Automated analysis and visualization of preclinical whole-body microCT data
Baiker, M.

Citation
Baiker, M. (2011, November 17). *Automated analysis and visualization of preclinical whole-body microCT data*. Retrieved from https://hdl.handle.net/1887/18101

5

Automated bone volume and thickness measurements in small animal whole-body MicroCT data

Chapter 5

Abstract

Purpose: Quantification of osteolysis is crucial for monitoring treatment effects in preclinical research and should be based on MicroCT data rather than conventional 2D radiographs, to obtain optimal accuracy. However, data assessment is greatly complicated in the case of three dimensional data. This paper presents an automated method to follow osteolytic lesions quantitatively and visually over time in whole-body MicroCT data of mice.

Procedures: This novel approach is based on a previously published method to coarsely locate user defined structures of interest in the data and present them in a standardized manner (Baiker et al. 2010, Kok et al. 2010). Here, we extend this framework by presenting a highly accurate way to automatically measure the volumes of individual bones, and demonstrate the technique by following the effect of osteolysis in the tibia of a mouse over time. Besides presenting quantitative results, we also give a visualization of the measured volume for qualitative assessment. In addition, we describe an approach to measure and visualize cortical bone thickness, which allows assessing local effects of osteolysis and bone remodeling. The presented techniques are fully automated and therefore allow obtaining objective results, which are independent of human observer performance variations. In addition, the time typically required to analyze whole-body data is greatly reduced.

Results: Evaluation of the approaches was performed using MicroCT follow-up datasets of fifteen mice ($n = 15$), with induced bone metastases in the right tibia. All animals were scanned three times: at baseline, after three weeks and after seven weeks. For each dataset, our method was used to locate the tibia and measure the bone volume. To assess the performance of the automated method, bone volume measurements were also done by two human experts. A quantitative comparison of the results of the automated method with the human observers showed, that there is a high correlation between the observers ($r = 0.9996$), between the first observer and the presented method ($r = 0.9939$) and also between the second observer and the presented method ($r = 0.9937$). In addition, Bland-Altman plots revealed excellent agreement between the observers and the automated method (Interobserver bone volume variability: $0.59 \pm 0.64\%$, Obs1 vs. Auto: $0.26 \pm 2.53\%$ and Obs2 vs. Auto: $-0.33 \pm 2.61\%$). Statistical analysis yielded no significant difference ($p = .10$) between the manual and the automated bone measurements and thus the method yields optimum results. This could also be confirmed visually, based on the graphical representations of the bone volumes. The performance of the bone thickness measurements was assessed qualitatively.

Conclusions: We come to the conclusion that the presented method allows to measure and visualize local bone volume and thickness in longitudinal data in an accurate and robust manner, proving that the automated tool is a fast and user friendly alternative to manual analysis.

Bone Volume and Thickness Measurements in MicroCT Data of Mice

5.1   Introduction

BREAST CANCER metastasizes preferentially to bone. Post mortem evaluation revealed that 70% of patients who died of breast cancer had bone metastases present in the skeleton [135]. Bone metastases cause severe morbidity in living patients such as bone pain, fracture, hypercalcaemia and nerve compression [136, 137]. As a result, quantification of osteolytic lesion size is pivotal in preclinical research of metastatic bone disease and treatment evaluation in small animal models.

Osteolysis is currently quantified using 2D radiographs [138, 139]. The scoring of these radiographs is performed manually by drawing a Region-of-Interest (ROI) around the lesion and measuring the bone area. The problem with this procedure is that lesions may be projected on top of each other and will therefore be underestimated when quantified, due to the flattening of the three dimensional structure [140] (Fig. 5.1). The same may happen for lesions on the side of bone. Furthermore, performing the analysis manually is prone to observer bias. MicroCT datasets provide spatial information, suitable for measurements of various bone parameters such as bone volume, bone thickness and bone mineral density. These measurements are potentially more informative than the radiographic analyses. Also, MicroCT enables the researcher to study the overall bone structure.

The use of MicroCT for quantitative measurements is not without difficulties. The shape and position of a Volume-of-Interest (VOI, the three dimensional counterpart of a ROI in 2D) in a 3D dataset greatly influence the measurement results. Therefore, it is crucial that the selection of a VOI is reproducible and not affected by the scan orientation or the observer who performs the procedure. We previously published a manual approach for the normalized selection of a region of interest in complex shapes [140]. This manual approach provides good and reproducible results but is very time consuming and requires well trained observers.

The comparison of whole-body datasets from longitudinal studies is even more difficult. Variation in posture of the animal during scans taken at different scan dates makes it nearly impossible to spot subtle disease induced differences between scans [17]. We previously published an approach to automatically align the skeletons of animals that were scanned at different points in time. The method can handle large postural differences between animals and as a result, specifically designed holders that are sometimes used to coarsely align animals [18], are not required. In addition, the user can select individual bones and generate side-by-side visualizations of these bones from multiple longitudinal datasets (Fig. 4.3). Such normalized visualizations greatly facilitate detailed qualitative assessment of structures in multiple complex and large datasets [16].

Here we describe an addition to this method, which enables the user to perform automated quantitative measurements of bone volume and thickness alongside the visual output. For evaluation, we applied the method to segment the femur and the tibia/fibula in whole-body follow-up MicroCT datasets and measured the bone volume and cortical thickness at three points in time: baseline, three weeks and seven weeks. To test whether this approach could be used to quantify biologically relevant changes in bone volume, breast cancer cells were injected into the right tibia after the baseline scan. The left tibia
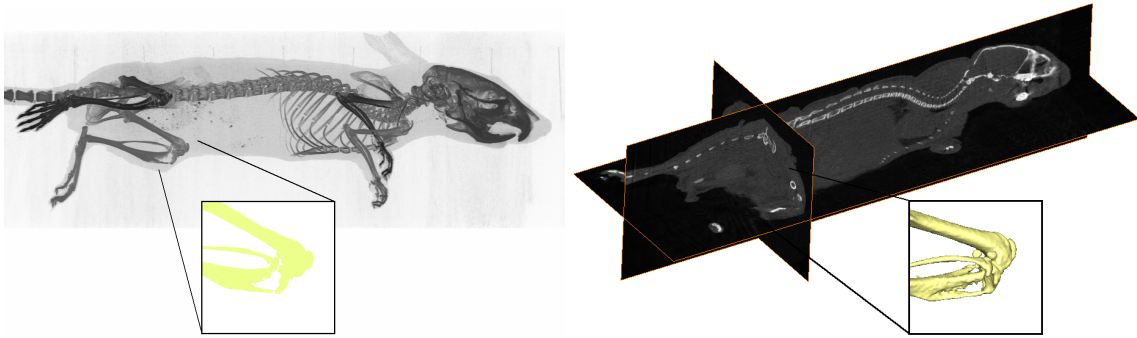
Figure 5.1: Examples of datasets for analysis of osteolysis based on a 2D radiograph (left) and a 3D MicroCT dataset (right). The insets are examples of the target structure to be analyzed. A ROI can easily be determined on the left whereas definition of a VOI on the right is not straightforward.

remained untreated and served as a reference. The results of the automated measurements are compared to manual measurements of two experts. We show that the automated segmentation and volume measurements perform equally accurate and reproducible as manual segmentation and volume measurements.

