Bone-Marrow Densitometry: Assessment of Marrow Space of Human Vertebrae by Single Energy High Resolution - Quantitative Computed Tomography

Authors: Jaime A. Peña[1], Felix Thomsen[2], Timo Damm[1], Graeme M. Campbell[1 3], Jan Bastgen[1], Reinhard Barkmann[1], Claus C. Glüer[1].

Target Journal: Medical Physics

1. Sektion Biomedizinische Bildgebung, Klinik für Radiologie und Neuroradiologie, Christian-Albrechts-Universität zu Kiel, Campus Kiel, Germany.

2. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional del Sur, Bahía Blanca, Argentina

3. Institut für Biomechanik, Technische Universität Hamburg-Harburg (TUHH), Hamburg, Germany.

Address for correspondence: Claus-C Glüer, Am Botanischen Garten 14, 24118-Kiel, Germany, glueer@rad.uni-kiel.de

Abstract

Purpose

Accurate non-invasive assessment of vertebral bone marrow fat fraction is important for diagnostic assessment of a variety of disorders and therapies known to affect marrow composition. Moreover, it provides a means to correct fat-induced bias of single energy quantitative computed tomography (QCT) based bone mineral density (BMD) measurements. We developed new segmentation and calibration methods to obtain quantitative surrogate measures of marrow-fat density in the axial skeleton.

Methods

We developed and tested two high resolution QCT (HR-QCT) based methods which permit segmentation of bone voids in between trabeculae hypothesizing that they are representative of bone marrow space. The methods permit calculation of marrow content in units of mineral equivalent marrow density (MeMD). The first method is based on global thresholding and a peeling (GTP) to define a volume of interest (VOI) away from the transition between trabecular bone and marrow. The second method, morphological filtering (MF), uses spherical elements of different radii (0.1-1.2 mm) and automatically places them in between trabeculae to identify regions with large trabecular interspace, the bone void space. To determine their performance data were compared *ex vivo* to high-resolution peripheral CT (HR-pQCT) images as the gold standard. The performance of the methods was tested on a set of excised human vertebrae with intact bone marrow tissue representative of an elderly population with low BMD.

Results

87% (MF) and 86% (GTP) of the voxels identified as true marrow space on HR-pQCT images were correctly identified on HR-QCT images and thus these volumes of interest can be considered to be representative of true marrow space. Within this volume MeMD was estimated with residual errors of 4.8 mg/cm³ corresponding to accuracy errors in fat fraction on the order of 5% both for MF and GTP methods.

Conclusions

The GTP and MF methods on HR-QCT images permit noninvasive localization and densitometric assessment of marrow fat with residual accuracy errors sufficient to study disorders and therapies known to affect bone marrow composition. Additionally, the methods can be used to correct BMD for fat induced bias. Application and testing *in vivo* and in longitudinal studies is warranted to determine the clinical performance and value of these methods.

Keywords: computed tomography; image analysis; bone marrow; marrow fat; bone densitometry.

1 Introduction

The development and function of adipocytes and osteoblasts is a tightly regulated process. Over time, the adipocytic cell lineage takes a more dominant role in the differentiation of stromal or mesenchymal cells at the expense of the osteoblastic cell lineage. As the body ages, fatty depots in the marrow space increase and replace red or hematopoietic bone marrow. The incidence and location of certain diseases like myeloma, Ewing sarcoma, and histiocytic lymphoma are associated to the location and distribution of red marrow. Moreover, certain lesions to the marrow like metastasis may occur in mixed red and yellow marrow but not in yellow marrow alone[1]. Yellow marrow largely consists of fat but a detailed assessment reveals a complex composition pattern[2]. In the appendicular skeleton of adults, more than 70% of the marrow space is occupied by fat[3, 4] but in the vertebrae changes in marrow composition occur somewhat later in life and thus marrow composition is more variable[1, 5] with red-marrow cellularity dropping from nearly 100% at birth to 60 % by adulthood.[5] In osteoporosis reduced levels of bone formation, owing to a reduction in the number and activity of osteoblasts are coupled with a corresponding increase in marrow adipose tissue. Such an increase from 40% at age 30 to 68% at 100 years in a cohort of healthy and osteoporotic patients has been observed.[6]

To assess fat non-invasively in the marrow, Magnetic Resonance Imaging (MRI) has been the standard technique.. There have been efforts to compare MRI and quantitative computed tomography (QCT) derived measures. Ito *et al* Ito et al investigated the relation between T1-relaxation times and QCT-derived BMD.[7] More recently, Hui *et al* have shown good correlations in the volumetric marrow-fat estimated with MRI and

marrow fat estimated with Dual-energy CT (DEQCT) in cancer patients (r = 0.77).[8] DEQCT offered a reduction in the fat-related error in estimating BMD compared to single-energy CT (SEQCT). Moreover, Goodsitt *et al.* used DEQCT to measure intraosseous fat *ex vivo*[9-11] and recently, with the introduction of calibration standards for fat volume fraction, have quantified errors in its estimation for radiotherapy treatment planning.[12] The earlier DEQCT applications had the disadvantage of delivering an increased radiation dose to the patient and having worse reproducibility compared to SEQCT. State of the art DECT devices may reduce these two drawbacks and iterative-reconstruction methods may reduce patient doses even further without a large detriment in image quality. These two technologies may become more widespread in the near future and therefore, may be implemented in large clinical osteoporosis and bone disorders related studies.

It is the purpose of the present study to develop a new segmentation and calibration method that yields a quantitative surrogate measure of marrow-fat content in the axial skeleton, which can be applied to SEQCT scans and also for future DEQCT studies. Data acquisition is based on a high resolution QCT (HR-QCT) imaging protocol which additionally permits assessment of vertebral bone microstructure[13, 14] and correction of BMD bias caused by marrow fat. The latter aspect can be best accomplished by calibrating the marrow background in the same way as done for bone, i.e. in mineral equivalent units. This calibration additionally permits estimation of the fat/water ratio of the bone marrow as a basis to study changes in marrow composition associated with aging, disease, and therapy.

We hypothesize that densitometry from sub-regions in the marrow devoid of bone can yield accurate representative measures of marrow density reflecting the fat/water ratio. In this study we present two image segmentation methods, one based on global thresholding plus peeling (GTP) and a second, a morphological filtering (MF) method,

with the goal of segmenting marrow sub-regions as free of bone as possible from CT-images, thereby minimizing the influence of the partial volume effect and blurring according to the modulation transfer function in the marrow phase. The GTP method erodes regions close to the bone phase by a fixed distance and excludes them from analysis. On the other hand the MF-method, termed granulometry in other image processing fields[15], has been previously implemented to estimate trabecular bone thickness[16] and resembles the Euclidean distance transform. We assess the performance of both methods by comparing them to High Resolution peripheral QCT (HR-pQCT) as ground truth imaging technique

2 Materials and Methods

2.1 Bone Specimens

The segmentation methods were evaluated for two different samples of specimens: "embedded vertebrae": human defatted vertebrae embedded in epoxy resin, and "excised vertebrae": fresh frozen vertebrae with intact bone marrow. Hereafter only results for the excised vertebrae are shown. Corresponding results for the embedded can be found in the online supplementary data.

