Processing and analysis of in-vivo high-resolution MR images of trabecular bone for longitudinal studies: reproducibility of structural measures and micro-finite element analysis derived mechanical properties

Osteoporos Int (2002) 13:278–287

Osteoporosis International

Original Article

Processing and Analysis of In Vivo High-Resolution MR Images of Trabecular Bone for Longitudinal Studies: Reproducibility of Structural Measures and Micro-Finite Element Analysis Derived Mechanical Properties

D. C. Newitt[1], B. van Rietbergen[2] and S. Majumdar[1]

[1]MRSC, Department of Radiology, University of California, San Francisco, CA, USA; and [2]University of Eindhoven, Eindhoven, The Netherlands

Abstract. The authors have developed a system for the characterization of trabecular bone structure from high-resolution MR images. It features largely automated coil inhomogeneity correction, trabecular bone region segmentation, serial image registration, bone/marrow binarization, and structural calculation steps. The system addresses problems of efficiency and inter- and intra-operator variability inherent in previous analyses. The system is evaluated on repetitive scans of 8 volunteers for both two-dimensional (2D) apparent structure calculations and three-dimensional (3D) mechanical calculations using micro-finite element analysis. Coil correction methods based on a priori knowledge of the coil sensitivity and on low-pass filtering of the high-resolution mages are compared and found to perform similarly. Image alignment is found to cause small but significant changes in some structural parameters. Overall the automated system provides on the order of a 3-fold decrease in trained operator time over previous manual methods. Reproducibility is found to be dependent on image quality for most parameters. For 7 subjects with good image quality, reproducibility of 2–4% is found for 2D structural parameters, while 3D mechanical parameters vary by 4–9%, with percent standardized coefficients of variation in the ranges of 15–34% and 20–38% respectively.

Correspondence and offprint requests to: David C. Newitt, MRSC, Box 1290, 1 Irving Street, San Francisco, CA 94143-1290, Tel: +1 (415) 476 9763. Fax: +1 (415) 476 8809. e-mail: newitt@mrsc.ucsf.edu

Keywords: Finite element analysis; High-resolution; MRI; Osteoporosis; Trabecular bone

Introduction

Determination of bone strength is of prime importance in the assessment of fracture risk and the measurement of efficacy of therapeutic interventions in bone disease. The propensity to fracture is determined by both extrinsic factors, such as the force of a fall, and intrinsic factors, including density, structure and composition, which determine the mechanical properties of the bone. Magnetic resonance imaging (MRI) provides unique capabilities for the study of the structural aspects of both trabecular and cortical bone, allowing in vivo, non-invasive, three-dimensional (3D) structure assessment without the use of ionizing radiation. Previous studies have demonstrated the usefulness of both high-resolution imaging (HR-MRI) [1–6] and MR relaxometry [7–11] for deriving structural parameters for the assessment of trabecular bone. For the current study we are concerned with HR-MRI, in which the trabecular structure is determined indirectly from a scan in which the bone marrow shows as a high intensity and the signal voids are assumed to be volumes of bone. Direct imaging of solid materials such as bone is not currently feasible in clinical imaging systems.

For in vivo studies in humans, HR-MRI of bone is generally limited to imaging of the extremities. This is due primarily to signal-to-noise ratio constraints, though

flow and motion artifacts also limit image resolution in the central skeleton. In this study we have used the distal radius site for imaging and analysis. This site has a large quantity of trabecular bone, is easily accessible with localized detection coils for high-resolution imaging, and is convenient for reasonably comfortable immobilization of the subject for the period required for high-resolution volumetric scans.

Analysis of HR-MRI of bone has generally been a very time-consuming and operator-intensive task. The typical process for the cross-sectional studies to date may be broken down into the following steps: (1) preprocessing, (2) segmentation to define the region of trabecular bone, (3) binarization into bone and marrow phases, and (4) structure parameter calculation. For longitudinal studies the added step of image registration would also be required. The segmentation and binarization steps have been particularly problematic, requiring manual tracing of trabecular regions of interest (ROIs) and an often subjective determination of a threshold value for binarization of the image into bone and marrow components. These steps are time-consuming and prone to inter-operator variations. In this paper we describe an analysis system that eliminates subjective thresholding errors and greatly reduces human interaction in the ROI definition step. The system also incorporates image preprocessing, specifically coil correction algorithms for inhomogeneous acquisitions using different coil geometries, and automatic image registration of serial examinations. We demonstrate the reproducibility and increased efficiency of the analysis system on serial scans of 8 normal volunteers. To evaluate the robustness of the automated image registration step we also applied it to 26 subjects with baseline and 12 month follow-up images acquired for a longitudinal drug efficacy study.