In summary, the goals of this work are to:

- Automate the task of measuring the volume of a user defined bone in whole-body in vivo MicroCT data and demonstrate the method by measuring the bone volume of the proximal tibia/fibula at several points in time.

- Compare the automated measurements with two human observers and show that the results are not significantly different.

- Present a way to assess the measurement quality visually, by providing proper visualization and

- Present a method to assess effects of osteolysis and bone remodeling locally (site-specific bone loss or gain) by automatically measuring and visualizing cortical bone thickness.

5.2 Materials and Methods

5.2.1 Animals

Fifteen ($n = 15$) female nude mice (BALB/c nu/nu, 6 weeks old) were acquired from Charles River (Charles River, L'Arbresle, France), housed in individually ventilated cages, food and water were provided ad libitum. Surgical procedures and MicroCT imaging were performed under injection anesthesia ($100mg/kg$ ketamine $+ 12.5mg/kg$ xylazine). Animals were sacrificed by cervical dislocation at the end of the experimental period. Animal experiments were approved by the local committee for animal health, ethics and research of Leiden University Medical Center.
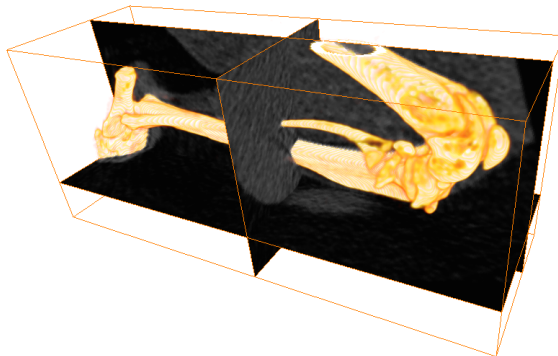
Figure 5.2: Example of an automatically determined subvolume, including the right tibia. The bone surface is shown together with the corresponding subvolume.

5.2.2   Cell Lines and Culture Conditions

The cell line MDA-231-B/Luc+ (hereafter MDA-BO2), a bone-seeking and luciferase-expressing subclone from the human breast cancer MDA-MB-231 [141,142], was cultured in DMEM (Invitrogen, Carlsbad, CA USA) containing 4.5g glucose/l supplemented with 10% fetal calf serum (FCS) (Lonza, Basel, Switzerland), $100units/ml$ penicillin, $50\mu g/ml$ streptomycin (Invitrogen) and $800\mu g/ml$ geneticin/G418 (Invitrogen). The cells were monthly checked for mycoplasma infection by PCR. The cells were donated by G. van der Pluijm (Leiden University Medical Center, Leiden, The Netherlands).

5.2.3   Experimental Setup

MDA-BO2 cells were injected into the right tibiae as described previously [142]. In brief, two holes were drilled through the bone cortex of the right tibia with a 25-gauge needle (25G 5/8, BD MicroFine, Becton Dickinson, Franklin Lakes, NJ USA) and bone marrow was flushed out. Subsequently, 250,000 MDA-BO2 cells/10μl PBS was injected into the right tibiae of the animals. MicroCT scans were made before the tumor cell inoculation (T0) in supine position, three weeks after tumor cell inoculation (T1) in prone position and seven weeks after tumor cell inoculation (T2) in supine position. The animals were scanned with arbitrary limb position.

5.2.4   MicroCT Data Acquisition

MicroCT scans were made using a SkyScan 1076 MicroCT scanner (SkyScan, Kontich, Belgium) using a source voltage and current set to 50kV and 200μA respectively, with an X-ray source rotation step size of 1.5° over a trajectory of 180°. Reconstructions were made using the nRecon V1.6.2.0 software (SkyScan) with a beam hardening correction set to 10%, a ring artifact correction set to 10 and the dynamic range set to -1000 - 4000 Hounsfield Units (HUs). The datasets were reconstructed with voxelsize $36.5 \times 36.5 \times 36.5\mu m^3$. Neither cardiac nor respiratory gating was used.

5.2.5   Manual Segmentation of the Tibia / Fibula

To assess the performance of the automated tibia volume measurements, two field experts were asked to segment the proximal part of the right tibia. To be able to use the data at full resolution, this was not based on the whole-body dataset but on a subvolume, corresponding to the right tibia, which was automatically determined following the procedure in Fig. 4.2. An example of such a subvolume is shown in Fig. 5.2. Starting with this subvolume, the experts were asked to segment the proximal part of the tibia/fibula, i.e. the part between the knee and the location where tibia and fibula separate. The manual segmentation was done using a tool that was developed in-house with MeVisLab V1.6 (MeVis Medical Solutions AG, Bremen, Germany) [140]. First, the user navigated through the subvolume by means of three orthogonal slices (see Fig. 5.2) and annotated the approximate bone centers at about twenty different locations in the femur and tibia, starting from the most proximal location in the femur and proceeding via the knee to the most distal location, in the tibia. Subsequently, cubic B-splines were fitted through these points, yielding a smooth curved line that runs approximately in the center of the femur and tibia, i.e. a bone 'centerline'. Subsequently, the user could navigate through the volume using slices that are perpendicular to this centerline.

Proceeding through the volume along the centerline, the expert first selected the slice (i.e. a flat plane) in the knee, which best separates the femur and the tibia, and second selected the slice, where the tibia and the fibula connect distally. To finally measure the bone volume in the region between the two slices, a three dimensional region grower was used. Starting from one or multiple seed point(s) set by the expert, this region grower decided, based on a user-defined threshold value, if the surrounding voxels are bone or not. This was iteratively repeated, until all bone voxels between the two slices were identified. Multiple seed points were necessary in case of fractured bone.

