Article:

Taylor, M., Viceconti, M., Bhattacharya, P. orcid.org/0000-0003-1109-1174 et al. (1 more author) (2021) Finite element analysis informed variable selection for femoral fracture risk prediction. Journal of the Mechanical Behavior of Biomedical Materials, 118. 104434. ISSN 1751-6161

https://doi.org/10.1016/j.jmbbm.2021.104434

Finite element analysis informed variable selection for femoral fracture risk prediction

Mark Taylor (1), Marco Viceconti (2,3), Pinaki Bhattacharya (4,5), Xinshan Li (4,5)

1. *Medical Device Research Institute, College of Science and Engineering, Flinders University, Tonsley, South Australia, Australia*
2. *Department of Industrial Engineering, Alma Mater Studiorum – University of Bologna*
3. *Medical Technology Lab, IRCCS Instituto Ortopedico Rizzoli, Bologna, Italy*
4. *Department of Mechanical Engineering, The University of Sheffield, Sheffield, UK*
5. *Insigneo Institute for in silico Medicine, The University of Sheffield, Sheffield, UK*

Corresponding Author:

Professor Mark Taylor
Medical Device Research Institute
College of Science and Engineering
Flinders University
1284 South Road
Clovelly Park
Adelaide
5042
South Australia
Australia

Author statement

Mark Taylor: Conceptulisation, methodology, investigation, formal analysis, writing – original draft

Marco Viceconti: Funding acquisition, data curation, writing - review and editing

Pinaki Bhattacharya: writing - review and editing

Xinshan Li: Data curation, writing - review and editing

Abstract

Logistic regression classification (LRC) is widely used to develop models to predict the risk of femoral fracture. LRC models based on areal bone mineral density (aBMD) alone are poor, with area under the receiver operator curve (AUROC) scores reported to be as low as 0.63. This has led to researchers investigating methods to extract further information from the image to increase performance. Recently, the use of active shape models (ASMs) and active appearance models (AAMs) have resulted in moderate improvements, but there is a risk that inclusion of too many modes will lead to overfitting. In addition, there are concerns that the effort required to extract the additional information does not justify the modest improvement in fracture risk prediction. This raises the question, are we reaching the limits of the information that can be extracted from an image? Finite element analysis was used in combination with active shape and appearance modelling to select variables to develop LRC models of fracture risk. Active shape and active appearance models were constructed based on a previously reported cohort of 94 post-menopausal Caucasian women (47 with and 47 without a fracture). T-tests were used to identify differences between the two groups for each mode of variation. Femur strength was predicted for two load cases, stance and a fall. Stepwise multi-variate linear regression was used to identify shape and appearance modes that were predictors of strength for the femurs in the training set. Femurs were also synthetically generated to explore the influence of the first 10 modes of the shape and appearance models. Identified modes of variation were then used to generate LRC models to predict fracture risk. Only 6 modes, 4 active appearance and 2 active shape mode, were identified that had a significant influence on predicted fracture strength. Of these, only two active appearance modes were needed to substantially improve the predictive mode performance (ΔAUROC = 0.080). The addition of 3 more modes (1 AAM and two ASM) further improved the performance of the classifier (ΔAUROC = 0.123). Further addition of modes did not result in any further substantial improvements. Based on these findings, it is suggested that we are reaching the limits of the information that can be extracted from an image to predict fracture risk.

Introduction

Fragility fractures at the hip are a major social-economic problem, particularly with the increasing size of the elderly population. Excess mortality after 1 year of hip fracture varies between 1 in 6 for women to 1 in 3 for men (Frost et al., 2013). Therefore, considerable efforts have been made to better understand the risk factors associated with fracture so that the right medical advice and intervention can be provided. Fracture risk is a function of femur geometry, bone density, microarchitecture, the applied loads and interaction with external environment at the time of a fall. The most widely used technique for assessing the risk of fracture is dual-energy X-ray absorptiometry, or DXA, which measures the areal bone mineral density (aBMD) in the femoral neck. It has been widely used in the clinic to assess bone mineral density status, especially for post-menopausal women, due to its low radiation dose and low cost (Griffith and Genant, 2008).

Logistic regression classifiers (LRC) are widely used for predicting femur fracture risk (Baker-Lepain et al., 2011; Bousson et al., 2011; Bredbenner et al., 2014; Carballido-Gamio et al., 2019; Cheng et al., 2007; Crabtree et al., 2002; Draper et al., 2012; Gnudi et al., 2002; Goodyear et al., 2013; Whitmarsh et al., 2012) and the predictive capability is commonly assessed using the area under the receiver operator curve (AUROC). LRC's built using DXA based aBMD have reported a wide range of AUROC values from 0.62 (Goodyear et al., 2013) to 0.84 (Carballido-Gamio et al., 2019). These reported AUROC values, particularly at the lower end of the range, suggest that the predictive capability of aBMD when used in isolation is, at best, modest. From a clinical perspective, this means that individuals may be misclassified as not being at risk of fracture and then not receive the appropriate care. In addition, a false positive will result in unnecessary intervention, exposing the patient to potential side effects of drug therapy and needless cost to the healthcare provider. This has resulted in a significant body of research investigating whether additional information can be extracted from an image in order to enhance the prediction of fracture risk. Various anatomical measurements, such as hip axis length, femoral neck width and femoral neck angle have been previously associated with fracture risk. However, the inclusion of discrete anatomical measures in LRC's have resulted in negligible improvement in the AUROC as compared to aBMD alone (Baker-Lepain et al., 2011, ∆AUROC = 0.016). Composite measures, which combine aBMD with anatomical measurements in an attempt to estimate strength, have also result in negligible improvements in AUROC (Leslie et al., 2009, ∆AUROC = 0.009; Li et al., 2013, ∆AUROC = 0.005). The use of quantitative computed tomography scans to assess the volumetric bone mineral density (vBMD) does not improve the reported AUROC values (Cheng et al., 2007, ∆AUROC = 0.005; Carballido-Gamio et al., 2019, ∆AUROC = -0.002). Moderate improvements in AUROC have been reported when cortex thickness related variables have been incorporated into the LRC's (Cheng et al., 2007, ∆AUROC = 0.024 to 0.05; Treece et al., 2015, ∆AUROC = 0.072). The greatest improvements have been seen when variables derived from active shape (ASM) and appearance (AAM) models have been used. Based on DXA images LRC's using ASM and AAMs have had mixed performance (Lu et al., 2017, ∆AUROC = 0.0; Goodyear et al., 2013, ∆AUROC = 0.03; Baker-Lepain et al., 2011, ∆AUROC = 0.160). When using ASMs and AAMs built from QCT images, there is a more consistent improvement in performance (Bredbenner et al., 2014, ∆AUROC = 0.12; Carballido-Gamio et al., 2019, ∆AUROC = 0.081). However, various authors have commented that the additional effort in extracting information from DXA and CT images (which are not routinely used in the diagnosis of osteoporosis) using increasingly sophisticated approaches does not significantly improve fracture prediction (Black et al., 2008; Treece et al., 2015). Therefore, are we reaching the limits of what can be achieved using information derived from an image alone?

The potential advantage of using ASM and AAMs is the ability to use all of the information contained within the image. Due to their nature, ASMs typically describe the variation in the external geometry of the femur. ASMs indirectly capture information related to discrete anatomical measures including femoral neck diameter, hip axis length and femoral neck-shaft angle within the shape modes. ASMs tend to be compact typically requiring approximately

10% of the available modes to describe 95% of the variations in shape (Bryan et al., 2010; Sarkalkan et al., 2014). AAMs are built using the bone density information within the image, with or without the shape component. AAMs describe the variation and distribution of bone density and therefore capture the variation in cortical bone thickness. AAMs are not compact and require a high proportion of the available modes to describe 95% of the variation in bone density (Bryan et al., 2010; Bonaretti et al., 2011; Sarkalkan et al., 2014). A significant number of the modes describe less than 1% of the variation and it is debatable whether these modes contain any meaningful information. ASMs and AAMs use principal component analysis to reduce the dimensionality of a much larger data set, however, it still remains a challenge to identify which modes of variation will be useful in building a logistic regression model to predict fracture risk. Traditionally, it has been recommended that the sample size should be at least 10 times greater than the number of predictive variables when developing a logistic regression model (van Smeden et al., 2019). Increasing the number of predictive variables beyond this threshold risks over fitting. ASM and AAMs generate $N - 1$ modes of variation, where N is the number of training sets, so generating many more potential variables than should be included in logistic regression. The most common approach to identify potential predictive variables is by performing a statistical test on the principal component (PC) scores for each mode between the fracture and non-fracture cohorts in the ASM and AAM training data. This could be as simple as performing a t-test (Goodyear et al., 2013), using Fisher linear discriminate analysis (Whitmarsh et al., 2011) through to more complex machine learning based techniques (Bredbenner et al., 2014; Carballido-Gamio et al., 2019;). For the first few modes, which capture a significant proportion of the variation, the differences in the PC scores are likely to be due to real differences in the shape or density distribution. However, statistical differences seen in later modes, particularly in AAMs, which explain less than 1% of the variation may not contain any meaningful information. Identifying predictive variables using this approach also gives no information about the relative importance of a selected mode and its contribution to the prediction of fracture risk.

