Predicting hip fracture type with cortical bone mapping (CBM) in the Osteoporotic Fractures in Men (MrOS) study

G.M. Treece, A.H. Gee, C. Tonkin, S.K. Ewing,
P.M. Cawthon, D. Black and K.E.S. Poole,
for the Osteoporotic Fractures in Men (MrOS) study

CUED/F-INFENG/TR 695

January 2015

Cambridge University Engineering Department
Trumpington Street
Cambridge CB2 1PZ
England

Corresponding e-mail: gmt11@eng.cam.ac.uk

Abstract

Hip fracture risk is known to be related to material properties of the proximal femur, but prospective studies investigating fracture prediction which add richer quantitative computed tomography (QCT) measures to more traditional dual energy X-ray (DXA)-based methods have shown only limited improvement. Though it is understood that fracture types have distinct relationships to predictors, few studies have sub-divided fracture into types, since this necessitates regional measurements and an increased number of fracture cases. We here report on a study which makes use of cortical bone mapping (CBM) to accurately assess, with no prior anatomical presumptions, the distribution of femoral properties related to fracture type. CBM makes use of QCT data to measure the cortical and trabecular properties, and is accurate even for very thin cortices below the imaging resolution. We use a prospective case-cohort of older men: the osteoporotic fractures in men (MrOS) study, in which we analyse 99 fracture cases (44 trochanteric and 55 femoral neck, by centralised review of radiology reports) compared to a cohort of 308, randomly selected from 5,994 men over 65. To our knowledge, this is the largest QCT-based prospective hip fracture study to date, as well as the first to incorporate CBM analysis into fracture prediction. We demonstrate that both cortical mass surface density, and endocortical trabecular BMD, show significant difference in fracture cases vs. cohort, in regions which are appropriate to fracture type. We incorporate these regions into predictive models using Cox proportional hazards regression (with accommodations for the case-cohort sampling) to estimate hazard ratios, and logistic regression to estimate the area under the receiver operating characteristic curve (AUC). Adding CBM to DXA-based BMD leads to a small but significant ($p < 0.005$) improvement in model prediction for any fracture, with AUC increasing from 0.78 to 0.79, when assessed using leave-one-out cross-validation. For specific fracture types, the improvement is more significant ($p < 0.0001$) and more marked, with AUC improving from 0.71 to 0.77 (trochanteric fractures) and 0.76 to 0.82 (femoral neck fractures). In contrast, adding DXA-based BMD to a predictive model already containing CBM regions does not result in any significant improvement.

1 Introduction

Hip fractures are the most common cause of acute orthopaedic hospital admission in older people (Parker and Johansen, 2006). Bone mineral density (BMD) is an important imaging marker which contributes to an individual's fracture risk, and is usually measured as areal BMD from dual energy X-ray (DXA), though volumetric BMD from quantitative computed tomography (QCT) is a viable surrogate. Although BMD is specific (Johnell et al., 2005; Kanis et al., 2008) it lacks sensitivity (Kanis et al., 2008; Kaptoge et al., 2008; Sanders et al., 2006), missing the majority who go on to fracture. There is now growing evidence that focal, structural weaknesses may predispose a hip to fracture (Mayhew et al., 2005; Poole et al., 2010; de Bakker et al., 2009). The distribution of both trabecular and cortical bone is critical in determining a femur's resistance to fracture (Holzer et al., 2009; Verhulp et al., 2008; Poole et al., 2012; Kopperdahl et al., 2014). Drug treatment and exercise regimes targeted at reducing fracture risk result in changes which are focused in particular regions rather than dispersed over the whole bone (Carballido-Gamio et al., 2013; Poole et al., 2011).

Cortical bone mapping (CBM) is a technique which allows measurement of cortical parameters from clinical QCT data, which has been presented and thoroughly validated in previous papers (Treece et al., 2010, 2012; Treece and Gee, 2015). Using CBM, it is possible to accurately measure cortical thickness (CTh, in mm), cortical mass surface density (CM, in mg/cm^2, the cortical mass per unit cortical surface area), cortical bone mineral density (CBMD, in mg/cm^3) and endocortical trabecular bone mineral density (ECTD, in mg/cm^3, the average density in the trabecular compartment close to the cortex). This is distinct from alternative measures in two important ways. Firstly, it provides a genuinely local measurement of cortical properties: these are independently measured at many thousand locations distributed over the surface of the proximal femur. Particular cortical regions which may be related to particular types of fracture are found from a statistical analysis of these measurements, rather than pre-grouping them into anatomical regions and testing

each region in turn for significance. Secondly, measurements are accurate even when the cortex is much thinner than the extent of the CT imaging blur. In contrast, in most alternative techniques, for thin cortices (< 3 mm) the measured 'thickness' when based on thresholds (either locally variable or fixed), whether or not refined by morphological operators, is often closer to the width of the CT imaging blur, and the measured 'density' (i.e. the recorded cortical CT value) is in fact closer to cortical mass surface density, rather than true cortical BMD.

Here we apply CBM to a prospective study of male subjects, some of whom have gone on to fracture their femur. QCT data from this study has been analysed before using both DXA-derived and QCT-derived regional bone quantities (Lang et al., 2014), though with a different randomly selected cohort. This present paper is the first application of CBM to a prospective fracture-risk study. It also reports on a much larger number of cases than previously published results: 99 rather than the previously reported 40. As a result, we consider these cases in two separate fracture groups (trochanteric and neck fractures) for the first time in a male, prospective cohort, although there have been previous discriminatory analyses of fracture type in retrospective female cohorts (Bousson et al., 2011; Poole et al., 2012). Using CBM we can show the different patterns of relationship between fracture type and cortical distributions and hence start to consider the assessment of not only *whether* a particular hip is at risk of fracture, but *where* that fracture is most likely to occur. This is of particular importance when considering that both exercise and drug regimes improve bone in specific areas: the eventual goal is to be able to match a particular individual at risk of fracture to the regime which will best ameliorate their specific fracture risks.

2 Method

2.1 Study design

The Osteoporotic Fractures in Men (MrOS) study (Orwoll et al., 2005; Blank et al., 2005) recruited 5994 men in the US from March 2000 until April 2002. Eligible subjects from six clinical sites were 65 or older, able to walk without assistance, and had not had bilateral hip replacement surgery. Various measurements at baseline included areal BMD at the hip by DXA (QDR 4500W; Hologic Inc., Bedford, MA, USA) and subject weight and height. Information about fractures was ascertained through questionnaires returned by the subjects every 4 months, however any reported hip fractures were validated by centralised physician review of the radiology report or radiographs. Fracture type was classified as either 'femoral neck', 'sub-trochanteric', 'inter-trochanteric', or 'other'.

A subset of 3684 participants (61% of the MrOS cohort) also had QCT of the hip at baseline. This subset consisted of approximately the first 650 subjects at each of the six centres, in addition to all those from minority backgrounds. QCT scans were performed on a variety of machines, but according to a protocol delivering 512 x 512 pixel slices at 3 mm separation, covering the femoral head to 3.5 cm below the lesser trochanter. All scans included a calibration phantom (Calcium Hydroxyapatite at 150, 75 and 0 mg/cm^3; Image Analysis Inc., Columbia, KY, USA) for converting from Hounsfield Units (HU) to BMD. 3572 of these scans were successfully transferred to the study centre for processing, and QCT-derived volumetric BMD were calculated from a further subset of 3358. Scanning and selection procedures, baseline statistics compared to the entire cohort and details of the QCT-derived measurements, have been described fully elsewhere (Marshall et al., 2006).