After segmentation, the number of bone voxels was determined using a threshold value to separate bone from background. To determine the optimum threshold for the in vivo datasets, the tibia of one of the animals was scanned ex vivo with high resolution ($9.125 \times 9.125 \times 9.125 \mu m^3$) after the follow-up experiment. Subsequently, the tibial bone volume was measured. To find the optimum threshold, for segmentation of bone from the background in the low-resolution data, the threshold was set such that the volume of the tibia of the same mouse in the low resolution data was the same as the volume of the tibia in the high resolution data. This threshold was kept constant for segmentation of all datasets. The result was a volume dataset with the same size as the initial subvolume with voxels labeled as relevant bone, i.e. the proximal tibia/fibula, and background (including irrelevant bone). Therefore, the bone volume of the proximal tibia/fibula could be determined by multiplying the total amount of bone voxels with the voxel volume i.e. in our case amount_of_voxels $* (36.5 \times 36.5 \times 36.5 \mu m^3)$. To be able to assess the quality of the segmentation visually, we provided a surface representation of the manually segmented subvolume. The tibia/fibula bone volume served as the reference for the automated method presented in the next subchapter.

5.2.6   Automated Segmentation of the Tibia / Fibula

An automated method should yield results that are as similar as possible to the results a human observer would obtain. Therefore, it should be designed such that it mimics the manual procedure as much as possible. Just as for the manual segmentation, presented in the previous subchapter, the automated segmentation was based on a subvolume as shown in Fig. 5.2 and the goal was to segment the proximal part of the tibia/fibula. First, a centerline was determined that runs through the center of the femur, the knee and the center of the tibia, based on the registration of the skeleton atlas to the MicroCT data. To this end, we defined 21 bone center locations (10 in the femur, 11 in the tibia) in the atlas. Subsequently, if the atlas bones are registered to the data (Fig. 4.2b), these atlas bone center locations are approximately in the bone centers of the femur and the tibia in the MicroCT data (the bone center locations do only have to be defined once for the atlas). Subsequently, a bone centerline was derived using cubic B-spline fitting through the bone centers. Next, the volume was segmented into bone and background, using global thresholding with the same threshold as was used for the manual segmentation (see previous subsection). Following the bone centerline from the knee towards the distal part of the tibia, the separation of the tibia and the fibula was determined using a hierarchical clustering technique with single linkage [143] that determined the number of bone clusters at regular spaced locations along the centerline. The Euclidean distance between points was chosen as the dissimilarity measure. The transition from two clusters (tibia and fibula) to one cluster identified the location of bone separation. Fig. 5.3 (right) shows a slice, perpendicular to the centerline, which is close to this point (tibia = large spot, fibula = small spot).

Separation of the tibia/fibula from the femur was done in a slightly different way as compared to the manual procedure, because it is very difficult to automatically determine a flat separation plane within the knee. Therefore, we chose to rely on a classifier that automatically separates all voxels labeled as 'bone' (i.e. after thresholding) into the two classes 'femur' and 'tibia/fibula'. The classifier was trained using volumetric (tetrahedral) meshes of the femur and tibia atlas after registration (Fig. 4.2b). Each node location of the meshes was weighted with a 3D Gaussian probability density function with width h (Parzen kernel density estimation [143]). Subsequently, all individual probability densities were summed up, yielding a bone dependent posterior probability density value within the entire data volume. A voxel labeled as 'bone' can thus be identified as 'femur' or 'tibia/fibula', depending on which of the two classes has the highest posterior probability at that voxel location. The parameter h was optimized using a leave-on-out test, based on the available datasets. Finally, the bone volume of the proximal tibia/fibula could be derived by counting the bone voxels classified as 'tibia/fibula' along the centerline, up to the tibia/fibula separation determined before and multiplying the total amount of bone voxels with the voxel volume. To assess the quality of the automated segmentation visually, we provided a surface representation of the result.
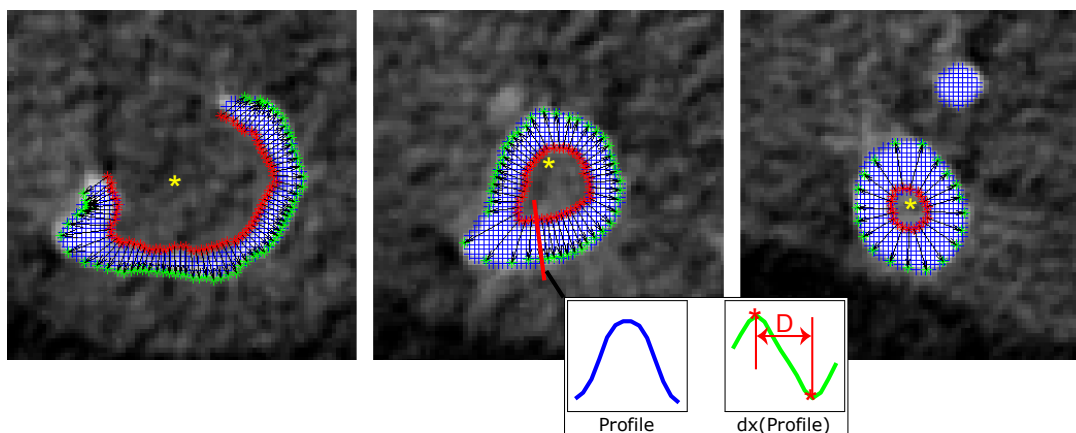
Figure 5.3: Demonstration of how the bone thickness D is determined automatically if osteolytic lesions are present. Shown are slices from the MicroCT subvolume that are orthogonal to the centerline, with an overlay of the voxels labeled 'bone' (blue net). Along the bone centerline (orange stars), grayvalue profiles are taken in axial direction at evenly spaced locations along the centerline. Shown are (left) a location close to the knee, (middle) a locations halfway between the knee and the tibia/fibula separation and (right) close to the tibia/fibula separation. Points on the inner boundaries are indicated by red stars, corresponding points on the outer boundaries by green stars. The black arrows indicate the directions, along which the grayvalue profiles for the bone thickness measurement are derived. An example of a profile path is shown in red (middle). The inset shows an example of a grayvalue profile in blue and its gradient values in green (dx symbolizes a mathematical derivation). The bone boundaries can be found where the gradients are maximum (red stars in the inset) and the bone thickness D is the distance between the boundaries.

