2015) 1597–1602, https://doi.org/10.1007/s00256-015-2206-5.

[40] Weaver AA, Beavers KM, Hightower RC, Lynch SK, Miller AN, Stitzel JD. Lumbar Bone Mineral Density Phantom-less Computed Tomography Measurements and Correlation with Age and Fracture Incidence. *Traffic Inj Prev.* 2015;16(02): 153–160. doi:https://doi.org/10.1016/j.physbeh.2017.03.040.

[41] H. Gudmundsdottir, B. Jonsdottir, S. Kristinsson, A. Johannesson, D. Goodenough, G. Sigurdsson, Vertebral bone density in Icelandic women using quantitative computed tomography without an external reference phantom, Osteoporos Int a J Establ as result Coop between Eur Found Osteoporos Natl Osteoporos Found USA. 3 (2) (1993) 84–89, https://doi.org/10.1007/BF01623378.

[42] A.S. Michalski, B.A. Besler, G.J. Michalak, S.K. Boyd, CT-based internal density calibration for opportunistic skeletal assessment using abdominal CT scans, Med. Eng. Phys. 78 (2020) 55–63, https://doi.org/10.1016/j.medengphy.2020.01.009.

[43] J. Aubrey, N. Esfandiari, V.E. Baracos, et al., Measurement of skeletal muscle radiation attenuation and basis of its biological variation, Acta Physiol. 210 (3) (2014) 489–497, https://doi.org/10.1111/apha.12224.

[44] J.T. Wilmink, J.G. Roukema, W. Van den Burg, Effects of IV contrast administration on intraspinal and paraspinal tissues: a CT study. 2. Visual assessment, Am. J. Neuroradiol. 9 (1) (1988) 191–193.

[45] White DR. Tissue substitutes in experimental radiation physics. *Med Phys.* 1978;5 (6).

[46] Hubbel JH, Seltzer SM. Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients (version 1.4). National Institute of Standards and Technology, Gaithersburg, MD. https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients. Published 2004. Accessed January 19, 2018.

[47] J.Y. Rho, M.C. Hobatho, R.B. Ashman, Relations of mechanical properties to density and CT numbers in human bone, Med. Eng. Phys. 17 (5) (1995) 347–355, https://doi.org/10.1007/978-94-010-0367-4_13.

[48] M.J. Ciarelli, S.A. Goldstein, J.L. Kuhn, D.D. Cody, M.B. Brown, Evaluation of orthogonal mechanical properties and density of human trabecular bone from the major metaphyseal regions with materials testing and computed tomography, J. Orthop. Res. 9 (5) (1991) 674–682, https://doi.org/10.1002/jor.1100090507.

[49] R.J. McBroom, W.C. Hayes, W.T. Edwards, R.P. Goldberg, A.A. White, Prediction of vertebral body compressive fracture using quantitative computed-tomography, J Bone Jt Surg. 67-A (8) (1985) 1206–1214.

[50] S.I. Esses, J.C. Lotz, W.C. Hayes, Biomechanical properties of the proximal femur determined in vitro by single-energy quantitative computed tomography, J. Bone Miner. Res. 4 (5) (1989) 715–722, https://doi.org/10.1002/jbmr.5650040510.

[51] Jeffrey Lotz, Tobin Gerhart, W. Hayes, Mechanical properties of trabecular bone from the proximal femur: a quantitative CT study, J. Comput. Assist. Tomogr. 14 (1) (1990) 107–114.

[52] T.S. Kaneko, J.S. Bell, M.R. Pejcic, J. Tehranzadeh, J.H. Keyak, Mechanical properties, density and quantitative CT scan data of trabecular bone with and without metastases, J. Biomech. 37 (4) (2004) 523–530, https://doi.org/10.1016/j.jbiomech.2003.08.010.

[53] E. Schileo, E. Dall'Ara, F. Taddei, et al., An accurate estimation of bone density improves the accuracy of subject-specific finite element models, J. Biomech. 41 (11) (2008) 2483–2491, https://doi.org/10.1016/j.jbiomech.2008.05.017.

[54] J.F. Vivanco, T.A. Burgers, S. Garcia-Rodriguez, et al., Estimating the density of femoral head trabecular bone from hip fracture patients using computed tomography scan data, Inst Mech Eng 228 (6) (2014) 616–626, https://doi.org/10.1177/0954411914540285.

[55] C.M. Ford, T.M. Keaveny, W.C. Hayes, The effect of impact direction on the structural capacity of the proximal femur during falls, J. Bone Miner. Res. 11 (3) (1996) 377–383, https://doi.org/10.1002/jbmr.5650110311.

[56] E.F. Morgan, H.H. Bayraktar, T.M. Keaveny, Trabecular bone modulus-density relationships depend on anatomic site, J. Biomech. 36 (7) (2003) 897–904, https://doi.org/10.1016/S0021-9290(03)00071-X.

[57] S.J. Lee, P.A. Anderson, P.J. Pickhardt, Predicting future hip fractures on routine abdominal CT using opportunistic osteoporosis screening measures: a matched case-control study, Am. J. Roentgenol. 209 (2) (2017) 395–402, https://doi.org/10.2214/AJR.17.17820.

[58] Software I. Mindways, Qct pro ™, 2013.

[59] C.C. Glüer, G. Blake, Y. Lu, B.A. Blunt, M. Jergas, H.K. Genant, Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques, Osteoporos. Int. 5 (4) (1995) 262–270, https://doi.org/10.1007/BF01774016.

[60] B. Helgason, E. Perilli, E. Schileo, F. Taddei, S. Brynjólfsson, M. Viceconti, Mathematical relationships between bone density and mechanical properties: a

literature review, Clin. Biomech. 23 (2) (2008) 135–146, https://doi.org/10.1016/j.clinbiomech.2007.08.024.

[61] Z. Altai, M. Qasim, X. Li, M. Viceconti, The effect of boundary and loading conditions on patient classification using finite element predicted risk of fracture, Clin. Biomech. 68 (June) (2019) 137–143, https://doi.org/10.1016/j.clinbiomech.2019.06.004.

[62] H.H. Bayraktar, E.F. Morgan, G.L. Niebur, G.E. Morris, E.K. Wong, T.M. Keaveny, Comparison of the elastic and yield properties of human femoral trabecular and cortical bone tissue, J. Biomech. 37 (1) (2004) 27–35, https://doi.org/10.1016/S0021-9290(03)00257-4.