The mechanical competency of the proximal femur, as measured through its strength, is a key factor in determining an individual's risk of fracture. Finite element (FE) analysis has been used extensively to assess proximal femoral strength. QCT based FE models (or biomechanical CT analysis) can predict between 80% and 94% of femoral strength in simulated fall or stance position (Dall'Ara et al., 2013; Hambli and Allaoui, 2013; Schileo et al., 2014; Zysset et al., 2015), whereas some papers showed a similar predictive accuracy between DXA-FE and DXA-aBMD (Amin et al., 2011; Yang et al., 2014). When used in combination with ASM and AAMs, FE models have the potential to investigate the contribution of individual modes to fracture strength. We hypothesised that modes which do not contribute to fracture strength are unlikely to contribute to the performance LRC models to predict fracture risk.

The aim of this study was to explore the use of FE analysis as an alternative method to identify meaningful predictive variables for use in the development of a logistic regression model of fracture risk. Independent ASM and AAMs were generated on a previously reported cohort of menopausal women with and without fractured femurs. FE analysis was used to predict the fracture strength of all femurs within the cohort. Stepwise multiple linear regression was used to identify the modes which contribute to femur strength. In addition, synthetic FE models were generated from the ASM and AAMs to explore how the primary modes of variation influence femur strength. Finally, the selected ASM and AAM modes were used to build logistic regression models to determine if sequential addition of the selected modes improved the prediction of fracture risk as compared to aBMD alone.

Methods

Cohort description:

Independent active shape and appearance models were constructed based on a previously reported cohort of Caucasian women (Qasim et al., 2016; Yang et al., 2014) and details of the cohort are briefly summarised. There were 100 women who were at least 5 years post menopause, 50 of whom had a low energy hip fracture and 50 of whom were selected to be pair matched in terms of age, height and weight. For fracture patients with a body mass index (BMI) between 16 and 34, the control was chosen to have an age ±5 years, height ± 5 cm and weight ± 5 kg. For fracture patients with BMI ≥34 or BMI ≤16, the control was chosen to have age ± 5 years and BMI ± 4 kg/m^2. Sheffield Local Research Ethics Committee approved the study and informed written consent was obtained for all participants. Bilateral QCT scans (LightSpeed 64 VCT, GE Medical Systems at 120 KVp/150mA) were obtained for each subject for a region which included from just above the femoral head to 3.5cm below the lesser trochanter. The European spine phantom was used to calibrate for bone density retrospectively. For subjects that had experience a fracture, the contralateral femur was used for analysis. A summary of the demographics of the subjects is included in Table 1. Note that Qasim et al., 2016 reported results for only 98 individuals. This was due to the fact that multiple high-density calcified areas were observed in the CT images of a fractured case. This femur and its paired control femur were excluded. In addition, a further four femurs were excluded from the present study as the femurs were too short, with the CT scan finishing close to the lesser trochanter rather than 3.5 cm below it.

	Fracture (n = 47)		**Control (n = 47)**	
	Mean (range)	**Std**	**Mean (range)**	**Std**
Age (years)	75.1 (54.8 – 88.7)	9.5	74.1 (55.9 – 91)	9.0
Weight (kg)	62.7 (31 – 101.1)	14.7	64.6 (42.8 – 92.7)	12.3
Height (cm)	158.7 (145 – 173)	6.6	157.5 (145.1 – 169.3)	5.9
T-score	-2.40	0.85	-1.62	0.85

	(-4.15 - -1.07)		(-3.07 – 0.85)	

Table 1: Subject demographics

Development of the active shape and appearance models:

The femur for each patient was segmented manually using ITK-Snap 2.0.0 (University of Pennsylvania) (Yushkevich et al., 2006) in order to extract the 3D bone geometry (Qasim et al 2016). Active shape and appearance models require correspondence across the training datasets and this was achieved by morphing a template mesh to each femur in the study using an established methodology (Grassi et al., 2011). A template mesh was generated for one of the femurs, which consisted of 295,589 tetrahedral elements, with an averaged finite element size of 3 mm. The morphing was performed in two steps. First surface morphing was performed, where the surface mesh was extracted from the template mesh and morphed to the target femur using a landmark based methodology. The landmarks acted as constraints and were used to interpolate the motion of all the nodes. Laplacian smoothing was implemented to maintain the quality of the surface mesh. In the second phase the template of the volumetric mesh was morphed to the target femur, using the nodes of the morphed surface mesh as the constraints. This was performed using a standard automatic meshing algorithm (ICEM CFD14, Ansys Inc, PA, USA). Elastic moduli were mapped from the CT scans to the morphed FE mesh of the femurs using Bonemat (http://www.bonemat.org/), using established relationships (Schileo et al., 2008a) between radiographic density and ash density ($\rho_{ash} = 0.877\rho_{QCT}+0.079$), ash density and wet apparent density ($\rho_{app} = \rho_{ash}/0.6$) and wet apparent density and elastic modulus ($E = 6590\ \rho_{app}^{1.49}$). Once correspondence was achieved, principal component analysis was performed on the nodal coordinates to generate the ASM and the apparent density assigned to each element to generate the AAM.

Finite element analysis:

FE analysis was performed on each femur contained within the training set. In addition, the active shape and appearance models were used to synthetically generate FE models in order to explore the influence of shape and bone density distribution on femoral bone strength. The mean and standard deviation of the PC scores were calculated for the fracture and control femurs, for each mode of the ASM and AAM. These values enabled the generation of synthetic femurs representative of the fracture and control groups. First, to explore the influence of shape, the active shape model was used to synthetically generate the mean femur geometry and femur geometries at ± 1 and ± 2 standard deviations of each of the first 10 shape modes for both the fracture and control groups. To isolate the influence of geometry, the mean bone density distribution for each group was applied to all models in that group. Second, to explore the influence of bone density distribution, the active appearance model was used to synthetically generate the mean bone density and bone densities at ± 1 and ± 2 standard deviations of each of the first 10 appearance modes for the fracture and control groups. To isolate the influence of bone density distribution, analyses were performed using the mean shape for each group.

Two loading conditions were simulated: the first replicating a stance load case with a vertically oriented joint contact force (0 degrees in both the frontal and sagittal planes) (Taylor et al., 2017). The distal end of the femur was rigidly constrained. The second load case represented a fall with the joint contact force applied at 90 and 0 degrees in the frontal and sagittal planes, respectively (Qasim et al., 2016). In addition to the distal femur being rigidly constrained, the most lateral node was constrained in the direction of loading, but was free to translate perpendicular to the direction of the applied load. The femoral neck fracture load was estimated using a modified version (Taylor et al., 2017) of Schileo's maximum principal strain criterion (Schileo et al., 2008b). The 90th percentile principal tensile and compressive strain, for all elements in the femoral neck, were expressed as a fraction of their tensile and compressive elastic limits. The maximum of these was taken as the femoral neck risk factor (RF). The femoral neck fracture load was then estimated by linearly scaling the applied forces until the risk factor reached unity. When the modes were studied independently, a mode was deemed to have a substantial, moderate or negligible influence on strength if the range of forces across two standard deviations of variation were greater than 1000 N (substantial), between 250 N and 1000 N (moderate) and less than 250 N (negligible).