5.2.7   Automated Segmentation of the Femur

As a proof of concept that the automated segmentation method can be applied to other skeletal elements besides the tibia as well, we demonstrate an automated segmentation of the femur. The femur is connected proximally to the pelvis and distally to the tibia. Following the procedure given in Section 5.2.6, the tibia was separated from the femur in a first step. Second, volumetric meshes of the atlas femur and the atlas pelvis after skeleton registration were used to derive a three dimensional posterior probability density function for these bones and to determine the separation of pelvis and femur, following the same procedure as described in in Section 5.2.6. The kernel width h was identical to the one used for the separation of the tibia and the femur. To assess the reproducibility of the volume measurements, the volume of the left femur of three animals was measured at all points in time and compared to the volume of the right femur over time. In addition, the bones were segmented manually, to assess measurement accuracy. To ensure that the influence of the induced cancer cells had a minimal effect on the femur bone volume, we chose three animals where osteolysis had only slightly progressed over time.

5.2.8 Automated Bone Thickness Measurements

Accurate knowledge of local bone thickness enables to follow the progress of osteolysis and bone remodeling over time. Therefore, a method is required to measure bone thickness in 3D and to relate the measurement to the exact location on the bone. Above that, the method should be able to handle severe structural changes over time, induced by osteolysis. There are mainly two approaches described in the literature to assess bone thickness in volumetric data: volume-based methods and surface (feature)-based methods [144]. These are focusing mainly on measuring trabecular bone and the approaches generally take the entire image domain into account. The advantage is that structures with very different shape can be analyzed. Although the approaches could be used for measuring cortical bone as well, the tube-like shape of long bones enables another approach. Since the registration of the skeleton atlas to the data yields a coarse segmentation of the skeleton, we can map a bone centerline, defined in the atlas femur and tibia, to the femur and tibia in the data. Subsequently, we can employ a technique similar to that presented in Van der Geest et al. [145], where the authors measure the diameter and wall thickness of blood vessels in MRA and CTA, based on slices that are orthogonal to the vessel centerline. The great advantage of relying on a centerline is that it is possible to determine exactly at which locations along the centerline the thickness should be measured. The main difference between analyzing vessels and potentially osteolytic bone is that vessels are continuous structures while bone can be highly fractured and contain holes.

The methods for trabecular thickness measurement generally take the entire image domain into account, which can be very time consuming, especially for large volumes or surfaces with a great amount of vertices. The proposed approach enables to greatly reduce computational burden. Above that, being able to define the thickness measurement based on a centerline allows to sample certain areas more densely than others, yielding more accurate measurements.

To determine the cortical bone thickness of the tibia automatically, we relied on the bone centerline presented in the previous section and the subvolume according to Fig. 5.2. At regularly spaced locations, following the centerline in distal direction, grayvalue profiles were extracted in axial direction, starting from the centerline and progressing outwards. In total, 360 profiles were taken per location, with 1° angle difference between them, covering an entire circle, oriented orthogonal to the centerline. Since the centerline lies in a low intensity area (bone marrow), the grayvalue profile will consist of low values at the beginning, high values within the bone and again low values outside the bone (muscle tissue). An example of such a profile is given in Fig. 5.3 (middle, inset). Subsequently, the inner boundary of the bone can be determined, using the highest positive gradient of the profile. Doing this for all 360 profiles, yielded 360 points located at the inner boundary of the bone. However, since the centerline may not always lie exactly in the center, these points are usually not evenly distributed along the boundary. Therefore, we applied an additional resampling step so that the points had a minimum distance of one voxel. Examples of resulting inner boundaries are shown in Fig. 5.3 (red stars). Next, again grayvalue profiles were taken, but this time orthogonal to the inner boundary of the bone, starting inside the bone and progressing outwards. An example path of such a profile is shown as a red line in Fig. 5.3 (middle). Finally, the bone thickness D could
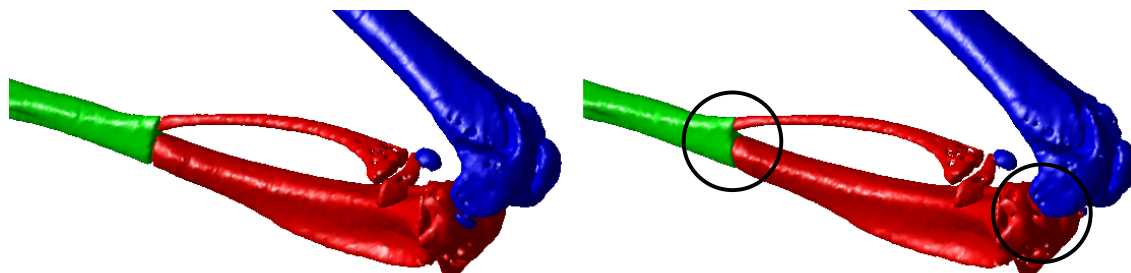
Figure 5.4: Bone surface visualization after manual (left) and after automated (right) segmentation of the proximal tibia/fibula. (femur = blue, proximal tibia/fibula = red, distal tibia/fibula = green). The circles highlight differences between the segmentations.

be determined using the highest positive and the highest negative gradient of the profile, demarcating the inner and the outer boundary of the bone. This is demonstrated in the inset in Fig. 5.3 (middle). Hence, our definition of bone thickness is the distance from the inner boundary to the outer boundary of the cortex, orthogonal to the inner boundary.

The bone thickness measurements can be uniquely related to the location on the bone, where they were derived. To be able to assess the bone thickness locally and still have the anatomical context information available, we present a visualization that is based on a surface representation of all bone in the subvolume (Fig. 5.2). To each location on the bone surface, we linked the corresponding bone thickness and assigned a value-dependent color. The result is a surface representation of the bone, on which the color indicates the bone thickness. The automated segmentations and bone thickness measurements and visualizations were performed using Matlab 2010b (The Mathworks, Natick, USA).

5.2.9 Quantitative Analysis of Measurement Results

To assess how similar the results of the automated method and the human experts are, Bland-Altman [146] plots as well as Pearson's correlation coefficients are presented. To investigate in detail the influence of the time point (i.e. baseline, first and second follow-up), the bone (i.e. healthy and pathologic) and the observer (i.e. automated, observer 1 and observer 2) on the bone volume measurement, we performed a statistical analysis using a three-way repeated measures analysis of variance (ANOVA) [147], with the bone volume as the dependent variable and observers, bone (i.e. healthy and pathologic) and time point as the independent variables ($3 \times 2 \times 3$ levels). A repeated measures design requires the variances of the differences between levels to be equal. Therefore, Mauchly's sphericity test should be non-significant if we are to assume that the condition of sphericity has been met. If the results of the test indicated that the assumption of sphericity was violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity [147]. To identify significant differences between group means for main and interaction effects, a Tukey HSD (honest significant difference) post hoc test was used. Effects were considered to be significant if $p < .05$. The statistical analysis was performed using Statistica 8.0 (StatSoft, Tulsa, USA).