14 excised human thoracic (T10) vertebrae, harvested in the framework of the BioAsset study, were deemed optimal (free of large air bubbles, non-fractured, no scan artifacts) for the present work. Details and results of that study have been published elsewhere.[17, 18]

2.2 Image acquisition

A standard HR-QCT protocol[13, 14, 19] (120 kVp, 360 mAs, 0.6 mm collimation, 0.188x0.188x0.3 mm³ voxel spacing, z-axis length ca. 70 mm was applied on a clinical CT scanner (Siemens Somatom 64, Siemens AG Erlangen, Germany). HR-pQCT

images were obtained on a XtremeCT I device (Scanco Medical AG, Brüttisellen, Switzerland) at the Institut für Osteologie und Biomechanik, Hamburg, Germany with a standard patient protocol (60 kVp, 0.9mA, isotropic voxel size: 0.082³mm³).

2.3 Density calibration

Calibration to BMD of the HR-pQCT images was performed using the approach implemented by the manufacturer. BMD calibration of the HR-QCT images was carried out by scanning a density calibration phantom (K_2HPO_4) (Mindways, Austin TX, USA) underneath the vertebral specimens. Two CT-reconstructions were obtained: One which comprised a field of view (FOV) large enough to include both the vertebra and calibration phantom and a second one with only the vertebra inside the FOV used for the actual density estimations. Both BMD and Mineral equivalent Marrow Density (MeMD) were assessed in the same way and were expressed using units of mineral density (mg CaHAP/cm^3 for HR-pQCT and mg K_2HPO_4/cm^3 for HR-QCT).

2.4 Image Pre-Processing

The cancellous bone in the HR-pQCT images was segmented with a Scanco VivaCT 80 system by manually placing ellipse-like VOIs in the case of the excised samples and using manufacturer scripts for semi-automatic contouring of the entire spongious volume for the embedded samples (Figure 1). Image noise was reduced with the edge-preserving Perona-Malik anisotropic diffusion[20] to improve the subsequent bone-void detection. This filter (anisodiff3D.m function; delta = 0.1429, kappa = 30, option = 2, iterations = 9 on HR-pQCT and 3 on HR-QCT) is available as community-shared code in MATLAB (The MathWorks Inc., Natick, MA, USA). Due to the presence of remnant air bubbles within the cancellous region, a global threshold setting (t_{air} = -150 for HR-pQCT and -200 for HR-QCT) was necessary to disregard those voxels in all density estimations.

2.5 Marrow Segmentation Methods

Marrow space volume can be either derived from MV=TV-BV, using an estimated BV=TV*BMD/1200mg/cm³ where a tissue mineral density (TMD) of 1200 mg/cm³ is assumed or segmented. We use two marrow segmentation approaches. The first one, a standard one, is a combination of a global threshold of 250 mg/cm³ (as used in our prior HR-QCT works[14, 21]) and a peeling of 0.3 mm to reduce the partial volume effect (GTP method). The alternative segmentation method was based on Morphological Filtering (MF) with spherical structuring elements of varying radius. The assessment of the method was based upon the following criteria: (1) Evaluation of a large fraction of the marrow space in order to assure that the measurements are representative for the vertebral body and (2) Optimal correlation between MeMD measured with HR-QCT and HR-pQCT aiming to minimize the residual errors in a regression analysis (RMS error). The MF approach contains spherical structuring elements that are reshaped by morphological operations. The resulting entities of connected spherical elements are denoted as “bone-voids”. The set of all bone voids identified represent the “bone-void space” used to estimate the properties of the true marrow space (i.e. the bone-void space of HR-pQCT, the reference technique). The bone-void space detection consisted of a series of morphological filtering operations using binary spherical structuring elements (*SE*) applied on a binary image and can be summarized as:

$$LT(Image, t) \rightarrow C\big(SE(r_1)\big) \rightarrow O\big(SE(r_2)\big) \rightarrow E\big(SE(r_3)\big) \quad (1)$$

We start from a binary mask obtained from global threshold binarization of the greyscale images, with *LT*: low-pass threshold operator and *t:* threshold setting. For the two specimens, embedded and excised, a threshold *t* of 160 mg/cm³ was used. Subsequently, a closing operation (*C*) excludes particles the size of one voxel from the mask, e.g. r_1 = *voxel-spacing (mm)*. After this, an opening operation (*O*) finds regions within the mask with spacing between trabeculae greater than or equal to that of the

sphere radius r_2 (range: 0.1 – 1.2 mm, in 0.1 mm steps). Finally, erosion (E) using a sphere radius r_3 of 60% of the sphere radius of the previous step (r_2) ensures that regions close to the bone phase are disregarded, thus further minimizing partial volume effects on the segmented image. The standard segmentations were compared to a gold standard HR-pQCT segmentation obtained with the MF-method using a radius of 0.2 mm to guarantee a substantial segmented bone-void space.

A sensitivity analysis of the MF-method with respect to the Anisotropic Diffusion "pre-filter" was conducted in the excised vertebrae by comparing the effect of filtering (filter=on) and not filtering (filter=off) the HR-QCT images. In a similar way the effect of the threshold setting (t) was tested on the HR-QCT by comparing t=100 mg/cm³ and t=160 mg/cm³. For this, the sphere-radius = 0.2 mm was used both for the HR-QCT and the gold standard HR-pQCT. In the online supplementary data we show results of these tests where the following metrics were evaluated: difference in Marrow Volume between techniques, $Coincidence_{HR\text{-}QCT}$ (see next section), regression analysis in MeMD (R^2 and RMSE).

2.6 Agreement of bone-void space location on HR-QCT and HR-pQCT

A multi-resolution image registration procedure (3D-rigid, ITK[22] implementation under Structural*Insight*, in-house software) was conducted to quantify the degree of overlapping of the bone-void masks segmented in the two imaging techniques. For this, the **coincidence** level of the bone-void space detected in HR-QCT and HR-pQCT images was defined as:

$$Coincidence_j[\%] = \frac{\left|X_{HR-QCT} \cap X_{HR-pQCT}\right|}{\left|X_j\right|} 100\% \quad (2)$$

Where $X_{HR\text{-}QCT}$ and $X_{HR\text{-}pQCT}$ represent the complete bone-void space detected in HR-QCT and HR-pQCT for a given specimen sample, respectively. The numerator in formula 2 takes into account only the bone-void space that was matched i.e. the

intersection of the bone-void spaces of the two techniques. The subscript j stands for either one of the imaging techniques, such that the denominator X_j will either be $X_{HR\text{-}QCT}$ or $X_{HR\text{-}pQCT}$. With these definitions $Coincidence_{HR\text{-}QCT}$ is a measure of the "true positive" rates of the bone void space detected in HR-QCT, with a perfect correspondence yielding 100% coincidence. On the other hand, $Coincidence_{HR\text{-}pQCT}$ is an indirect measure of the "false negative" rates reflecting "true" bone-void space that was missed or not detected in the HR-QCT technique. High values in both of these figures of merit indicate a good performance of the segmentation method.

