journal of the mechanical behavior of biomedical materials 139 (2023) 105664

Contents lists available at ScienceDirect

Journal of the Mechanical Behavior of Biomedical Materials

journal homepage: www.elsevier.com/locate/jmbbm

Research paper

The predictive ability of a QCT-FE model of the proximal femoral stiffness under multiple load cases is strongly influenced by experimental uncertainties

Morteza Amini [a,*], Andreas Reisinger [b], Alexander Synek [a], Lena Hirtler [c], Dieter Pahr [b,a]

[a] Institute of Lightweight Design and Structural Biomechanics, TU Wien, Getreidemarkt 9, 1060, Vienna, Austria
[b] Division Biomechanics, Karl Landsteiner University of Health Sciences, Dr.-Karl-Dorrek-Straße 30, 3500 Krems an der Donau, Austria
[c] Center for Anatomy and Cell Biology, Medical University of Vienna, Währinger Straße 13, 1090, Vienna, Austria

ARTICLE INFO

Keywords:
Biomechanics
Femur
Finite element modeling
Parametric study
Experimental uncertainties

ABSTRACT

Despite significant improvements in terms of the predictive ability of Quantitative Computed Tomography based Finite Element (QCT-FE) models in estimating femoral strength (fracture load and stiffness), no substantial clinical adoption of this method has taken place to date. Narrowing the wide variability of FE results by standardizing the methodology and validation protocols, as well as reducing the uncertainties in the FEA process have been proposed as routes towards improved reliability. The aim of this study was to: First, validate a QCT-FE model of proximal femoral stiffness in multiple stance load cases, and second, using a parametric approach, determine the influence of select experimental and modeling parameters on the predictive ability of our model. Ten fresh frozen human femoral samples were tested in neutral stance, 15° adducted and 15° abducted load cases. Voxel-based linear-elastic QCT-FE models of the samples were generated to predict the models' stiffness values in all load cases. The base FE models were validated against the experimental results using linear regression. Thirty six deviated models were created using the minimum and maximum values of experiment-based "plausible range" for 18 parameters in 4 categories of embedding, loading, material, and segmentation. The predictive ability of the models were compared in terms of the coefficient of determination (R^2) of the linear regression between the measured and predicted stiffness values in all load cases. Our model was capable of capturing 90% of the variation in the experimental stiffness of the samples in neutral stance position (R^2 = 0.9, concordance correlation coefficient (CCC) = 0.93, percent root mean squared error (RMSE%) = 8.4%, slope and intercept not significantly different from unity and zero, respectively). Embedding and loading categories strongly affected the predictive ability of the models with an average percent difference in R^2 of 4.36% ± 2.77 and 2.96% ± 1.69 for the stance-neutral load case, respectively. The performance of the models were significantly different in adducted and abducted load cases with their R^2 dropping to 71% and 70%, respectively. Similarly, off-axes load cases were affected by the parameters differently compared to the neutral load case, with the loading parameter category imposing more than 10% difference on their R^2, larger than all other categories. We also showed that automatically selecting the best performing plausible value for each parameter and each sample would result in a perfectly linear correlation (R^2 > 0.99) between the "tuned" model's predicted stiffness and experimental results. Based on our results, high sensitivity of the model performance to experimental parameters requires extra diligence in modeling the embedding geometry and the loading angles since these sources of uncertainty could dwarf the effects of material modeling and image processing parameters. The results of this study could help in improving the robustness of the QCT-FE models of proximal femur by limiting the uncertainties in the experimental and modeling steps.

* Corresponding author.
E-mail addresses: morteza.amini@tuwien.ac.at (M. Amini), andreas.reisinger@kl.ac.at (A. Reisinger), asynek@ilsb.tuwien.ac.at (A. Synek), lena.hirtler@meduniwien.ac.at (L. Hirtler), dieter.pahr@kl.ac.at (D. Pahr).

https://doi.org/10.1016/j.jmbbm.2023.105664
Received 26 October 2022; Received in revised form 3 January 2023; Accepted 5 January 2023
Available online 9 January 2023

1. Introduction

Hip fracture is one of the most consequential musculoskeletal events with 50% of the patients unable to walk after the fracture and 30% morbidity rate after one year (Brauer et al., 2009). Added to the high prevalence of osteoporosis (1 in 3 women and 1 in 5 men) and the aging demography of the developed countries, it has been a main target for biomechanical studies to improve the clinical fracture risk assessment criteria using novel methods in the past couple of decades (Lee et al., 2019; Bouxsein et al., 2020; Falcinelli and Whyne, 2020). In the absence of any direct in situ bone strength assessment techniques, surrogate metrics are employed to formulate fracture risk predictions (Bouxsein and Delmas, 2008). The current clinical standard is based on DXA-BMD measurements. It has been shown that image-based finite element models can outperform this metric significantly (Falcinelli and Whyne, 2020; Viceconti et al., 2018). However, there are still no clinically approved FE models being used in daily practice.

FE-assisted prediction of hip fracture risk in comparison with methods based on purely morphological parameters requires substantially more endeavor, and consequently, is exposed to more sources of uncertainty and error. Several imaging modalities including micro computed tomography (μCT), high resolution peripheral quantitative CT (HR-pQCT), QCT, dual-energy X-ray absorptiometry (DXA), and magnetic resonance imaging (MRI) have been used for FE modeling of bone (Falcinelli and Whyne, 2020; Rajapakse et al., 2020). High resolution variants (μCT and HR-pQCT) are not clinically available and have very restricted in vivo use cases (Panyasantisuk et al., 2018). DXA images, due to their 2-D nature, are limited in terms of the amount of data they can provide for 3D FE models (Dall'Ara et al., 2016). Despite its non-radiating nature, and conceptually proven ability to provide adequate resolution and accuracy in measuring bone mineral density and structural properties, substantially high costs and very limited clinical availability of MRI machines have prevented them to play a significant role in bone biomechanics studies (Zhang et al., 2021; Haque et al., 2018; Chou and LeBoff, 2017). With wide clinical accessibility and adequate in vivo image quality, QCT scans have been the main basis for image-based FE modeling of bone (Sas et al., 2020). Since the initial pioneering works in this field a few decades ago, the accuracy and precision of FE models in mimicking the physiological behavior of bone has been significantly improved. In particular, recent works have reported high predictive ability values for fracture load (R^2 = 0.64–0.96) and overall stiffness (R^2 = 0.55–0.82) of femur (Sas et al., 2022; Rossman et al., 2016; Miura et al., 2017; Helgason et al., 2016). Despite these advances, there are almost no clinical applications of FE models to date. The failure to transition from lab to the bed side has been attributed to lack of standards and substantial complexity in every aspect of the FE modeling process and hence the wide variability in results (Bouxsein et al., 2020; Falcinelli and Whyne, 2020).

For an FE model to be deemed valid, the gold standard is to assess its results against in vitro mechanical experiments. A general work flow for FE model development and validation is depicted in Fig. 1. In the experimental phase, fresh-frozen samples go through preparation and imaging steps with probable storage intervals in between. Specimens are then embedded and fitted in accordance to the data collection methods (applying strain gauges, painted with speckles, fitted with markers, etc.). Specific test setups are utilized to achieve the target loading configuration. For the FE modeling phase, the model geometry, mesh, and material properties are generated and assigned based on the sample-specific QCT image data. The boundary conditions and loading are prescribed according to the experimental protocol. Models are validated by comparing their predicted outcomes to the experimentally measured values. Every step of the way, choices are to be made from a multitude of available options with some level of flexibility and uncertainties are introduced into the final results. Since the FE models are developed in accordance with the experimental process, the extent to which their predictions can be generalized is limited. A study targeting the predictive ability of established FE models across multiple laboratories for surface strain measurements of a specific unseen femur revealed that although each model might have achieved a high score in their respective validation study, a significantly large range of errors were present in their predictions of the unseen strains (Kluess et al., 2019).