Materials and Methods

HR-MRI was performed on a Signa echo-speed system (General Electric, Milwaukee, WI) using a quadrature wrist coil (Medical Advances, Milwaukee, WI) for detection. Subjects were positioned supine with their wrist placed in the coil at their side. As motion during the 12–17 min HR scan is a critical problem, a custom-designed wrist/hand holder was used to immobilize the arm being scanned. Following a coronal locator scan, a 3D fast gradient recalled echo (FGRE) sequence was used to obtain HR axial images of the distal radius with an in-plane resolution of 0.156 mm and a slice thickness of 0.5 mm. For each HR scan a 3 cm (60 slice) section beginning near the distal endplate was processed. Typical HR images are shown in Fig. 1, along with a binarized image used for analysis.

Images were transferred to a Sun workstation for image processing with software developed on-site using the IDL (Research Systems, Boulder CO) and C programming languages. The image processing consisted of five steps, described individually below: coil sensitivity correction, trabecular region segmentation,

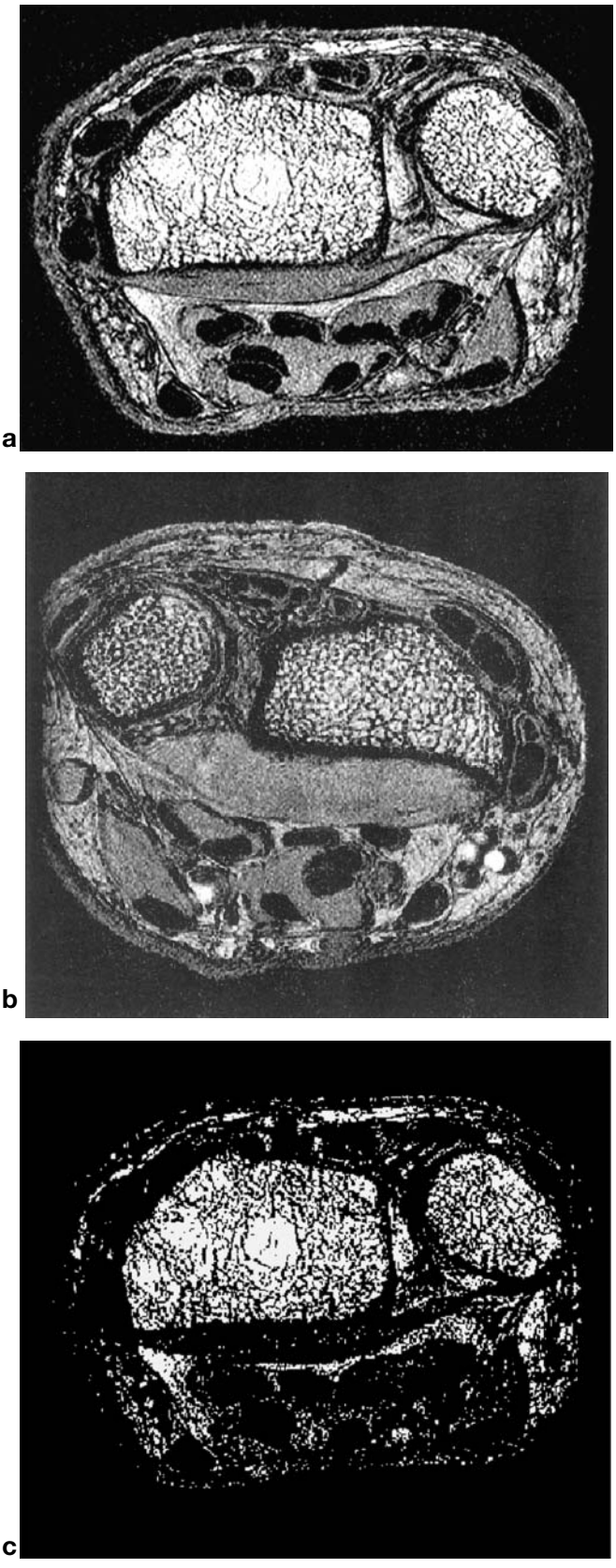


Fig. 1a–c. Typical images from 2 subjects. High-resolution images from: **a** a 52-year-old woman and **b** a 39-year-old man. **c** The slice shown in **a** after thresholding.

manual region adjustment, registration and bone/marrow thresholding. The processed images were then analyzed slice by slice for two-dimensional (2D) apparent structure parameters. In addition, cubic volumes of interest within the processed images were converted to micro-finite element (μFE) models that were used to calculate the apparent mechanical properties of the trabecular bone [12].