2.7 Conversion of MeMD to marrow fat fraction

In order to convert the results of MeMD, i.e. data expressed in mineral equivalents, into marrow fat fraction, a limited experiment was performed. We evaluated the CT scans of a set of density calibration standards with different nominal marrow fat fractions (CT simulator inserts for bone mineral analysis Model 004, CIRS Medical, Norfolk, VA, USA). Three inserts with reference densities of 50, 100, 150 mg CaHAP/cm³ have marrow-fat equivalent fractions of 15%, 30% and 15%, respectively. Each insert was scanned (three times without repositioning) with the same CT-scanner, acquisition protocol and calibration method as the excised samples. The average density reading (*Dens*) was obtained using an ellipse-like VOI for each insert. The actual measurement of the 100 insert in [mg K_2HPO_4/cm³] was compared to the value obtained from a calibration line of the reference density (*RefDens*) against *Dens* using the 50 and 150 inserts. The difference reflects the impact of a change in marrow fat-fraction content from 15% to 30%. These test inserts differ from the custom made inserts employed in Goodsitt *et al*[12] in that those included a broader range of bone, fat and red-marrow volume fractions. Yet both kinds of inserts were manufactured as a mixture that mimics the x-ray attenuation of the cancellous bone constituents as a whole but without considerations about the network nature of trabecular bone.

2.8 Statistical analysis

Linear regressions models and Pearson product-moment correlations were calculated to study the association of HR-QCT and HR-pQCT. A level of $p<0.05$ was considered significant. Error bars represent one standard deviation from the mean unless otherwise stated. All statistics were performed using JMP (JMP®, Version 9.03, SAS Institute Inc., Cary, NC, USA) and Matlab Statistics Toolbox.

3 Results

Table 1 shows characteristics of the samples under study (mean ± 95% confidence interval). The excised vertebrae represent an osteoporotic cohort. For the BMD a regression analysis yielded: $BMD_{HR\text{-}pQCT} = 10.88 + 0.80 \times BMD_{HR\text{-}QCT}$ ($r=0.98$, $R^2=0.95$, RMSE=4.03 mg/cm³).

Table 1 Descriptive statistics for the excised specimen samples. Results are expressed as mean values (95% confidence interval) for each imaging technique, HR-pQCT and HR-QCT.

	Excised	
Variable	**HR-pQCT**	**HR-QCT**
Number	14	14
BMD_{trab} [mg/cm³]*	59.2 (48.8,69.6)	60.2 (47.6,72.8)
$BV/TV_{segmented}$ [1]†	0.071 (0.054,0.089)	0.062 (0.039,0.084)
$BV/TV_{estimated}$ [1]‡	0.049 (0.040,0.057)	0.05 (0.039,0.06)
$MV/TV_{estimated}$ [1]#	0.95 (0.942,0.959)	0.95 (0.939,0.96)

*Calibrated to mg CaHAP/cm³ for HR-pQCT and mg K_2HPO_4/cm³ for HR-QCT, †with threshold of 250 mg/cm³, ‡Calculated as BV = (TV x BMD)/1200 assuming a tissue mineral density of 1200 mg CaHAP/cm³, #Calculated as (1-$BV/TV_{estimated}$).

3.1 Identification of representative bone-void space

Application of the MF method resulted in the identification of bone-void space. Figure 1 depicts axial reformatted images of the HR-pQCT and HR-QCT scans of an excised

vertebra (bottom row) with the correspondingly identified bone-void space in the HR-QCT images. The marrow cavities identified as bone-voids are color-coded according to the sphere-radius that fits in between bone structures. Red colors indicate locations of larger bone-voids, i.e. regions of larger trabecular separation.

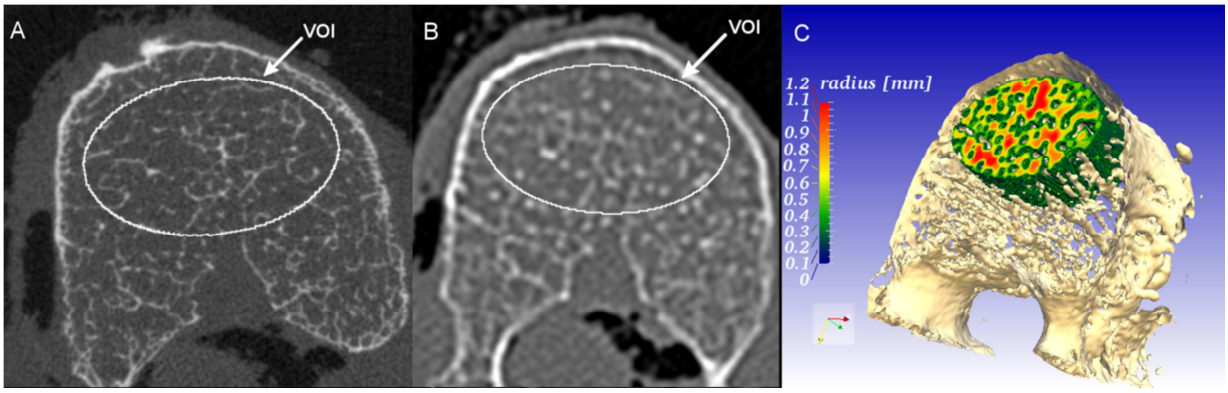

Figure 1 (A-C) Segmentation results for an excised vertebral specimen. (A) Transversal view of HR-pQCT and (B) HR-QCT scans. The volume of interest (VOI) where the search of bone-void regions takes place is marked in white. (C) 3D-depiction of the detected bone-void regions in the HR-QCT image using the Morphological Filtering (MF) method color-coded according to the sphere-radius (mm) that fits between the trabecular bone structures.