The analysis of this study follows a case-cohort design (Barlow et al., 1999). The MrOS study team took the 3572 participants with hip QCT scans and removed any with previous hip fracture or hip replacement leaving 3515. The cases were all 104 participants from this set with hip fractures as of February 2012. Three times this number (312) were taken as a random sample from the whole set to form the cohort, of which 10 had fractured and hence were also in the cases, leaving a total sample size of 406. Eight participants were

excluded from the analysis due to problems with the QCT data (two incomplete scans, four severely misregistered scans, one incomplete phantom scan and one extremely noisy reconstruction), leaving 398 in the final analysis. Of these 398, 308 were in the cohort and 99 were fracture cases. Fracture types were loosely grouped into two, giving 44 'trochanteric' (37 inter-trochanteric, 4 sub-trochanteric and 3 other) and 55 'neck' fractures.

An additional set of 38 QCT scans was also supplied, for analysing CBM reproducibility. This consisted of two independent scans at visits three months apart from each of 19 men.

2.2 Cortical bone mapping

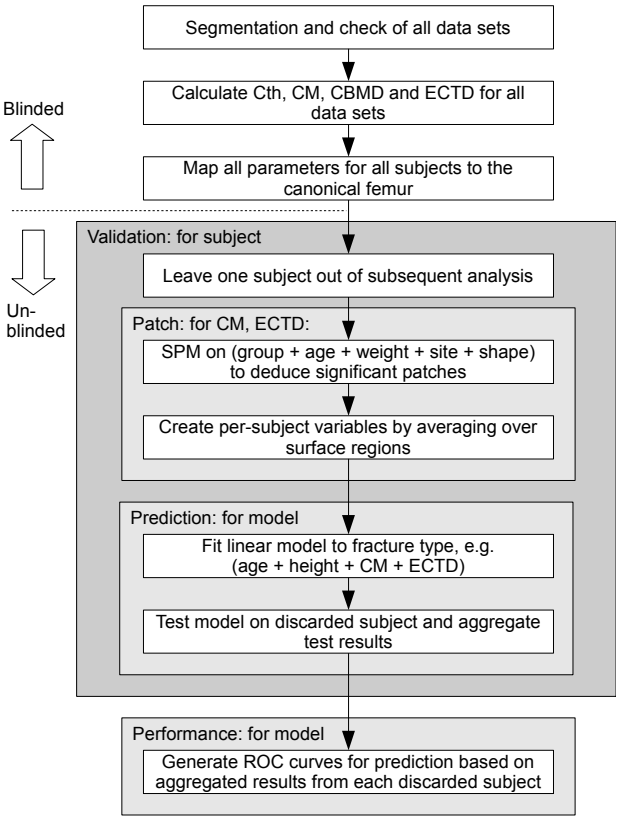


Figure 1: Process for creating and validating predictive linear models based on CBM parameters.

The first step in CBM is an approximate segmentation of each proximal femur from the QCT data for each subject. This segmentation is performed using in-house developed software Stradwin[1], and results in a triangulated surface mesh with between 5 and 15 thousand vertices distributed uniformly over the proximal femoral surface. This is the only manual step in the CBM process. These vertices are used to establish the location and directions at which more accurate, fully automated, measurements of cortical parameters (including more precise surface location) are made. Hence the manual segmentation is only required to be accurate to within 2 mm of the real femoral surface.

CBM measurements are then made at each vertex, again using in-house Stradwin software. This is followed by non-rigid registration of each femur from each subject to a canonical femur surface and mapping

[1]Free to download from http://mi.eng.cam.ac.uk/~rwp/stradwin

of the CBM data to that surface, using another in-house package wxRegSurf[2]. Statistical parametric mapping (SPM) is then applied, using the SurfStat (Worsley et al., 2009) package, to identify patches of differences in any of these parameters associated with each fracture type (trochanteric or neck). Where both hips were available (in 358 out of 398 subjects) results were averaged over the left and right femur, otherwise data was used from whichever hip was available. We have previously demonstrated that (at least in older women), although there are some differences between left and right femurs from the same subject, these are relatively small and do not appear in areas associated with fracture (Turmezei et al., 2012).

The general linear model variables used in SPM to discover patches for each of CTh, CM, CBMD or ECTD, were group, fracture, age, weight, clinical site, and shape. Group was bi-modal (case or cohort) whereas fracture was tri-modal (none, trochanteric, neck). Clinical site was used to model any variations in scanner calibration and regional demographics. To model shape we used the five most significant shape modes, not including size. These are required as confounding variables to account for any systematic mis-registration during the mapping process (Gee and Treece, 2014). For each cortical parameter, we identified two patches (i.e. the regions where there was significant contrast between the cohort and the specific fracture type): one patch for trochanteric fractures, and another for femoral neck fractures. Having identified these patches, the cortical data was averaged within the corresponding patch. This gives single values per patch, cortical parameter and subject. As a result of a prior retrospective study on a different data set (Poole et al., 2012, 2013), it was decided in advance of this study only to use the CM and ECTD patches for analysis of hazard and odds ratios, giving four parameters in all to carry forward into the next phase of statistical analysis. CTh and CBMD patches are nevertheless also reported for completeness.

All stages previous to SPM analysis were carried out blinded to case and fracture status, and all the analysis was according to a study plan which had been previously submitted to, and approved by, the MrOS study group. The process is summarised in the top section of Fig. 1.

2.3 Predictive models

Parameters of interest were age, height, DXA-based areal BMD (total hip, ThBMD, and femoral neck, FnBMD), QCT-based volumetric BMD (total hip and femoral neck) as well as the CM and ECTD patches based on CBM analysis for trochanteric and neck fracture. Outcomes were either any fracture, or trochanteric fracture, or femoral neck fracture. QCT-based BMD measurements were provided for 384 of the 398 subjects.

Modelling of the time to first incident hip fracture was performed using Cox proportional hazards regression with the Barlow weighting method and robust variance estimation, necessitated by the case-cohort design. Hazard ratios were estimated for hip fracture for a 1 SD change in each parameter, adjusted for age, height and clinical site. According to Barlow's method for weighting study participants in the pseudo-likelihood, cases were weighted by 1 and sub-cohort controls were weighted by the inverse of the sampling fraction α, where $\alpha = 308/3515 = 8.8\%$. Case-cohort analyses for hazard ratios were conducted in SAS version 9.1 (SAS Institute, Cary, NC, USA).

In addition, we examined the ability of each group of parameters to predict ten-year fracture incidence, by performing either binomial (any fracture) or trichotomous multinomial (specific fracture type) logistic regression for various models, allowing for age, height and clinical site. Logistic models were created using MATLAB R2014a (The MathWorks, Inc., Natick, MA, USA). We calculated odds ratios for each parameter individually. In addition, we investigated predictive models including no imaging parameters, DXA-derived BMD values only, QCT-derived BMD values only, CBM patches only, DXA and CBM patches, and QCT and CBM patches. Receiver operating characteristic (ROC) curves were calculated with leave-one-out cross-validation. Since the extent of the CBM patches were also dependent on the data, SPM for calculating the

[2]Free to download from http://mi.eng.cam.ac.uk/~ahg/wxRegSurf

Table 1: Baseline values for the study. Significant difference compared to no fracture is given for $^{\dagger}p < 0.05$, $^{\ddagger}p < 0.005$ or $^{\dagger\dagger}p < 0.0001$. *QCT values were given for a subset of the data, with n = 288, 96, 43 and 53, respectively.

quantity	no fracture (n = 299)* mean (SD)	fractures all (n = 99)* mean (SD)	fractures trochanteric (n = 44)* mean (SD)	fractures neck (n = 55)* mean (SD)
age (years)	73.4 (5.7)	76.8†† (5.8)	75.5† (5.7)	77.8†† (5.7)
weight (kg)	84.6 (14.1)	80.7† (13.0)	78.8† (11.4)	82.3 (14.1)
height (cm)	174.4 (7.3)	174.3 (6.3)	174.1 (6.0)	174.4 (6.5)
DXA ThBMD (g/cm^2)	0.956 (0.132)	0.837†† (0.131)	0.827†† (0.120)	0.844†† (0.140)
DXA FnBMD (g/cm^2)	0.786 (0.119)	0.675†† (0.105)	0.679†† (0.107)	0.672†† (0.104)
QCT ThBMD (g/cm^3)	0.278 (0.049)	0.236†† (0.049)	0.233†† (0.047)	0.238†† (0.051)
QCT FnBMD (g/cm^3)	0.286 (0.057)	0.242†† (0.053)	0.244†† (0.052)	0.240†† (0.053)

location of the patches was also included in the leave-one-out process. Hence, for each subject, fracture probability for each type of fracture was predicted from a model based on the remaining subjects, and also (where contained in the model) from CBM patches based on the remaining subjects. This process is summarised in the lower section of Fig. 1.

