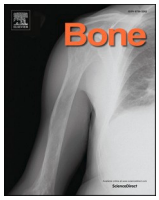
Bone 206 (2026) 117830

Contents lists available at ScienceDirect

Bone

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



Full Length Article

DXA-derived three-dimensional finite element models of the femur: validation against CT-based models

Alessandra Aldieri [a], Pinaki Bhattacharya [b], Margaret Paggiosi [c], Richard Eastell [c], Lorenzo Grassi [d,*]

[a] Polito[BIO]Med Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Torino, Italy
[b] Insigneo Institute and School of Mechanical, Aerospace and Civil Engineering, University of Sheffield, Sheffield, UK
[c] Division of Clinical Medicine, University of Sheffield, Sheffield, UK
[d] Department of Biomedical Engineering, Lund University, Lund, Sweden

ARTICLE INFO

Keywords:
DXA
FE models
Three-dimensional models
Fracture risk prediction

ABSTRACT

Three-dimensional finite element (FE) models derived from Computed Tomography (CT) images predict hip fractures better than areal bone mineral density measurements from Dual-energy X-ray Absorptiometry (DXA). Yet, these results have not justified the adoption of CT in clinical practice, and only 2D DXA images are clinically available. Statistical shape and appearance models can be used to reconstruct three-dimensional FE models from 2D DXA images. While *ex vivo* validations have been performed on 3D reconstructed DXA-based FE models, it is not clear how well 3D reconstructed DXA-based FE models can predict fractures compared to CT-based models. The aim of this study was thus to evaluate the ability of one such methodology, namely DXA2FEM, to predict fractures in a clinical cohort of pair-matched fractured and control subjects, for whom both DXA and CT images were available. 3D FE models of the femur were built from both DXA and CT, and FE simulations were run reproducing a sideways fall in 28 different femoral configurations. An absolute risk of fracture (ARF0) was then computed based on the FE-predicted femoral strength values. DXA- and CT-derived models were compared with respect to geometry, density distribution, and FE-predicted proximal femoral strength. DXA-derived 3D FE models had an average point-to-surface distance of −2 mm from CT-based models, whereas the Young's moduli were 29% higher. ARF0 by CT reported statistically significantly better diagnostic accuracy (0.83, 95% CI 0.75 to 0.91) than standard hip DXA (0.69, 95% CI 0.6 to 0.8) or FRAX (0.69, 95% CI 0.57 to 0.81). The diagnostic accuracy of ARF0 by DXA was between ARF0 by CT and standard hip DXA/FRAX (0.74, 95% CI 0.62 to 0.86), albeit neither difference was statistically significant in the analysed cohort.

1. Introduction

Osteoporosis is responsible for more than 37 million fragility fractures worldwide each year [1]. Not only do such fractures represent a considerable socio-economic burden, but the reported numbers are expected to increase in the coming years due to the progressive ageing of the population [2]. The total direct cost of osteoporotic fractures in 2019 amounted to 56.9 billion euros [3]. Hip fractures, in particular, represent the most serious type of fragility fractures, as they are associated with an increased mortality in the first year after fracture occurrence, and in general with loss of mobility and thus a decreased quality of life [3,4].

An accurate estimate of hip fracture risk would therefore be pivotal, enabling preventive actions to be implemented to prevent fractures. In current clinical practice, bone fragility is estimated primarily using areal bone mineral density (aBMD) obtained from dual-energy X-ray absorptiometry (DXA) two-dimensional images [5]. The risk of a fragile hip fracture can also be assessed using FRAX [6], which combines aBMD with other aBMD-independent clinical risk factors. Despite being the standard surrogate for hip fracture risk, aBMD is only a moderate predictor of proximal femur strength and of incident hip fractures [8,9], with only half of subjects who experience a fracture being diagnosed as at high risk [7].

Aiming to improve the limitations of the clinically assessed standard, promising results have been obtained by quantitative computed tomography (CT)-based subject-specific finite element (FE) models of the

* Corresponding author at: Department of Biomedical Engineering, Lund University, Box 118, 221 00, Lund, Sweden.
E-mail address: lorenzo.grassi@bme.lth.se (L. Grassi).

https://doi.org/10.1016/j.bone.2026.117830
Received 1 December 2025; Received in revised form 4 February 2026; Accepted 8 February 2026
Available online 9 February 2026

femur, which were shown to overcome aBMD by a significant amount in predicting femoral strength and hip fracture risk [8–11]. Among these, the Bologna Biomechanical Computed Tomography (BBCT-hip) methodology has shown promising results in both *in vitro* and *in vivo* validation studies [8,9,12], improving aBMD stratification accuracy in a retrospective clinical cohort. Yet, the promising performance of such methodologies is hindered by the fact that CT images are not routinely acquired for osteoporosis diagnosis, and the evidence gathered so far on the improved prediction accuracy of CT-based computational methodologies has not justified its adoption yet [7]. Recently, a pipeline (hereby referred to as DXA2FEM) has been developed to reconstruct three-dimensional FE models from DXA images using a statistical shape and appearance model trained on a population of Finnish subjects [13]. This pipeline was able to stratify fractured and control cases better than aBMD [14,15] in 2 prospective clinical cohorts for which DXA images and history of fractures were available.

Despite their promising performance, both methodologies still require additional clinical validation across different ethnicities and scanners before they can be progressively translated into clinical practice. For BBCT-hip like technologies, any improvement in hip fracture prediction should be significant enough to justify the increase in radiation dose and operational cost that come with CT imaging. For DXA2FEM, it would be crucial to understand how well it could predict hip fracture risk in a clinical cohort compared not only to the clinical standard aBMD, but also to the engineering-standard CT-based FE models. Additionally, an *in vivo* validation on cohorts from different countries or with different ethnicities to the one used for its development would assess its generalisability.

Therefore, this study aimed to validate the DXA2FEM pipeline with respect to its stratification accuracy in separating fracture from non-fracture cases compared to BBCT-hip. We used a retrospective cohort of Caucasian women [16] which includes DXA and CT images acquired on the same day. The same cohort had previously been employed to validate BBCT-hip [8,9]. We hypothesize that DXA2FEM-based FE models will provide a superior stratification accuracy compared to aBMD, albeit inferior to that of BBCT-hip models due to the 2D-to-3D reconstruction errors.

2. Materials and methods

2.1. Subjects and available data

The study was based on a cohort of post-menopausal (>5 years) Caucasian women originally comprising 50 subjects (55–89 years old) who had sustained a proximal femur fracture and 50 control subjects who were pair-matched in terms of age, height, and weight to the fractured subjects recruited at a single site (Sheffield, UK). CT (LightSpeed 64 VCT, GE Medical Systems, Milwaukee, WI, USA, 120 kVp, 170–200 mA, 0.625 mm slice thickness) and DXA (Discovery scanner, Hologic, Inc. Bedford, MA, USA) images had been acquired for each subject on the same day. For the fractured subjects, the images were recorded at a median of 67 days after the hip fracture event [16]. The left femur was analysed for all the healthy control subjects, whereas the contralateral intact femur was analysed for fractured subjects. Further details about the cohort and clinical images acquisition are provided in [16]. Subjects with missing data for either CT or DXA were excluded, leading to a total of 88 subjects analysed in this study (43 fracture cases and 45 control cases). A two-sample *t*-test confirmed that no statistical difference ($p < 0.05$) in age, height, weight was present between the two groups in the analysed cohort. Table 1 summarises the main clinical information about the included subjects.

Table 1
Clinical information about the fractured and control groups considered in this study.