The relationship between sphere radius and bone-void space parameters is displayed quantitatively in Figure 2. The total volume evaluated amounts to (2.82 ± 0.24) cm^3. For a radius of 0.2 mm the HR-QCT bone-void space amounts to (2.36 ± 0.31) cm^3 and 87.2% of all voxels identified by HR-QCT coincide with the marrow-volume of HR-pQCT. Coincident HR-QCT voxels sample 91.7% of the estimated HR-pQCT based MV for radius 0.2 mm, (Figure 2B, dark blue bars), corresponding to an accurately sampled bone-void space of (2.06 ± 0.33) cm^3. For a radius of 0.6 mm the HR-QCT bone-void space amounts to (1.18 ± 0.39) cm^3. 78% of all voxels identified by HR-QCT coincide with the marrow-volume of HR-pQCT with a radius 0.6 mm corresponding to an accurately sampled bone-void space of (0.94 ± 0.38) cm^3. For the GTP-method HR-QCT samples a volume of 2.12 cm^3. The level of $Coincidence_{HR\text{-}QCT}$ of 91.1% represents

86.0% of the sampled MV in HR-pQCT (Coincidence$_{HR\text{-}pQCT}$), corresponding to an accurately sampled bone-void space of (1.93 ± 0.34) cm³ (Table 2).

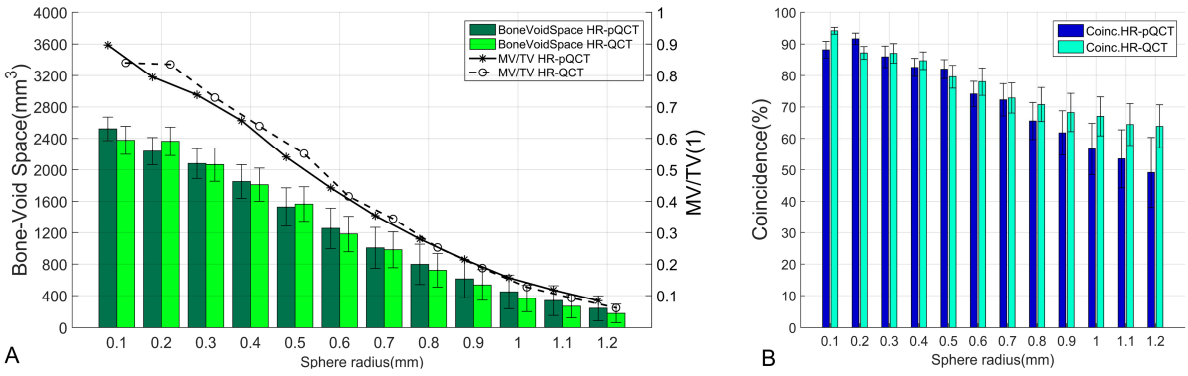

Figure 2 Bone-void statistics for the excised vertebrae obtained at different sphere radii (x-axis) of the MF method (A) Segmented bone void space obtained at different radii. Bone void space can also be expressed as marrow volume fraction (MV/TV), right-hand axis, (B) coincidence levels in bone-void space between HR-QCT and HR-pQCT segmented volumes (see formula 2). Error bars are the 95% confidence interval of the mean.

3.2 MeMD estimation

Figure 3 depicts the overall MeMD distribution as a function of the sphere-radii for the embedded and the excised datasets for the MF segmentation method. The highest MeMD values (23.7 and 38.3 mg/cm³ for HR-QCT and HR-pQCT, respectively) are obtained at a radius of 0.1 mm. The best agreement of the mean values of both techniques is observed for a radius of 0.2 mm with MeMD results of 23.2 mg/cm³ for HR-QCT and 28.9 mg/cm³ for HR-pQCT. The lowest level for HR-QCT, -8.3 mg/cm³, is reached at a radius of 0.7 mm. For HR-pQCT MeMD continues to drop to a radius of 1.2 mm. Not all samples had a detectable bone-void space for sphere-radii larger 1.2 mm. Regarding the GTP-method, average MeMD values of 24.8 and 20.4 mg/cm³ were obtained for HR-QCT and HR-pQCT, respectively.

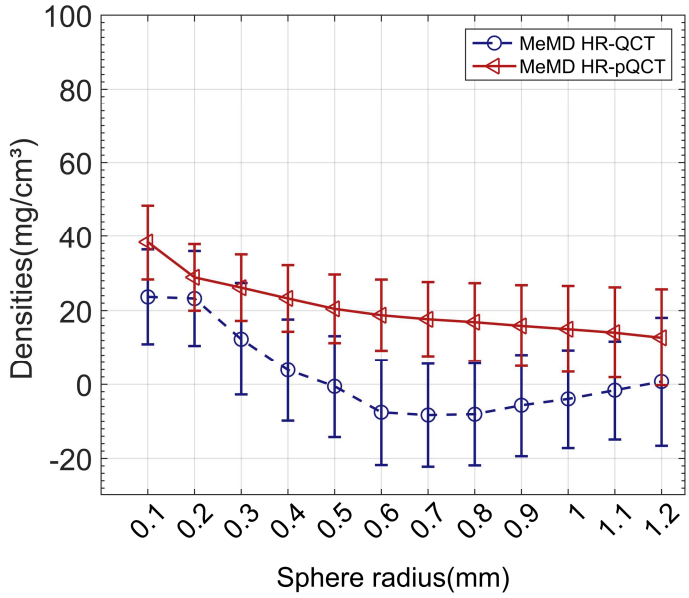

Figure 3 Densities (MeMD) obtained for different sphere-radii of the MF method for HR-QCT and HR-pQCT for the excised vertebrae.

Table 2 shows the results of the MF (radius = 0.2 mm) and the GTP methods for a regression analysis of MeMD HR-QCT and the gold standard HR-pQCT. Adjusted R^2 coefficients of around 0.6 and RMS errors below 5 mg/cm³ for both methods were observed. The models obtained for the MF method with a sphere radius of 0.2 mm and GTP, both unadjusted for BMD, were given by $MeMD_{HR\text{-}pQCT} = 14.9 + 0.6 \times MeMD_{HR\text{-}QCT}$ and by $MeMD_{HR\text{-}pQCT} = 16.0 + 0.52 \times MeMD_{HR\text{-}QCT}$, respectively. Here, both intercept and MeMD terms were significantly different from zero ($p<0.0001$) in both models.

The dependency of the MeMD measurements on BMD is highlighted by the Pearson correlation (r) listed in Table 2. For HR-QCT if evaluated with GTP, and with MF at radii 0.2 mm, and 0.6 mm, the r values were 0.45, 0.54 and -0.03. Similarly for HR-pQCT the r values were 0.55, 0.55 and 0.3. The impact of adjacent trabecular bone was not reduced by adjusting the estimates of $MeMD_{HR\text{-}pQCT}$ for BMD. MF at a radius of 0.2 mm did not show a reduced RMS error (similar levels of 4.8 to 5.0 mg/cm³) after adjusting. Similar results were observed for GTP. In contrast, for the embedded specimens a

reduction in RMSE (4.6 to 1.3 mg/cm³ for MF radius = 0.2 mm) and (4.7 to 1.6 mg/cm³ for GTP) after adjusting the models for BMD was observed (See Table A2 in supplementary data).