2.4 Precision

Precision of the CBM parameters was assessed with an additional set of QCT data which consisted of 38 scans of 19 subjects, with two independent scans of each subject. CBM parameters were derived from these scans in the same way as for the other analyses above and mapped to the canonical femur model. Each scan was processed blinded to subject. The variance of the measurement error was assumed to be half the variance of the measurement differences from each pair of scans of the same subject[3]. The measurement standard deviation (SD) was calculated for each CBM parameter, at each location on the canonical femur, since the precision was expected to vary with anatomical location. We also calculated the SD of the CBM parameters after averaging these over the patches found in the SPM analysis.

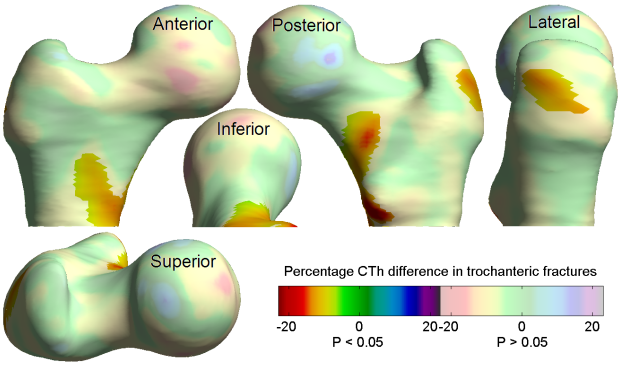
3 Results

Baseline data for no fracture and for the fracture cases is given in Table 1.

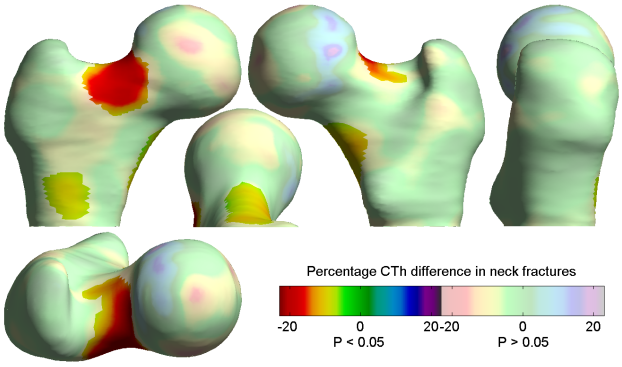
Figure 2 shows the result of running the SPM analysis over all subjects[4]. Results are shown for all CBM parameters, though only CM and ECTD were subsequently used in calculating hazard ratios and in the predictive models. These results are shown as percentage difference with respect to the mean value at each point. This is slightly problematic for ECTD, since the average trabecular bone mineral density can be very near to zero, and hence a small absolute change may lead to a large percentage change. Nevertheless,

[3]Assuming the measurement error to be normally distributed with zero mean, the difference between two samples of the same measurement has twice the variance of each of the samples alone.

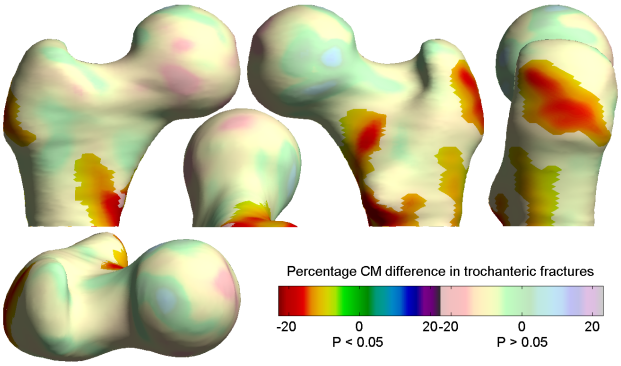
[4]The patches used in the statistical models are very slightly different, since each of these leaves one subject out, however the effect of this variation is almost invisible in this figure.