Statistical analysis:

i) The PC scores of the fracture and non-fracture groups were compared for each mode in both the active shape and appearance models (Daruwalla et al., 2010). Student T-tests were used to identify statistical differences ($p<0.05$).

ii) Stepwise multivariate linear regression (MLR) was performed with the predicted femoral neck strength as the output variable. Three MLR models were built using the first 20 modes (10 ASM and 10 AAM), the first 40 modes (20 ASM and 20 AAM) and the modes required to describe 95% of the variation. Only variables that achieved significance ($p<0.05$) value were included in the final MLR models. This was performed for both the stance and fall load cases. The coefficient of regression (R^2) and the root mean square (RMS) error were reported.

iii) A series of logistic regression models were built, using variables identified from (i) and (ii). In each case, a five-fold cross validation methodology was employed. The training set was randomly divided into 5 equal folds. The first 4 folds were used to train the logistic regression classifier and the fifth fold was used to validate the classifier. This process was repeated five times and the area under the receiver operator curve (AUROC) and the precision, as defined by the percentage of correct true and false classifications, were calculated. In order to assess the robustness of each classifier, the entire process was repeated 100 times. The mean and range for the resulting AUROC and precision have been reported. As a reference, a logistic regression classifier was also built using aBMD using the same methodology.

Results

The ASM was compact, with more than 95% of the variation in shape (Table 2) explained by the first 12 modes (out of a total of 93 modes). ASM Mode 1 (figure 1) was a scaling mode, as well as capturing the variation in length of the femur below the lesser trochanter contained within the original images, and explains over 55% of the variation in shape. Analysis of the PC scores for the fracture and non-fracture groups shows that statistical differences were only observed for modes 4, 5, 12, 35 and 67. Modes 4 and 5 combined explained 8.5% of the observed variation, and the other 3 modes combined explained less than 1% of the variation. Mode 4 captures the variation in the geometry of the femoral neck, with variations in both neck length and diameter (figure 1). Mode 5 captures the variation in geometry of the greater trochanter and to a lesser extent, the geometry of superior femoral neck.

	Active shape model		Active appearance model	
Mode	**Percentage of explained variation**	**P value**	**Percentage of explained variation**	**P value**
1	55.0	0.99	39.8	**$1x10^{-6}$**
2	11.5	0.32	6.6	0.28
3	7.5	0.72	5.1	**0.01**
4	4.9	**0.02**	4.3	0.40
5	3.8	**0.02**	3.3	**0.02**
6	2.3	0.96	2.5	0.94
7	2.1	0.99	2.0	0.92
8	1.6	0.18	1.8	0.14
9	1.2	0.95	1.8	0.60
10	1.1	0.49	1.4	0.70

Table 2: Comparison of the percentage of explained variation for the first 10 modes of the active shape and active appearance models for the entire cohort. The P value is reported for the difference in the mode weights between the fracture and non-fracture groups. Statistically significant differences (P<0.05) are shown in bold.

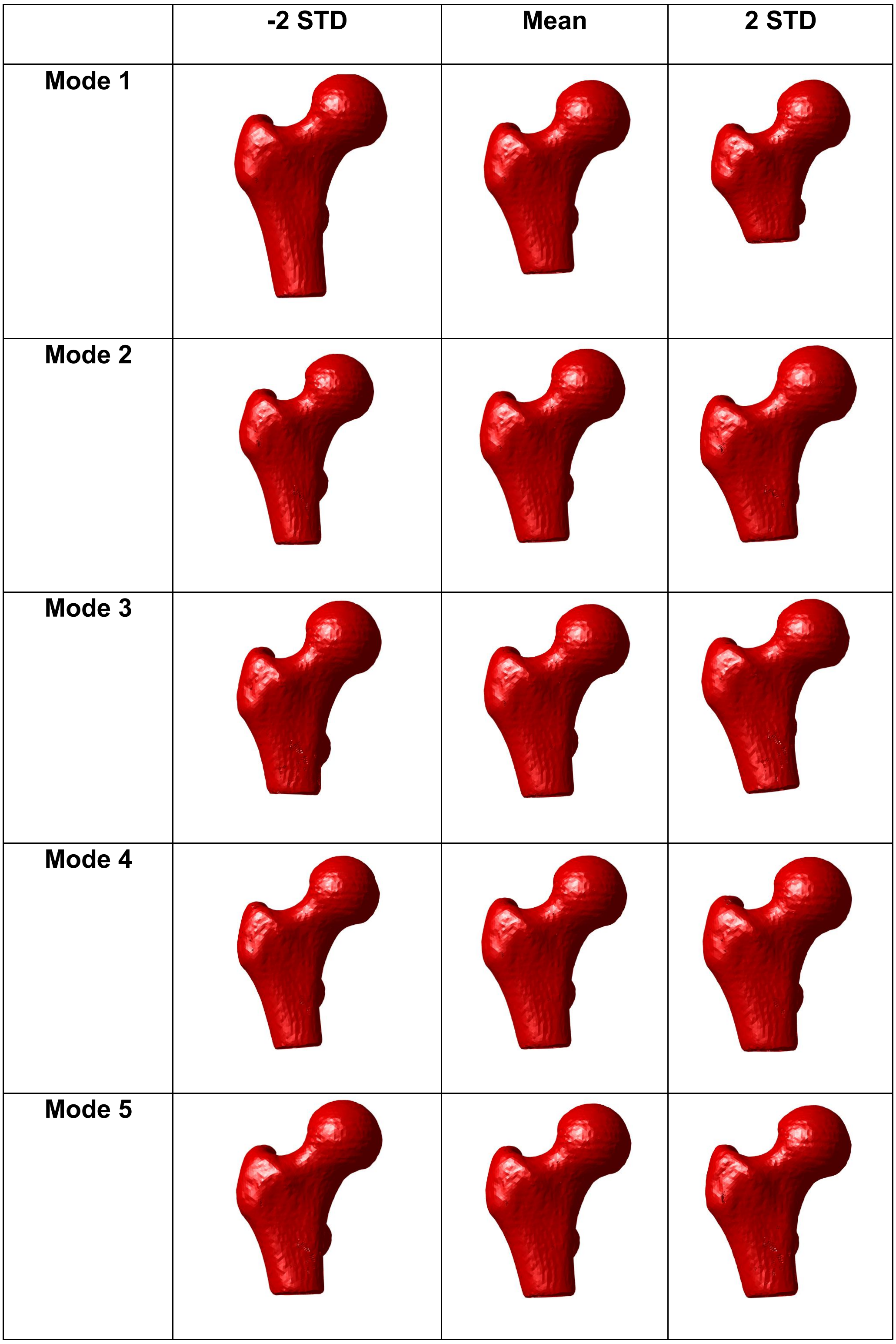


Figure 1: The first 5 modes of variation (STD, standard deviation) of the active shape model

The AAM was less compact, with the first 62 modes required to explain 95% of the variation in the proximal femoral bone density distribution. The mean femur in the fracture group had noticeably lower bone density as compared to the mean femur in the non-fracture group (figure

2). AAM mode 1 captures 39.8% of the variation (Table 2) and captures the overall variation in bone mass, as well as variation in cortex thickness. In particular, the cortex of the superior femoral neck appears to be thinner and less dense in the fracture group as compared to the control group. Statistically significant differences were seen between the mode PC scores of the fracture and non-fracture groups for modes 1, 3, 5 (table 2) and 26. These explained 39.8, 5.1, 3.3 and 0.6 percent of the variation of the bone density distribution respectively.