3.3 MeMD conversion to marrow-fat fraction

The average density readings (*Dens*) of each of the CIRS inserts obtained from the three HR-QCT scans were 50.09, 81.93 and 141.70 K_2HPO_4/cm³ for the 50, 100 and 150 mg/cm³ inserts, respectively. Using the 50 and 150 inserts (*RefDens*), which had the same nominal fat fraction content, a calibration equation was found to be *Dens*[mg K_2HPO_4/cm³] = 4.276 + 0.916x*RefDens*[mg CaHA/cm³]. With this calibration equation the density reading of the 100 insert was 95.88 mg/cm³. Thus, comparing this value with the actual measurement yielded a density difference of 13.95 mg K_2HPO_4/cm³ reflecting an increase in fat content of 15%. In units of [mg CaHA/cm³] this difference amounts to 10.6 mg CaHA/cm³.

Table 2 Results for the Morphological Filtering (MF) and Global Threshold plus Peeling (GTP) segmentation methods for the excised vertebra samples and both CT techniques.

		Mean (95% CI)			Regression MeMD*		Correlation (MeMD, BMD)‡	Regression BMDadj*	MeMD
	radius [mm]	MV/TV [1]	MeMD [mg/cm³]	Coincidence [%]†	RMSE	adj R^2	r	RMSE	adj R^2
GTP (HR-QCT)	NA	0.749(0.703,0.794)	24.8(16.30,33.23)	91.1(89.3,92.9)	4.8	0.72	0.45	4.8	0.71
MF (HR-QCT)	0.2	0.836(0.797,0.874)	23.2(15.9,30.59)	87.2(85.2,89.2)	4.8	0.71	0.54#	5.0	0.69
GTP (HR-pQCT)	NA	0.583(0.516,0.650)	20.4(13.86,26.87)	86.0(83.4,88.7)	-	-	0.55#	-	-
MF (HR-pQCT)	0.2	0.795(0.757,0.834)	28.9(23.74,34.04)	91.7(89.8,93.5)	-	-	0.55#	-	-

*For MF at r=0.2 mm and GTP the gold standard was HR-pQCT from MF at r=0.2 mm. †Calculated according to formula 2, see text. ‡Pearson product moment correlation coefficient. #$p < 0.05$

To study the MeMD relationships in more detail Bland Altman like analyses of MeMD (HR-QCT) – MeMD (HR-pQCT) as a function of MeMD (HR-pQCT) analyzed with MF(radius=0.2 mm) and GTP for HR-QCT and with MF(radius=0.2 mm) for HR-pQCT as gold standard were performed. Results for the mean offset and the slope are presented in Table 3.

Table 3. Bland-Altmann like analysis (Difference in MeMD between techniques vs MeMD HR-pQCT) for the two segmentation methods MF(radius=0.2mm) and GTP. SEM: Standard Error of the Mean, SE: Standard Error.

	Mean offset ± SEM	Slope ± SE
MF	-5.7 ± 1.8 mg/cm^3, p < 0.01	0.22 ± 0.21, p = 0.32
GTP	-4.1 ± 2.2 mg/cm^3, p < 0.09	0.41 ± 0.24, p = 0.11

Finally, one could argue that selection of the same radius for HR-QCT and HR-pQCT might not lead to optimal results. In the online supplementary data (Figure C1) we present contour plots of RMSE obtained from a linear model with $MeMD_{HR\text{-}pQCT}$ as the dependent variable versus $MeMD_{HR\text{-}QCT}$ adjusted for $BMD_{HR\text{-}QCT}$ for different combinations of sphere-radii in HR-QCT and HR-pQCT. It could be observed that similar errors to the ones in Table 2 were obtained for the embedded (RMSE = 1.3 mg/cm^3) and the excised (RMSE = 4.4 mg/cm^3 for HR-QCT radii of 0.1, 0.2 and 1.0 mm and HR-pQCT radius of 0.1 mm) vertebrae, respectively.

3.4 Sensitivity to Segmentation Parameters

Results from the sensitivity tests indicated a limited impact of the pre-filter (see Table B1, Supplementary Data). With regard to the threshold selection, we opted for the 160-160 threshold combination because (i) the coincidence level was good, (ii) the difference in MV between techniques was smaller and (iii) the bone-void space identified for this

threshold level was larger than when using the lower threshold (100 mg/cm³) thus being more representative. A further optimization might be necessary to account for different scanner characteristics.

4 Discussion

Comparison of the two Segmentation Methods

We developed, tested and compared two image segmentation methods, GTP and MF, for the assessment of MeMD in human vertebrae, a viable surrogate for the evaluation of bone marrow fat/water ratio. The bone specimens used feature a low BMD and represent a typical osteoporotic population with large marrow space. These specimens were used to test the methods of estimating marrow content in a realistic setting.

For the segmentation parameters deemed optimal for MF (radius = 0.2 mm) we achieved an accuracy error of 4.8 mg/cm³ for MeMD for the set of excised vertebrae. The same level of accuracy was achieved for the second segmentation method (GTP) plus slightly higher levels of coincidence compared to MF at a radius of 0.2 mm. Here, MeMD values of the two methods were very similar at 23.2 and 24.8 mg/cm³ (Table 2). Bland Altman like analyses showed significant offsets when comparing the two HR-QCT methods with the gold standard result by HR-pQCT for MF. These offsets are caused by differences in calibration of the HR-QCT and the HR-pQCT methods and bias due to limited spatial resolution Table 1. However, we did not observe significant slopes in the Bland Altman like analyses, indicating that the offset is not dependent on the level of MeMD. Therefore, the bias of HR-QCT in the assessment of MeMD is only affected by the density of the trabecular bone network, not by varying fat/water content, no matter whether assessed by MF or GTP. Given these results both MF and GTP proved to be similarly suited for marrow fat fraction assessment via MeMD.

HR-pQCT as the Gold-Standard

HR-pQCT was chosen as the gold standard approach to develop our method. Its spatial resolution with a voxel size of 82^3 µm³ is substantially better than that of HR-QCT. Still, it also shows partial volume effects. We strived to reduce this while keeping a large as possible segmented bone-void space. If we regard true marrow space as 95% of the total volume of bone and marrow (Table 1) a good compromise was achieved by both segmentation techniques. The MF method measured a marrow space that encompasses about 87% at a radius of 0.2 mm and still around 50% for radius of 0.6 mm of the true marrow space (Figure 2A, right hand axis). Similarly GTP measured 79% of the true marrow space. Furthermore, the MF with radius = 0.2 mm showed the best average MeMD-agreement between techniques (Figure 3), best and second-best coincidence levels with respect to HR-pQCT and HR-QCT, respectively (Figure 2B), a large sampled marrow volume (Figure 2A) and second lowest RMS errors (Figure C1, supplementary online data). The alternative GTP segmentation method showed similar levels of performance, likely because the volume sampled is close to the one sampled by MF at a radius of 0.2 mm (Table 2) and thus bone bias effects should be similar.