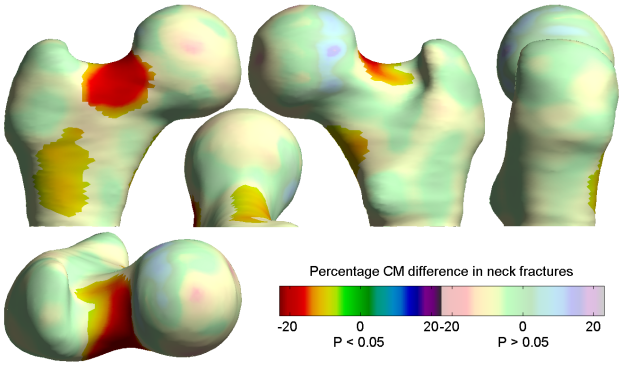
(a) CTh trochanteric patch



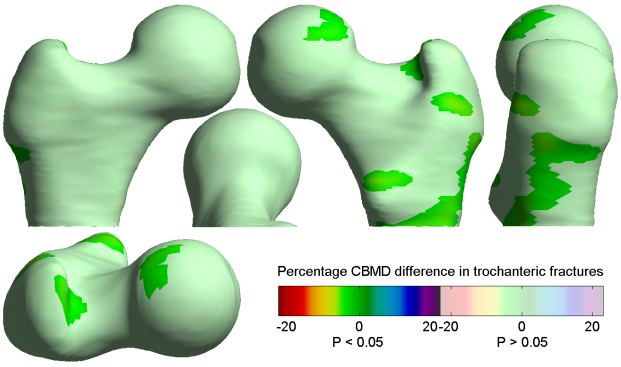
(b) CTh neck patch



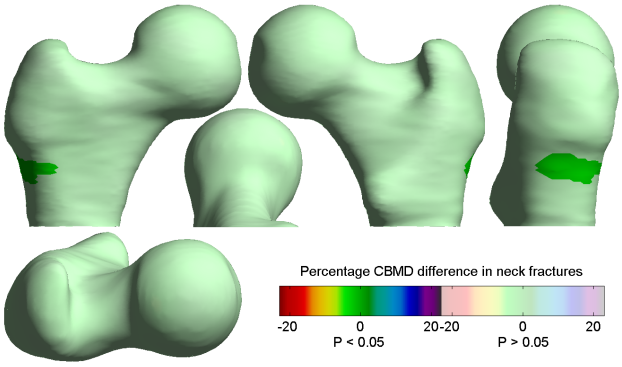
(c) CM trochanteric patch



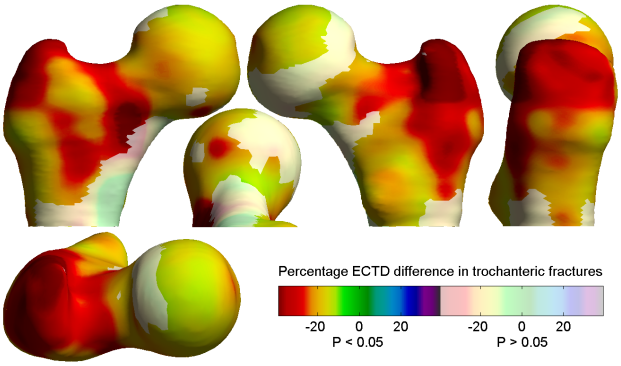
(d) CM neck patch



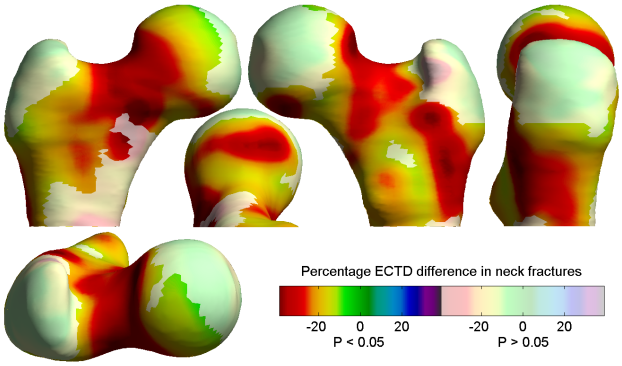
(e) CBMD trochanteric patch



(f) CBMD neck patch



(g) ECTD trochanteric patch



(h) ECTD neck patch

Figure 2: CBM effects related to each fracture type, shown as percentage differences between fracture cases and cohort. Paler colours indicate no significant relationship between fracture type and the CBM quantity.

Table 2: Hazard ratios, for each 5 year increase (age), 1 SD increase (height) or 1 SD decrease (all others) in the quantity, calculated from an unadjusted model (for age and height) or age + height + site + quantity (for all others). Significance is given for $^{\dagger}p < 0.05$, $^{\ddagger}p < 0.005$ or $^{\dagger\dagger}p < 0.0001$.

quantity	all fractures		trochanteric fracture		neck fracture	
	hazard	95% CI	hazard	95% CI	hazard	95% CI
age	$1.81^{\dagger\dagger}$	(1.47, 2.23)	$1.52^{\dagger}$	(1.15, 2.02)	$2.10^{\dagger\dagger}$	(1.62, 2.73)
height	0.97	(0.79, 1.19)	0.94	(0.71, 1.23)	0.99	(0.76, 1.28)
DXA ThBMD	$2.86^{\dagger\dagger}$	(1.98, 4.12)	$3.52^{\dagger\dagger}$	(2.24, 5.54)	$2.56^{\dagger\dagger}$	(1.63, 4.01
DXA FnBMD	$3.65^{\dagger\dagger}$	(2.30, 5.78)	$3.86^{\dagger\dagger}$	(2.07, 7.22)	$3.70^{\dagger\dagger}$	(2.05, 6.69)
QCT ThBMD	$3.28^{\dagger\dagger}$	(2.11, 5.11)	$3.94^{\dagger\dagger}$	(2.22, 7.01)	$2.94^{\ddagger}$	(1.70, 5.09)
QCT FnBMD	$2.80^{\dagger\dagger}$	(1.83, 4.28)	$2.88^{\ddagger}$	(1.71, 4.82)	$2.82^{\ddagger}$	(1.63, 4.88)
CM trochanter patch	$2.34^{\dagger\dagger}$	(1.67, 3.28)	$3.45^{\dagger\dagger}$	(2.13, 5.58)	$1.80^{\ddagger}$	(1.24, 2.62)
CM neck patch	$3.00^{\dagger\dagger}$	(2.06, 4.38)	$2.80^{\dagger\dagger}$	(1.80, 4.35)	$3.32^{\dagger\dagger}$	(1.98, 5.58)
ECTD trochanter patch	$3.70^{\dagger\dagger}$	(2.39, 5.72)	$4.63^{\dagger\dagger}$	(2.59, 8.30)	$3.25^{\dagger\dagger}$	(1.86, 5.66)
ECTD neck patch	$4.87^{\dagger\dagger}$	(2.91, 8.14)	$4.52^{\dagger\dagger}$	(2.51, 8.13)	$5.36^{\dagger\dagger}$	(2.57, 11.18)

we prefer to visualise percentage changes, since they are more closely linked to the statistically significant regions: a larger absolute change is generally required in an area with larger mean value to attain significance.

Hazard ratios (from Cox regression) and odds ratios (from binomial and multinomial logistic regression) are shown in Tables 2 and 3 respectively. These contain ratios for fractures of any type, as well as for specific (trochanteric or neck) fracture types. Predictive models were assessed using leave-one-out cross validation, and subsequent ROC curves are contained in Fig. 3. Areas under these curves (AUCs) were calculated, with confidence bounds determined by the bootstrap technique, and these are contained in Table 4. The models were assessed for significant difference by comparing the deviances of nested models, using a χ^2 test.

The precision of CBM parameters was also assessed and the estimated measurement SD given in Table 5. This is reported as an absolute value, percentage of the mean value, and also as a percentage of the SD of that parameter in the full cohort. This latter value gives a better indication of how useful this measurement is as a predictor. The minimum detectable difference for a single subject, at 95% confidence, is approximately twice this value. The distribution of measurement error is also given in Fig. 4.

4 Discussion

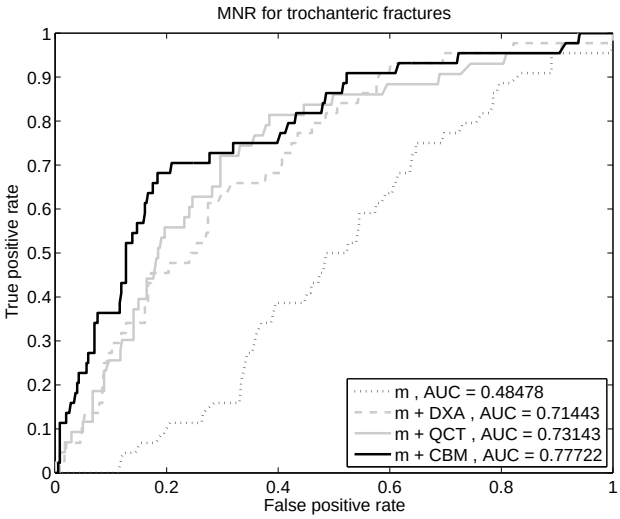
Looking at the baseline figures in Table 1, the results for DXA- and QCT-based BMD largely confirm what has already been widely reported: that all of these measures are clearly predictive of fracture, though it is interesting to note that this is the case for either class of fractures, irrespective of whether BMD is measured over total hip or femoral neck. Weight, height and age results agree well with a previous retrospective study on women (Bousson et al., 2011), confirming that age is a stronger predictor, as well as discriminator, of fracture than either weight or height. However, age shows much more significance for neck fracture than for trochanteric fracture, which may well be related to the particular regions associated with age-related bone loss (Marshall et al., 2006) and how these compare to those regions associated with each fracture type in Fig. 2. In contrast, reduced weight shows significance for trochanteric but not neck fractures, though at a

Table 3: Odds ratios for ten-year fracture incidence, for each 5 year increase (age), 1 SD increase (height) or 1 SD decrease (all others) in the quantity, calculated from a model of age + height + site (for age and height) or age + height + site + quantity (for all others). Significance is given for $^{\dagger}p < 0.05$, $^{\ddagger}p < 0.005$ or $^{\dagger\dagger}p < 0.0001$.