Several studies have investigated the sensitivity of femur's strength measures (fracture load (F_u) and stiffness (K)) to different parameters. Age and sex was shown to have significant effect on the F_u and K (Patton et al., 2019; Miura et al., 2017). It was shown that CT scanning parameters and conditions such as calibration methods, scanner settings, scanner filters, and sample environment could alter accuracy of bone FE models, specifically FE predictions of femoral strength and stiffness (Benca et al., 2020; Dragomir-Daescu et al., 2015). Different loading directions and configurations (i.e. stance and fall) have resulted in different strength predictions of femur (Bessho et al., 2009; Wakao et al., 2009). Loading rate has also been shown to be a predictive factor for femoral strength (Askarinejad et al., 2019). In both stance and fall configurations, definitions of the boundary conditions have been shown to significantly affect the outcomes of the FE models of femoral strength (Altai et al., 2019; Rossman et al., 2016). When it comes to simulating the material behavior of the models, different material mapping techniques and density-modulus relationships have led to different predictions of the femoral strength (Helgason et al., 2016; Wille et al., 2016). Other parameters such as mesh type and size, partial volume effect, pathological conditions, etc. have also been investigated in different studies (Falcinelli and Whyne, 2020; Bouxsein et al., 2020; Lee et al., 2019; Mishra et al., 2022). The dominant approach of these studies has been to report the predictive ability of their model and move on to measure the resulted change in the predicted strength metric due to the altered parameter choice. There are only a few examples in which the predictive ability of the model was selected as the outcome variable (Patton et al., 2019; Panyasantisuk et al., 2018; Alomari et al., 2018).

In our previous study, we tried to provide a comparison for effect of experimental parameters on the measured stiffness of proximal femoral sample (Amini et al., 2021). As the next step, we wanted to develop and validate a QCT-FE model based on the same samples and experimental data. However, there was a lack of studies systematically addressing the discrepancies between the superb level of performance of such FE models during their validation phase and their subpar outcomes when applied to unseen samples. Such information is necessary for making progress in terms of defining standards in this field and increasing the possibility of clinical use cases of these models. This study aimed to validate a voxel-based QCT-FE model of the proximal femur for three stance load cases and examine the influence of various sources of uncertainty on its predictive ability of the overall sample stiffness through parametric analysis. The selected parameters fall within four categories of embedding, loading, material properties, and segmentation.

2. Methods & material

2.1. Experiments

An extensive description of the experimental procedure and setup used in this study has been previously reported elsewhere (Amini et al., 2021). Briefly:

Samples. Ten paired fresh frozen human femora were obtained and stored at −23° in accordance with relevant guidelines (Center for Anatomy and Cell Biology, Medical University of Vienna). The specimens originated from voluntary body donations for scientific and teaching purposes to the Center and were screened for lack of any pathological disease (According to protocol approved by the ethics committee of Karl Landsteiner University of Health Sciences) (Table 1).

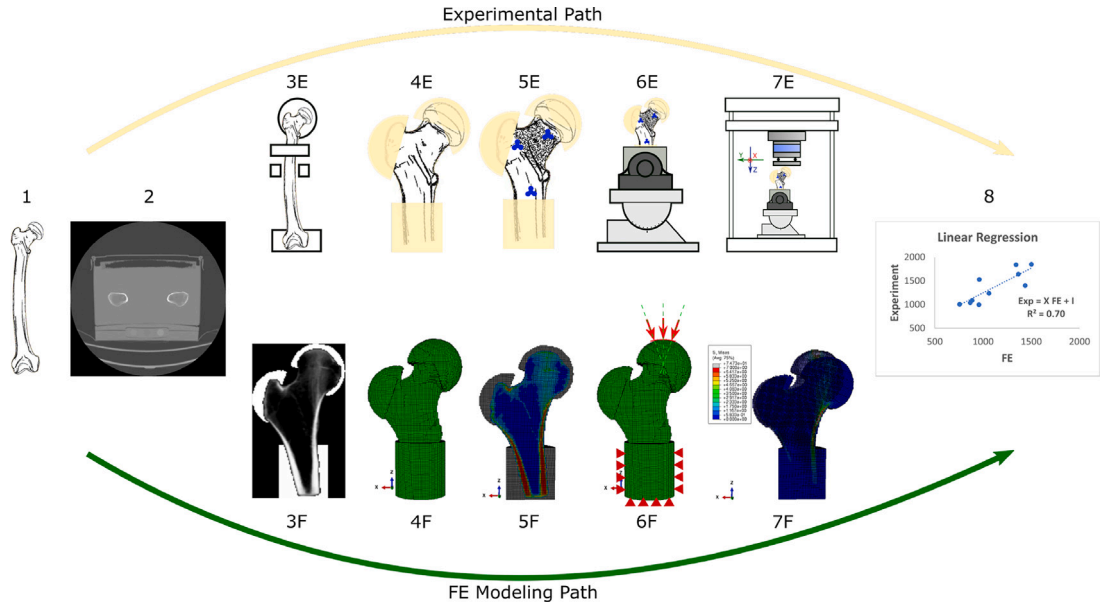


Fig. 1. (1) Fresh frozen samples are acquired from donor banks, (2) QCT images are taken from the samples in physiologically relevant conditions, (3E) Samples are prepared by cleaning, alignment, and cutting, (4E) Samples are embedded to provide support sections and help with controlled BC and loading definition, (5E) Required means of data collection are applied (SG, Speckles, Markers, etc.), (6E) Samples are fixed into the defined positions and configurations, (7E) Mechanical testing is performed, (3F) The model geometry is extracted from the image data through various image processing steps, (4F) Generated volume is meshed, (5F) Material properties are mapped from the images to the elements using certain material mapping algorithms and selected density-modulus relationships, (6F) Boundary conditions corresponding to the experimental setup or physiological phenomenon are defined, (7F) Analysis is done according to defined material behavior, (8) Experimental (measured) and FE (predicted) results are compared to assess the validity of the model.

Table 1
Information on donors of the samples. Both femora from each donor were tested (10 samples). Only samples with no pathological conditions were included in the study.

Donor ID	Gender	Age	Height (m)	Weight (kg)
#1	F	63	1.53	65.4
#2	F	69	1.67	49.3
#3	F	70	1.67	48.1
#4	M	78	1.71	72.0
#5	M	78	1.77	66.8

Prior to the experiments, samples were defrosted in room temperature 0.9% PBS (Phosphate-buffered saline) bath, removed of the soft tissue and periosteum in the proximal region, aligned in the neutral stance position using a custom-made rig, and cut to get a 15-cm sample length. Finally, they were embedded in Polyurethane (PU) resin at the head, trochanter, and shaft locations using potting setup (Fig. 2).

Imaging. QCT images of the samples were taken before the mechanical tests (Toshiba Aquilion Prime, Toshiba Medical Systems Corp., Tokyo, Japan). Custom-made scan chambers capable of maintaining the alignment of the samples and keeping them in paired position while submerged in PBS solution were used (Fig. 3. Submerged samples were placed in a vacuum chamber for 30 min to remove the air bubbles (Trivac D8B; OC Oerlikon Management AG, Pfäffikon, Switzerland). Scans were taken with 0.625 mm in-plane resolution and 0.25 mm slice thickness (voltage: 120 kV, current: 150 mA). A three-rod hydroxyapatite (HA) phantom (HA0, HA100, HA200) (QRM GmbH, Germany) was placed underneath the chamber for calibration purposes . After each sample was tested, μCT scans of it were taken using another custom scan chamber which kept the sample in vertical alignment while maintaining the sample hydration throughout the scanning span (Skyscan 1173, Bruker, Belgium) (field of view: 120 × 150 mm, resolution: 30 μm, voltage: 130 kV, current: 60 mA, exposure: 580 ms, filter: Al 1.0 mm) (Fig. 3).