Correlations between MeMD and BMD

In the excised specimens, MF at a radius of 0.2 mm yielded a correlation level of $r = 0.55$ ($p = 0.04$) in the gold-standard HR-pQCT. Some part of this positive correlation may well reflect real physiological relationships in the vertebral specimens, i.e. higher bone mineral density being associated with a lowered marrow adiposity which in turn is reflected in a higher MeMD measurement according to our methods and definitions. Hence a positive correlation coefficient between these two measures can be expected. In HR-QCT for MF with a radius of 0.2 mm the residual correlations observed between MeMD and BMD were reduced to $r = 0.54$ ($p<0.05$). GTP yielded a similar reduction of the correlation to $r = 0.45$ (n.s.). Some part of this positive correlation may well reflect the physiological relationship noted above. However, at a radius of 0.2 mm, i.e. the level of closest agreement with HR-pQCT, the HR-QCT results are still affected by partial

volume effects as can be seen by the substantial drop in MeMD depicted in Figure 3 when increasing the radius of MF from 0.2 to 0.6 mm. This is consistent with a stronger partial volume effect in HR-QCT compared to HR-pQCT.

We could envision two different causes for this decrease in MeMD with increasing radius: aside of limited spatial resolution fatty deposits may be more concentrated at greater distance from the trabeculae, on the other hand the fact that an MeMD decline with the radius was also observed for the set of epoxy embedded vertebrae (Figure A2 , supplementary online data) evaluated on parallel to this work hints towards spatial resolution bias i.e. MTF and partial volume effects as the underlying reason for this. At a radius of 0.6 mm the lowest level of MeMD is achieved. Further increases in the radius do not improve the ability to assess MeMD as noise in the estimations goes up as the evaluated volume of interest shrinks. Thus MeMD measured with HR-QCT by MF with a **radius of 0.6mm** can be considered as being largely unbiased by BMD (R^2 = 0.001). However, the volume of marrow space evaluated at this radius may be too small to be representative for an assessment on an individual basis. Therefore, an analysis at a radius of 0.2 mm was chosen, but a correction for residual bone bias may be warranted. In the supplementary online data the reader can find corresponding results for the embedded samples. The idea of including these samples in the study was that they are stable over time and thus can be used for testing the segmentation and calibration method on different CT-scanners, extending our techniques to studies in multiple centers and populations. In those bone specimens, that featured much denser trabecular structure and a largely homogeneous epoxy-filled marrow space, the higher partial volume effect lead to a higher correlation between MeMD and BMD in HR-pQCT (r=0.75) than in HR-QCT (r=0.42). However, for a radius of 0.6 mm the correlation of r = 0.45 was no longer significant (p = 0.31) and sufficiently small like in the excised vertebrae. This suggests that for a non-osteoporotic vertebral cohort a different gold-standard may be necessary to achieve a necessary reduction of partial volume effects.

The entangled relationships between MeMD and BMD might also explain why after adjusting for the confounder BMD, the RMS errors in MeMD improved drastically (from 4.6 to 1.3 mg/cm³) in those embedded vertebrae (not affected by the physiological marrow condition) but not in the (clinically more relevant) excised vertebrae (4.8 and 5.0 mg/cm³, Table 2). Therefore, completely adjusting for the confounding effect of BMD established here in clinical studies aimed at estimating MeMD would be incorrect: by doing this one would artificially reduce apparent variations in marrow composition which in reality reflect the physiological process of opposing changes in bone and fat content of the vertebrae. On the other hand, the magnitude of the partial volume effect will be larger for subjects with higher BMD, as evidenced by a stronger drop in MeMD in the embedded vertebrae. Thus completely ignoring the confounding effects of BMD would also lead to a bias in MeMD. A somewhat more accurate adjustment can be expected by choosing BV/TV instead of BMD since biasing contributions of variations in tissue mineral density would be eliminated.

MeMD and Fat Volume Fraction Conversion

Our study was aimed at developing a method for the assessment of marrow content suited also for correction of the resulting bias of BMD measurements. The results of MeMD thus are expressed in units of mineral equivalent. To document how estimates of marrow fat fraction can be derived from MeMD our limited cross-calibration study with the separate test inserts provides some guidance. Our results showed that an increase in marrow fat fraction by 15% is associated with a decrease in measured BMD of 10.62 mg CaHAP/cm³ or 14.0 mg K_2HPO_4/cm³. Therefore, each increase of 1% in marrow fat fraction approximately results in a reduction in measured BMD by approximately 1 mg K_2HPO_4/cm³. These results are in line with other experimental reports where an increase in the accuracy error of 0.7% due to an increase of 1% of the marrow fat content by weight in QCT-measurements of the femur had been reported.[23]

Results from our study suggest that the methods developed (both MF and GTP) permit assessment of MeMD with residual accuracy errors of around 5 mg/cm^3, equivalent to approximately 5% marrow fat fraction, in agreement with earlier SEQCT studies[24] reporting a 10.8 mg/cm^3 change in BMD for a 10% fat by volume change at an x-ray tube voltage of 120 kVp. The magnitude of this error is sufficiently small to allow addressing interesting physiological or pathological differences in marrow composition. An increase in vertebral marrow fat content of 14%, typical for age-related changes over a decade in the postmenopausal period, can lead to a bias in BMD of around 14 mg/cm^3 according to our results, in agreement with reported levels of 6 - 13 mg K_2HPO_4/cm^3 [25]. Thus such age-related differences could be distinguished with the methods developed. In parallel, an age associated bone loss of 20 - 25 mg/cm^3 can occur in normal women between 40 and 50 years of age[26] and it is obvious that paralleling changes in marrow fat have to be considered to obtain accurate estimates of true BMD changes.

Clinical Aspects

Distinguishing marrow induced bias of BMD results (technical causes) from direct associations of marrow composition (detrimental role of bone marrow fat) would be important to understand the mechanism and relevance of these associations. Our study was cross-sectional and thus we did not investigate the precision error of repeated measurements of MeMD over time. Therefore, additional work is required to determine whether the precision of our measures would be sufficient to investigate longitudinal fat-related changes of treatment. Treatments might include both therapies that affect marrow composition, e.g. radiation therapy[8], and others aimed at increasing BMD in osteoporosis cohorts[27].