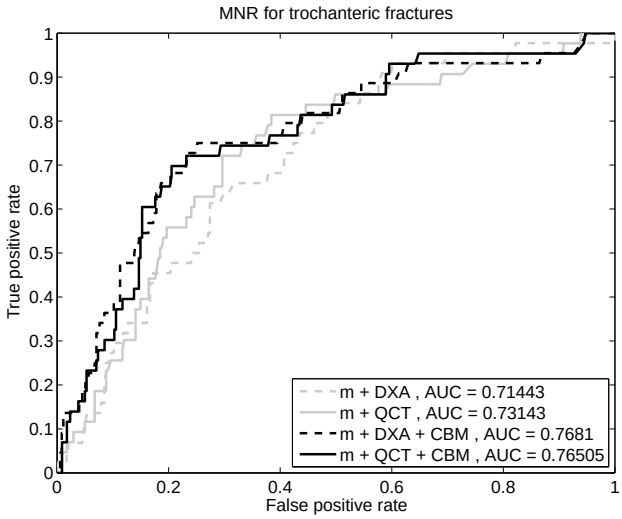
quantity	all fractures		trochanteric fracture		neck fracture	
	odds	95% CI	odds	95% CI	odds	95% CI
age	$1.75^{\dagger\dagger}$	(1.40, 2.18)	$1.41^{\dagger}$	(1.05, 1.89)	$2.11^{\dagger\dagger}$	(1.58, 2.81)
height	1.06	(0.83, 1.37)	1.00	(0.71, 1.40)	1.12	(0.82, 1.54)
DXA ThBMD	$2.45^{\dagger\dagger}$	(1.83, 3.29)	$2.85^{\dagger\dagger}$	(1.91, 4.24)	$2.18^{\dagger\dagger}$	(1.53, 3.10)
DXA FnBMD	$2.94^{\dagger\dagger}$	(2.10, 4.12)	$3.01^{\dagger\dagger}$	(1.93, 4.68)	$2.89^{\dagger\dagger}$	(1.91, 4.37)
QCT ThBMD	$2.67^{\dagger\dagger}$	(1.93, 3.69)	$3.18^{\dagger\dagger}$	(2.06, 4.91)	$2.32^{\dagger\dagger}$	(1.58, 3.40)
QCT FnBMD	$2.33^{\dagger\dagger}$	(1.70, 3.21)	$2.44^{\dagger\dagger}$	(1.61, 3.71)	$2.24^{\dagger\dagger}$	(1.51, 3.31)
CM trochanter patch	$2.20^{\dagger\dagger}$	(1.65, 2.94)	$3.20^{\dagger\dagger}$	(2.09, 4.90)	$1.71^{\ddagger}$	(1.23, 2.39)
CM neck patch	$2.47^{\dagger\dagger}$	(1.82, 3.33)	$2.38^{\dagger\dagger}$	(1.61, 3.50)	$2.54^{\dagger\dagger}$	(1.74, 3.69)
ECTD trochanter patch	$3.00^{\dagger\dagger}$	(2.17, 4.16)	$3.47^{\dagger\dagger}$	(2.22, 5.42)	$2.69^{\dagger\dagger}$	(1.82, 3.96)
ECTD neck patch	$3.33^{\dagger\dagger}$	(2.35, 4.72)	$3.07^{\dagger\dagger}$	(1.97, 4.78)	$3.59^{\dagger\dagger}$	(2.32, 5.55)

Table 4: Cross-validated AUCs for various predictive models, calculated using binomial (all fractures) or multinomial (specific fractures) logistic regression. The base model **m** includes age + height + site. Significance (based on the deviance DEV) is given for $^{\ddagger}p < 0.0001$ compared to **m**, $^{\dagger}p < 0.005$ compared to the same model without CBM parameters or $^{\dagger\dagger}p < 0.0001$ compared to the same model without CBM parameters.

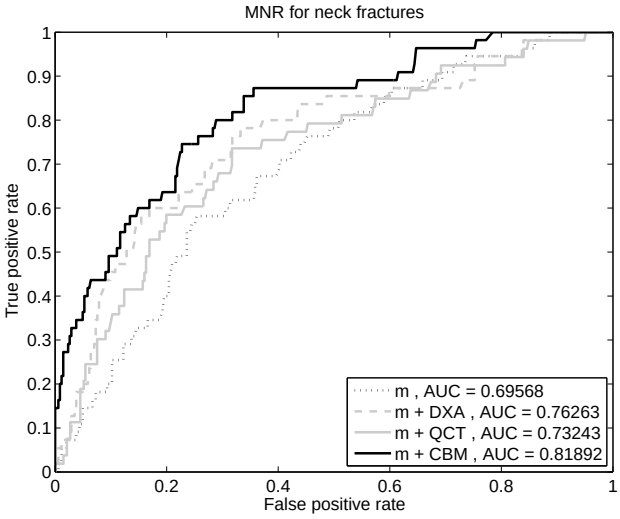
model	binomial			multinomial				
	all fractures			trochanteric fracture		neck fracture		
	AUC	95% CI	DEV	AUC	95% CI	AUC	95% CI	DEV
m	0.652	(0.59, 0.71)	415.1	0.485	(0.40, 0.56)	0.693	(0.62, 0.76)	545.1
m + DXA	0.782	(0.72, 0.83)	$354.7^{\ddagger}$	0.714	(0.63, 0.78)	0.761	(0.68, 0.82)	$482.4^{\ddagger}$
m + QCT	0.764	(0.70, 0.82)	$350.6^{\ddagger}$	0.731	(0.63, 0.79)	0.730	(0.64, 0.80)	$473.3^{\ddagger}$
m + CBM	0.787	(0.73, 0.84)	$343.7^{\ddagger}$	0.777	(0.68, 0.85)	0.818	(0.76, 0.87)	$433.1^{\ddagger}$
m + DXA + CBM	0.790	(0.73, 0.84)	$339.7^{\dagger}$	0.767	(0.68, 0.83)	0.816	(0.74, 0.87)	$428.3^{\dagger\dagger}$
m + QCT + CBM	0.794	(0.73, 0.85)	$318.9^{\dagger\dagger}$	0.766	(0.68, 0.84)	0.832	(0.77, 0.89)	$400.3^{\dagger\dagger}$



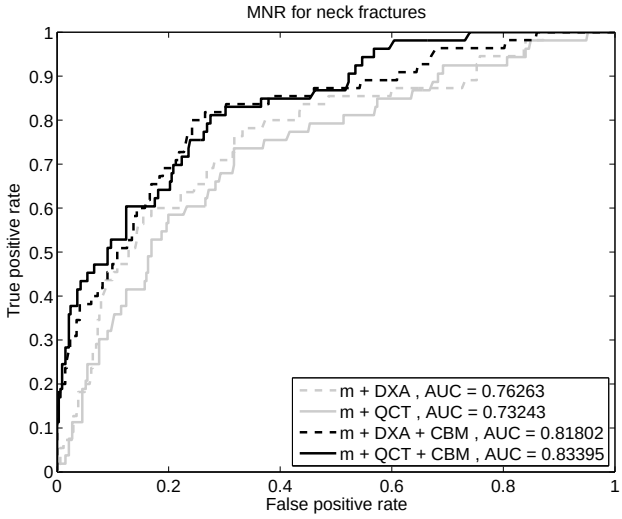
(a) trochanteric fracture on individual models