	Fractured subjects (n = 43)		Control subjects (n = 45)	
	Mean	SD	Mean	SD
Age	75.8	8.98	78.8	9.12
Mass (kg)	63.1	14.70	64.0	12.21
Height (m)	1.59	0.07	1.58	0.05
aBMD (g/cm^2)	0.59	0.12	0.63	0.10

2.2. DXA2FEM pipeline: from DXA to a subject-specific 3D FE model

Subject-specific 3D reconstructions of the proximal femur were obtained from a single anteroposterior DXA scan of the hip by means of a 2D-to-3D reconstruction. The method has been previously presented [13] and validated against both mechanical tests [17,18] and clinical cohorts [14,15] data, and is only briefly reported here. A mesh-based statistical shape and appearance model (SSAM) was trained for the proximal femur (40 men, 19 women, median age 58 years, range 18–88 years) and the hemipelvis (14 women, median age 74 years, age range 69–78 years). The SSAM, incorporating three-dimensional geometric and densitometric information, was then used to reconstruct 3D anatomies from 2D DXA images. First, eight anatomical landmarks were automatically identified on each subject's DXA image to register the previously built SSAM onto the input DXA image. A digitally reconstructed radiograph (DRR) was then obtained by projecting the femur and hemipelvis SSAM three-dimensional instances on the frontal plane. The final reconstructed 3D proximal femur was obtained by running an optimization process that identifies the SSAM instances that minimises the sum of the absolute differences between the pixel-wise aBMD of its DRR and that of the target DXA image. The so obtained femur reconstruction was in the format of a 3D FE model (19532 nodes, 103761 tetrahedral elements, average element size 1.8 mm) with element-specific volumetric BMD information. Element-specific linear elastic material properties were thus assigned based on the volumetric BMD information using the density–Young's modulus relationship proposed by [19] and tetrahedral elements were converted to 2nd order elements. An anatomical reference system was calculated [20] and boundary conditions replicating a fall to the side were applied. The force was distributed over the 10 most medial points on the femoral head surface, whilst a no-friction slider was applied to the most inferior point on the greater trochanter. A hinge was placed 80 mm distal from the minor trochanter and only allowed adduction/abduction [21]. In the following, the models obtained according to this approach will be referred to as *DXA-based Models*.

2.3. BBCT-hip pipeline: from CT to a subject-specific FE model

For the construction of the CT-based FE model, the subject-specific 3D proximal femur geometries were extracted from the CT images *via* manual segmentation (ITK-Snap 2.0.0, University of Pennsylvania) [22]. The segmented femur models were then meshed with ten-node tetrahedral elements with an element size of 2 mm following a convergence analysis [23]. Linear elastic heterogeneous material properties were mapped [24] from the CT to the FE models using the same apparent density–Young's modulus relationship that was adopted for the DXA2-FEM pipeline [19]. Anatomical landmarks (centre of femur head, distal femur epicondyles most medial and lateral points, distal femur medial and lateral epicondyles most posterior points) were identified for the definition of an anatomical reference system used in the application of the boundary conditions replicating a fall. The impact load was applied at the femoral head centre, and a rigid frictionless contact plane perpendicular to the load direction was created at the greater trochanter to mimic the ground. A hinge was placed at the knee centre, leaving the femur free to rotate in adduction/abduction (Fig. 1). These FE models built from CT images will be referred to as *CT-based Models*.

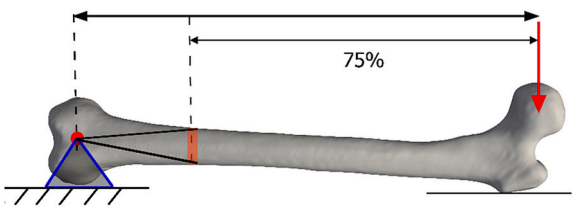


Fig. 1. Schematic of the applied boundary conditions and loading for a representative femur orientation. The red arrow represents the load, applied at femur head centre, the black one the biomechanical length. In orange, the 4 mm thick bands of nodes located at 75% of the biomechanical length from the femur head is highlighted, which was constrained to only rotate around knee centre. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. FE simulations

For each DXA-based and CT-based model, a total of 28 different FE simulations were solved by varying the impact load direction from 0° to 30° in the frontal plane (adduction) and −30° to +30° in transverse plane (internal–external rotation) in steps of 10° in the anatomical reference system and according to the specific boundary conditions adopted by the two pipelines as explained in the previous sections. Static linear elastic simulations were run in Abaqus/Standard (v2023, Dassault Systèmes). The failure load (hereby referred to as proximal femoral strength), *i.e.*, the load which would cause the femoral fracture, was identified for each simulated configuration using a previously validated principal strain-based failure criterion [25]. The criterion consists of an asymmetric principal strain limit (0.73% tensile, 1.04% compressive limit), where FE-predicted failure load was determined as the load needed to cause one node on the surface to reach either principal strain limit value.