Limitations

Our HR-QCT methods have limitations such as radiation dose levels (approximately 2-3 mSv for imaging one vertebra) which precludes whole body data acquisition. Moreover, there are remaining limits of spatial resolution, clearly visible when comparing image

data with those obtained with HR-pQCT. Voxel sizes and the level of anisotropy between in-plane and z-axis will have an impact on the results achievable. The performance achieved here was based on a particular but widely used clinical scanner. Similar voxel sizes can be achieved on virtually all clinical scanners and thus we would expect similar performance. Still, *In vivo* additional errors need to be recognized in the desired skeletal region, the axial skeleton. For example we did not assess the influence of patient size in our measurements. Large body sizes will reduce the CT image quality with an increase in the image noise level. This can make segmentation using either MF or GTP methods more challenging and may demand restriction to larger bone-voids within the vertebral body, i.e. larger radii for the MeMD evaluation. Similarly, beam hardening effects may differ in the *in vivo* setting. Furthermore, although the two HR-pQCT scanners used in this study stem from the same manufacturer we did not cross-calibrate them. Histological comparisons were not carried out in the excised vertebral specimens. It is to notice that a detailed analysis of marrow fat content may require more sophisticated cross calibration methods but our preliminary data should document how to proceed if cross-calibration of results from units of MeMD into other units is required, e.g. by means of calibration inserts like the ones used in this or in similar works.[12]

Parallel to similar works

Marrow fat content can be assessed by other methods including DEQCT and MRI. In the 1980s a number of SEQCT and DEQCT analysis approaches were compared in various studies.[23, 28] The spatial resolution of these approaches was lower compared to state of the art HR-QCT imaging.[21, 29, 30] Quite recently Goodsitt *et al*[12] used a DEQCT method to determine region specific variations in cancellous bone composition for improved red marrow dose estimation in radionuclide therapy.[12] The authors used phantoms and reported RMS errors between around 3% and 9%, acknowledging that errors in real vertebrae and under *in vivo* conditions may be higher. The accuracy error of 5% reported for our approach was obtained on real vertebrae and thus would be well within

the range they reported for their method. Also the radiation exposure of our method of approximately 3 mSv is in the range of exposure settings they tested. Our HR-QCT method additionally spatially separates bone and marrow compartments. This may facilitate to study and exclude biases caused by compositional changes not only with regard to bone marrow but also collagen.[31] In our approach we restrict the density estimation problem to only two components, namely fat (encompassing contributions from both yellow- and red-marrow) and water (encompassing non-fatty marrow constituents like water, red-blood cells and proteins). Further studies are required to compare the performance of the various SEQCT and DEQCT approaches with our segmentation methods and also with those obtained on the new dual source CT scanners. Compared to MRI our method may be advantageous if not only the marrow but also the bone phase is of importance, for example in the assessment of bone quality status of multiple myeloma patients[32] or in treatment monitoring against the osteoporosis.

Outlook and Conclusion

In conclusion, in this work we have developed and tested HR-QCT-based segmentation methods that allow assessment of marrow fat fraction on SEQCT scans with accuracy errors of approximately 5%. The development of these methods aimed at limiting the partial volume effect while keeping large volumes of segmented marrow volume. Our calculation of MeMD provides a direct link to correcting fat induced biases in SEQCT based assessment of BMD. This is relevant for a large number of diseases in which marrow composition is altered along with more accurate estimates of true BMD changes of treatment agents against osteoporosis and other disorders. HR-QCT also provides insight into bone microstructure and fat-corrected BMD should yield more accurate estimates of QCT based measurements.

5 Acknowledgements

The authors would like to acknowledge the contributions from Sonja Waldhausen and Beata Hoffman (*Sektion Biomedizinische Bildgebung* and *Klinik für Radiologie und Neuroradiologie*, *Kiel, Germany*) for performing the HR-QCT scans, Matthias Krause (*Institut für Osteologie und Biomechanik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany*) for the XtremeCT scans of the excised specimens, Isolde Frieling (*Osteoporosepraxis Neuer-Wall*, *Hamburg, Germany*) for the XtremeCT scans of the embedded specimens, and Sanjay Tiwari and Robert Tower (*Sektion Biomedizinische Bildgebung, Klinik für Radiologie und Neuroradiologie*, *Christian-Albrechts-Universität zu Kiel*, *Campus Kiel, Germany*) for proof reading and helpful discussions. The *BioAsset* consortium is funded by a grant of the *Bundesministerium für Bildung und Forschung (BMBF), Germany, Förderkennzeichen 01EC1005.*

6 References

[1] M.E. Kricun, "Red-yellow marrow conversion: its effect on the location of some solitary bone lesions," Skeletal radiology **14**, 10-19 (1985).

[2] M.M. Goodsitt, P. Hoover, M.S. Veldee, S.L. Hsueh, "The composition of bone marrow for a dual-energy quantitative computed tomography technique. A cadaver and computer simulation study," Investigative radiology **29**, 695-704 (1994).

[3] C.J. Rosen, C. Ackert-Bicknell, J.P. Rodriguez, A.M. Pino, "Marrow fat and the bone microenvironment: developmental, functional, and pathological implications," Critical reviews in eukaryotic gene expression **19**, 109-124 (2009).

[4] P.K. Fazeli, M.C. Horowitz, O.A. MacDougald, E.L. Scheller, M.S. Rodeheffer, C.J. Rosen, A. Klibanski, "Marrow fat and bone--new perspectives," The Journal of clinical endocrinology and metabolism **98**, 935-945 (2013).

[5] J.S. Blebea, M. Houseni, D.A. Torigian, C. Fan, A. Mavi, Y. Zhuge, T. Iwanaga, S. Mishra, J. Udupa, J. Zhuang, R. Gopal, A. Alavi, "Structural and functional imaging of normal bone marrow and evaluation of its age-related changes," Seminars in nuclear medicine **37**, 185-194 (2007).

[6] J. Justesen, K. Stenderup, E.N. Ebbesen, L. Mosekilde, T. Steiniche, M. Kassem, "Adipocyte tissue volume in bone marrow is increased with aging and in patients with osteoporosis," Biogerontology **2**, 165-171 (2001).

[7] M. Ito, K. Hayashi, M. Uetani, Y. Kawahara, M. Ohki, M. Yamada, H. Kitamori, M. Noguchi, M. Ito, "Bone mineral and other bone components in vertebrae evaluated by QCT and MRI," Skeletal radiology **22**, 109-113 (1993).

[8] S.K. Hui, L. Arentsen, T. Sueblinvong, K. Brown, P. Bolan, R.G. Ghebre, L. Downs, R. Shanley, K.E. Hansen, A.G. Minenko, Y. Takhashi, M. Yagi, Y. Zhang, M. Geller, M. Reynolds, C.K. Lee, A.H. Blaes, S. Allen, B.B. Zobel, C. Le, J. Froelich, C. Rosen, D. Yee, "A phase I feasibility study of multi-modality imaging assessing rapid expansion of marrow fat and decreased bone mineral density in cancer patients," Bone **73**, 90-97 (2015).