Mechanical testing. The samples were fixed at the shaft region on a 5 degrees of freedom (DOF) (X, Y, Rx, Ry, Rz) tilting table keeping

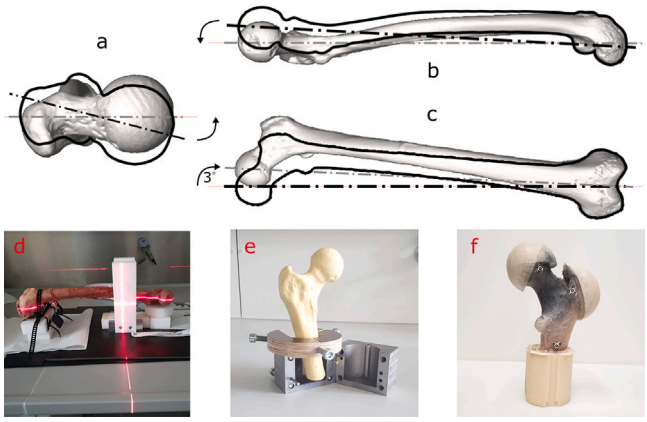


Fig. 2. To achieve a neutral stance alignment, (a) the shaft axis was brought into horizontal position by rotating the intact specimen around its shaft axis, (b) the femoral head center was elevated until leveled with the distal mid-condylar line, and (c) the mechanical axis was adducted 3° beyond the reference line. (d) A custom-made alignment setup with two cross lasers were used. (e) The alignment was fixed on the proximal portion by fixing a ring a 100 mm below the superior surface of the femur and cut 50 mm below the ring. (f) PU resin was used to embed the samples at the shaft, head, and trochanter regions.

them in the desired stance alignment. The table was fixed on a 30 kN electro-mechanical axial testing machine (Z030, ZwickRoell Ulm, Germany). The loading plate was free to move in horizontal directions against a flat ball-bearing support. A non-destructive loading regime with 0.75 of the donor's body weight at a 5 mm/sec rate, with no pre-loading, was applied to the samples (Fig. 4). Each sample was tested in neutral-stance, 15°-abducted and 15°-adducted orientations, in a random order, to avoid systematic errors (Further details and justifications for the experimental protocol could be found in a previous publication (Amini et al., 2021))

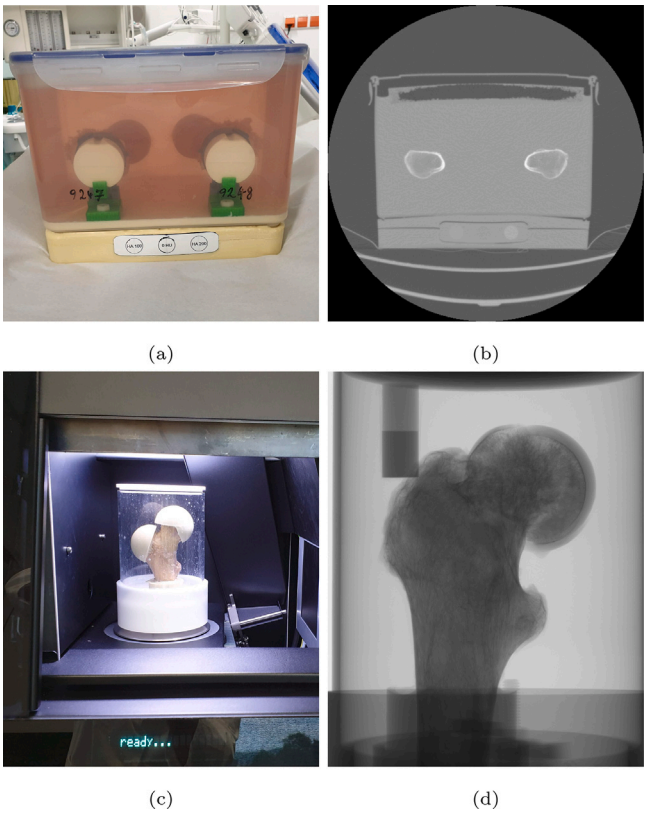


Fig. 3. (a) QCT: Samples were clamped in an aligned and paired configuration while submerged in saline solution. (b) A sample slice of the QCT scan. (c) μCT: Each sample was clamped in upright position in a custom chamber keeping it hydrated. (d) A sample μCT scan projection.

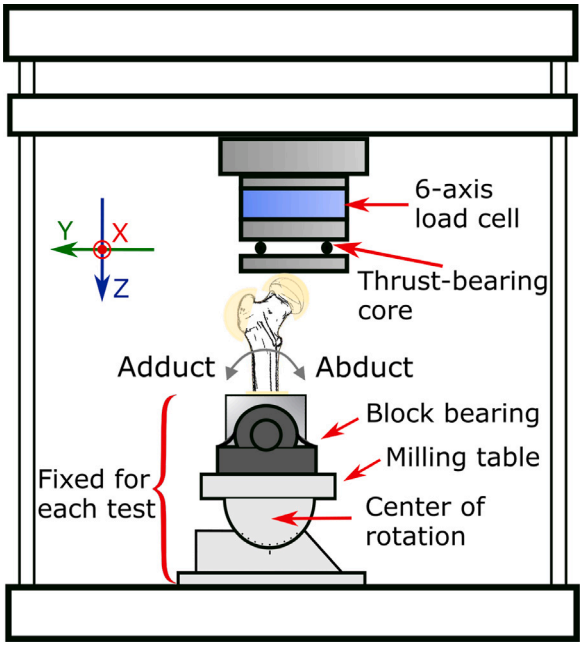


Fig. 4. Different parts of the mechanical testing setup. This setup was capable of testing the samples in stance and fall configurations and with various alignments.

Data. The load–displacement data was captured directly from the testing machine and the 6-axis load cell (Hottinger Baldwin Messtechnik (HBM) GmbH, Germany). The transverse load data was used to monitor the integrity of the boundary conditions at the load introduction point. No substantial peak shear forces (< 2 N) were recorded across all tests.

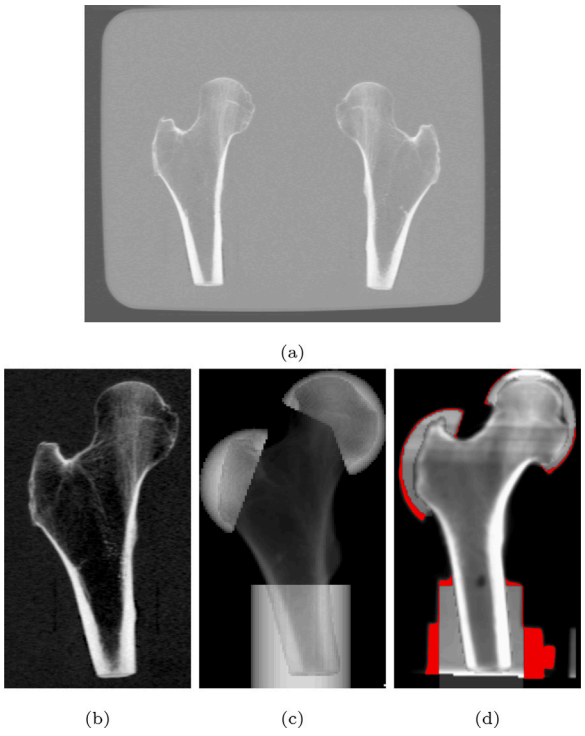


Fig. 5. (a) CT images of the paired samples submerged in physiological solution, (b) Each sample was cropped, scaled to 1 mm voxel size, calibrated using the phantom linear equation, and segmented from the surrounding fluid and soft tissue, (c) Embedding volumes were added using a combination of geometrical steps based on the experimental protocol, (d) μCT images of the embedded samples were registered with the CT-based volumes to fine-tune the embedding locations.

2.2. QCT-FE modeling

Image processing. The Hounsfield unit (HU) scale was converted to equivalent bone mineral density (BMD) scale using the linear calibration equation from the phantom. The most dense cortical bone was assumed to have a 1200 mgHA/cc density (Chiang et al., 2018). Samples were isolated from the paired scans. Using an automated segmentation algorithm based on global threshold, image enhancement and a morphological closing operations (Pahr and Zysset, 2009), samples were segmented from the surrounding fluid and chamber (Medtool 4.2, Dr. Pahr Ingenieurs e.U., Austria) (Fig. 5). Images were re-scaled to acquire a 1 mm voxel size. Since the submerged embedding material was invisible in our CT images, embedding volumes were added to the segmented bones using geometrical transformations based on the experimental protocol using medtool. The location and geometry of the embedding volumes were cross-checked with down-scaled μCT images of each embedded sample using 3D volume registration, and any location/dimension inaccuracies were corrected accordingly (3D slicer, slicer.org) (Fedorov et al., 2012) (Fig. 5).

Model generation. Images were converted to 8-bit gray-level format with values between 0 and 250, keeping the remaining 251–255 gray values for the embedding material. The Image voxels were directly converted into linear hexahedron elements. The Young's modulus (E (MPa)) of the elements was directly mapped from the gray value (GV) of the corresponding voxels using the power-law density-elasticity relationship below, where E_0 is assigned to the voxels with 100% bone density (Daszkiewicz et al., 2016; Synek et al., 2015; Gross et al., 2012).