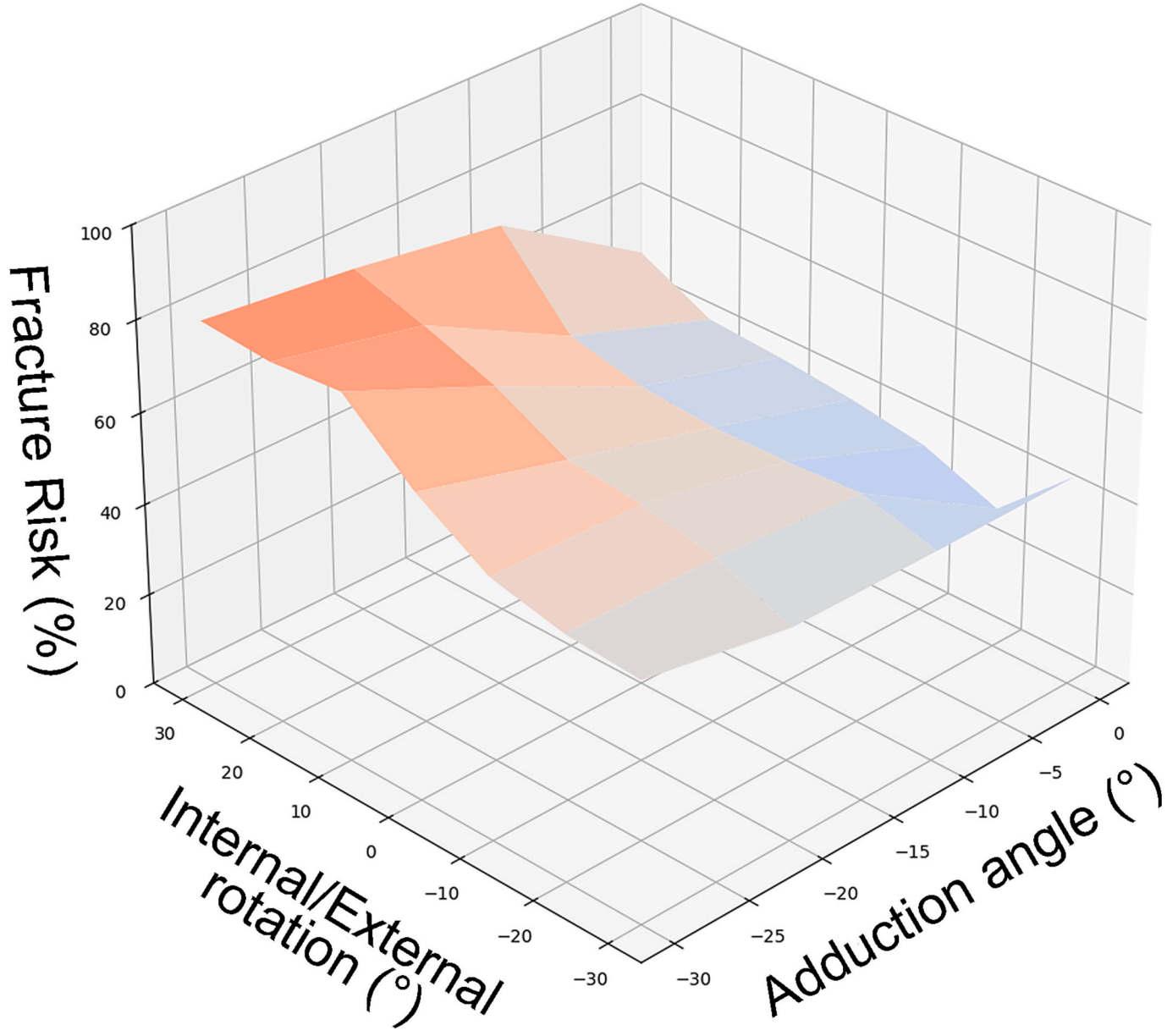


Fig. 2. An example of response curve obtained by interpolating the risk of fracture at each simulated configuration. ARF0 value was obtained through integration of the response surface.

2.5. ARF0 computation

ARF0, the absolute risk of fracture at the hip at time 0 upon falling, could be derived by combining the FE models with a physics-based stochastic mathematical model. This latter computes possible individual impact forces due to a fall, whilst the first calculates proximal femoral strength [9]. The physics-based stochastic model idealises the fall event as an inverted pendulum: the rotation occurs around a spherical joint (hinge) fixed to the floor and located near the foot on the side of impact. The body mass is considered to be concentrated at the moving end of the inverted pendulum, the static end of which is located at the hinge. In such a way, 1,000,000 falls of a body of the height and weight equal to that of the subject can be estimated by sampling through inverse Latin hypercube over 6 stochastic input variables contained in the model: patient initial and final position, initial speed and acceleration, damping coefficients due to active and passive soft tissues.

Each of the so estimated one million impact forces computed were then compared to the 28 femoral strength values previously obtained, so that 28 values for the risk of fracture were computed as the ratio between NF, the number of forces exceeding the femoral strength, and the total number of forces (*i.e.*, 1,000,000). A response surface (Fig. 2) could then be built and integrated to obtain ARF0. This approach has been previously presented in [8,9,12] where its accuracy in correctly identifying subjects at risk of a hip fracture was also compared to the clinical standard T-score on a clinical cohort of Caucasian post-menopausal women [16].

2.6. Comparison metrics

2.6.1. Geometry

After registering the DXA-based models to the CT-based models *via* an iterative-closest-point algorithm, the node-to-surface distance between the outer shape of the two models was retrieved. The Hausdorff (Hd) and Chamfer (Cd) distances were also computed according to the following equations:

$$Hd(A,B) = max(Ed(A,B)\,, Ed(B,A)\,), \tag{1}$$

$$Cd(A,B) = mean(Ed(A,B)\,) + mean(Ed(B,A)\,), \tag{2}$$

where A and B refer to the compared shapes, and Ed to the Euclidean distances computed between each point in A and its closest point in B ($Ed(A,B)$) or *vice versa*.

Only the nodes of the proximal part of the femur geometry (defined as the nodes included in a 60 mm radius sphere centred at the femoral head centre) were considered for this analysis.

2.6.2. Local material properties distribution

As the three-dimensional meshes of the CT-based and DXA-based FE models were not isotopological, a mapping algorithm was adopted to identify, for each element of the CT-based mesh, the elements of the DXA-based mesh intersecting it and the percentage of the total volume occupied by each intersecting element. Hence, an element-by-element material properties comparison could be carried out comparing the element-wise Young's modulus values mapped on the CT-based models with the corresponding Young's modulus values E_R coming from the DXA-based models according to the following mixture equation:

$$E_R^i = \sum_{j=1}^{k} E_R^j \frac{V^j}{V^i} \tag{3}$$

where, for each i^{th} element of the original model, the reconstructed Young's modulus E_R^i is computed as the sum of the Young's modulus values of the j^{th} element of the DXA-based mesh weighted by the relative volume occupied by the j^{th} element inside the i^{th} element, for all k elements of the DXA-based model intersecting i^{th} element of the CT-based model.

2.6.3. Loads to failure and stratification accuracy

Linear regressions were computed for all the failure load and ARF0 values to compare the predictions of DXA-based and CT-based models. The coefficient of determination was calculated, together with slope and intercept of the linear regression, and the root mean square error (RMSE). The statistical significance of the correlations was assessed with an F-test on the regression model.

The stratification accuracy of DXA-based models and CT-based models was assessed by logistic regression using the ARF0 values as predictor variable. A 10-fold cross validation strategy was adopted to avoid overfitting. An identical analysis was performed with aBMD and FRAX as predictor variables. Receiver Operating Characteristics (ROC) curves were generated and the area under the curve (AUC) calculated to compare the stratification accuracy of DXA-based models, CT-based models, aBMD and FRAX. Statistical significance in the difference between AUCs was assessed using DeLong method (statistical significance $p < 0.05$) [26].

3. Results

3.1. Geometry

Considering the surfaces comparison between the reconstructed DXA-based shapes and those segmented from the CT scans for all the subjects considered in this study, the Chamfer distance was found to be below 10 mm for all the subjects, while the highest Hausdorff distance turned out to be 16.6 mm (Fig. 3). The mean point-to-surface distance computed between the CT- and the DXA-based shapes was -2.0 ± 2.2 mm, showing a slight underestimation of the proximal femur volume by the DXA-based reconstruction algorithm (Fig. 3, right panel).

Fig. 4 displays the local distributions of the point to surface distances for the two subjects presenting the highest (left panel) and the lowest (right panel) average point-to-surface distance values, equal to 4.6 and 0.9 mm respectively. As for the former (left), the highest distances can be identified at the head and at the region between the greater and the lesser trochanter. For the latter subject (right), a slight overestimation of the shape of the outer surface can be identified in the proximal part of the femur, specifically by the greater trochanter, while an underestimation of the outer surface by the DXA-based reconstruction is visible in the most distal part shown.

3.2. Local material properties distribution

The Young's modulus values distributions extracted from the CT-based and DXA-based FE models, displayed in Fig. 5 (left panel), do not highlight considerable discrepancies between the CT-based and DXA-based Young's moduli. By analysing the element-wise relative differences between the CT- and DXA-based Young's moduli (Fig. 5, right panel), reported in absolute value, a median value of 29% was obtained, with the 25th and 75th percentiles settling at 9% and 84%, respectively. The signed element-wise relative differences instead yielded a median equal to -6%, the 25th percentile equal to -60.2% and the 75th percentile equal to 16.1% (Fig. S1, Electronic Supplementary Material). Absolute errors in the Young's modulus between the CT-based and DXA-based models yielded a median value of 1.12 GPa and a 75th percentile value equal to 3 GPa.

In Fig. 6 the Young's modulus local distributions for the DXA-based and CT-based FE models are compared on the two subjects characterised by the lowest (left panel) and highest (right panel) average relative differences, equal to 46% and 99% respectively. A general overestimation of the Young's moduli in the cortical bone by the DXA-based models can be observed in both models. Such Young's moduli distributions are depicted in both cases on the CT-based mesh.