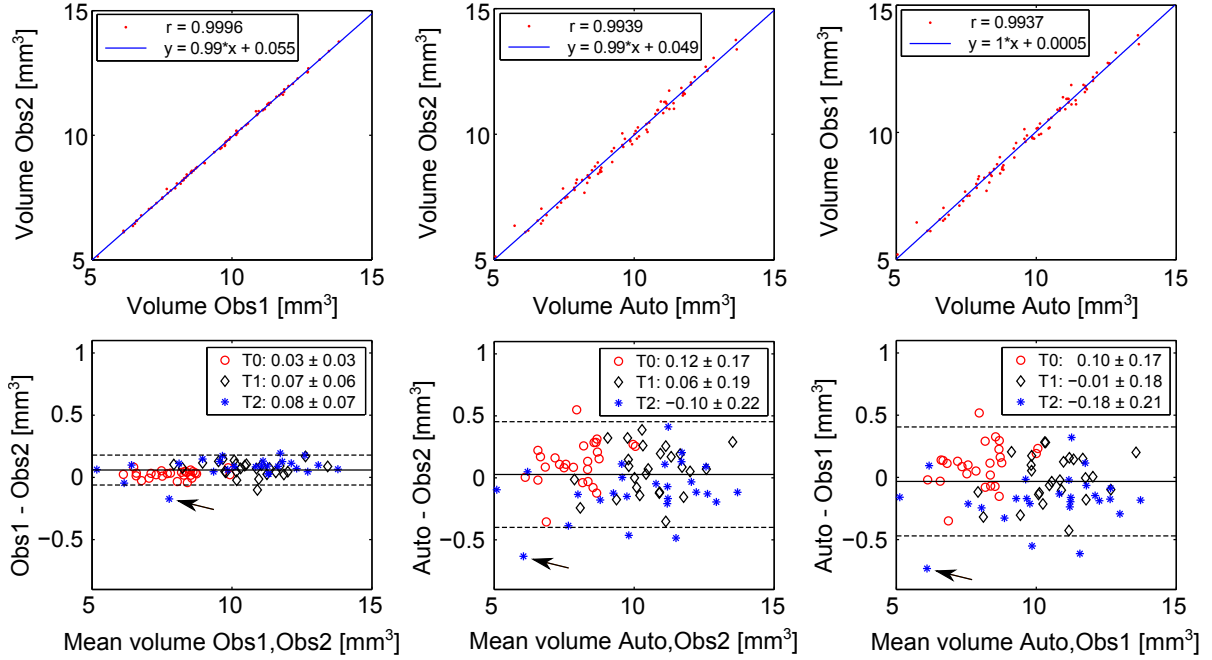## Bone Volume and Thickness Measurements in MicroCT Data of Mice

Figure 5.5: (Top row) Correlation between the measurements (in mm^3) of the two human observers and the automated method. Shown are Obs1 vs. Obs2, Auto vs. Obs2 and Auto vs. Obs1. The blue line represents a linear best fit, defined by the function in the legend. The Pearson correlation r, based on the data (red), is also shown in the legend. (Bottom row) Bland-Altman plots representing the measurement agreement between the two human observers and the automated method. The black lines indicate the grand means (–) ± 1.96 times the standard deviation (- -), which are $0.06 \pm 0.12 mm^3$, $0.03 \pm 0.43 mm^3$ and $-0.03 \pm 0.44 mm^3$ respectively. The arrows indicate the measurement with maximum disagreement between the observers. To assess, if the agreement is dependent on the time point when the data was acquired, these are shown in different colors (Baseline or T0 = red circles, T1 = black diamonds, T2 = blue stars). Note that the values in the legends are the means ± 1 time the standard deviation.

5.3 Results

To be able to assess the accuracy of a manual and an automated segmentation of the proximal tibia/fibula, surface visualizations are generated after the measurements. Examples are shown in Fig. 5.4.

The results of the correlation tests are shown in the top row of Fig. 5.5 and the measurement agreements are presented in the bottom row of Fig. 5.5. To assess a possible influence of the time point on the agreement, the data is shown for each time point individually (see legends).