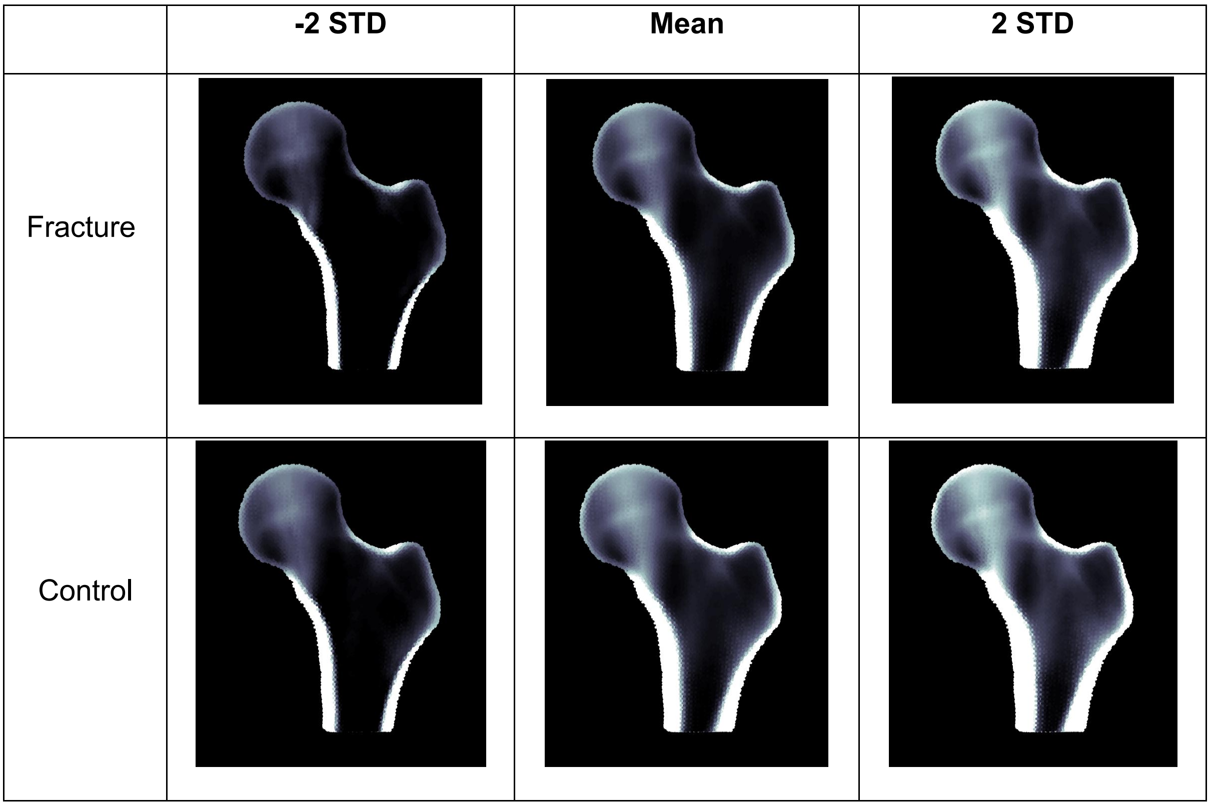


Figure 2 – Variation in mode 1 of the active appearance model for the fracture group (top row) and the controls (bottom row). The average (middle) and -2 std’s (left) +2 std’s (right) are shown.

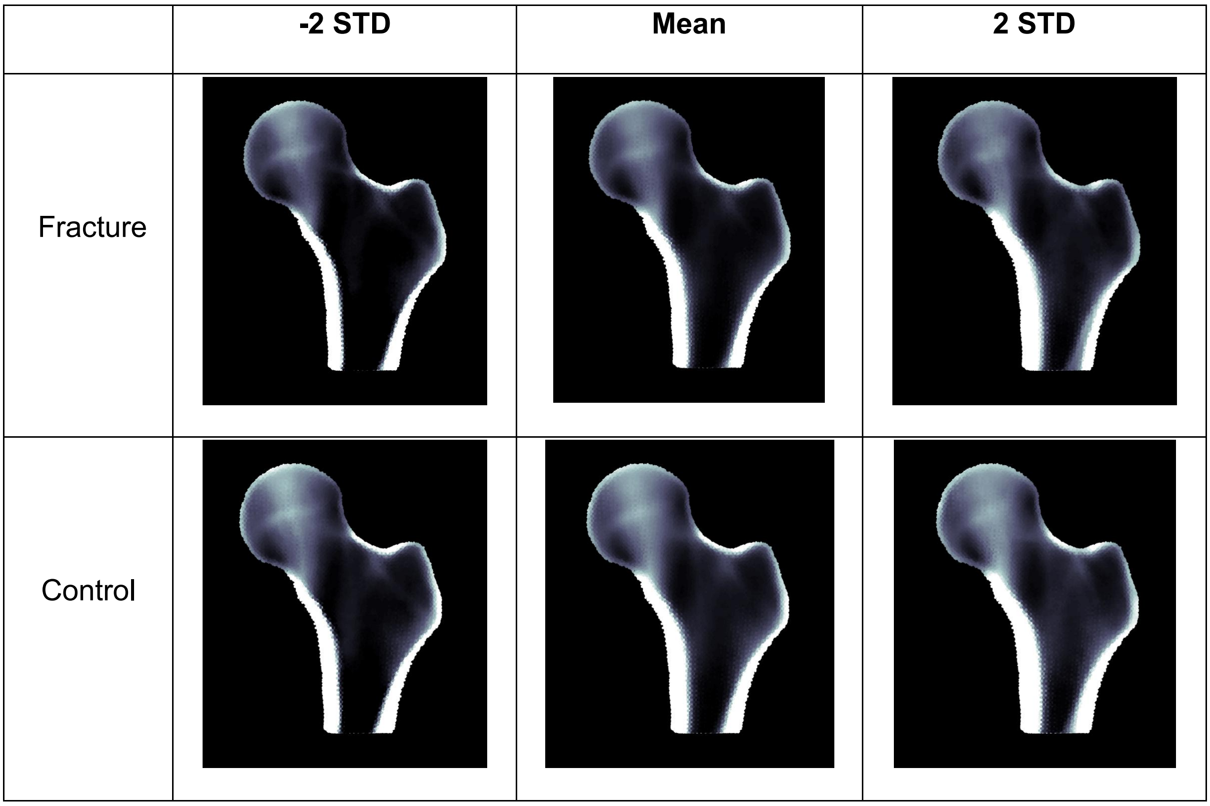


Figure 3 – Variation in mode 2 of the active appearance model for the fracture group (top row) and the controls (bottom row). The average (middle) and -2 std's (left) +2 std's (right) are shown.

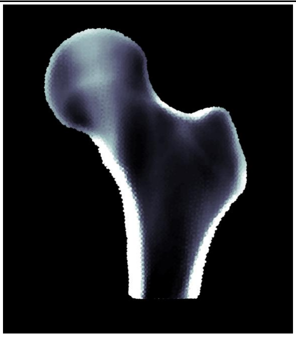
	-2 STD	Mean	2 STD
Fracture			
Control			

Figure 4 – Variation in mode 3 of the active appearance model for the fracture group (top row) and the controls (bottom row). The average (middle) and -2 std's (left) +2 std's (right) are shown.

The predicted fracture strength for the average femur was 3380 N and 5552 N for the stance load case, and 1221 N and 2256 N for the fall load case for the fracture and control groups respectively. For the fall load case, independently varying the first 10 modes of the ASM resulted in a maximum difference of 339 N and 705 N for the fracture and control groups respectively (figure 5). The greatest change in the predicted fracture load was associated with ASM mode 4, followed by mode 2 and then mode 1. The remaining modes had no significant influence on femoral neck strength. Independently varying the first 10 modes of the AAM resulted in maximum differences of 2269 N and 2564 N from the fracture and control groups respectively. AAM mode 1 was associated with the greatest change in the fracture load, followed by mode 3, mode 8 and then mode 2. AAM modes 7 and 9 also influenced the femoral neck strength more than was observed for any of the ASM modes.

Three MLR models were built using 20 modes, 40 modes and all the modes required to describe 95% of the variation (12 ASM and 62 AAM modes) for the fall load case. Using stepwise MLR to elimination of the trivial modes resulted in models based on 6, 11 and 17 predictive variables. These had R^2 values of 0.83, 0.88 and 0.92 and RMSE values of 372 N, 316 N and 265 N respectively. The first 6 predictive variables were the same in each of the MLR models and were, in order of importance, AAM 1, AAM 3, AAM 2, AAM 8, ASM 4 and ASM 5. Mode 1 of the AAM was the most important term, yielding an R^2 value of 0.466. The addition of mode 3, mode 2 and then mode 8 of the AAM improved the R^2 value to 0.792 and reduced the RMS error from 642 N to 407 N. The addition of ASM modes 4 and 5 resulted in a minor increase in the R^2 value from 0.792 to 0.830 and a minor reduction in the RMSE from 407 N to 372 N. A similar trend was observed for the stance loadcase. Four predictive variables were identified, AAM modes 1, 3 and 8 and ASM mode 4, producing an R^2 value of 0.781 and an RMS error of 1032 N.