Eventually, if the DXA- and CT-based average Young's modulus values computed on each subject are compared (Fig. 7), a good correlation ($R^2 > 0.5$) can be observed, although the general overestimation

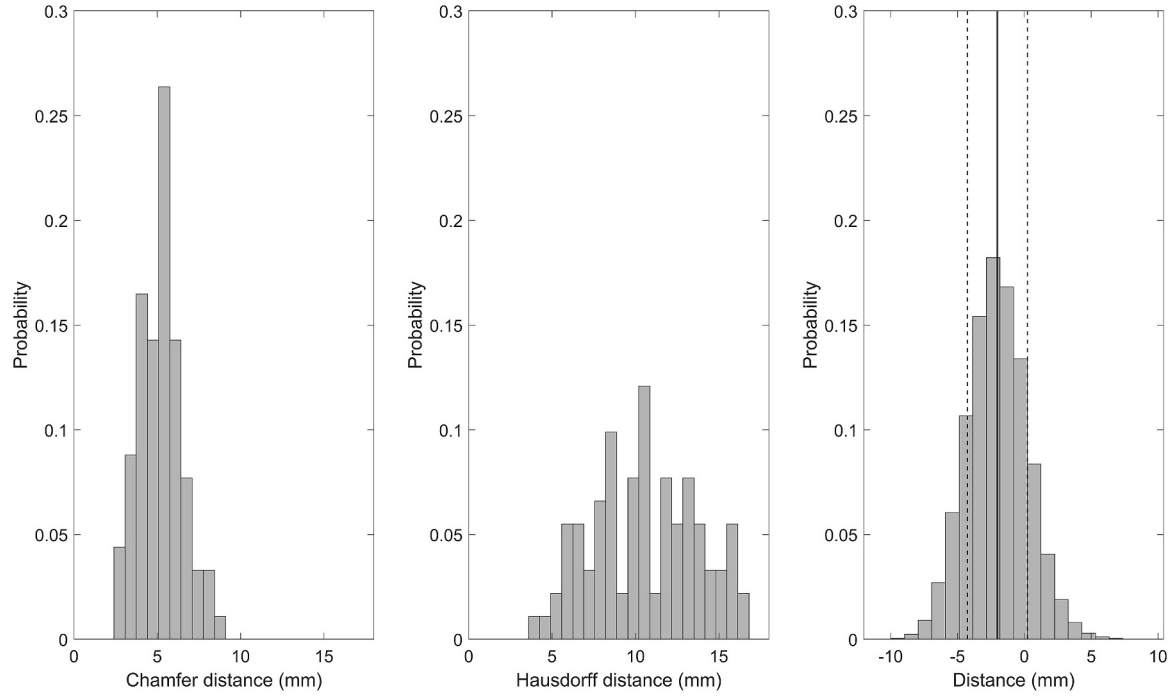


Fig. 3. Histograms of the distance metrices computed between the CT-based and the reconstructed DXA-based shapes. The left panel shows the Chamfer distance, the central panel the Hausdorff distance and the right panel the point to surface distance between the CT- and DXA-based shapes outer surface. The average point to surface distance is displayed as a black straight line, with the $\pm\sigma$ range highlighted through dotted lines.

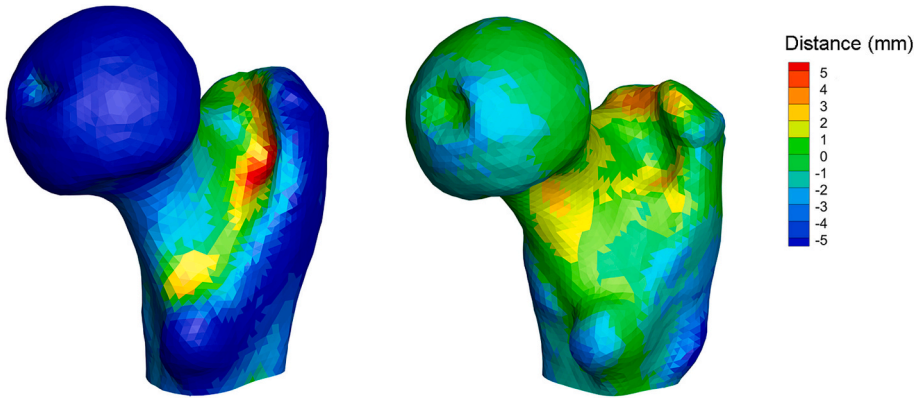


Fig. 4. Local distributions of the point to surface distances for the subjects showing the highest (left panel) and the lowest (right panel) average distance values.

of the Young's modulus values by the DXA-based models emerges again.

3.3. FE simulations outcomes and stratification accuracy

The agreement between the femoral strength values predicted by the DXA-based and CT-based FE simulations turned out to be modest but always statistically significant ($p < 0.001$): if all the 28 failure loads computed for all the subjects are analysed together (Fig. 8), a coefficient of determination R^2 equal to 0.41 was obtained, while looking at each simulation configuration separately (Electronic Supplementary Material, Fig. S2), R^2 values ranged from 0.18 to 0.58. The comparison between the average and the lowest femoral strength values across the 28 configurations for all the subjects instead is displayed in Fig. 9. These values are also compared to the subject-specific aBMD and FRAX values in Figs. S3 and S4 in the Electronic Supplementary Material.

The ARF0 values derived from the DXA-based simulations could explain 40% of the variability observed in the values derived from the CT-based FE models (Fig. 10, $R^2 = 0.4, p < 0.001$).

By comparing the distributions of predicted ARF0 values for fracture and control cases, significant differences ($p < 0.05$) were identified between the two groups for both DXA- and CT-derived quantities (Fig. 11, upper panel). Significant differences between the fracture and control groups also emerged considering the both DXA- and CT-based average loads to failure distributions (Fig. 11, upper panel) but not looking at the minimum strength values as obtained from the DXA-based simulations (Fig. 11, upper panel).

Furthermore, comparison of DXA- and CT-derived ARF0 values with the clinical standards aBMD and FRAX indicates that neither provides a complete discrimination between fracture and control case (Fig. S5, Electronic Supplementary Material). Yet, considering the corresponding distributions (Fig. 11, lower panel), in all cases significant differences ($p < 0.05$) between the fracture and control subjects medians were identified in all cases. As highlighted in Fig. 12, ARF0 values from the CT-based models achieved the best stratification performances (AUC = 0.83, 95% CI 0.75–0.91), followed by ARF0 values from DXA-based models (AUC = 0.74, 95% CI 0.62–0.86). aBMD and FRAX yielded