Mauchly's test indicated a violation of the sphericity assumption and therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (see Tab. 5.1 in the 'Appendix' for details). The results show that there are significant differences in measured bone volume for the main effect Time, $F(1.39, 16.73) = 28.80$, $p < .001$, as well as the interaction effects Method * Time, $F(1.63, 19.59) = 16.71$, $p < .001$, and Bone * Time, $F(1.08, 12.93) = 12.75$, $p < .05$. The Tuckey HSD post hoc tests revealed a
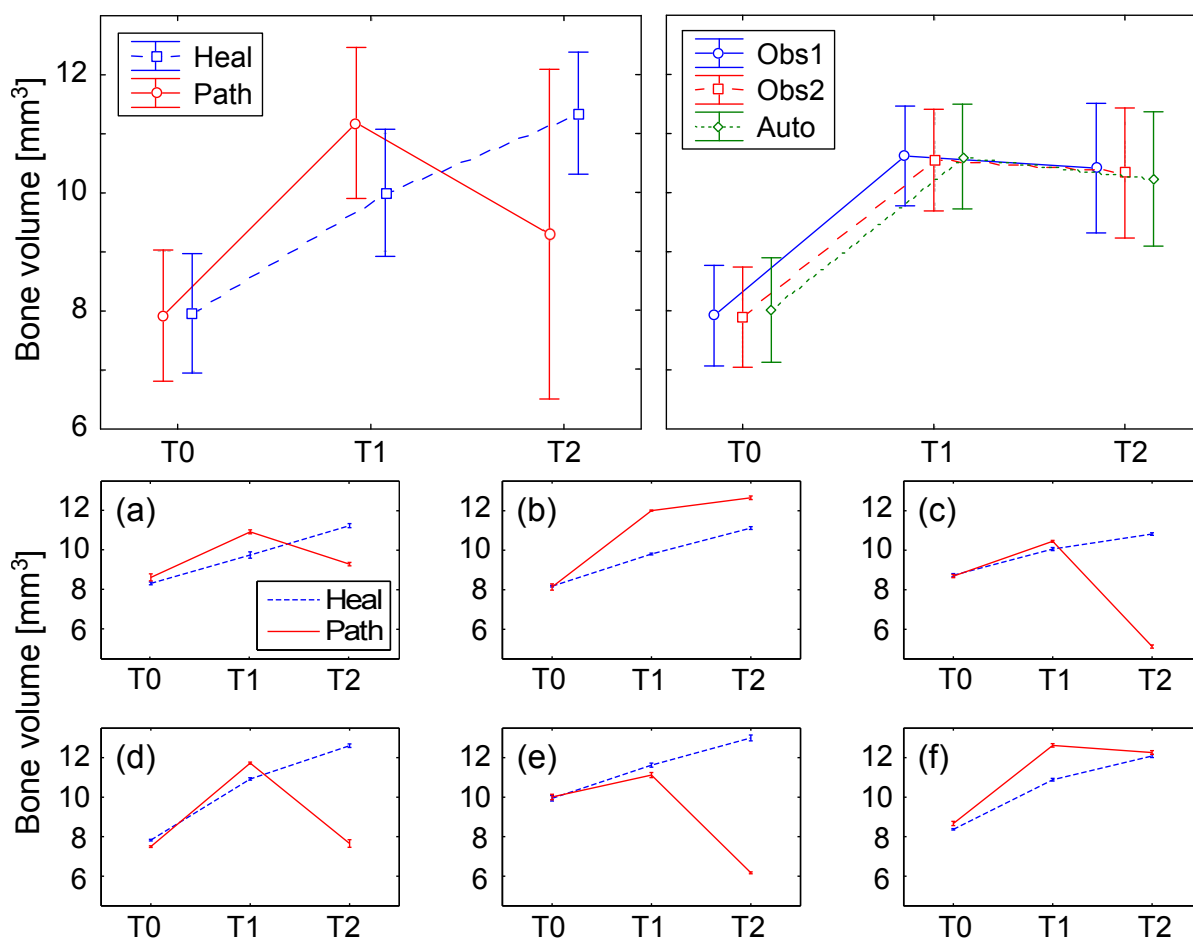
*Figure 5.6: (Top row) Mean bone volume (mm³) over time for the pathologic (Path) and the healthy (Heal) bones respectively, Bone * Time interaction (left), and bone volume over time for the two human observers (Obs1, Obs2) and the automated method (Auto), Observer * Time interaction (right). The results are based on including all mice. Error bars indicate 95% confidence intervals. (Middle and bottom rows) Mean bone volume (mm³) and the standard deviation of the healthy (Heal) and pathologic (Path) bones for six different mice (a-f) over time, averaging the measurements of the automated method and the two human observers.*

significant difference in bone volume between T0 and T1 ($p < .001$) as well as T0 and T2 ($p < .001$). There was no significant difference between T1 and T2 ($p > .05$).

For the Bone * Time interaction effect (Fig. 5.6, top left), relevant significant effects were present for healthy vs. pathologic bone at T2 ($p < .001$), but not at T0 and T1 (both $p > .05$). For the Method * Time interaction effect (Fig. 5.6, top right), relevant significant effects were present for Obs1 vs. Auto and Obs2 vs. Auto at T0 ($p < .05$ and $p < .001$) but not for Obs1 vs. Obs2 at T0 ($p > .05$). Furthermore there were significant effects for Obs1 vs. Auto and Obs2 vs. Auto at T2 ($p < .001$ and $p < .05$) but not for Obs1 vs. Obs2 at T2 ($p > .05$). There were no significant effects at T1.

The results of the comparison of the difference in bone volume between healthy and pathologic bone for six different mice are given in Fig. 5.6 (middle and bottom rows).

The results of the femur segmentation and subsequent volume measurements are

shown in Fig. 5.7. The average volume of the right and the left femur was $0.89 \pm 0.64\%$ when measured manually and $0.83 \pm 0.53\%$ when measured automatically. To see if there is a significant difference between the human observer and the automated method, a similar statistical analysis as presented in section 5.2.9. was performed, this time including one human observer instead of two. Mauchly's test indicated no violation of the sphericity assumption ($p > .05$). The results show that the main effect Method is significant $F(1, 2) = 92.894$, $p < .05$, and the mean difference between the automated and the manual method is $-2.15 \pm 0.75\%$. This means that the automated method results in lower measured volumes than the manual method.

A comparison of the development of the bone thickness over time for a healthy and a pathologic bone are given in Fig. 5.8 by means of bone surface visualizations where color indicates the bone thickness.

5.4   Discussion

In this article, we described a fully automated approach to analyze skeletal changes in rodent whole-body MicroCT scans. The automated approach is capable to 1.) align scans of the same animal, taken at different time points, 2.) automatically segment a subvolume (VOI) in these scans, 3.) measure the bone volume, 4.) measure cortical thickness and 5.) visualize it by means of assigning thickness-dependent colors. In addition, the user can visually check the segmentation performance using 3D bone surface representations and can generate normalized sections of identical sectioning planes in longitudinal scans for side-by-side comparison.

Conventional analysis of radiographs involves identifying osteolytic lesions manually. The procedure of manually drawing a region of interest is prone to observer bias and small changes in thickness or multiple lesions projected on top of each other are easily overlooked [140]. Manual analysis of MicroCT data is a better alternative, but is very labor intensive [140].

An automated method for MicroCT analysis has several advantages over manual analysis. The risk of non-objectivity and interobserver variability are greatly reduced by minimizing the active manual input of the researcher. Only an automated approach can be purely objective and handle every dataset in exactly the same manner. Additionally, an automated analysis method is much faster than any manual procedure. Thus, by automating the analysis, a relatively larger number of scans can be evaluated, compared to a human observer.