The ability to predict fracture risk was assessed using logistic regression models. The reference fracture risk model was built using aBMD resulted in an AUROC score of 0.719 and a precision of 65.2% (Table 4). A fracture risk model built using just AAM mode 1 yielded similar results to the one built using aBMD. The addition of AAM mode 3 improved the AUROC score to 0.799 and the precision to 75.2%. Fracture risk models built using the other modes identified by MLR did not improve their performance (Table 4). The best performing fracture risk model, with an AUROC of 0.865 and a precision of 80.5% was constructed using all of the modes identified using T-tests. Removal of the trivial modes (those contributing less that 1% of the explained variation) produced a similar level of performance (AUROC = 0.842, precision = 80.6%). This fracture risk model consisted of AMM modes 1, 3 and 5 and ASM modes 4 and 5.

Variables used in building MLR model	**R squared**	**RMSE (N)**
AAM 1	0.466	642
AAM 1+ AAM 3	0.646	525
AAM 1+ AAM 3 + AAM 2	0.722	468
AAM 1+ AAM 3 + AAM 2 + AAM8	0.792	407
AAM 1+ AAM 3 + AAM 2 + AAM8 + ASM4	0.816	385
AAM 1+ AAM 3 + AAM 2 + AAM 8 + ASM 4 +ASM 5	0.830	372

Table 3: Stepwise multivariate linear regression to predict fracture load for the fall load case based on the first 20 modes (10 ASM and 10 AAM modes).

Variables used in building logistic regression classifier	**Mean AUROC (min – max)**	**Mean precision (%) (min -max)**
aBMD	0.719 (0.688 – 0.737)	65.2 (59.6 – 68.1)
AAM 1	0.761 (0.732 – 0.782)	68.7 (64.9 – 73.4)
AAM 1+ AAM 3	0.799 (0.762 – 0.825)	75.2 (68.1 – 79.8)
AAM 1+ AAM 2 + AAM 3 + AAM8 + ASM 4 + ASM 5[1]	0.794 (0.748 – 0.825)	78.1 (67.0 – 77.7)
AAM 1+ AAM 2 + AAM 3 + AAM 7 + AAM8 + AAM 12 + AAM 15 + AAM 17+ ASM 3 + ASM 4 + ASM 5[2]	0.801 (0.740 – 0.847)	72.5 (64.8 – 79.8)
AAM 1+ AAM 3 + AAM 5 + AAM 26 + ASM 4 + ASM 5 + ASM 12 + ASM 35 + ASM 67[3]	0.865 (0.793 – 0.905)	80.5 (75.5 – 85.1)
AAM 1+ AAM 3 + AAM 5 + ASM 4 + ASM 5[4]	0.842 (0.781 – 0.870)	80.6 (75.5 – 85.1)

Table 4: Summary of the performance of the logistic regression classifiers to predict fracture risk. [1]Features identified based on multivariate linear regression of the first 10 ASM and AAM modes. [2]Features identified based on multivariate linear regression of the first 20 ASM and AAM modes. [3]Features identified based t-testing for significant differences between the modes for the fracture and control groups. [4]Same as 3 but with the minor modes removed. This was also the highest performing classifier based on any combination of 5 variables chosen from 7 candidate variables identified from t-testing and MLR.

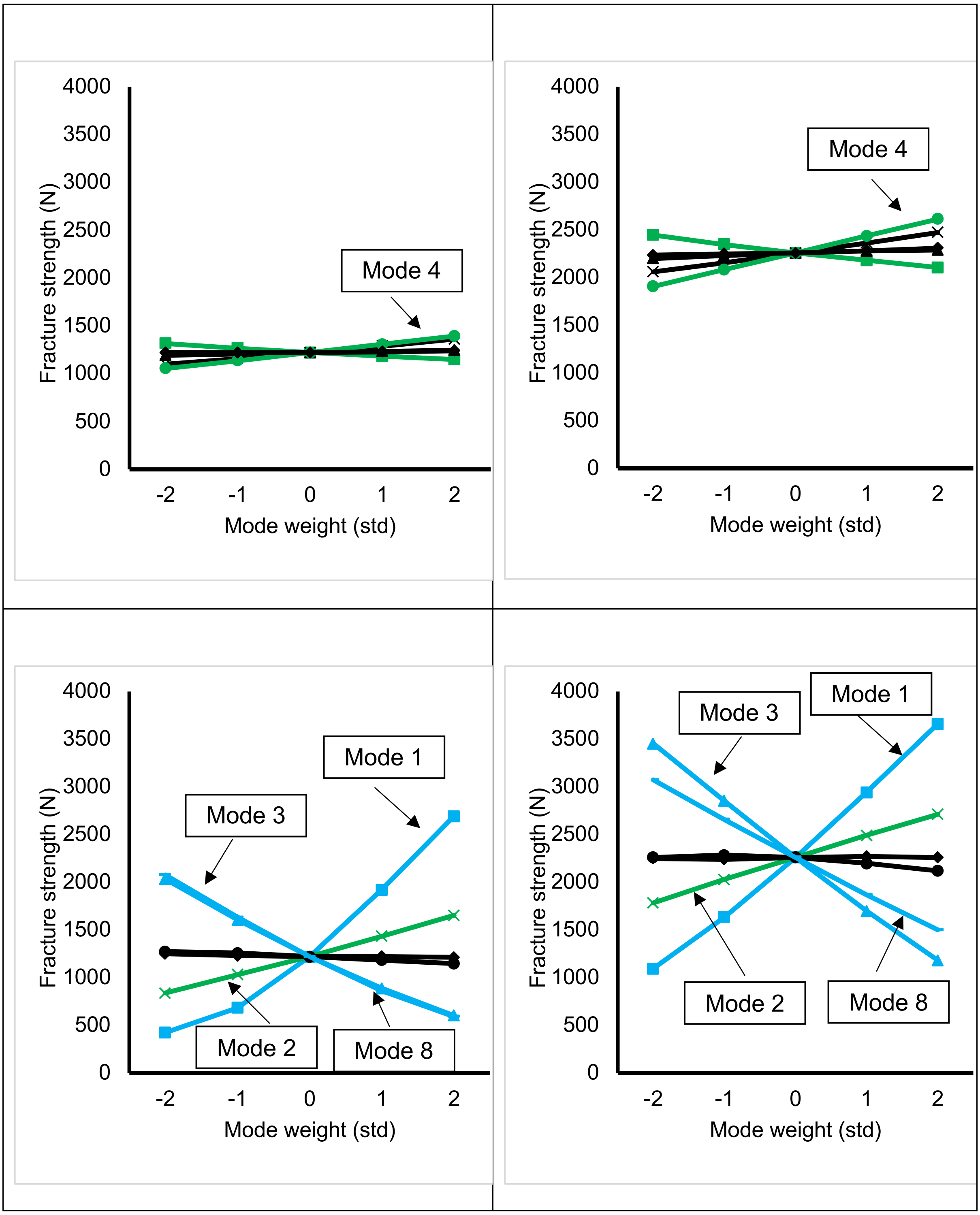


Figure 5 – Influence of shape (top row) and density (bottom row) on the predicted fracture strength of the fractured femurs (left) and non-fractured femurs (right) subjected to the fall load case. Data labels: mode 1 – square; mode 2 – cross, mode 3 – triangle, mode 4 – circle and mode 5 – diamond, mode 8 - dash. Modes which have a substantial influence on strength are

highlighted in blue, a moderate influence in green and negligible influence in black. For clarity only the first 5 modes are shown, plus mode 8. The other modes have negligible influence on the fracture strength and are not shown.

Discussion

The current benchmark for prediction of femoral neck fracture risk is based on aBMD. Researchers have included discrete geometric measures of anatomy to try and improve predictions with limited success (Baker-Lepain et al., 2011; Cody et al., 2000a; Gnudi et al., 2002; Gregory et al., 2004; Michelotti and Clark, 1999; Partanen et al., 2001; Pulkkinen et al., 2004). Statistical regression and classification models built using variables derived from ASM and AAMs in 2D (Goodyear et al., 2013) and 3D (Bredbenner et al., 2014; Carballido-Gamio et al., 2019) have the potential to utilise all of the information contained within the image. ASMs have been shown to be compact, however, AAMs, either based on density alone or in combination with shape, required a high percentage of the modes to describe the variation in the training set. When using these modes as variables in a predictive statistical model, there is a risk that using too many will lead to over-fitting and, as a consequence, an over-estimation in the model's ability to predict fracture risk. FE analysis, used in combination with ASM and AAMs, provides a mechanism to explicitly explore and identify the meaningful modes that contribute to the prediction of fracture strength, which can then be used to inform the selection of appropriate variables for use in predictive models of fracture risk. In this study, FE was used to explore which modes were useful by: (i) identifying potential variables using stepwise MLR to predict fracture strength and (ii) examining the influence of each mode in isolation on the predicted fracture strength.