Image Processing

MR Detection Coil Inhomogeneity Correction. Correction for the spatial variations in the detection coil sensitivity is required for accurate quantitative analysis of all images acquired with local coils. For trabecular structure analysis of the radius we have implemented two different coil correction schemes for the following types of acquisition coils [13]. For quadrature or birdcage coils with satisfactory in-plane homogeneity a phantom-based, single-axis (longitudinal) correction is used. For all other acquisition coils a modified low-pass filter (LPF)-based correction scheme is used. Images from the reproducibility study were processed with both the phantom-based and the LPF-based correction schemes for comparison.

For phantom-based correction, data from a uniform cylindrical phantom were acquired using an extended-coverage, moderate-resolution, 3D FGRE sequence. One hundred twenty-four 1.0 mm thick slices were acquired at a field of view (FOV) of 8 cm and an acquisition matrix of 256×256. The middle 100 slices were used to obtain the longitudinal sensitivity profile of the coil for a range of ±5 cm relative to the coil center, and a third-order polynomial fit of the sensitivity profile was used to define a longitudinal correction factor versus position function. Correction of the high-resolution bone images is completely automatic, using position information from the image file headers, and is dependent only on accurate placement of the scanner landmark at the center of the coil.

The LPF-based correction is an extension of that described by Wald et al. [14] for correction of surface-coil-acquired brain images, modified as follows. A 3D filter is used to preserve volumetric structure characteristics and to allow a single value thresholding in the binarization process. The LPF algorithm is subject to edge artifacts due to sharp intensity changes and to aliasing artifacts. To minimize edge effects the LP image is generated from a masked image containing only the trabecular bone ROIs, with the background on each slice filled with a constant value. For slices containing an ROI the fill value is taken as the mean intensity of the ROI, while other slices are filled with the mean value of the nearest-slice ROI. While in-plane aliasing effects are not significant, since the radius is in all cases located near the center of the image, aliasing in the longitudinal (SI) direction can cause significant intensity variations in the end slices. These are removed by padding the image with 8 slices at either end, with all pixels set to the mean intensity of the nearest end slice. The LPF correction is a fully automatic part of the image processing requiring no operator interaction.

Automatic Trabecular Region Segmentation. For segmentation purposes, the image was first resampled down by a factor of 2 in each direction, to a resolution of 0.312 mm in plane and 1 mm slice thickness. Initially, the background noise level, found automatically from a region outside the wrist, was used to segment the entire wrist region on all slices. From these regions, a cropped, volume data file was created to increase speed and memory efficiency in the later processing steps. A combination of gray-level erosion, dilation and median filtering operations was used to eliminate the trabecular structure while preserving the cortical shell, resulting in

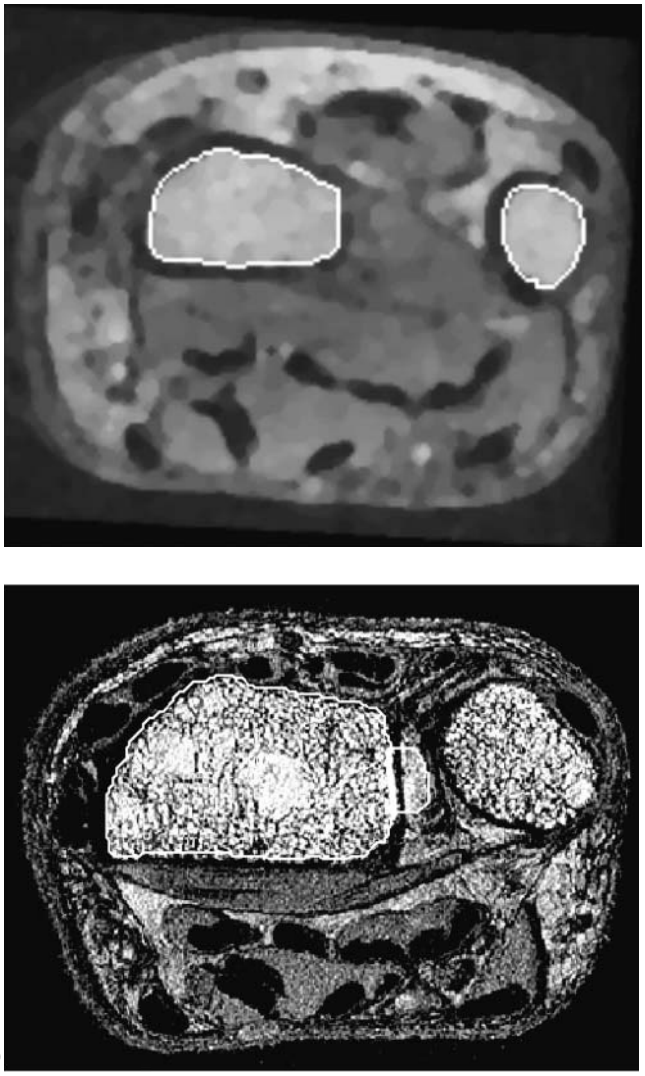


Fig. 2a,b. Trabecular ROI segmentation. **a** Initial, automatically defined trabecular regions for the radius and ulna. **b** Automatically defined region with typical manual adjustment (*thick line*) for ultradistal slice. The image in **a** has been resampled and filtered to remove the trabecular bone structure.