Researchers want to know exactly how quantified data is generated and tend to dislike automated 'black-box' approaches. To enable the researcher to check every step along the way, the automated method generates visualizations of the segmented volume. These visualizations can be evaluated after the analysis is complete. The automatic segmentation can be overruled manually or some datasets can be excluded from further analysis. Moreover, the cortical thickness maps enable the researcher to directly pinpoint where structural changes of the cortical bone occurred. This way, the cortical thickness maps help to identify areas of interest in the original scan data and in other modalities such as histological sections. Assessment of trabecular bone is not possible with the proposed
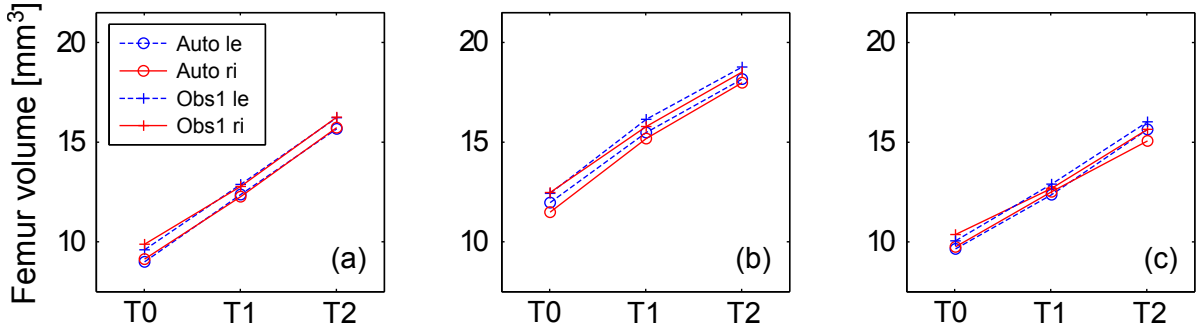
Figure 5.7: Result of the automated (auto) and manual (Obs1) volume measurement for the right (ri) and left (le) femur for three different mice (a-c) over time.

method, because the relatively low resolution of the in vivo data ($36.5 \times 36.5 \times 36.5 \mu m^3$) renders measuring the trabecular thickness accurately very difficult [148].

We validated the presented automated method by comparing it to the 'best available' method, namely manual bone segmentation and bone volume measurements. Therefore, we acquired datasets of 15 mice ($n = 15$) with induced bone metastases in the tibia at three points in time. The volume measurement results show that there is an excellent correlation between the human observers and the automated method: $r_{Obs1Obs2} = 0.9996$, $r_{AutoObs2} = 0.9939$ and $r_{AutoObs1} = 0.9937$. The Bland-Altman plots (Fig. 5.5, bottom row) based on all data indicate excellent agreement among the two human observers (interobserver variability) as well as the observers and the automated method. There is no obvious relation between the difference and the mean. Residual disagreement can therefore be explained by the bias and the deviation, which is very low in all cases, namely $0.59 \pm 0.64\%$, $0.26 \pm 2.53\%$ and $-0.33 \pm 2.61\%$ respectively. The residual errors are the result of mainly two factors that may influence the measurement outcome: the registration accuracy, and subsequently the segmentation accuracy, and the chosen threshold to separate bone from the background. The registration accuracy has the largest influence on the result and therefore, improving the accuracy would require a modification of the registration method. Special attention should be paid to the robustness of potential methods with respect to bone resorption. The thresholding procedure also influences the measured volume, because both values are inversely related i.e. if the threshold value increases, the volume decreases and vice versa. We chose a global threshold since the resolution of the in vivo data does not allow to reliably segment the trabecular bone [148] but methods including local thresholds may be more accurate, if data resolution increases.

Ideally, the automated measurements are identical to the manual measurements. The ANOVA revealed no significant difference between observers (Method, $p = .10$), which means that the automated method is performing equally well as the two human observers. However, the low p value indicates that significant interaction effects may be present. It appears that there is a dependency of the performance of the automated method on the time point since the automated method is significantly different from the human observers at T0 and T2. Visual inspection of Fig. 5.6 (top right) suggests overestimation of the volume at T0 and underestimation of the volume at T2. There is no significant difference at T1. This is supported by the Bland-Altman plots (Fig. 5.5, bottom row) in which

Bone Volume and Thickness Measurements in MicroCT Data of Mice

the mean difference in measurement is close to zero at T1. However, the differences are borderline and probably due to the very small variation between the human observers.

The bone volumes of pathologic bones were significantly decreased compared to the healthy bones at T2 (Fig. 5.6, top left). There are no significant differences at T0 and T1. There are two explanations why there is no volume decrease at earlier time points. Firstly, the bone marrow is partially flushed out of the bone during the intra osseous injection of tumor cells. This partial bone marrow ablation has profound anabolic effects on local bone turnover. Bone formation induced by bone marrow ablation reaches a maximum one week after the intervention. After this initial week the bone volume normalizes gradually over time as the bone recovers from the procedure, a process that can take weeks [149,150]. Secondly, starting osteolytic lesions around the tumor create weak areas in the bone. The mechanical stress on other, healthy parts of the bone will increase due to these weak areas. Both the anabolic effects due to the partial bone marrow ablation and due to the increased mechanical stress result in a local increase of bone volume alongside osteolytic lesions. Combined, these anabolic and osteolytic processes influence the volume measurements as can be seen in Fig. 5.6 (middle and bottom rows, a-d and f). The cortical thickness maps provide an excellent tool to see exactly where the volume changes occur in relation to the osteolytic lesion site (Fig. 5.8).

The presented segmentation method is not restricted to the tibia, but can be applied to any bone of the skeleton in whole-body MicroCT scans, as long as it is contained in the MOBY mouse atlas [16, 17, 59]. We are currently implementing the volume measurements of every segmented skeletal element using the same principle. We segmented the femur as preliminary proof of concept. Several conclusions can be drawn from the results in Fig. 5.7. The volumes of the right and the left femur are very similar for the manual and the automated measurement, meaning that measuring the femur is highly reproducible. The automated method however, underestimates the volume compared the manual method. This underestimation is to be expected since the femur included in the MOBY mouse atlas does not include the femoral head and neck. Therefore the segmentation result 'cuts' the femoral neck approximately in the middle and the amount of underestimated volume thus corresponds to the volume of the femoral head and part of the femoral neck. Note that this is a systematic error and only leads to inaccurate results if the femoral head and neck are of particular interest within a study. The same type of measurement error may occur for other bones as well, since most of the bones in the MOBY atlas are simplified versions of the real bone shape. However, as is the case for the femur, this should not lead to problems because the error is systematic. If higher segmentation accuracies are required in a particular part of the bone that is simplified, another animal model with more details could be employed. One should however bear in mind that using simplified bone shapes has the advantage that the influence of e.g. differences in strain or animal size can be minimized by leaving out the fine details.