Similar to other studies (Baker-Lepain et al., 2011; Bredbenner et al., 2014; Bryan et al., 2010; Goodyear et al., 2013; Sarkalkan et al., 2014), the ASM was compact with the first 12 modes (of 93) explaining more than 95% of variation in geometry. The AAM required a higher percentage of modes (62 out of 93) to explain 95% of the variation in bone density distribution in the training cohort. The poorer performance of the AAMs as compared to the ASMs has been commonly reported (Bonaretti et al., 2014; Bredbenner et al., 2014; Bryan et al., 2010; Sarkalkan et al., 2014).

Based on the empirical recommendation of the sample size being at least 10 times the number of predictive variables (van Smeden et al., 2019), up to 9 predictive variables would be appropriate for this study. Selection based on differences in the PC scores for each mode between the fracture and non-fracture groups identified 5 ASM modes and 4 AAM modes as potential predictive variables. Depending on the number of modes used to develop an MLR model of fracture strength, between 6 (2 ASM and 4 AAM modes) and 17 modes (4 ASM and 13 AAM modes) were identified as predictive variables for the fall load case. Increasing the number of predictive variables improved the R^2 score and reduced the RMS error in the MLR models. In theory, the additional information contained by including more modes has the potential to improve the performance of logistic regression model of fracture risk, but this was found not to be the case. Although stepwise MLR regression has identified a particular mode as a predictive variable, the question is whether it truly contributes to femur strength and hence to fracture risk. Synthetically generating and analysing FE models for each of the ASM and AAM modes in isolation allows us to quantify its contribution to fracture strength. Bredbenner et al. (Bredbenner et al., 2014) built a coupled statistical shape and intensity model. Based on a larger training set than used in this study (N = 450), they identified 20 modes which were used to build a logistic regression classifier. The first 3 modes described 46% of the variation and the remaining 17 models explained a further 7.1% of the variation. Based on our results, in the interest of determining causality, it is recommended that separate ASMs and AAMs are built in order to identify their relative contributions to strength. In addition, it is likely that modes that only account for a small percentage of variation do not contribute to femur strength and therefore are questionable in their role in predicting fracture risk.

The first 6 predictive variables in each of the MLR models were, in order of importance, AAM modes 1, 3, 2 and 8 and ASM modes 4 and 5. AAM modes 1, 3 and 8 all had a substantial influence of femoral neck strength for the fall load case. AAM mode 1, which accounted for 39.9% of the variation, captures two main features, the global distribution of bone density and the cortex thickness, particularly of the superior femoral neck. When examined independently, AAM mode 1 has the greatest influence on femoral neck strength of any of the appearance or shape modes (Figure 5). AAM mode 3 accounts for just 5% of the variability in the bone density distribution and appears to capture the architecture of the head-neck region (figure 4), controlling the width of the medial column stretching from the femoral head to the medial cortex, as well as the lateral arch spanning from the lateral cortex through the superior femoral neck and into the distal medial femoral head. When studied independently, AAM mode 3 also has a substantial influence on femoral neck strength, which varied from 602 N to 2033 N and from 1183 N to 3455 N for the fracture and non-fracture cohorts respectively.

The next two predictive variables identified by MLR were AAM modes 2 and 8. When the influence of AAM mode 8 was studied independently it was found to have a substantial influence on femoral neck strength (figure 5), similar to AAM mode 3 (ranging from 592 N to 2077 N and 1502 N to 3074 N for the fracture and non-fracture cohorts respectively). In comparison, AAM mode 2 (figure 3) only had a moderate influence of neck strength. The higher ranking of AAM mode 2 by the MLR models may be due to the percentage of variability explained by this mode, which was 6.6% as compared to 1.8% by AAM mode 8. AAM modes 7 and 9 were also found to have a moderate effect on neck strength and were identified by the MLR models which include more initial modes, whereas AAM modes 4, 5, 6, and 10 had negligible influence.

Only ASM modes 4 and 5 were shown to be of importance both through the statistical comparison of the PC scores and through the MLR. ASM mode 4 only accounts for 5% of the variation, but appears to describe the geometry of the femoral neck in terms of its length and diameter (figure 1). ASM mode 4 was shown to have a moderate influence on femoral neck strength. ASM mode 5 was identified as a predictive variable through MLR but when studied in isolation was found to have negligible influence on femoral neck strength. ASM modes 1 and 2 both have a moderate influence on femoral neck strength, but their contribution was small in comparison to the AAM modes with a moderate effect (AAM 2, 7 and 9). ASM modes 1 and 2 were not selected through MLR as predictive variables. This is noteworthy, particularly with respect to ASM mode 1. This is a scaling mode and essentially describes the overall size of the femur. Hence the MLR suggests that size is not a determinant of femur strength. The remaining shape modes (3, 6, 7, 8, 9 and 10) all had a negligible influence on strength.

This study has demonstrated, through independent analysis of the ASM and AAM modes and MLR of fracture strength, that the density distribution contributes more to fracture strength than the external size and morphology of the femur. Both Whitmarsh et al. (Whitmarsh et al., 2011) and Carballido-Gamio et al. (Carballido-Gamio et al., 2019) have reported similar findings. Whitmarsh et al. (Whitmarsh et al., 2011) found that the first 3 AAM modes were most important in the development of a Fisher linear discriminant model to differentiate between fracture and non-fractured femurs. Carballido-Gamio et al. (Carballido-Gamio et al., 2019) reported that inclusion of AAM based variable always improved fracture risk classification. Carballido-Gamio et al. (Carballido-Gamio et al., 2019) noted that AAMs capture cortical bone thickness and that these may be surrogates measures of bone strength. Through this study, we have been able to demonstrate that this is indeed the case. Only ASM mode 4 was identified as a predictive variable by MLR and found to have a moderate influence on strength, perhaps explaining why studies that have incorporated discrete anatomical measurements into logistic regression models have resulted in no or minor improvements in the prediction of fracture risk (Baker-Lepain et al., 2011; Cody et al., 2000a; Gnudi et al., 2002; Gregory et al., 2004; Michelotti and Clark, 1999; Partanen et al., 2001; Pulkkinen et al., 2004).

Logistic classifiers were built using the predictive variables. The AUROC score for the aBMD model (0.719) and the best predictive model (AUROC = 0.865) are within the ranges reported in the literature (Carballido-Gamio et al., 2019; Whitmarsh et al., 2012) and the improvement in the AUROC score of 0.146 is of a similar order of magnitude. Using just 5 modes (ASM modes 4 and 5, AAM modes 1, 3 and 5) resulted a similar level of performance as the best performing fracture risk model. Four of these variables (AAM modes 1 and 3 and ASM modes 4 and 5) were identified by MLR and shown to have moderate to substantial influence on femoral neck strength. Only AAM mode 5 was not identified as a predictor of femoral neck strength. A logistic regression model built using just AAM mode 1 (table 4) resulted in an AUROC score better than that generated by aBMD, similar to that reported by Whitmarsh et al. (Whitmarsh et al., 2012). Areal BMD is only a measure of bone density, whereas AAM mode 1 captures both the variation in bone density as well as the cortex thickness (figure 2), leading to a better prediction of fracture risk. The addition of AAM mode 3 further enhances the AUROC (0.799) and precision (75.2) of the fracture risk model.