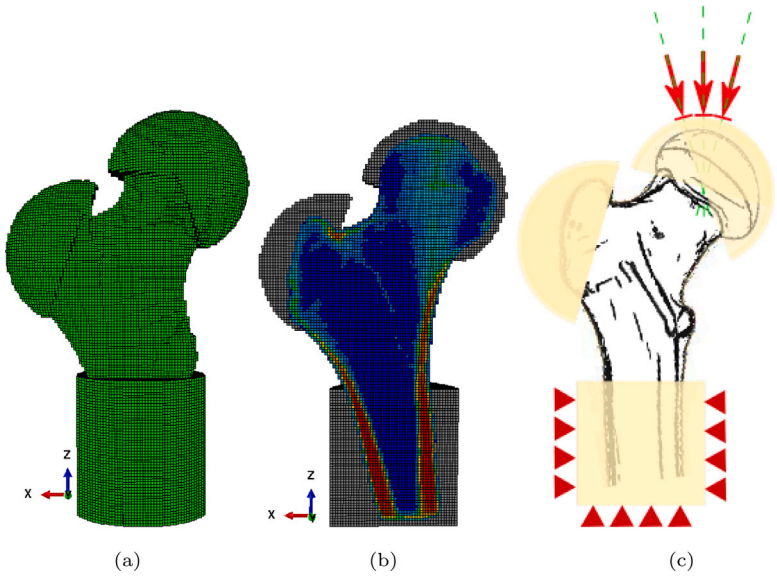


Fig. 6. (a) A meshed volume with 1-mm linear brick elements. (b) Material was mapped directly from each voxel to the corresponding element using Eq. (1). (c) Boundary conditions were defined according to the testing setup. Load was applied through a reference node coupled with contact nodes on the femoral head embedding sphere.

The Poisson's ratio was set to 0.3 for all elements (Fig. 6):

$$E = E_0 \times \rho^{\alpha} = 10000 \times (\frac{GV}{250.0})^{1.63} \quad (1)$$

Boundary conditions. The surface nodes of the shaft embedding cylinder, except for the top surface, were fixed in all translational and rotational degrees of freedom. To apply the load, a contact area on the femoral head embedding segment was defined. Preliminary contact FE analyses of composite hemispheres (representing the embedded femoral head) with 50 to 60 mm diameter and a steel plate resulted in a contact radius of 1.6 to 1.8 mm. Given the 1-mm voxel size of the model, a radius of 2 mm was chosen. To apply the load, a reference node was defined at the center of the contact area. The contact nodes were selected by positioning the contact area according to the sample orientation and selecting the enclosed surface nodes of the femoral head embedding sphere. The load was applied to the reference node perpendicular to the contact area and equal to 0.75 of the donor's body weight. A kinematic coupling was defined between the reference and contact nodes. The 15° adducted and abducted (tilted) load cases were created by rotating the contact surface around the femoral head center (Fig. 6).

Analysis. Input files were automatically generated using python scripts in Medtool. Linear FE analyses were done in Abaqus (Simulia, Dassault Systèmes, France) using 8 cores on an Linux server. Each job took approximately 5 min.

Model validation. The measured and predicted values for the overall proximal femoral stiffness of the 10 samples were compared to each other in order to test the validity of the FE models. The experimental stiffness was defined as the slope of the linear section of the load–displacement curve recorded by the mechanical testing machine. Based on preliminary experiments, the linear section was defined between 200 N and 400 N. Since the FE analyses were done using a linear elastic material model, the numerical stiffness was predicted by dividing applied load by the vertical displacement of the reference node:

$$K_{FE} = \frac{F}{\Delta Z} \quad (2)$$

where K_{FE} stand for the numerical stiffness, F is the applied load to the reference node equal to 75% of each donor's body weight, and ΔZ represents the vertical displacement of the reference node.

2.3. Parametric structure

Outcome variable. The outcome variable of the study was the predictive ability of the models. Predictive ability was defined as the coefficient of determination (R^2) of the FE predicted vs experimentally measured overall proximal femoral sample stiffness regression analysis.

Parameters. The parameters for which the sensitivity of the R^2 was tested are (Table 2):

- *Embedding Location:* Spatial positioning of the embedding spheres and cylinder with respect to the sample's coordinate system.
- *Embedding Thickness:* Thickness of embedding material along the loading direction between bone and surface of the embedding segment.
- *Embedding Dimension:* Radius of the embedding spheres and radius and height of the shaft cylinder.
- *Embedding Elastic Modulus:* Young's modulus of the embedding material.
- *Loading Angle:* Angle of the applied load with respect to the vertical axis in the stance neutral position.
- *Contact Area:* Radius of the contact area at the load introduction site.
- *Reference Node Location:* Spatial positioning of the reference node with respect to the sample's coordinate system.
- E_0*:* Bone tissue elastic modulus according to density-elasticity relationship in Eq. (1).
- α*:* Calibration constant according to density-elasticity relationship in Eq. (1).
- *Calibration threshold:* The density value assumed for the most dense cortical bone in the calibration step.
- *Segmentation threshold:* The bone gray value threshold in the segmentation step within the 0–250 range.

Parameters were grouped under categories and subcategories based on relevance. For each parameter, a plausible deviation range was defined based on the experience gained during the experimental process. In each iteration, only one parameter was deviated from its baseline and the subsequent FE models for all samples were solved. Hence, for each parameter, in addition to the baseline model, two predicted stiffness values for positive deviation and negative deviation were generated per sample per load case. This resulted in 37 models per sample per load case with a total of 1110 FE analyses. The R^2 for predicted-measured

Table 2
Parameter descriptions.

Category	Parameter	Deviation	Notes
Embedding	Embedding location	±2 mm	X,Y,Z directions
	Embedding Thickness	±2 mm	Z direction
	Embedding Dimensions	±2 mm	Diameter
	Embedding Elastic Modulus	±10%	MPa
Loading	Loading Angle	±2 degrees	Abduction, Adduction
	Contact Area	±2 mm	Superior surface
	Reference Node Location	±2 mm	X,Y,Z directions
Material	E_0 [a]	±10%	MPa
	α [a]	±10%	–
	Calibration threshold [b]	±10%	–
Segmentation	Segmentation threshold	±2 ($\simeq \pm 12HU$)	Within the re-sampled gray value range of 0–250.

[a] E_0 and α based on Eq. (1).
[b] The density value assumed for the most dense cortical bone in the calibration step.

stiffness regression line of the deviated model was calculated for each load case and for the pooled data of all three load cases.

Tuning effect. To investigate the effect of tuning the model on its performance, we formed a 37 (models per sample) by 30 (10 samples in 3 load cases) table of absolute differences between the measured and predicted values of overall stiffness. For each sample and load case, we selected the deviated model corresponding to the smallest value in the table to create a "tuned" model. This allowed us to evaluate the extent to which our model's performance could be manipulated by our set of parameters, while staying within their plausible range of values with respect to the experimental results.

2.4. Statistical analysis

Linear regression was used to compare the predicted and measured overall sample stiffness values. The slope, intercept, coefficient of determination (R^2), root mean squared error (RMSE) of the prediction , and RMSE normalized by the mean predicted stiffness (RMSE%) are reported. A general linear hypothesis was used to test if the slope is different from unity and intercept is different from zero, with p-values < 0.05 (using regression analysis tool and Student t-test function in Excel, Microsoft Inc., USA). Adjusted R^2 was reported to account for the difference in sample size between the pooled and isolated load cases. To provide a more comprehensive evaluation of the agreement between the model and experiment outcomes in the validation phase, Bland–Altman plots and concordance correlation coefficient (CCC) were reported as well. Bland–Altman plots allow visualization of the agreement and identification of any trends or biases in the data, while the CCC provides a summary statistic that can be easily compared across studies. The effect of parameters was reported as the average percentage of absolute difference between R^2 of the two deviated models and the baseline. Results are reported for each of the three load cases as well as the pooled data.