The increased radiation dose of MicroCT compared to radiographs has always been a major concern limiting its use in cancer research. This is not a problem anymore as modern MicroCT scanners can perform whole-body scans in less than a minute [151]. The delivered radiation dose during these scans is well below a dose that would affect tumor growth, even during longitudinal follow-up studies [151–153].
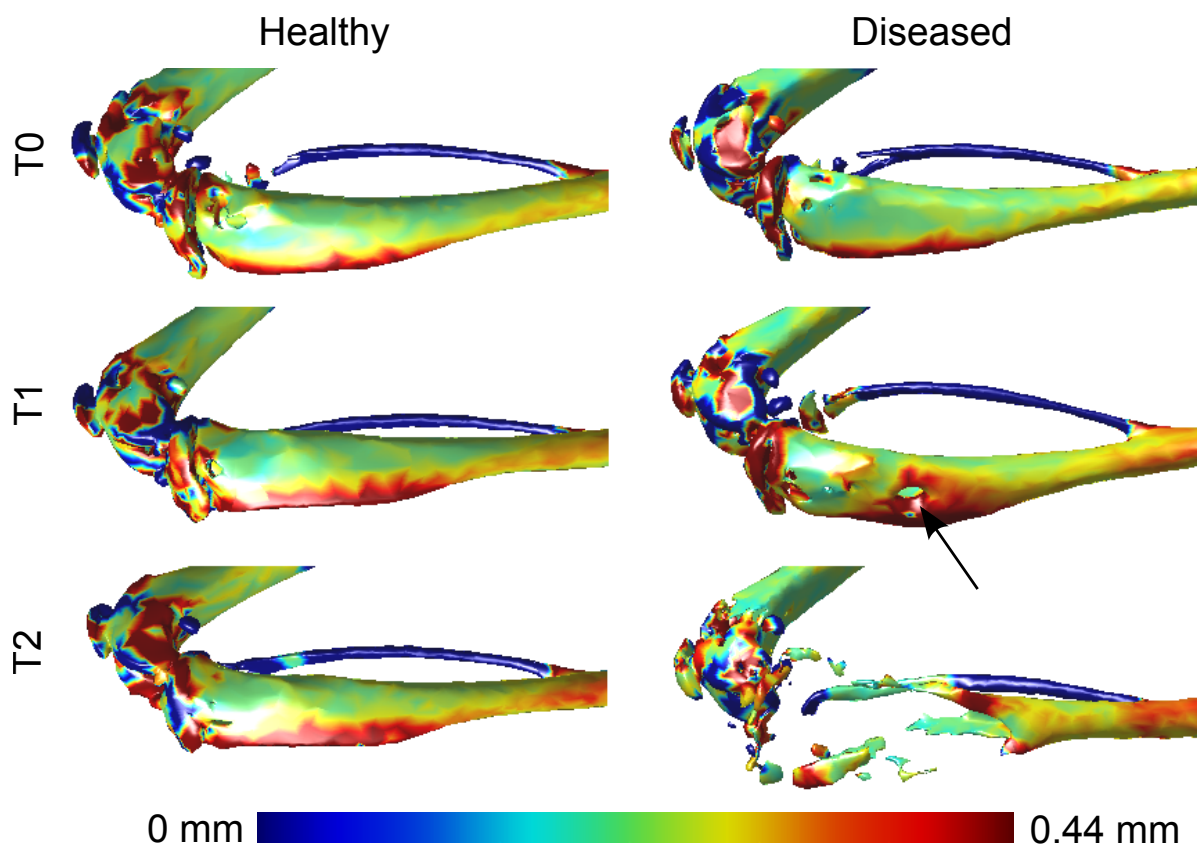
Figure 5.8: Comparison of the bone thickness development over time for a healthy and a pathologic (diseased) bone. Shown are bone surface representations. The colors indicate the bone thickness at each location on the bone. The bone marrow was partially flushed out of the bone during the intraosseous inoculation used to induce bone metastases. This partial bone marrow ablation leads to a local increase in bone volume preceding cancer-induced osteolysis [140]. The arrow indicates this local increase in bone thickness around the site of early osteolysis. Note that the measurements at the distal end of the femur and the proximal end of the tibia are not meaningful, because at these locations, a substantial amount of trabecular bone is present. However, bone thickness measurements are only meaningful for cortical bone.

All datasets used in this article have been generated with a standard scanning protocol using the Skyscan 1076 MicroCT. However, the described methods can be performed on any other whole-body MicroCT dataset acquired on a different machine and with a different protocol. Other scans might require an adjustment of threshold values and the initial scan resolution will always be a limiting factor during further analysis.

Finally we want to stress, that the described method is general and can be applied to others species as well as long as an anatomical skeleton atlas is available.

5.5   Conclusion

We suggested a new MicroCT analysis paradigm based on the combined approach of previously published methods for animal posture correction, normalized visualization of

follow-up data and the quantification and visualizations discussed in this paper. Together, this results in a fast and automated workflow, in which the user can easily compare whole-body MicroCT scans on the whole-body level, zoom in to the level of a single bone or bone segment of choice and gain qualitative and quantitative data of that segment. The animals can be scanned in any posture. Normalized and interactive side-by-side visualizations of the exact same section of skeletal elements at different time points can be generated from longitudinal scans in which one animal is scanned multiple times over time. The detailed side by side visualizations greatly help the researcher to identify changes in the skeleton. The researcher can then identify and zoom in on the bone or bone segment of interest and automatically generate quantitative volumetric data alongside visualizations of the segmented volume and visualizations of the cortical thickness of that specific skeletal element. This new workflow greatly reduces analysis time, aids the handling of complicated scan data and improves the overall qualitative and quantitative assessment of MicroCT scans. The method was validated by quantification of osteolytic effects over time in the tibia but can easily be adapted to other bones of the skeleton. In addition, the approach can be used for other species as well, given that an animal skeleton atlas exists for that animal.

Acknowledgements

Appendix

5.6 Results of Mauchly's test and Greenhouse-Geisser correction

Table 5.1: Results of Mauchly's test and Greenhouse-Geisser correction.

	Mauchly's test				Greenhouse-Geisser correction							
	W	χ^2	df	p	df M	df E	F	p	ϵ	Adj. df1	Adj. df2	Adj. p
Method	.05	33.72	2	.00	2	24	3.18	.06	.51	1.02	12.29	.10
Bone	1				1	12	0.69	.42	1	1.00	12	.42
Time	.57	6.26	2	.04	2	24	28.80	.00	.70	1.39	16.73	.00
Method * Bone	.18	18.9	2	.00	2	24	0.80	.46	.55	1.10	13.18	.40
Method * Time	.03	35.3	9	.00	4	48	16.71	.00	.41	1.63	19.59	.00
Bone * Time	.14	21.3	2	.00	2	24	12.75	.00	.54	1.08	12.93	.00
Method * Bone * Time	.05	31.3	9	.00	4	48	2.81	.04	.55	2.18	26.18	.07

Bone Volume and Thickness Measurements in MicroCT Data of Mice