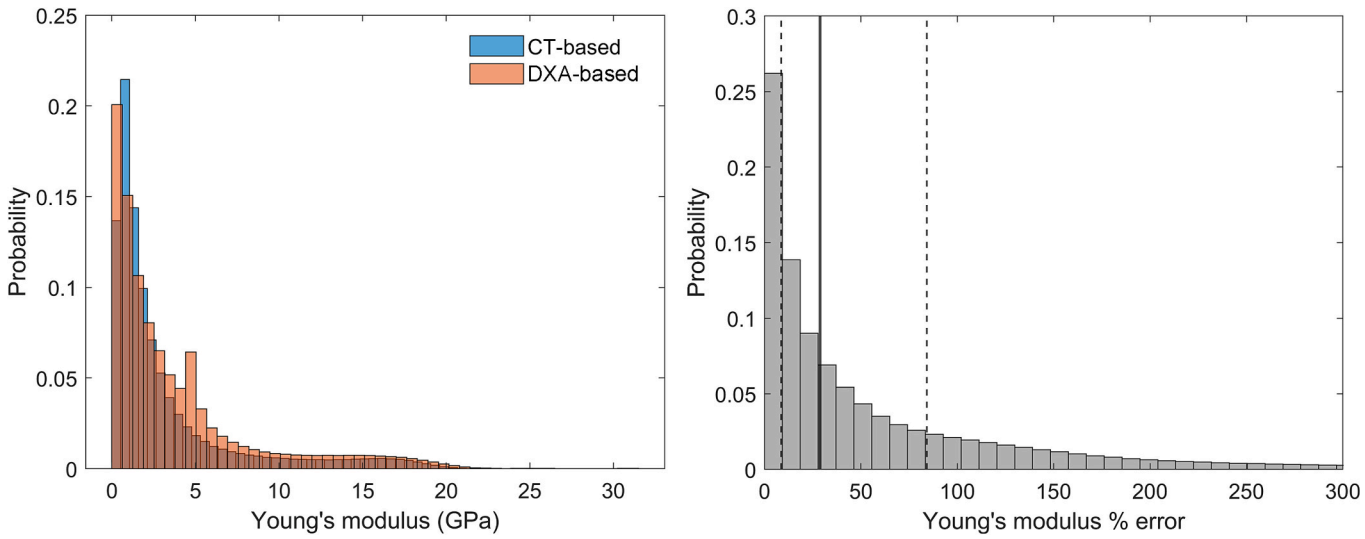


Fig. 5. Left panel: comparison between the DXA- and CT-based FE models Young's modulus values distributions. Right panel: distribution of the element-wise Young's modulus relative difference in absolute value.

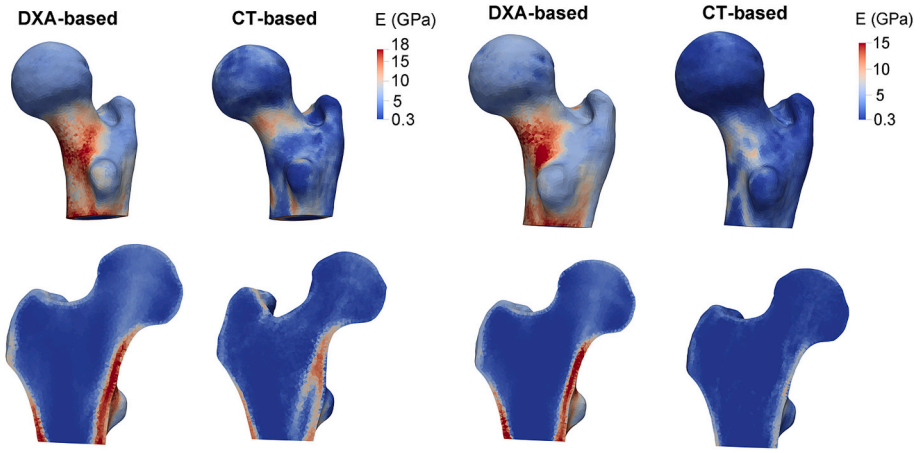


Fig. 6. DXA-based and CT-based element-wise Young's modulus values distributions depicted for the subject showing the lowest average relative error (left panel) and the subject showing the highest average relative error (right panel) computed on the Young's modulus values. In both cased the local distributions are depicted on the CT-based FE meshes.

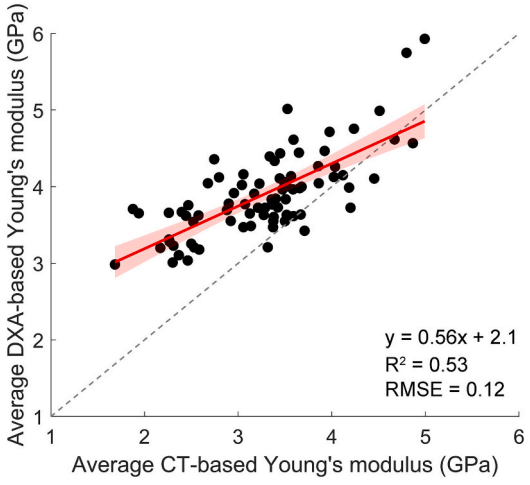


Fig. 7. Comparison between the DXA- and CT-based subject-specific average Young's modulus values. The regression line is superimposed onto the data. The regression line equation, R^2 value and normalised RMSE are also reported.

analogous stratification accuracies (AUC = 0.69, 95% CI 0.6–0.8 and AUC = 0.69, 95% CI 0.57–0.81, respectively). No statistically significant differences emerged between the DXA- and CT-derived ARF0 ROC curves ($p < 0.05$). Statistically significant differences were found between the AUC values yielded by aBMD and FRAX compared to CT-derived models ($p < 0.05$) but not compared to DXA-derived models ($p < 0.05$).

4. Discussion

To improve the prediction of femoral osteoporotic fracture risk in a clinically attainable way, exploiting the information contained in 2D DXA is paramount, as DXA is the current standard imaging technique for osteoporosis diagnosis and fracture risk assessment. CT-based subject-specific FE models can be seen as the research gold standard by the biomechanics community, showing superior accuracy in predicting femoral strength [8,11,27]. However, their adoption in clinical practice has been hindered by the additional costs associated with CT imaging and the construction of FE models.

In this study, we compared the stratification ability of 3D FE models reconstructed from 2D DXA images (DXA-based models) and CT-based

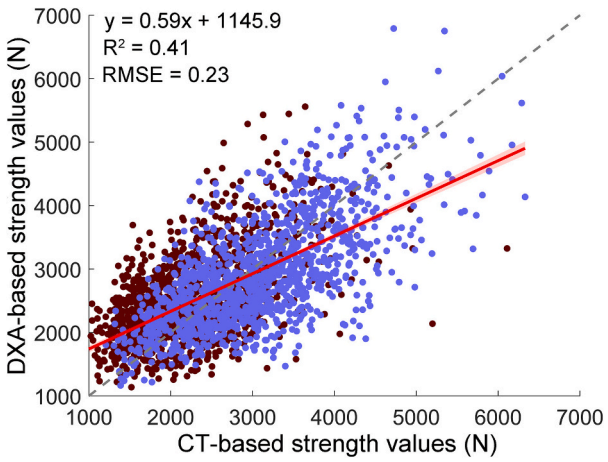


Fig. 8. Comparison between the subject-specific femoral strength values obtained from the DXA-based and CT-based FE simulations for all the 28 different femur configurations simulated. Fracture cases (F) are depicted in red, while control cases (C) are depicted in lilac. The regression line is superimposed onto the data. The regression line equation, R^2 value and normalised RMSE are also reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

FE models to that of the current clinical standards, aBMD from DXA and the FRAX scores. The research hypothesis was that both CT-based models and DXA-based models should provide a superior stratification ability compared to aBMD and FRAX, and that CT-based models should be superior to DXA-based models: while CT can ensure greater fidelity in reproducing the actual subject-specific anatomy and density distribution, a DXA-derived model is expected to carry a higher degree of uncertainty. We performed our analyses on a unique clinical cohort that included post-menopausal women equally divided in fractured and control subjects. DXA and CT images acquired during the same day were available. This allowed to compare CT-based and DXA-based three-dimensional FE models in terms of 1) geometry and material properties as well as 2) stratification accuracy.