3. Results

Validation. Example experimental load–displacement plots are shown in Fig. 7. Linear regression analysis for the measured and predicted values of the overall proximal femoral stiffness of the 10 samples using the base model showed strong correlation for all load cases as well as the pooled data. The strongest correlation was achieved in the neutral stance load case where the model was able to account for 90% of the variation in measured stiffness of the samples (R^2: 0.90, RMSE%: 8%, slope and intercept of the regression line not significantly different from one and zero, respectively.) (Table 3). The tilted load cases showed

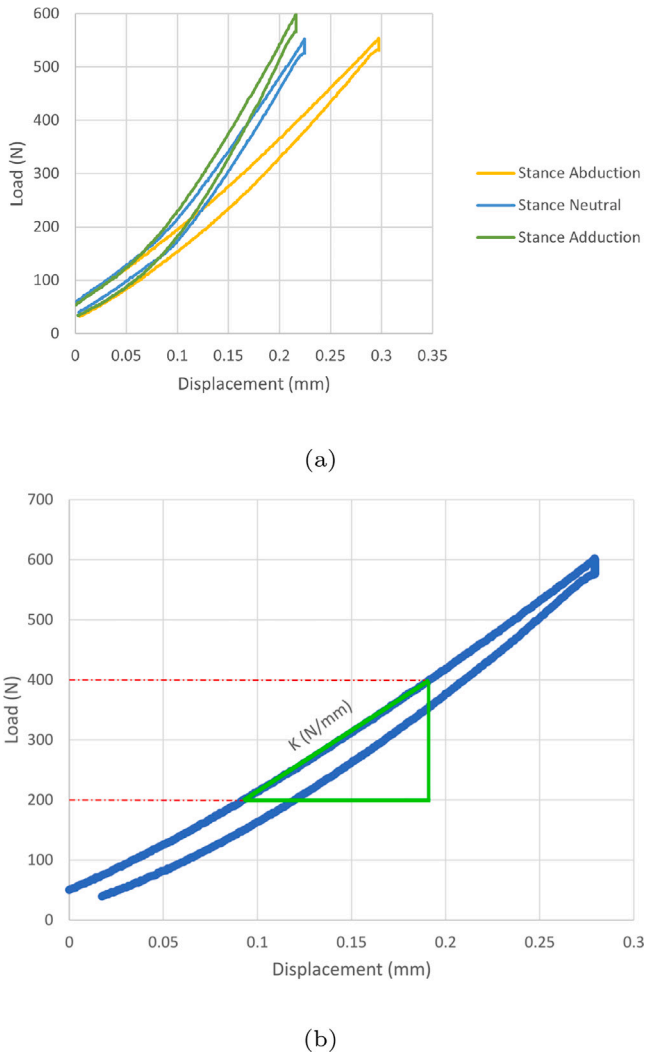


Fig. 7. (a) Load–displacement plots for a sample in all three load cases based on experimental results. (b) Overall stiffness of the samples were calculated based on the linear section of the plots.

almost similar predictive ability with $R^2 = 0.71$ & 0.70 for the adducted and abducted cases, respectively. Pooling the data points for all three load cases resulted in a high R^2 of 0.87, and a regression line with slope and intercepts significantly different from one and zero, respectively. Fig. 8 shows the linear regression and Bland–Altman plots for all cases. The concordance correlation analysis showed high accuracy and precision for the stance neutral and pooled data, with moderate results for the adducted and abducted load cases (CCC: 0.59–0.93) (Table 3).

Parametric analysis. Table 4 shows the effect of each parameter on the overall stiffness of the samples in terms of the average absolute percentage difference of the R^2 between the base model and the deviated models. In the neutral stance case, the embedding dimension imposed the largest effect on the predicted stiffness values with 7.2% difference between the R^2 of deviated and baseline models. Loading angle and contact area were the most influential parameters in case of the adducted and abducted load cases, with 12.71% & 24.87% average absolute percentage difference of the R^2s, respectively. Averaging the results for all three load cases per parameter resulted in the embedding dimension having the largest effect on the R^2 of the models equal to 13.22%. An alternative representation of these outcomes based on the absolute difference in R^2 between the base model and the deviated models is available in the Appendix A.

Table 3
Results of linear regression and concordance correlation analyses for the experimentally measured and FE predicted overall sample stiffness for the 10 samples in three different stance load cases. The p values are reported in parentheses, with 0.05 as the significance threshold.

	Neutral	15° adducted	15° abducted	Pooled
R^2	0.90 ($p < 0.01$)	0.71 ($p < 0.01$)	0.70 ($p < 0.01$)	0.87 ($p < 0.01$)
Adjusted R^2	0.89	0.67	0.66	0.87
Slope	0.86 ($p = 0.19$)	0.65 ($p = 0.0501$)	1.04 ($p = 0.87$)	0.81 ($p = 0.01$)
Intercept (N/mm)	354.63 ($p = 0.11$)	821.83 ($p = 0.03$)	203.86 ($p = 0.48$)	471.23 ($p = 0.002$)
CCC	0.93	0.79	0.59	0.89
RMSE (N/mm)	161.49	280.76	306.56	257.48
RMSE%	8.41	13.74	27.53	15.22

Table 4
Sensitivity of the FE model's predictive ability to select parameters. The values are the average of absolute percentage difference for all deviations under each parameter subcategory.

Category	Parameter	Loading direction			
		Neutral	15° adducted	15° abducted	Pooled
Embedding	Embedding location	2.91% ± 5.74	7.94% ± 13.11	9.33% ± 14.29	4.74% ± 9.97
	Embedding Thickness	6.1% ± 7.1	7.41% ± 8.1	8.61% ± 7.5	1.44% ± 2.18
	Embedding Dimensions	7.2% ± 8.37	6% ± 7.68	17.57% ± 20.5	13.22% ± 15.72
	Embedding Elastic Modulus	1.23% ± 0.17	3.4% ± 1.18	0.21% ± 0.3	1.15% ± 0
Loading	Loading Angle	1.11% ± 1.57	12.71% ± 7.63	1.43% ± 1.42	4.02% ± 4.06
	Contact Area	3.33% ± 3.84	12.24% ± 8.97	24.87% ± 34.44	11.49% ± 13.94
	Reference Node Location	4.43% ± 3.26	6.71% ± 6.14	6.1% ± 5.78	5.17% ± 3.45
Material	E_0 [a]	0% ± 0	1.41% ± 2	0.72% ± 1.01	0.57% ± 0.81
	α [a]	1.11% ± 0	0.71% ± 1	0.72% ± 1.01	0.57% ± 0.81
	Calibration threshold [b]	5.54% ± 0	2.82% ± 0	10.04% ± 0	4.6% ± 0
Segmentation	Segmentation threshold	3.33% ± 4.7	0.71% ± 1	2.87% ± 2.03	0.57% ± 0.81

[a] E_0 and α based on Eq. (1).
[b] The density value assumed for the most dense cortical bone in the calibration step.

Table 5
The average effect of parameter categories on the model's predictive ability.

Category	Loading Direction			
	Neutral	15° adducted	15° abducted	Pooled
Embedding	4.36% ± 2.77	6.19% ± 2.03	8.93% ± 7.09	5.14% ± 5.63
Loading	2.96% ± 1.69	10.55% ± 3.34	10.8% ± 12.4	6.9% ± 4.02
Material	2.22% ± 2.93	1.65% ± 1.08	3.83% ± 5.38	1.92% ± 2.32
Segmentation	3.33% ± 4.7	0.71% ± 1	2.87% ± 2.03	0.57% ± 0.81

Table 5 lists the parameter categories and their effects on the R^2. The embedding category presented the highest effect compared to the rest of the parameter categories with an average of 4.36% difference incurred in the R^2 of the predicted and measured stiffness of the 10 samples in neutral case. The embedding dimension followed by the thickness sub categories contributed the most. Between the head and shaft locations, the head embedding was more influential across all parameters. In tilted load cases however, the loading category of parameters dominated the rest of the parameters with the loading angle having the largest effect in both cases. The pooled data showed the highest sensitivity to the loading category followed by the embedding. In all four groups, the material and segmentation parameter categories fell far behind the other two in terms of influencing the outcome results (Table 5).

The predictive ability of the FE model was different based on the loading direction (Fig. 8). Similarly, the parameter categories affect the R^2 of predicted vs measured stiffness differently based on the loading direction (Fig. 9). There was an increase in the average percent difference of R^2 from the stance neutral alignment to stance-adducted and stance-abducted alignments.