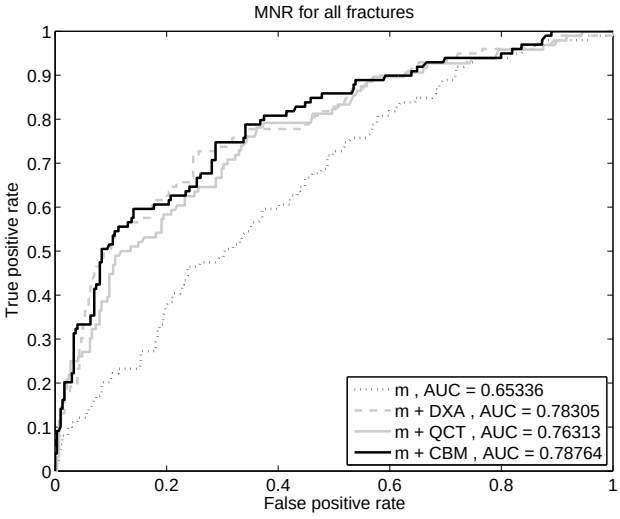
(b) trochanteric fracture on combined models



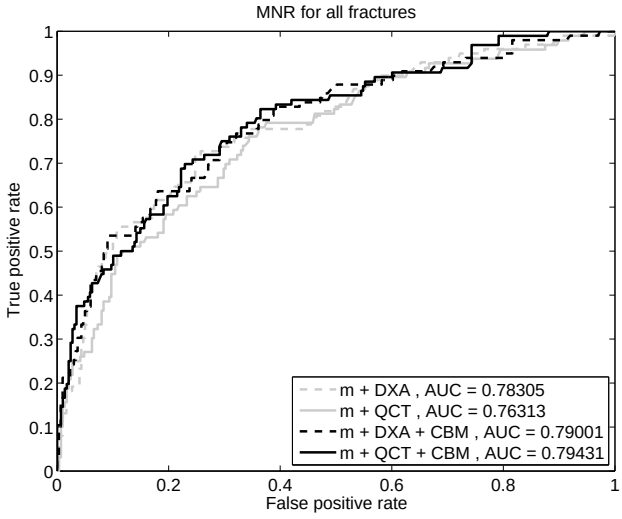
(c) neck fracture on individual models



(d) neck fracture on combined models



(e) any fracture on individual models

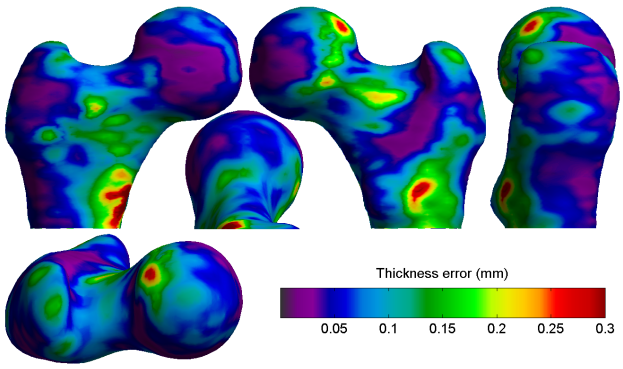


(f) any fracture on combined models

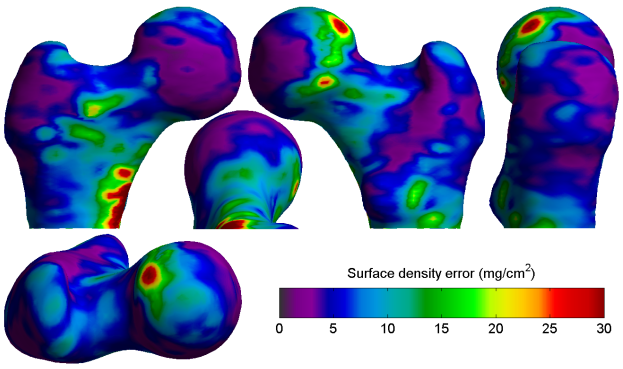
Figure 3: ROC curves for fracture prediction based on leave-one-out cross-validation of multinomial logistic regression models.

Table 5: Estimated precision (SD) of measurement repeatability. CBM precision is shown both for an individual measurement, and for all measurements aggregated within each of the trochanteric and neck patches. The right-hand column shows the precision as a percentage of the SD seen in the entire cohort.

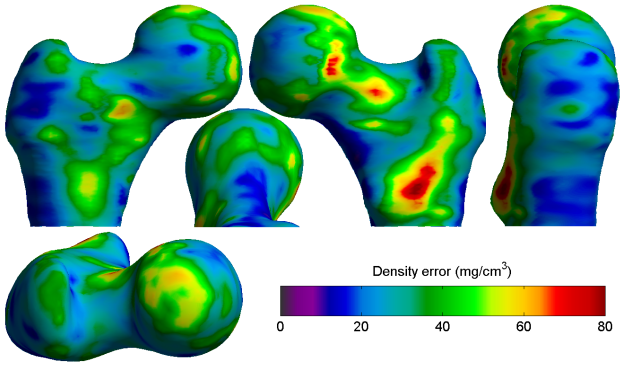
quantity	**precision**		
	absolute	**% of mean**	**% of SD in cohort**
CTh (mm)	0.099	6.22	9.8
CTh trochanter patch (mm)	0.033	1.06	2.97
CTh neck patch (mm)	0.027	1.07	2.41
CM (mg/cm^2)	9.14	5.20	7.27
CM trochanter patch (mg/cm^2)	3.98	1.32	2.46
CM neck patch (mg/cm^2)	3.08	1.15	2.20
CBMD (mg/cm^3)	34.7	3.17	42.3
CBMD trochanter patch (mg/cm^3)	17.5	1.63	26.6
CBMD neck patch (mg/cm^3)	19.3	1.76	26.5
ECTD (mg/cm^3)	15.0	8.81	21.2
ECTD trochanter patch (mg/cm^3)	2.47	1.49	3.72
ECTD neck patch (mg/cm^3)	2.63	1.55	4.20



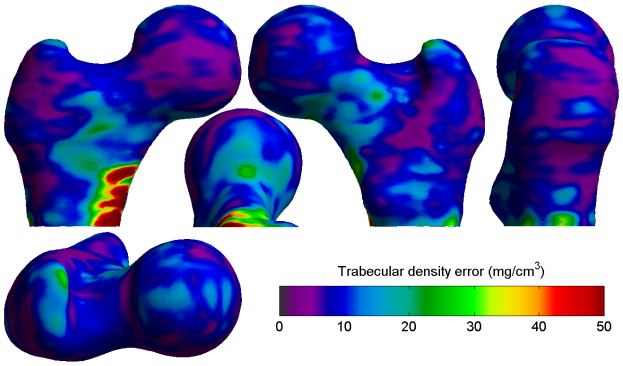
(a) CTh



(b) CM



(c) CBMD



(d) ECTD

Figure 4: Precision of local CBM measurements.

fairly low significance level.

4.1 Cortical bone mapping

The first step in assessing how CBM might contribute to fracture risk is to detect over which parts of the proximal femur each CBM parameter is associated with fracture, and this is shown in Fig. 2, for CTh, CM, CBMD and ECTD. The distribution and percentage changes in cortical thickness and cortical mass surface density show a striking similarity to those found by the same method in a previously published independent retrospective study in women (Poole et al., 2012, 2013). In both cases, particularly with CM, there is a clear patch at the supero-lateral side of the trochanter associated with trochanteric fracture, and an even clearer patch at the superior femoral neck associated with neck fracture. This might not be a surprising result in general; in fact it would be worrying if CBM parameters were not located in regions associated with that fracture type; but nevertheless it is striking that the shape and location of these patches is so consistent between studies.