In terms of geometry, the DXA-based reconstruction did not differ substantially from the CT-derived femoral shapes: the median Hausdorff distance was 1 cm, and the highest point-to-surface distance was <1 cm. Such errors can be considered acceptable given the 2D nature of the DXA images and the reported Hausdorff distances of 8 mm for intra- and inter-operator variability in segmenting CT images [28]. Besides, the average node-to-surface distance and the more comprehensive Chamfer distance metrics, which explicitly consider bidirectional distances between the two surfaces, confirmed results found in the literature [29,30], settling below 2 mm when comparing CT-based 3D shapes with the corresponding 3D reconstructions from 2D imaging techniques. By looking at the element-wise Young's modulus values differences, although the median value for the relative error was 29%, the highest relative error value exceeded 200%, which, however, might have been related to low-density regions inside the femur. In absolute terms, the median error computed on the Young's modulus turned out to be 1.12 GPa, with the 90th percentile settling at 7.8 GPa. In this respect, since the femoral strength calculation from the FE simulations considered only strain values on the surface of the femur model, high internal errors on Young's moduli might not have impacted the subsequent outcomes. Accordingly, significant linear relations could be established between

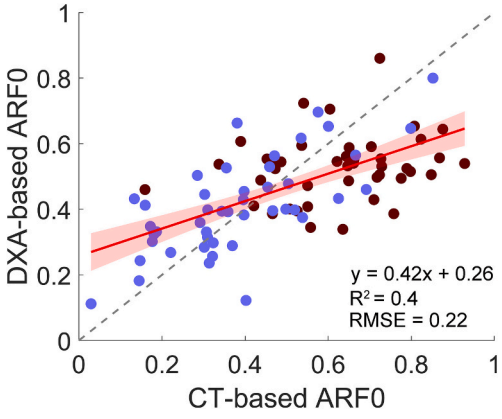


Fig. 10. Comparison of the subject-specific ARF0 values coming from the DXA-based and the CT-based FE simulations. Fracture cases (F) are depicted in red, while control cases (C) are depicted in lilac. The regression line is superimposed onto the data. The regression line equation, R^2 value and normalised RMSE are also reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

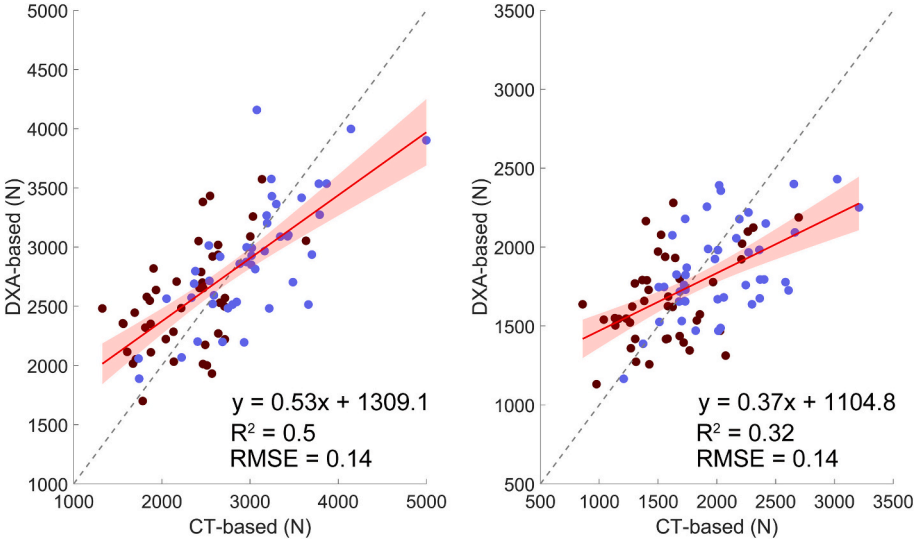


Fig. 9. Comparison between the DXA-based and CT-based average (left panel, $R^2 = 0.5.p < 0.001$) and minimum (right panel, $R^2 = 0.32.p < 0.001$) subject-specific femoral strength values across the 28 simulated femur configurations simulated. Fracture cases (F) are depicted in red, while control cases (C) are depicted in lilac. The regression line is superimposed onto the data. The regression line equation, R^2 value and normalised RMSE are also reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

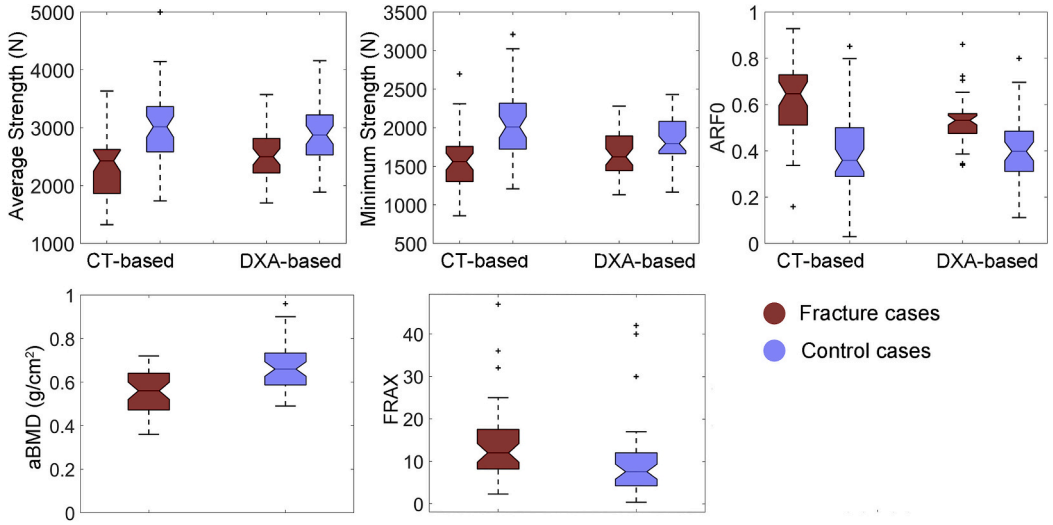


Fig. 11. Boxplots depicting the distributions of average femoral strength (upper panel, left), minimum femoral strength (upper panel, centre) and ARF0 (upper panel, right) as computed from the DXA-based and CT-based FE models compared to aBMD (lower panel, left) and FRAX (lower panel, right) for fracture and control subjects. Fracture cases are depicted in red, while control cases are depicted in lilac. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

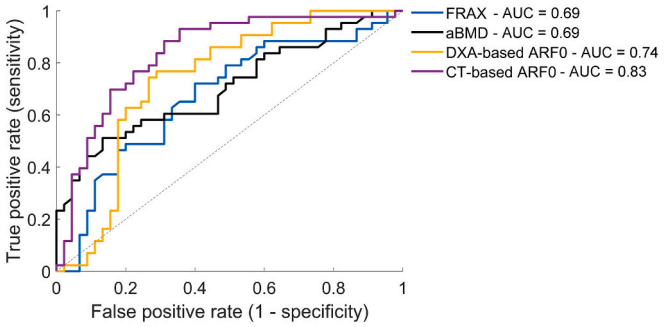


Fig. 12. ROC curves comparing the stratification accuracy of aBMD and FRAX with the stratification accuracy achieved by the DXA- and CT-derived ARF0 values.

the DXA- and CT-derived loads to failure in all 28 configurations and ARF0 values. However, the DXA-derived outcomes could only partially explain the variability of the CT-derived results in all cases, as highlighted by the moderate coefficients of determination. Considering the outcomes obtained for the fracture and control groups separately, statistically significant differences in the median values of the two emerged for the clinical standards aBMD and FRAX, as well as for ARF0 values derived from both the CT- and DXA-based FE models. By comparing the 10-fold cross-validation stratification accuracy (AUC) yielded by the DXA- and CT-derived ARF0 values in comparison to the clinical standard aBMD and FRAX however, ARF0 was in any case able to outperform the latter, despite DXA-derived ARF0 values yielded an AUC value lower than the CT-derived ones. In addition, the AUC values here observed for the DXA-derived ARF0 turned out to be comparable with the 0.74–0.79 AUC values range presented in recent studies employing as classifier the femur fall strength and using the same DXA-based reconstruction algorithm adopted herein [14,15].