The R^2 for the tuned model was 0.99 for all load cases with the slope and intercept not significantly different from 1 and 0, respectively ($p < 0.01$) (Fig. 10). The same as for the parametric results, embedding and loading categories were the determinant choices for the tuned model with the most accurate stiffness predictions for each sample coming from the deviated models in the loading and embedding categories.

4. Discussion

In this study, a QCT-based FE model of proximal femoral stiffness was validated in three different stance loading direction. Using a parametric approach, four categories of uncertainty sources were investigated in terms of their effects on the predictive ability of the model. The model's performance and sensitivity to the parameters was dependent on the loading direction, with the neutral stance case showing the most robust results.

Based on the presented results, our voxel-based linear-elastic FE model was able to capture 90% of the variation in the measured proximal femoral samples' stiffness in neutral stance loading direction. The results of the linear regression analysis, Bland–Altman plots, and concordance correlation analysis all pointed at strong accuracy and precision of the model for this load case. The predictive ability of our model is in the higher range of the stiffness prediction results in the literature (Patton et al., 2019; Askarinejad et al., 2019; Panyasantisuk et al., 2018; Miura et al., 2017; Rossman et al., 2016; Helgason et al., 2016; Enns-Bray et al., 2016; Dragomir-Daescu et al., 2015; Dall'Ara et al., 2013).

The parameters under the Embedding group induced the highest average difference of 4.36% in the predictive ability of the neutral model. Our results indicate that the QCT-FE models of proximal femoral samples are most sensitive to the relative location of the embedding volumes (specially the head semi-sphere), the thickness of the embeddings along the loading axis, and the dimensions of the embedding. Understanding the uncertainties in geometrical modeling of the embedding volumes is particularly important when the embedding is numerically added to the image-based FE model (Sas et al., 2020; Patton et al., 2019; Miura et al., 2017; Rossman et al., 2016; Dall'Ara et al., 2013). Since the typical embedding materials are much softer than the cortical bone (PU resin: 1.36 GPa, PMMA: 3 GPa), these sections highly influence the overall sample stiffness. Inaccurate modeling of the embeddings could ultimately result in discarding the stiffness measurements all together due to poor correlation with the experiments (Haider et al., 2018). The option to omit the embedding

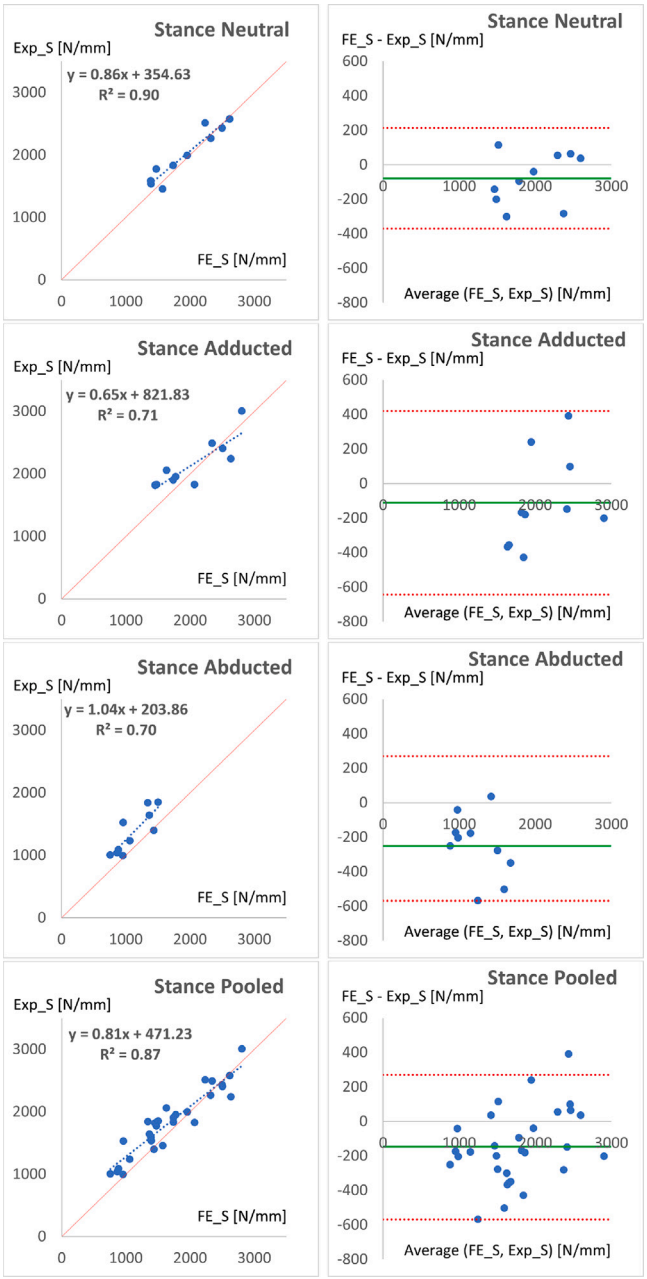


Fig. 8. Linear regression and Bland–Altman plots for the measured and predicted overall sample stiffness of the 10 samples in three stance load cases and pooled.

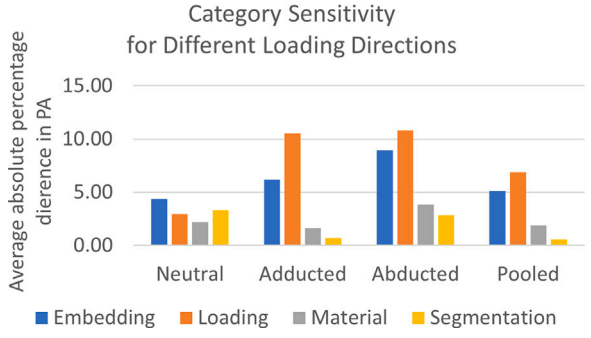
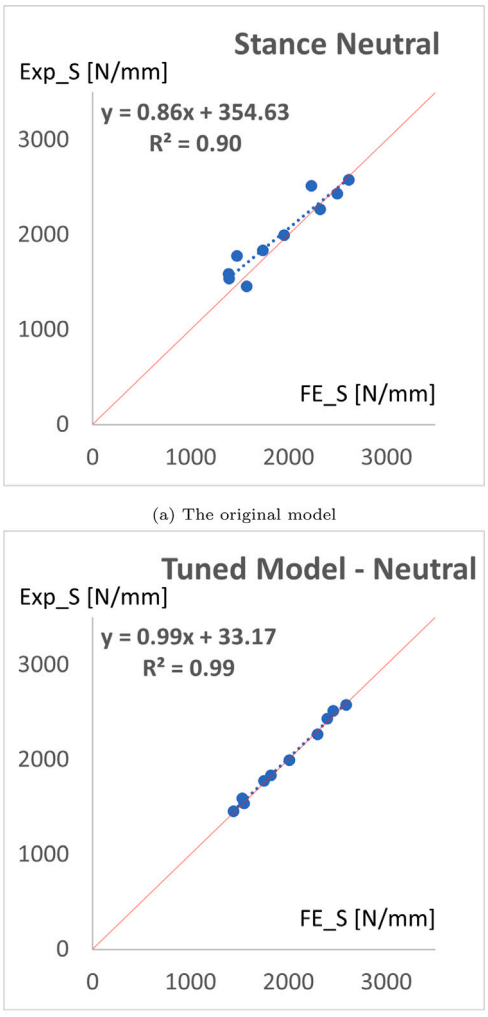


Fig. 9. Loading direction influences the sensitivity of the model to different parameter categories.



(a) The original model

(b) The tuned model

Fig. 10. Automatically picking the plausible option for the parameter which resulted in the best stiffness prediction for each sample generated a model with almost perfect predictive ability. The same results were observed in all load cases. Here, the neutral stance plot is shown as an example.

sections and focus on modeling only the bone portions, despite presence of embedding segments on the samples during experimental measurements, has shown to have significant effects on the model predictions as well (Kluess et al., 2019).