Looking at CBMD distribution in Fig. 2(e) and (f), this does appear to be associated with trochanteric fracture, and in regions which are appropriate to that fracture type, however the percentage difference is much smaller between cases and cohort. When this is considered in light of the substantially poorer precision of this measurement as seen in Table 5 and Fig. 4, then it would seem unlikely that this will perform well as a predictor, even if there are some detectable differences in fracture cases. In addition, the relatively low variation of CBMD between cases and cohort underlies the similarity between the CTh and CM distributions. Hence our prior decision to include neither CTh nor CBMD in the predictive models seems justified: in the former case there is similar information and at slightly better precision in CM, and in the latter case the CBMD changes are much too small compared to the measurement precision.

The ECTD distribution in Fig. 2(g) and (h), on the other hand, shows quite dramatic differences between fracture cases and cohort, and between the fracture types. Whilst this difference is significant over much of the proximal femur, the distributions are quite distinct between trochanteric and neck fractures. As with CM, CBM patches for trochanteric fracture are focussed on the supero-lateral trochanter, and neck fractures at the superior femoral neck, with substantially less involvement on the inferomedial side. This is a particularly interesting result, given that it is known that the inferomedial femoral neck is largely preserved with age (Mayhew et al., 2005), and may go some way to explaining why this preservative effect does not prevent neck fracture risk increasing with age. It should be noted here that, although ECTD is clearly very significant, the percentages in Fig. 2(g) and (h) (which have an increased scale compared to the others) need to be interpreted with care given that the mean ECTD can approach a value close to zero.

We have already discussed the relatively poor performance of the CBMD measurement compared to the other CBM parameters. Looking at the other precision values in Table 5 clarifies that this is largely due to the small variation of this parameter in the cohort: all point-wise errors are at around 5% of the mean value, but CBMD has a much smaller cohort variation. The distribution of these errors in Fig 4 reveals that there is a repeatable pattern to where they tend to occur. For CTh and CM, there are larger errors around the femoral head and also at the medial side of the lesser trochanter. The former is due to the presence of the acetabulum, which is very close to the femur at this point and makes it harder to separate out the respective cortices in this region. In the latter area, the cortex is often not well modelled as a single layer and the measurements are less precise as a result. Fortunately, most of the patches in Fig. 2(a) to (d) are not coincident with these high-error regions, and hence the CBM precision after aggregation over these patches is substantially better, at about 1% of the mean. ECTD is not affected by the acetabulum but suffers similar imprecision at the medial side of the lower trochanter.

4.2 Fracture prediction

As expected, hazard ratios in Table 2 are significant for all BMD measures, with QCT-based ThBMD the most significant for trochanteric fracture (and for any fracture), but DXA-based FnBMD the quantity of choice for neck fracture. In contrast to QCT, the DXA-derived areal BMDs contain some measure of size as well as volumetric density, and this may contribute to the improved performance for neck fracture from what is otherwise a less direct measurement of BMD. Nevertheless, hazard ratios between DXA and QCT BMD are largely similar, supporting the general results from previous studies which fail to show any significant benefit from adding such QCT measurement to fracture predictors which already include DXA measurements (Bousson et al., 2011; Black et al., 2008; Lang et al., 2014). Hazard ratios for CM are similar to those for BMD, which would indicate that this is about as good at predicting fracture as the BMD measurements. However, the CM trochanter patch is strongly discriminative of fracture type; the CM neck patch is less so, but still shows more variation between fracture types than FnBMD.

ECTD hazard ratios are somewhat larger, a similar result to that found by (Lang et al., 2014), where their trabecular volumetric BMD measurements also tended to be the best predictors. This is not a surprising result, since their work was based on data from the same study, but it confirms that the result holds true for our larger sample size, and indicates that our measurement of endocortical trabecular density is capturing the important information contained within the trabecular compartment. As with CM, the ECTD trochanter patch shows the largest difference in hazard ratio between fracture types. All CBM patches show greater hazard ratios for the fracture type on which they were based.

The odds ratios for ten-year fracture incidence in Table 3 are somewhat smaller than the hazard ratios. This is to be expected, since the predictive model to which these lead is designed only to detect *whether* a subject will fracture, not *how long* it will be before they fracture. Hence there is some loss of information: it may be presumed that those who fracture a long time after the initial baseline scan had relatively thicker or stronger bones at baseline than those who fractured soon afterwards. The predictive model does not take advantage of this information, but nevertheless is necessary in order to evaluate how well we can predict fracture outcome. Although the results here for DXA and QCT BMD are again quite similar, it is the DXA-derived FnBMD which has the highest odds ratio for all fractures, as well as for neck fractures: QCT-based ThBMD is still the highest for trochanteric fracture. CBM odds ratios show the expected pattern in that the corresponding ratio is always higher for the fracture type on which the patch is based. The CM trochanter patch seems to show particularly good contrast between trochanteric and neck fractures, though the ECTD values, as with the hazard ratios, are larger.

Of more importance is how well these quantities perform when combined into predictive models, and performance is assessed using ROC curves in Fig. 3 and summarised in Table 4. The ROC curves for just age and height in the left hand graphs of Fig. 3 show that any useful fracture prediction that these might give is purely due to neck fractures: these variables have little power to predict trochanteric fracture. Adding DXA-based BMD to these values greatly improves performance, although it is clear that ability to predict trochanteric fracture is still much more limited than for neck fracture, a similar result to a recent study on the use of finite element analysis in fracture prediction (Nishiyama et al., 2014). AUCs for these models in Table 4 are somewhat smaller than have occasionally been recorded in the literature. Similar models, including DXA parameters, have reported AUCs for any fracture type of 0.86 (Lang et al., 2014) and 0.80 (Bousson et al., 2011), rather than our value of 0.78. The former was based on the same study as this current work, though with a different cohort and a smaller number of cases. Hence we conclude that our lower result is probably due to the increased number of cases in this study, or the different use of cross-validation techniques. Neither of these previous publications specify how the AUC results were validated, and it is worth noting that, had we not used leave-one-out cross-validation in this present work, all of our AUC results would have increased by between 0.02 and 0.03, bringing them more in line with previous results.

Looking at Fig. 3, it can be seen that any of the models which are based on CBM results are more capable of distinguishing between fracture types, mainly due to the significantly increased AUCs for neck fracture recorded in Table 4, though there is also an improvement in trochanteric fracture prediction. It is clear that CBM significantly improves the ability to distinguish fractures in prediction, and this improvement remains very significant when adding CBM quantities to models which already have either DXA- or QCT-based BMD values in them.

Improvements to the prediction of any fracture, irrespective of type, are rather smaller, but adding CBM to either DXA- or QCT-based BMD still results in an improvement which is statistically significant in both cases, albeit at a slightly lower level of $p < 0.005$ for the DXA-based model. In fact, whilst adding CBM to a DXA-model results in a significant improvement, adding DXA to a CBM-based model does not, which indicates that, if CBM data is available, there is no need to combine it with any information from DXA.

Acknowledgements

The Osteoporotic Fractures in Men (MrOS) Study is supported by National Institutes of Health funding. The following institutes provide support: the National Institute on Ageing (NIA), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the National Center for Advancing Translational Sciences (NCATS), and NIH Roadmap for Medical Research under the following grant numbers: U01 AG027810, U01 AG042124, U01 AG042139, U01 AG042140, U01 AG042143, U01 AG042145, U01 AG042168, U01 AR066160, and UL1 TR000128.

References

Barlow, W. E., Ichikawa, L., Rosner, D., Izumi, S., Dec. 1999. Analysis of case-cohort designs. Journal of Clinical Epidemiology 52 (12), 1165–1172.

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. Journal of Bone and Mineral Research 23 (8), 1326–1333.

Blank, J. B., Cawthon, P. M., Carrion-Peterson, M. L., Harper, L., Johnson, J. P., Mitson, E., Delay, R. R., Oct. 2005. Overview of recruitment for the osteoporotic fractures in men study (MrOS). Contemporary Clinical Trials 26 (5), 557–568.