There are a number of limitations associated with this study. The examined cohort is based on an all-female dataset. Cody et al. (Cody et al., 2000b) reported that there were differences in the variables associated with fracture risk between men and women. Therefore, further work is required to establish if similar trends are maintained in a mixed gender cohort. The study is based on a sample of 94 subjects, evenly divided between fracture cases and matched controls. The ratio of fractured femurs to controls used in comparable studies developing logistic regression models varies from 0.53:1 (Carballido-Gamio et al., 2019b) to 1:10 (Bredbenner et al., 2014). Regardless of the ratio of fractured femurs to controls, clinical studies have routinely demonstrated that the bone mineral density is statistically significantly lower in the fractured cohort (Baker-Lepain et al., 2011a; Cody et al., 2000b; Gnudi et al., 2002; Gregory and Aspden, 2008; Pulkkinen et al., 2004). In this study, mode 1 of the AAM captures the global magnitude and distribution of bone density and clear differences were observed between the fracture and control groups. Therefore, it is unlikely that changing the ratio of fractured femurs to controls will change this in our study. ASM mode 4 was shown to be a minor contributor to the prediction of femoral neck strength. Increasing ratio of control to fractured femurs may further reduce its contribution. In terms of developing ASMs and AAMs, a larger training set is always desirable. The active shape model is compact and extending the size of the training cohort is unlikely to improve its quality. In comparison, the compactness of the active appearance model is poor. Although the number of training sets does have an influence on the quality of an active appearance model, PCA based methodologies do not appear to be able to generate a compact model (Bonaretti et al., 2014; Bredbenner et al., 2014; Bryan et al., 2010; Sarkalkan et al., 2014). Alternative statistical approaches are needed to better describe the variation in bone density distribution. This may also require the separate segmentation of cortical and cancellous bone and of the bone marrow cavity. A number of logistic regression models were built and the key findings summarised in Table 4. An exhaustive search was not performed. There may be a combination of variables that result in a minor improvement, but it is unlikely there will be a significant improvement in prediction of fracture risk. Only two load cases were investigated, simulating stance and a fall. Finite element analysis using these load cases successfully identified 4 out of the 5 variables necessary to predict fracture risk. Expanding the study to explore more load cases may help to identify the contribution of AAM mode 5 to the prediction of fracture strength and hence, fracture risk. Finally, the logistic regression models have not been adjusted for age, body weight or other factors. However, previous studies have shown that this only marginally improves the prediction of femoral neck fracture risk (Bredbenner et al., 2014; Carballido-Gamio et al., 2019).

The findings of this study further demonstrate that bone density and cortex thickness are the primary determinants of femoral neck strength and fracture risk and bone size and morphology are secondary factors. Depending on the number of input variables, MLR identified between 6 and 17 variables which could be included in the development of a logistic regression model. Of the variables finite element identified as predictors of femoral neck strength, 4 contributed to the performance of the best fracture risk model. A reasonable level of performance can be achieved using just two AAM modes as input variable.

To address the question of whether we are reaching the limits of what can be extracted from an image, the findings of this study would suggest that we are. The power of ASMs and AAMs is their ability to capture all of the available information within the image. As has been shown, a substantial proportion of the modes do not contribute to femoral neck strength and hence, from a causality standpoint, the inclusion of these modes is unlikely to improve the prediction of fracture risk. There are many other external factors that cannot be accounted for with the limited information contained within an image, including comorbidities; activity levels; propensity to fall; type of fall etc. In the absence of this data, it appears that we are reaching the plateau of what can be achieved from an image alone.

Acknowledgements

This study was funded, in part, by the EPSRC Frontier Engineering Awards, MultiSim and MultiSim2 projects (Grant Reference Numbers: EP/K03877X/1 and EP/S032940/1), European Commission H2020 programme through the CompBioMed and CompBioMed2 Centres of Excellence and the SANO European Centre for Computational Medicine (Grants N. H2020-EINFRA-2015-1/675451, H2020-INFRAEDI-2018-1/823712 and H2020-WIDESPREAD-2018-01/857533).

References

Amin, S., Kopperdhal, D.L., Melton, L.J., Achenbach, S.J., Therneau, T.M., Riggs, B.L., Keaveny, T.M., Khosla, S., 2011. Association of hip strength estimates by finite-element analysis with fractures in women and men. J. Bone Miner. Res. 26, 1593–1600. https://doi.org/10.1002/jbmr.347

Baker-Lepain, J.C., Luker, K.R., Lynch, J.A., Parimi, N., Nevitt, M.C., Lane, N.E., 2011a. Active shape modeling of the hip in the prediction of incident hip fracture. J. Bone Miner. Res. 26, 468–474. https://doi.org/10.1002/jbmr.254

Baker-Lepain, J.C., Luker, K.R., Lynch, J.A., Parimi, N., Nevitt, M.C., Lane, N.E., 2011b. Active shape modeling of the hip in the prediction of incident hip fracture. J. Bone Miner. Res. 26, 468–474. https://doi.org/10.1002/jbmr.254

Black, D.M., Bouxsein, M.L., Marshall, L.M., Cummings, S.R., Lang, T.F., Cauley, J.A., Ensrud, K.E., Nielson, C.M., Orwoll, E.S., 2008. Proximal femoral structure and the prediction of hip fracture in men: A large prospective study using QCT. J. Bone Miner. Res. 23, 1326–1333. https://doi.org/10.1359/jbmr.080316

Bonaretti, S., Seiler, C., Boichon, C., Reyes, M., Büchler, P., 2014. Image-based vs. mesh-based statistical appearance models of the human femur: Implications for finite element simulations. Med. Eng. Phys. 36, 1626–1635. https://doi.org/10.1016/j.medengphy.2014.09.006

Bousson, V.D., Adams, J., Engelke, K., Aout, M., Cohen-Solal, M., Bergot, C., Haguenauer, D., Goldberg, D., Champion, K., Aksouh, R., Vicaut, E., Laredo, J.D., 2011. In vivo discrimination of hip fracture with quantitative computed tomography: Results from the prospective European Femur Fracture Study (EFFECT). J. Bone Miner. Res. 26, 881–893. https://doi.org/10.1002/jbmr.270

Bredbenner, T.L., Mason, R.L., Havill, L.M., Orwoll, E.S., Nicolella, D.P., 2014. Fracture risk predictions based on statistical shape and density modeling of the proximal femur. J.

Bone Miner. Res. 29, 2090–2100. https://doi.org/10.1002/jbmr.2241

Bryan, R., Mohan, P.S., Hopkins, A., Galloway, F., Taylor, M., Nair, P.B., 2010. Statistical Modelling of the Whole Human Femur Incorporating Geometric and Material Properties. Med. Eng. Phys. 32, 57–65. https://doi.org/10.1016/j.medengphy.2009.10.008

Carballido-Gamio, J., Yu, A., Wang, L., Su, Y., Burghardt, A.J., Lang, T.F., Cheng, X., 2019a. Hip Fracture Discrimination Based on Statistical Multi-parametric Modeling (SMPM). Ann. Biomed. Eng. 47, 2199–2212. https://doi.org/10.1007/s10439-019-02298-x

Carballido-Gamio, J., Yu, A., Wang, L., Su, Y., Burghardt, A.J., Lang, T.F., Cheng, X., 2019b. Hip Fracture Discrimination Based on Statistical Multi-parametric Modeling (SMPM). Ann. Biomed. Eng. https://doi.org/10.1007/s10439-019-02298-x

Cheng, X., Li, J., Lu, Y., Keyak, J., Lang, T., 2007. Proximal femoral density and geometry measurements by quantitative computed tomography: Association with hip fracture. Bone 40, 169–174. https://doi.org/10.1016/j.bone.2006.06.018

Cody, D.D., Divine, G.W., Nahigian, K., Kleerekoper, M., 2000a. Bone density distribution and gender dominate femoral neck fracture risk predictors. Skeletal Radiol. 29, 151–161. https://doi.org/10.1007/s002560050585

Cody, D.D., Hou, F.J., Divine, G.W., Fyhrie, D.P., 2000b. Femoral structure and stiffness in patients with femoral neck fracture. J. Orthop. Res. 18, 443–448. https://doi.org/10.1002/jor.1100180317

Crabtree, N.J., Kroger, H., Martin, A., Pols, H.A.P., Lorenc, R., Nijs, J., Stepan, J.J., Falch, J.A., Miazgowski, T., Grazio, S., Raptou, P., Adams, J., Collings, A., Khaw, K.T., Rushton, N., Lunt, M., Dixon, A.K., Reeve, J., 2002. Improving risk assessment: Hip geometry, bone mineral distribution and bone strength in hip fracture cases and controls. The EPOS study. Osteoporos. Int. 13, 48–54. https://doi.org/10.1007/s198-002-8337-y

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, 27–38. https://doi.org/10.1016/j.bone.2012.09.006

Daruwalla, Z.J., Courtis, P., Fitzpatrick, C., Fitzpatrick, D., Mullett, H., 2010. An application of principal component analysis to the clavicle and clavicle fixation devices. J. Orthop. Surg. Res. 5, 1–8. https://doi.org/10.1186/1749-799X-5-21