The second most influential category was loading, with reference load location and contact area having the largest effect on the models' predictive ability (Table 5). In our model, the size of the loading area was assigned based on preliminary simplified FE contact models using a two-material hemisphere representing the femoral head and embedding layer on a flat steel disk. Given the range of applied loads, head embedding diameter and the composite structure of the sphere in contact with the loading plate, a contact area diameter range between 3.2 and 3.6 mm was achieved. A 2 mm radius was assigned for all models considering the 1 mm voxel size. The high sensitivity of the models' predictive ability to loading area is in line with the previous studies on effect of boundary conditions on femoral strength predictions (Rossman et al., 2016). Yet plenty of ambiguity in loading area definition and reference node location assignment is present in different modeling examples (Hennicke et al., 2022; Altai et al., 2019; Haider et al., 2018).

According to our results, the predictive ability and sensitivity of the FE models were affected by the loading direction. Linear regression analysis and concordance correlation results were inferior for both

adducted and abducted load cases almost to the same extent but only slightly lower for the pooled data ($R^2 \simeq 0.71$ and 0.87, respectively). In the same fashion, the tilted load cases suffered from higher sensitivity in almost all parameter categories. Although a R^2 of 0.7 is still within the range of reported models in the literature (References mentioned above), the drop from the 0.9 value for neutral stance load case is of significance. One major contributing factor was the voxel-based surface of the model. As described, the tilted load cases were a product of rotating the contact area around the femoral head's embedding sphere. As a result, in the tilted load cases, contact circle is moved from the superior flat surface of the model to the jagged areas on the edges of the hexahedral elements. This was confirmed upon inspection of the models and the location of the contact area in all three load cases and in the deviated models for the loading angle parameter which is the driver of the increased sensitivity of the tilted models to the loading category rather than the embedding parameters. A more coherent predictive ability among all three load cases and less sensitivity to the loading angle uncertainty might be achieved in the smooth-mesh models.

Our results show that by systematically selecting the most accurate model for each sample and load case through a process of sample-based tuning, it is possible to achieve almost perfect model performance with an R^2 of 0.99 (Fig. 10). This highlights the importance of examining the model generation protocols and carefully considering the experimental parameters that have the greatest impact on performance of the model. In our study, these parameters were found to be primarily in the embedding and loading categories. However, it is important to note that the "tuning" process used in this study involved selecting the most accurate model from a set of models that were generated using plausible values for the experimental parameters. This demonstrates the importance of carefully reporting and considering these parameters in finite element studies to avoid unrealistic outcomes.

There are limitations in this work that are worth discussing: First, in this study, the baseline value and range of variation for each parameter were mainly according to our experience during the experimental phase. However, a different set of starting points or ranges could effectively alter the outcomes in terms of their relevant importance. For a fundamentally different experimental study, testing a more relevant set of parameters and variation characteristics is be recommended. Second, the presented results are based on the overall stiffness of the whole samples rather than the more frequently used metric of failure load. Using the same samples in our previous study (Amini et al., 2021), we were also able to show that the overall sample stiffness was highly correlated with the bone stiffness measured via marker tracking. Third, stiffness values are affected by the embedding compartments as well as the testing setup. Isolating the stiffness of the bone from that of the embedding and machine has been done in studies using optical measurement techniques such as landmark/marker tracking (Dall'Ara et al., 2013) or DIC (Katz and Yosibash, 2020). However, in the same studies, it has been shown that strains/stiffness measurements at the femoral neck region still suffers from much larger errors compared to the shaft. We were also able to decompose different terms of the measured overall stiffness in a sub group of our experiments and showed that the machine stiffness was on average 10 times that of the sample stiffness and that the inter-specimen bone stiffness variation was four times higher than those of the sample or machine stiffness (Amini et al., 2021). In terms of the relevance of this metric, using sample stiffness values (including the embedding sections) in experimental validation studies of QCT-FE holds an ongoing presence in the literature (Patton et al., 2019; Panyasantisuk et al., 2018; Miura et al., 2017; Rossman et al., 2016; Dall'Ara et al., 2013). Finally, the amount of variation per parameter is not normalized. The geometrical parameters could only take discrete values with 1 mm intervals (due to the voxel size), and to keep the number of model variations in a reasonable range, only the minimum and maximum of the defined plausible ranges were considered for modeling. The deviation range for each parameter was based on preliminary experiments and experience. As an example, the average adjustment in the 3D-registration of the μCT and CT-based volumes was in the range of ± 2 mm.

In conclusion, a voxel-based linear-elastic QCT-FE model of proximal femoral sample stiffness was developed and validated using 10 samples in three stance load cases. It was shown that embedding parameters affect the predictive ability of the neutral model more than the loading, material, or segmentation parameters. The predictive ability and sensitivity of the model was dependent on the load case. We also showed that using a process to select the best-fit plausible values for the parameters resulted in an almost perfect linear correlation between the measured and predicted values with predictive ability of 0.99. The results of this study could help improve the robustness of QCT-FE models of femoral stiffness.

5. Declarations

Ethics approval and consent to participate

The specimens originated from voluntary body donations for scientific and teaching purposes to the Center for Anatomy and Cell Biology of Medical University of Vienna with written informed consent of the donors, and according to protocol accepted by the ethics committee of Karl Landsteiner University of Health Sciences. All procedures were performed in accordance with relevant guidelines.

CRediT authorship contribution statement

Morteza Amini: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andreas Reisinger:** Writing – review & editing, Methodology, Conceptualization. **Alexander Synek:** Writing – review & editing, Methodology, Investigation. **Lena Hirtler:** Writing – review & editing, Resources. **Dieter Pahr:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We would like to acknowledge the contribution of all donors to the life sciences. The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

Funding

This study was co-funded by the Karl Landsteiner University of Health Sciences and Lower Austrian Research and Education Corporation (NFB, ID: SC16-009).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmbbm.2023.105664. As an alternative outcome variable to the reported "average of absolute percentage difference", the supplementary table is based on the "average of absolute difference" between the R^2 values for each sub category and category of the parameters.

References

Alomari, A.H., Wille, M.-L., Langton, C.M., 2018. Bone volume fraction and structural parameters for estimation of mechanical stiffness and failure load of human cancellous bone samples; in-vitro comparison of ultrasound transit time spectroscopy and X-ray μCT. Bone 107, 145–153.

Altai, Z., Qasim, M., Li, X., Viceconti, M., 2019. The effect of boundary and loading conditions on patient classification using finite element predicted risk of fracture. Clin. Biomech. 68, 137–143.

Amini, M., Reisinger, A., Hirtler, L., Pahr, D., 2021. Which experimental procedures influence the apparent proximal femoral stiffness? A parametric study. BMC Musculoskelet. Disord. 22 (1), 815. http://dx.doi.org/10.1186/s12891-021-04656-0.

Askarinejad, S., Johnson, J.E., Rahbar, N., Troy, K.L., 2019. Effects of loading rate on the of mechanical behavior of the femur in falling condition. J. Mech. Behav. Biomed. Mater. 96, 269–278.

Benca, E., Amini, M., Pahr, D.H., 2020. Effect of CT imaging on the accuracy of the finite element modelling in bone. Eur. Radiol. Exp. 4, 51.

Bessho, M., Ohnishi, I., Matsumoto, T., Ohashi, S., Matsuyama, J., Tobita, K., Kaneko, M., Nakamura, K., 2009. Prediction of proximal femur strength using a CT-based nonlinear finite element method: differences in predicted fracture load and site with changing load and boundary conditions. Bone 45 (2), 226–231.

Bouxsein, M.L., Delmas, P.D., 2008. Considerations for development of surrogate endpoints for antifracture efficacy of new treatments in osteoporosis: a perspective. J. Bone Miner. Res. 23 (8), 1155–1167.

Bouxsein, M., Zysset, P., Glüer, C., McClung, M., Biver, E., Pierroz, D., Ferrari, S.L., 2020. Perspectives on the non-invasive evaluation of femoral strength in the assessment of hip fracture risk. Osteoporos. Int. 31 (3), 393–408.

Brauer, C.A., Coca-Perraillon, M., Cutler, D.M., Rosen, A.B., 2009. Incidence and mortality of hip fractures in the United States. JAMA 302 (14), 1573–1579.