To the authors' best knowledge, this is the first time that CT-based models and 3D DXA-based FE models reconstructed from DXA have been compared in terms of fracture risk prediction accuracy in a clinical cohort. This is also one of the most extensive datasets onto which the reconstruction ability of 2D-to-3D reconstruction techniques has been validated using *in vivo* images. Recently, the stratification accuracy of 3D FE models developed starting from DXA images in separating fractured subjects from control subjects was demonstrated to outperform aBMD [14,15,31], highlighting the great potential of such an approach which is based on clinically attainable data. However, such studies did not have access to CT data of the analysed subjects. Conversely, other studies could compare DXA-based and CT-based 3D FE models outcomes *ex vivo* [29] and *in vivo* [32], even though the fracture status of the subjects was not available. In these latter studies a better agreement ($R^2 > 0.8$) between DXA-based and CT-based FE outcomes was obtained with respect to the here obtained results. Reasons may lie in the exact same FE modelling pipelines adopted for DXA-derived and CT-based models, which were instead slightly different in this study, and in the DXA scanner employed to develop and validate the DXA2FEM pipeline. The DXA2FEM pipeline [13] had indeed been developed based on a cohort from a different country and on DXA images acquired with a different scanner (Lunar Prodigy, GE Healthcare). In this light, the current results might also be interpreted as a valuable proof of the clinical applicability and generalisability of the DXA2FEM pipeline.

This study is characterised by some limitations which are worth mentioning. First, a relatively low number of subjects coming from the same population was analysed ($n = 88$). On the other hand, the investigated cohort presented unique features, such as the inclusion of fracture history (50% of fractured cases and 50% of matched controls) and the acquisition of both DXA and CT on the same day for all the analysed subjects. Moreover, as the DXA2FEM framework was developed using healthy subjects, its reconstruction accuracy may be reduced when applied to fractured individuals. In addition, the DXA2FEM algorithm was developed and trained on DXA images acquired with GE Healthcare scanners, whereas the DXA scans used here were obtained with Hologic scanners. This might have influenced the obtained outcomes, since each company has its own proprietary method for obtaining aBMD measurements. Cross-validation studies have indeed reported non-negligible differences in aBMD measured with DXA devices from different manufacturers [33]. Besides, such analyses are limited to global aBMD measurements at the total hip or femoral neck region, since pixel-by-pixel aBMD information are usually not disclosed. From this perspective, the results of this study may also, as previously mentioned, contribute to assessing the generalisability of the DXA2FEM pipeline. Future studies might investigate the effects of the DXA scanner on the reconstructed model density values and thus determine whether a scaling factor is needed to correct them. Another limitation is that the presented methodology is relying on proprietary software and is methodologically

advanced, which can be a hinder in terms of applicability in a clinical scenario. However, DXA2FEM is fully automated, and BBCT-hip is in the process of becoming fully automated; therefore, technical expertise is not expected to be required for their execution in the future. Future work will aim at replacing proprietary software with open-source alternatives to further enhance the clinical applicability of the method.

In conclusion, this comparative study enabled an extensive comparison between three-dimensional FE models reconstructed from DXA images using the DXA2FEM pipeline and those built from CT images. Both reconstruction accuracy against relevant *in vivo* clinical CT images and the ability to discriminate fracture cases from controls were analysed. DXA-based ARF0 diagnostic accuracy was lower than CT-derived ARF0, which is in line with the assumptions needed to reconstruct a 3D subject-specific femur FE model from a single clinical DXA scan. DXA-based ARF0, however, yielded an enhanced AUC compared to aBMD and FRAX. Neither the improvement relative to standard clinical methods nor the detriment relative to CT-derived ARF0 were statistically significant, despite this study using one of the largest clinical cohorts of this kind, which included both DXA and CT. By providing a detailed comparison at multiple levels between 3D finite element models based on CT and DXA images, this work opens the way for the development of *in silico* frameworks capable of integrating current clinical gold standards with the most informative model-based indicators attainable from available clinical data. In scenarios where CT imaging is not available, the predictions derived from either a single 3D model reconstructed from a subject's DXA image or from multiple 3D models consistent with that DXA may still be employed to complement aBMD-based predictions.

CRediT authorship contribution statement

Alessandra Aldieri: Writing – original draft, Methodology, Formal analysis, Conceptualization, Visualization. **Pinaki Bhattacharya:** Writing – review & editing, Methodology. **Margaret Paggiosi:** Writing – review & editing, Resources. **Richard Eastell:** Writing – review & editing, Resources. **Lorenzo Grassi:** Writing – review & editing, Resources, Project administration, Funding acquisition, Conceptualization, Supervision.

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.

Acknowledgements

The computation of the 2D-to-3D reconstructions was enabled by resources provided by LUNARC (the center for scientific and technical computing at Lund University), as well as by resources provided by SNIC (the Swedish National Infrastructure for Computing), partially funded by the Swedish Research Council [no. 2018-05973]. The study received funding from Swedish Research Council [no. 2022-03474], the Mats Paulsson's foundation and the EU Erasmus+ Programme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bone.2026.117830.

Data availability

Data will be made available on request.