Draper, C.E., Fredericson, M., Gold, G.E., Besier, T.F., Delp, S.L., Beaupre, G.S., Quon, A., 2012. Patients with patellofemoral pain exhibit elevated bone metabolic activity at the patellofemoral joint. J. Orthop. Res. 30, 209–213. https://doi.org/10.1002/jor.21523

Frost, S.A., Nguyen, N.D., Center, J.R., Eisman, J.A., Nguyen, T. V., 2013. Excess mortality attributable to hip-fracture: A relative survival analysis. Bone 56, 23–29. https://doi.org/10.1016/j.bone.2013.05.006

Gnudi, S., Ripamonti, C., Lisi, L., Fini, M., Giardino, R., Giavaresi, G., 2002. Proximal femur geometry to detect and distinguish femoral neck fractures from trochanteric fractures in postmenopausal women. Osteoporos. Int. 13, 69–73. https://doi.org/10.1007/s198-002-8340-2

Goodyear, S.R., Barr, R.J., McCloskey, E., Alesci, S., Aspden, R.M., Reid, D.M., Gregory, J.S., 2013. Can we improve the prediction of hip fracture by assessing bone structure using shape and appearance modelling? Bone 53, 188–193. https://doi.org/10.1016/j.bone.2012.11.042

Grassi, L., Hraiech, N., Schileo, E., Ansaloni, M., Rochette, M., Viceconti, M., 2011. Evaluation of the generality and accuracy of a new mesh morphing procedure for the human femur. Med. Eng. Phys. 33, 112–120. https://doi.org/10.1016/j.medengphy.2010.09.014

Gregory, J.S., Aspden, R.M., 2008. Femoral geometry as a risk factor for osteoporotic hip fracture in men and women. Med. Eng. Phys. 30, 1275–1286. https://doi.org/10.1016/j.medengphy.2008.09.002

Gregory, J.S., Testi, D., Stewart, A., Undrill, P.E., Reid, D.M., Aspden, R.M., 2004. A method for assessment of the shape of the proximal femur and its relationship to osteoporotic hip fracture. Osteoporos. Int. 15, 5–11. https://doi.org/10.1007/s00198-003-1451-y

Griffith, J.F., Genant, H.K., 2008. Bone mass and architecture determination: state of the art. Best Pract. Res. Clin. Endocrinol. Metab. 22, 737–764. https://doi.org/10.1016/j.beem.2008.07.003

Hambli, R., Allaoui, S., 2013. A robust 3D finite element simulation of human proximal femur progressive fracture under stance load with experimental validation. Ann. Biomed. Eng. 41, 2515–2527. https://doi.org/10.1007/s10439-013-0864-9

Leslie, W.D., Pahlavan, P.S., Tsang, J.F., Lix, L.M., 2009. Prediction of hip and other osteoporotic fractures from hip geometry in a large clinical cohort. Osteoporos. Int. 20, 1767–1774. https://doi.org/10.1007/s00198-009-0874-5

Li, G.W., Chang, S.X., Xu, Z., Chen, Y., Bao, H., Shi, X., 2013. Prediction of hip osteoporotic fractures from composite indices of femoral neck strength. Skeletal Radiol. 42, 195–201. https://doi.org/10.1007/s00256-012-1473-7

Lu, R.-S., Dennison, E., Denison, H., Cooper, C., Taylor, M., Bottema, M.J., 2017. Texture analysis based on Gabor filters improves the estimate of bone fracture risk from DXA images. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 1163, 1–12. https://doi.org/10.1080/21681163.2016.1271726

Michelotti, J., Clark, J., 1999. Femoral neck length and hip fracture risk. J. Bone Miner. Res. 14, 1714–1720. https://doi.org/10.1359/jbmr.1999.14.10.1714

Partanen, J., Jämsä, T., Jalovaara, P., 2001. Influence of the upper femur and pelvic geometry on the risk and type of hip fractures. J. Bone Miner. Res. 16, 1540–6. https://doi.org/10.1359/jbmr.2001.16.8.1540

Pulkkinen, P., Partanen, J., Jalovaara, P., Jämsä, T., 2004. Combination of bone mineral density and upper femur geometry improves the prediction of hip fracture. Osteoporos. Int. 15, 274–280. https://doi.org/10.1007/s00198-003-1556-3

Qasim, M., Farinella, G., Zhang, J., Li, X., Yang, L., Eastell, R., Viceconti, M., 2016. Patient-specific finite element estimated femur strength as a predictor of the risk of hip fracture: the effect of methodological determinants. Osteoporos. Int. 27, 2815–2822. https://doi.org/10.1007/s00198-016-3597-4

S, B., C, S., C, B., P, B., M, R., 2011. Mesh-based vs. Image-based Statistical Appearance Model of the Human Femur: a Preliminary Comparison Study for the Creation of Finite Element Meshes. Mesh Process. Med. Image Anal. https://doi.org/978-3-642-15948-0

Sarkalkan, N., Weinans, H., Zadpoor, A.A., 2014. Statistical shape and appearance models of bones. Bone 60, 129–140. https://doi.org/10.1016/j.bone.2013.12.006

Schileo, E., Ara, E.D., Taddei, F., Malandrino, A., Schotkamp, T., Ã, M.B., Viceconti, M., 2008a. An accurate estimation of bone density improves the accuracy of subject-specific finite element models 41, 2483–2491. https://doi.org/10.1016/j.jbiomech.2008.05.017

Schileo, E., Balistreri, L., Grassi, L., Cristofolini, L., Taddei, F., 2014. To what extent can linear finite element models of human femora predict failure under stance and fall loading configurations? J. Biomech. 47, 3531–3538. https://doi.org/10.1016/j.jbiomech.2014.08.024

Schileo, E., Taddei, F., Cristofolini, L., Viceconti, M., 2008b. Subject-specific finite element models implementing a maximum principal strain criterion are able to estimate failure risk and fracture location on human femurs tested in vitro 41, 356–367. https://doi.org/10.1016/j.jbiomech.2007.09.009

Taylor, M., Perilli, E., Martelli, S., 2017. Development of a surrogate model based on patient weight, bone mass and geometry to predict femoral neck strains and fracture loads. J. Biomech. 55, 121–127. https://doi.org/10.1016/j.jbiomech.2017.02.022

Treece, G.M., Gee, A.H., Tonkin, C., Ewing, S.K., Cawthon, P.M., Black, D.M., Poole, K.E.S., 2015. Predicting Hip Fracture Type with Cortical Bone Mapping (CBM) in the Osteoporotic Fractures in Men (MrOS) Study. J. Bone Miner. Res. 30, 2067–2077. https://doi.org/10.1002/jbmr.2552

van Smeden, M., Moons, K.G.M., de Groot, J.A.H., Collins, G.S., Altman, D.G., Eijkemans, M.J.C., Reitsma, J.B., 2019. Sample size for binary logistic prediction models: Beyond events per variable criteria. Stat. Methods Med. Res. 28, 2455–2474. https://doi.org/10.1177/0962280218784726

Whitmarsh, T., Fritscher, K.D., Humbert, L., Del Rio Barquero, L.M., Roth, T., Kammerlander, C., Blauth, M., Schubert, R., Frangi, A.F., 2011. A statistical model of shape and bone mineral density distribution of the proximal femur for fracture risk assessment. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 6892 LNCS, 393–400. https://doi.org/10.1007/978-3-642-23629-7_48

Whitmarsh, T., Fritscher, K.D., Humbert, L., Luis, M., Barquero, R., Roth, T., Kammerlander, C., Blauth, M., Schubert, R., Frangi, A.F., 2012. Hip fracture discrimination from dual-energy X-ray absorptiometry by statistical model registration. Bone 6–11. https://doi.org/10.1016/j.bone.2012.08.114

Yang, L., Palermo, L., Black, D.M., Eastell, R., 2014. Prediction of incident hip fracture with the estimated femoral strength by finite element analysis of DXA scans in the study of osteoporotic fractures. J. Bone Miner. Res. 29, 2594–2600. https://doi.org/10.1002/jbmr.2291

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage 31, 1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015

Zysset, P., Pahr, D., Engelke, K., Genant, H.K., McClung, M.R., Kendler, D.L., Recknor, C., Kinzl, M., Schwiedrzik, J., Museyko, O., Wang, A., Libanati, C., 2015. Comparison of proximal femur and vertebral body strength improvements in the FREEDOM trial using an alternative finite element methodology. Bone 81, 122–130. https://doi.org/10.1016/j.bone.2015.06.025