Chiang, C.Y., Zebaze, R., Wang, X.-F., Ghasem-Zadeh, A., Zajac, J.D., Seeman, E., 2018. Cortical matrix mineral density measured noninvasively in pre- and postmenopausal women and a woman with vitamin D–dependent rickets. J. Bone Miner. Res. 33 (7), 1312–1317. http://dx.doi.org/10.1002/jbmr.3415, arXiv:https://asbmr.onlinelibrary.wiley.com/doi/pdf/10.1002/jbmr.3415, URL https://asbmr.onlinelibrary.wiley.com/doi/abs/10.1002/jbmr.3415.

Chou, S.H., LeBoff, M.S., 2017. Vertebral imaging in the diagnosis of osteoporosis: a clinician's perspective. Curr. Osteoporos. Rep. 15 (6), 509–520. http://dx.doi.org/10.1007/s11914-017-0404-x.

Dall'Ara, E., Eastell, R., Viceconti, M., Pahr, D., Yang, L., 2016. Experimental validation of DXA-based finite element models for prediction of femoral strength. J. Mech. Behav. Biomed. Mater. 63, 17–25.

Dall'Ara, E., Luisier, B., Schmidt, R., Kainberger, F., Zysset, P., Pahr, D., 2013. A nonlinear QCT-based finite element model validation study for the human femur tested in two configurations in vitro. Bone 52 (1), 27–38.

Daszkiewicz, K., Maquer, G., Zysset, P.K., 2016. The effective elastic properties of human trabecular bone may be approximated using micro-finite element analyses of embedded volume elements. Biomech. Model. Mechanobiol. 16 (3), 731–742. http://dx.doi.org/10.1007/s10237-016-0849-3.

Dragomir-Daescu, D., Salas, C., Uthamaraj, S., Rossman, T., 2015. Quantitative computed tomography-based finite element analysis predictions of femoral strength and stiffness depend on computed tomography settings. J. Biomech. 48 (1), 153–161.

Enns-Bray, W., Ariza, O., Gilchrist, S., Widmer Soyka, R., Vogt, P., Palsson, H., Boyd, S., Guy, P., Cripton, P., Ferguson, S., Helgason, B., 2016. Morphology based anisotropic finite element models of the proximal femur validated with experimental data. Med. Eng. Phys. 38 (11), 1339–1347. http://dx.doi.org/10.1016/j.medengphy.2016.08.010, URL http://www.sciencedirect.com/science/article/pii/S1350453316301874.

Falcinelli, C., Whyne, C., 2020. Image-based finite-element modeling of the human femur. Comput. Methods Biomech. Biomed. Eng. 23 (14), 1138–1161.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al., 2012. 3D slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imaging 30 (9), 1323–1341.

Gross, T., Pahr, D.H., Zysset, P.K., 2012. Morphology–elasticity relationships using decreasing fabric information of human trabecular bone from three major anatomical locations. Biomech. Model. Mechanobiol. 12 (4), 793–800. http://dx.doi.org/10.1007/s10237-012-0443-2.

Haider, I.T., Goldak, J., Frei, H., 2018. Femoral fracture load and fracture pattern is accurately predicted using a gradient-enhanced quasi-brittle finite element model. Med. Eng. Phys. 55, 1–8.

Haque, S., Lau, A., Beattie, K., Adachi, J.D., 2018. Novel imaging modalities in osteoporosis diagnosis and risk stratification. Curr. Treat. Options Rheumatol. 4 (2), 133–141. http://dx.doi.org/10.1007/s40674-018-0099-x.

Helgason, B., Gilchrist, S., Ariza, O., Vogt, P., Enns-Bray, W., Widmer, R., Fitze, T., Pálsson, H., Pauchard, Y., Guy, P., et al., 2016. The influence of the modulus–density relationship and the material mapping method on the simulated mechanical response of the proximal femur in side-ways fall loading configuration. Med. Eng. Phys. 38 (7), 679–689.

Hennicke, N., Saemann, M., Kluess, D., Bader, R., Sander, M., 2022. Subject specific finite element modelling of periprosthetic femoral fractures in different load cases. J. Mech. Behav. Biomed. Mater. 126, 105059.

Katz, Y., Yosibash, Z., 2020. New insights on the proximal femur biomechanics using digital image correlation. J. Biomech. 109599.

Kluess, D., Soodmand, E., Lorenz, A., Pahr, D., Schwarze, M., Cichon, R., Varady, P.A., Herrmann, S., Buchmeier, B., Schröder, C., et al., 2019. A round-robin finite element analysis of human femur mechanics between seven participating laboratories with experimental validation. Comput. Methods Biomech. Biomed. Eng..

Lee, Y., Ogihara, N., Lee, T., 2019. Assessment of finite element models for prediction of osteoporotic fracture. J. Mech. Behav. Biomed. Mater. 97, 312–320.

Mishra, R.N., Singh, M.K., Kumar, V., 2022. Biomechanical analysis of human femur using finite element method: A review study. Mater. Today: Proc..

Miura, M., Nakamura, J., Matsuura, Y., Wako, Y., Suzuki, T., Hagiwara, S., Orita, S., Inage, K., Kawarai, Y., Sugano, M., et al., 2017. Prediction of fracture load and stiffness of the proximal femur by CT-based specimen specific finite element analysis: cadaveric validation study. BMC Musculoskelet. Disord. 18 (1), 1–8.

Pahr, D.H., Zysset, P.K., 2009. From high-resolution CT data to finite element models: development of an integrated modular framework. Comput. Methods Biomech. Biomed. Eng. 12 (1), 45–57. http://dx.doi.org/10.1080/10255840802144105, PMID: 18839383.

Panyasantisuk, J., Dall'Ara, E., Pretterklieber, M., Pahr, D., Zysset, P., 2018. Mapping anisotropy improves QCT-based finite element estimation of hip strength in pooled stance and side-fall load configurations. Med. Eng. Phys. 59, 36–42.

Patton, D.M., Bigelow, E.M.R., Schlecht, S.H., Kohn, D.H., Bredbenner, T.L., Jepsen, K.J., 2019. The relationship between whole bone stiffness and strength is age and sex dependent. J. Biomech. 83, 125–133.

Rajapakse, C.S., Farid, A.R., Kargilis, D.C., Jones, B.C., Lee, J.S., Johncola, A.J., Batzdorf, A.S., Shetye, S.S., Hast, M.W., Chang, G., 2020. MRI-based assessment of proximal femur strength compared to mechanical testing. Bone 115227.

Rossman, T., Kushvaha, V., Dragomir-Daescu, D., 2016. QCT/FEA predictions of femoral stiffness are strongly affected by boundary condition modeling. Comput. Methods Biomech. Biomed. Eng. 19 (2), 208–216.

Sas, A., Ohs, N., Tanck, E., van Lenthe, G.H., 2020. Nonlinear voxel-based finite element model for strength assessment of healthy and metastatic proximal femurs. Bone Rep. 12, 100263.

Sas, A., Sermon, A., van Lenthe, G.H., 2022. Experimental validation of a voxel-based finite element model simulating femoroplasty of lytic lesions in the proximal femur. Sci. Rep. 12 (1), 1–10.

Synek, A., Chevalier, Y., Baumbach, S., Pahr, D., 2015. The influence of bone density and anisotropy in finite element models of distal radius fracture osteosynthesis: Evaluations and comparison to experiments. J. Biomech. 48 (15), 4116–4123. http://dx.doi.org/10.1016/j.jbiomech.2015.10.012.

Viceconti, M., Qasim, M., Bhattacharya, P., Li, X., 2018. Are CT-based finite element model predictions of femoral bone strengthening clinically useful? Curr. Osteoporos. Rep. 16 (3), 216–223.

Wakao, N., Harada, A., Matsui, Y., Takemura, M., Shimokata, H., Mizuno, M., Ito, M., Matsuyama, Y., Ishiguro, N., 2009. The effect of impact direction on the fracture load of osteoporotic proximal femurs. Med. Eng. Phys. 31 (9), 1134–1139.

Wille, H., Ruess, M., Rank, E., Yosibash, Z., 2016. Uncertainty quantification for personalized analyses of human proximal femurs. J. Biomech. 49 (4), 520–527.

Zhang, L., Wang, L., Fu, R., Wang, J., Yang, D., Liu, Y., Zhang, W., Liang, W., Yang, R., Yang, H., et al., 2021. In vivo assessment of age-and loading configuration-related changes in multiscale mechanical behavior of the human proximal femur using MRI-based finite element analysis. J. Magn. Reson. Imaging 53 (3), 905–912.