References

[1] GBD 2019 Fracture Collaborators, Global, regional, and national burden of bone fractures in 204 countries and territories, 1990-2019: a systematic analysis from the global burden of disease study 2019, Lancet Healthy Longev. 2 (9) (2021) e580–e592, https://doi.org/10.1016/S2666-7568(21)00172-0.
[2] J.-Y. Reginster, N. Burlet, Osteoporosis: a still increasing prevalence, Bone 38 (2 Suppl 1) (2006) S4–S9, https://doi.org/10.1016/j.bone.2005.11.024.
[3] J.A. Kanis, N. Norton, N.C. Harvey, T. Jacobson, H. Johansson, M. Lorentzon, E. V. McCloskey, C. Willers, F. Borgström, SCOPE 2021: a new scorecard for osteoporosis in Europe, Arch. Osteoporos. 16 (1) (2021) 1–82, https://doi.org/10.1007/s11657-020-00871-9.
[4] S.R. Cummings, L.J. Melton, Epidemiology and outcomes of osteoporotic fractures, Lancet 359 (9319) (2002) 1761–1767, https://doi.org/10.1016/S0140-6736(02)08657-9.
[5] J.A. Kanis, E.V. McCloskey, H. Johansson, A. Oden, L.J. Melton, N. Khaltaev, A reference standard for the description of osteoporosis, Bone 42 (3) (2008) 467–475, https://doi.org/10.1016/j.bone.2007.11.001.
[6] C. De Laet, A. Odén, H. Johansson, O. Johnell, B. Jönsson, J.A. Kanis, The impact of the use of multiple risk indicators for fracture on case-finding strategies: a mathematical approach, Osteoporos. Int. 16 (3) (2005) 313–318, https://doi.org/10.1007/s00198-004-1689-z.
[7] M. Viceconti, M. Qasim, P. Bhattacharya, X. Li, Are CT-based finite element model predictions of femoral bone strengthening clinically useful? Curr. Osteoporos. Rep. 16 (3) (2018) 216–223, https://doi.org/10.1007/s11914-018-0438-8.
[8] A. Aldieri, M. Terzini, A.L. Audenino, C. Bignardi, M. Paggiosi, R. Eastell, M. Viceconti, P. Bhattacharya, Personalised 3D assessment of trochanteric soft tissues improves HIP fracture classification accuracy, Ann. Biomed. Eng. 50 (3) (2022) 303–313, https://doi.org/10.1007/s10439-022-02924-1.
[9] P. Bhattacharya, Z. Altai, M. Qasim, M. Viceconti, A multiscale model to predict current absolute risk of femoral fracture in a postmenopausal population, Biomech. Model. Mechanobiol. 18 (2) (2019) 301–318, https://doi.org/10.1007/s10237-018-1081-0.
[10] A. Aldieri, M. Terzini, G. Osella, A.M. Priola, A. Angeli, A. Veltri, A.L. Audenino, C. Bignardi, Osteoporotic hip fracture prediction: is T-score-based criterion enough? A hip structural analysis-based model, J. Biomech. Eng. 140 (2018) 111004, https://doi.org/10.1115/1.4040586.
[11] I. Fleps, H. Pálsson, A. Baker, W. Enns-Bray, H. Bahaloo, M. Danner, N.B. Singh, W. R. Taylor, S. Sigurdsson, V. Gudnason, S.J. Ferguson, B. Helgason, Finite element derived femoral strength is a better predictor of hip fracture risk than aBMD in the AGES Reykjavik study cohort, Bone 154 (2022) 116219, https://doi.org/10.1016/j.bone.2021.116219.
[12] A. Aldieri, C. Curreli, J.A. Szyszko, A.A. La Mattina, M. Viceconti, Credibility assessment of computational models according to ASME V&V40: application to the Bologna biomechanical computed tomography solution, Comput. Methods Prog. Biomed. 240 (2023) 107727, https://doi.org/10.1016/j.cmpb.2023.107727.
[13] S.P. Väänänen, L. Grassi, G. Flivik, J.S. Jurvelin, H. Isaksson, Generation of 3D shape, density, cortical thickness and finite element mesh of proximal femur from a DXA image, Med. Image Anal. 24 (1) (2015) 125–134, https://doi.org/10.1016/j.media.2015.06.001.
[14] L. Grassi, S.P. Väänänen, L. Jehpsson, Ö. Ljunggren, B.E. Rosengren, M.K. Karlsson, H. Isaksson, 3D finite element models reconstructed from 2D dual-energy X-ray absorptiometry (DXA) images improve hip fracture prediction compared to areal BMD in osteoporotic fractures in men (MrOS) Sweden cohort, J. Bone Miner. Res. 38 (9) (2023) 1258–1267, https://doi.org/10.1002/jbmr.4878.
[15] L. Grassi, S.P. Väänänen, A. Voss, T. Nissinen, R. Sund, H. Kröger, H. Isaksson, DXA-based 3D finite element models predict hip fractures better than areal BMD in elderly women, Bone 195 (2025) 117457, https://doi.org/10.1016/j.bone.2025.117457.
[16] L. Yang, W.J.M. Udall, E.V. McCloskey, R. Eastell, Distribution of bone density and cortical thickness in the proximal femur and their association with hip fracture in postmenopausal women: a quantitative computed tomography study, Osteoporos. Int. 25 (1) (2014) 251–263, https://doi.org/10.1007/s00198-013-2401-y.
[17] L. Grassi, S.P. Väänänen, M. Ristinmaa, J.S. Jurvelin, H. Isaksson, Prediction of femoral strength using 3D finite element models reconstructed from DXA images: validation against experiments, Biomech. Model. Mechanobiol. 16 (3) (2017) 1–12, https://doi.org/10.1007/s10237-016-0866-2.
[18] L. Grassi, I. Fleps, H. Sahlstedt, S.P. Väänänen, S.J. Ferguson, H. Isaksson, B. Helgason, Validation of 3D finite element models from simulated DXA images for biofidelic simulations of sideways fall impact to the hip, Bone 142 (2021) 115678, https://doi.org/10.1016/j.bone.2020.115678.
[19] 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.
[20] M.C.M. Fischer, S.A.G.A. Grothues, J. Habor, M. de la Fuente, K. Radermacher, A robust method for automatic identification of femoral landmarks, axes, Planes and bone coordinate systems using surface models, Sci. Rep. 10 (1) (2020) 20859, https://doi.org/10.1038/s41598-020-77479-z.
[21] L. Grassi, E. Schileo, F. Taddei, L. Zani, M. Juszczyk, L. Cristofolini, M. Viceconti, Accuracy of finite element predictions in sideways load configurations for the proximal human femur, J. Biomech. 45 (2) (2012) 394–399, https://doi.org/10.1016/j.jbiomech.2011.10.019.
[22] M. Qasim, G. Farinella, J. Zhang, X. Li, L. Yang, R. Eastell, M. Viceconti, 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.

[23] B. Helgason, F. Taddei, H. Pálsson, E. Schileo, L. Cristofolini, M. Viceconti, S. Brynjólfsson, A modified method for assigning material properties to FE models of bones, Med. Eng. Phys. 30 (4) (2008) 444–453, https://doi.org/10.1016/j.medengphy.2007.05.006.

[24] F. Taddei, E. Schileo, B. Helgason, L. Cristofolini, M. Viceconti, The material mapping strategy influences the accuracy of CT-based finite element models of bones: an evaluation against experimental measurements, Med. Eng. Phys. 29 (9) (2007) 973–979, https://doi.org/10.1016/j.medengphy.2006.10.014.

[25] 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.

[26] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (3) (1988) 837–845.

[27] C. Falcinelli, C. Whyne, Image-based finite-element modeling of the human femur, Comput. Methods Biomech. Biomed. Engin. 23 (14) (2020) 1138–1161, https://doi.org/10.1080/10255842.2020.1789863.

[28] A. Aldieri, R. Biondi, A.A. La Mattina, J.A. Szyszko, S. Polizzi, D. Dall'Olio, N. Curti, G. Castellani, M. Viceconti, Development and validation of a semi-automated and unsupervised method for femur segmentation from CT, Sci. Rep. 14 (1) (2024) 7403, https://doi.org/10.1038/s41598-024-57618-6.

[29] A. Dudle, Y. Gugler, M. Pretterklieber, S. Ferrari, K. Lippuner, P. Zysset, 2D-3D reconstruction of the proximal femur from DXA scans: evaluation of the 3D-shaper software, Front. Bioeng. Biotechnol. 11 (2023) 1111020, https://doi.org/10.3389/fbioe.2023.1111020.

[30] L. Steiner, A. Synek, D.H. Pahr, Femoral strength can be predicted from 2D projections using a 3D statistical deformation and texture model with finite element analysis, Med. Eng. Phys. 93 (2021) 72–82, https://doi.org/10.1016/j.medengphy.2021.05.012.

[31] C. Ruiz Wills, A.L. Olivares, S. Tassani, M. Ceresa, V. Zimmer, M.A. González Ballester, L.M. del Río, L. Humbert, J. Noailly, 3D patient-specific finite element models of the proximal femur based on DXA towards the classification of fracture and non-fracture cases, Bone 121 (2019) 89–99, https://doi.org/10.1016/j.bone.2019.01.001.

[32] M. Qasim, M. López Picazo, C. Ruiz Wills, J. Noailly, S. Di Gregorio, L.M. Del Río Barquero, J. Malouf Sierra, L. Humbert, 3D-DXA based finite element modelling for femur strength prediction: evaluation against QCT, J. Clin. Densitom. 27 (2) (2024) 101471, https://doi.org/10.1016/j.jocd.2024.101471.

[33] E. Reitshamer, K. Barrett, K. Shea, B. Dawson-Hughes, Cross-calibration of prodigy and horizon a densitometers and precision of the horizon A densitometer, J. Clin. Densitom. 24 (3) (2021) 474–480, https://doi.org/10.1016/j.jocd.2021.02.003.