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., Apr. 2011. In vivo discrimination of hip fracture with quantitative computed tomography: results from the prospective femur fracture study (EFFECT). Journal of Bone and Mineral Research 26 (4), 881–893.

Carballido-Gamio, J., Harnish, R., Saeed, I., Streeper, T., Sigurdsson, S., Amin, S., Atkinson, E. J., Therneau, T. M., Siggeirsdottir, K., Cheng, X., III, L. J. M., Keyak, J., Gudnason, V., Khosla, S., Harris, T. B., Lang, T. F., Mar. 2013. Proximal femoral density distribution and structure in relation to age and hip fracture risk in women. Journal of Bone and Mineral Research 28 (3), 537–546.

de Bakker, P. M., Manske, S. L., Ebacher, V., Oxland, T. R., Cripton, P. A., Guy, P., Aug. 2009. During sideways falls proximal femur fractures initiate in the superolateral cortex: Evidence from high-speed video of simulated fractures. Journal of Biomechanics 42 (12), 1917–1925.

Gee, A. H., Treece, G. M., Feb. 2014. Systematic misregistration and the statistical analysis of surface data. Medical Image Analysis 18 (2), 385–393.

Holzer, G., von Skrbensky, G., Holzer, L. A., Pichl, W., 2009. Hip fractures and the contribution of cortical versus trabecular bone to femoral neck strength. Journal of Bone and Mineral Research 24 (3), 468–474.

Johnell, O., Kanis, J. A., Oden, A., Johansson, H., Laet, C. D., Delmas, P., Eisman, J. A., Fujiwara, S., Kroger, H., Mellstrom, D., Meunier, P. J., 3rd, L. J. M., O'Neill, T., Pols, H., Reeve, J., Silman, A., Tenenhouse, A., 2005. Predictive value of BMD for hip and other fractures. Journal of Bone and Mineral Research 20 (7), 1185–1194.

Kanis, J. A., Burlet, N., Cooper, C., Delmas, P. D., Reginster, J. Y., Borgstrom, F., Rizzoli, R., 2008. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporosis International 19 (4), 399–428.

Kaptoge, S., Beck, T. J., Reeve, J., Stone, K. L., Hillier, T. A., Cauley, J. A., Cummings, S. R., 2008. Prediction of incident hip fracture risk by femur geometry variables measured by hip structural analysis in the study of osteoporotic fractures. Journal of Bone and Mineral Research 23 (12), 1892–1904.

Kopperdahl, D. L., Aspelund, T., Hoffmann, P., Sigurdsson, S., Siggeirsdottir, K., Harris, T. B., Gudnason, V., Keaveny, T. M., Mar. 2014. Assessment of incident spine and hip fractures in women and men using finite element analysis of CT scans. Journal of Bone and Mineral Research 29 (3), 570–580.

Lang, T. F., Saeed, I. H., Streeper, T., Carballido-Gamio, J., Harnish, R. J., Frassetto, L. A., Lee, S. M. C., Sibonga, J. D., Keyak, J. H., Spiering, B. A., Grodinsky, C. M., Bloomberg, J. J., Cavanagh, P. R., accepted in 2014. Spatial heterogeneity in the response of the proximal femur to two lower-body resistance exercise regimens. Journal of Bone and Mineral Research.

Marshall, L. M., Lang, T. F., Lambert, L. C., Zmuda, J. M., Ensrud, K. E., Orwoll, E. S., Aug. 2006. Dimensions and volumetric BMD of the proximal femur and their relation to age among older U.S. men. Journal of Bone and Mineral Research 21 (8), 1197–1206.

Mayhew, P. M., Thomas, C. D., Clement, J. G., Loveridge, N., Beck, T. J., Bonfield, W., Burgoyne, C. J., Reeve, J., 2005. Relation between age, femoral neck cortical stability, and hip fracture risk. The Lancet 366 (9480), 129–135.

Nishiyama, K. K., Ito, M., Harada, A., Boyd, S. K., Aug. 2014. Classification of women with and without hip fracture based on quantitative computed tomography and finite element analysis. Osteoporosis International 25 (2), 619–626.

Orwoll, E., Blank, J. B., Barrett-Connor, E., Cauley, J., Cummings, S., Ensrud, K., Lewis, C., Cawthon, P. M., Marcus, R., Marshall, L. M., McGowan, J., Phipps, K., Sherman, S., Stefanick, M. L., Stone, K., Oct. 2005. Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study — a large observational study of the determinants of fracture in older men. Contemporary Clinical Trials 26 (5), 569–585.

Parker, M., Johansen, A., 2006. Hip fracture. British Medical Journal 333 (7557), 27–30.

Poole, K., Treece, G., Blesic, K., Mayhew, P., Turmezei, T., Johannesdottir, F., Vindlacheruvu, M., Donell, S., Vaculik, J., Dungl, P., Horak, M., Stepan, J., Gee, A., Oct. 2013. Precise 3D localisation of cortical defects associated with subcapital, transcervical and trochanteric hip fractures in life. In: Proceedings of the American Society for Bone and Mineral Research. Baltimore, USA.

Poole, K. E. S., Mayhew, P. M., Collette, M., Brown, J. K., Bearcroft, P. J., Loveridge, N., Reeve, J., 2010. Changing structure of the femoral neck across the adult female lifespan. Journal of Bone and Mineral Research 25 (3), 482–491.

Poole, K. E. S., Treece, G. M., Mayhew, P. M., Vaculik, J., Dungl, P., Horák, M., Štěpán, J. J., 2012. Cortical thickness mapping to identify focal osteoporosis in patients with hip fracture. PLoS ONE 7 (6), e38466.

Poole, K. E. S., Treece, G. M., Ridgway, G. R., Mayhew, P. M., Borggrefe, J., Gee, A. H., 2011. Targeted regeneration of bone in the osteoporotic human femur. PLoS ONE 6 (1), e16190.

Sanders, K. M., Nicholson, G. C., Watts, J. J., Pasco, J. A., Henry, M. J., Kotowicz, M. A., Seeman, E., 2006. Half the burden of fragility fractures in the community occur in women without osteoporosis. When is fracture prevention cost-effective? Bone 38 (5), 694–700.

Treece, G. M., Gee, A. H., Jan. 2015. Independent measurement of femoral cortical thickness and cortical bone density using clinical CT. Medical Image Analysis 20, 249–264.

Treece, G. M., Gee, A. H., Mayhew, P. M., Poole, K. E. S., Jun. 2010. High resolution cortical bone thickness measurement from clinical CT data. Medical Image Analysis 14 (3), 276–290.

Treece, G. M., Poole, K. E. S., Gee, A. H., Jul. 2012. Imaging the femoral cortex: thickness, density and mass from clinical CT. Medical Image Analysis 16 (5), 952–965.

Turmezei, T. D., Treece, G. M., Gee, A. H., Tonkin, C. J., Vindlacheruvu, M., Blesic, K., Poole, K. E. S., Oct. 2012. Proximal femoral cortical thickness in postmenopausal women shows highly localised significant asymmetry. In: Proceedings of the American Society for Bone and Mineral Research. Minneapolis, USA.

Verhulp, E., van Rietbergen, B., Huiskes, R., 2008. Load distribution in the healthy and osteoporotic human proximal femur during a fall to the side. Bone 42 (1), 30–35.

Worsley, K. J., Taylor, J. E., Carbonell, F., Chung, M. K., Duerden, E., Bernhardt, B., Lyttelton, O., Boucher, M., Evans, A. C., 2009. Surfstat: A Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory. NeuroImage 47 (Supplement 1), S102–S102, Organization for Human Brain Mapping, 2009 Annual Meeting.