[9] M.M. Goodsitt, R. Murano, M.L. Richardson, "A DPA technique for simultaneously measuring bone, soft tissue, and fat content," Investigative radiology **24**, 762-767 (1989).

[10] M.M. Goodsitt, D.I. Rosenthal, "Quantitative computed tomography scanning for measurement of bone and bone marrow fat content. A comparison of single- and dual-energy techniques using a solid synthetic phantom," Investigative radiology **22**, 799-810 (1987).

[11] M.M. Goodsitt, D.I. Rosenthal, W.R. Reinus, J. Coumas, "Two postprocessing CT techniques for determining the composition of trabecular bone," Investigative radiology **22**, 209-215 (1987).

[12] M.M. Goodsitt, A. Shenoy, J. Shen, D. Howard, M.J. Schipper, S. Wilderman, E. Christodoulou, S.Y. Chun, Y.K. Dewaraja, "Evaluation of dual energy quantitative CT for determining the spatial distributions of red marrow and bone for dosimetry in internal emitter radiation therapy," Medical physics **41**, 051901 (2014).

[13] C. Graeff, W. Timm, T.N. Nickelsen, J. Farrerons, F. Marín, C. Barker, C.C. Glüer, "Monitoring Teriparatide-Associated Changes in Vertebral Microstructure by High-Resolution CT In Vivo: Results from the Eurofors Study," Journal of Bone and Mineral Research **22**, 1426-1433 (2007).

[14] A. Krebs, C. Graeff, I. Frieling, B. Kurz, W. Timm, K. Engelke, C.C. Gluer, "High resolution computed tomography of the vertebrae yields accurate information on trabecular distances if processed by 3D fuzzy segmentation approaches," Bone **44**, 145-152 (2009).

[15] P. Soille, *Morphological image analysis : principles and applications*, 2nd ed. (Springer, Berlin ; New York, 2003).

[16] R. Moreno, M. Borga, Ö. Smedby, presented at the Proc. SPIE Medical Imaging 2012: Image Processing2012 (unpublished).

[17] C.C. Glüer, M. Krause, O. Museyko, B. Wulff, G. Campbell, T. Damm, M. Daugschies, G. Huber, Y. Lu, J. Peña, S. Waldhausen, J. Bastgen, K. Rohde, S. Breer, I. Steinebach, F. Thomsen, M. Amling, R. Barkmann, K. Engelke, M. Morlock, J. Pfeilschifter, K. Püschel, "New horizons for the in vivo assessment of major aspects of bone quality. Microstructure and material properties assessed by Quantitative Computed Tomography and Quantitative Ultrasound methods developed by the BioAsset consortium," Osteologie **22**, 223-233 (2013).

[18] M. Krause, M. Soltau, E.A. Zimmermann, M. Hahn, J. Kornet, A. Hapfelmeier, S. Breer, M. Morlock, B. Wulff, K. Puschel, C.C. Glueer, M. Amling, B. Busse, "Effects of long-term alendronate treatment on bone mineralisation, resorption parameters and biomechanics of single human vertebral trabeculae," European cells & materials **28**, 152-163; discussion 163-155 (2014).

[19] C. Graeff, F. Marin, H. Petto, O. Kayser, A. Reisinger, J. Pena, P. Zysset, C.C. Gluer, "High resolution quantitative computed tomography-based assessment of trabecular microstructure and strength estimates by finite-element analysis of the spine, but not DXA, reflects vertebral fracture status in men with glucocorticoid-induced osteoporosis," Bone **52**, 568-577 (

[20] P. Perona, J. Malik, "Scale-Space and Edge-Detection Using Anisotropic Diffusion," Ieee T Pattern Anal **12**, 629-639 (1990).

[21] C. Graeff, Ph.D. Thesis, Hamburg-Harburg Techn. Univ.: Germany, 2010.

[22] T.E. Yoo, *Insight into Images*. (A.K. Peters, 2004).

[23] J.W. Kuiper, C. van Kuijk, J.L. Grashuis, A.G. Ederveen, H.E. Schutte, "Accuracy and the influence of marrow fat on quantitative CT and dual-energy X-ray absorptiometry measurements of the femoral neck in vitro," Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA **6**, 25-30 (1996).

[24] T. Ikeda, K. Sakurai, "Influence of bone marrow fat on the determination of bone mineral content by QCT," Nihon Igaku Hoshasen Gakkai zasshi. Nippon acta radiologica **54**, 886-896 (1994).

[25] C.C. Gluer, H.K. Genant, "Impact of marrow fat on accuracy of quantitative CT," Journal of computer assisted tomography **13**, 1023-1035 (1989).

[26] J.E. Block, "Models of spinal trabecular bone loss as determined by quantitative computed tomography," JBMR **4**1989).

[27] Y. Yang, X. Luo, F. Yan, Z. Jiang, Y. Li, C. Fang, J. Shen, "Effect of zoledronic acid on vertebral marrow adiposity in postmenopausal osteoporosis assessed by MR spectroscopy," Skeletal radiology **44**, 1499-1505 (2015).

[28] M. Tanno, T. Horiuchi, M. Ogihara, T. Kishino, Y. Mashima, K. Endoh, S. Karube, M. Igarashi, H. Yamada, "Comparative study of bone mineral density estimated by various methods of single- and dual-energy quantitative computed tomography: the capability of the four-equation four-unknown method.," Bone **18**, 239-247 (1996).

[29] J. Borggrefe, C. Graeff, T.N. Nickelsen, F. Marin, C.C. Gluer, "Quantitative computed tomographic assessment of the effects of 24 months of teriparatide treatment on 3D femoral neck bone distribution, geometry, and bone strength: results from the EUROFORS study," Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research **25**, 472-481 (2010).

[30] C. Graeff, Y. Chevalier, M. Charlebois, P. Varga, D. Pahr, T.N. Nickelsen, M.M. Morlock, C.C. Gluer, P.K. Zysset, "Improvements in vertebral body strength under teriparatide treatment assessed in vivo by finite element analysis: results from the EUROFORS study," Journal of Bone and Mineral Research **24**, 1672-1680 (2009).

[31] M.M. Goodsitt, R.F. Kilcoyne, R.A. Gutcheck, M.L. Richardson, D.I. Rosenthal, "Effect of collagen on bone mineral analysis with CT," Radiology **167**, 787-791 (1988).

[32] J. Borggrefe, S. Giravent, G. Campbell, F. Thomsen, D. Chang, M. Franke, A. Gunther, M. Heller, A. Wulff, "Association of osteolytic lesions, bone mineral loss and trabecular sclerosis with

prevalent vertebral fractures in patients with multiple myeloma," European journal of
640 radiology2015).