a fairly homogeneous marrow space surrounded by a darker boundary. Using a slice located far enough from the distal endplate to ensure a clear cortical shell, initial regions for the radius and ulna were defined using intensity contouring and pattern matching. Specifically, a contour intensity level was found that resulted in a pair of roughly elliptical regions appropriately located within the wrist outline which satisfied set criteria for size, relative positions, area to boundary length ratio and boundary intensity gradient (Fig. 2a). From these two regions the radius was then chosen using the wrist laterality obtained from the GE image file header. The initial regions were then projected and refined on successive slices in the SI direction using an adaptive intensity threshold level, to define the full 3D trabecular region.

Manual Region Adjustment. A generalized ROI definition program was used by an operator to adjust the regions on slices where the automatic segmentation did not correctly identify the full trabecular bone region (Fig. 2b). The operator also added new regions on slices where the automatic segmentation process failed to identify any trabecular region, usually due to poor signal-to-noise ratio, even though a region was discernible. While these regions were not included in the trabecular structure analysis, they were useful for improving serial image registration.

Image Registration. The coronal scout image on which the HR scan was prescribed was used to provide a longitudinal anatomic reference point for comparing results in similar regions between subjects. An intensity profile was calculated along the axial center of the bone using the centroids of two trabecular ROIs for reference points. The position of the distal endplate was taken to be the point at which the intensity fell to half the peak intensity in the trabecular bone region (Fig. 3).

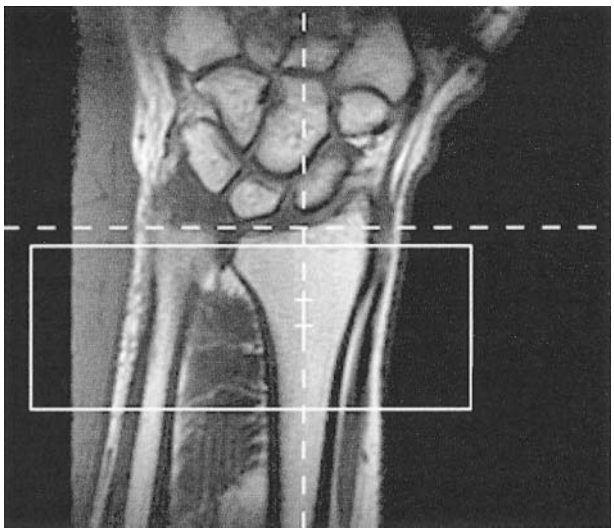


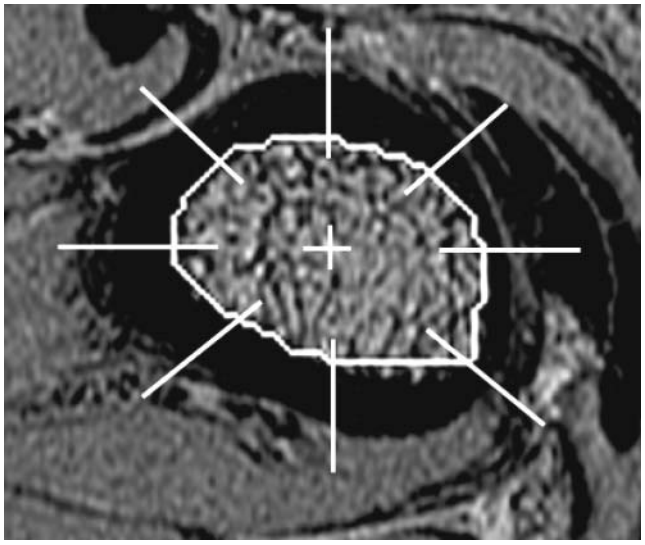
Fig. 3. A typical slice from a coronal locator scan used for definition of the longitudinal reference point. The *white rectangle* indicates the volume selected for the high-resolution (HR) scan, with the two *small crosses* showing centroid points of the radius taken from two HR slices. The *dashed cross* indicates the chosen endplate position, where the intensity falls to 50% of the maximum inside the bone. Note that the profile is actually taken in the 3D volume, and thus may include data from more than one locator slice.

For follow-up scans on the same subject, the manually adjusted ROIs were used to register the images from serial examinations, using a volume and surface matching registration program [15]. This registration was done using only rigid body translations and rotations, and the aligned image was generated using nearest-neighbor sampling as interpolation was found to significantly degrade the images.

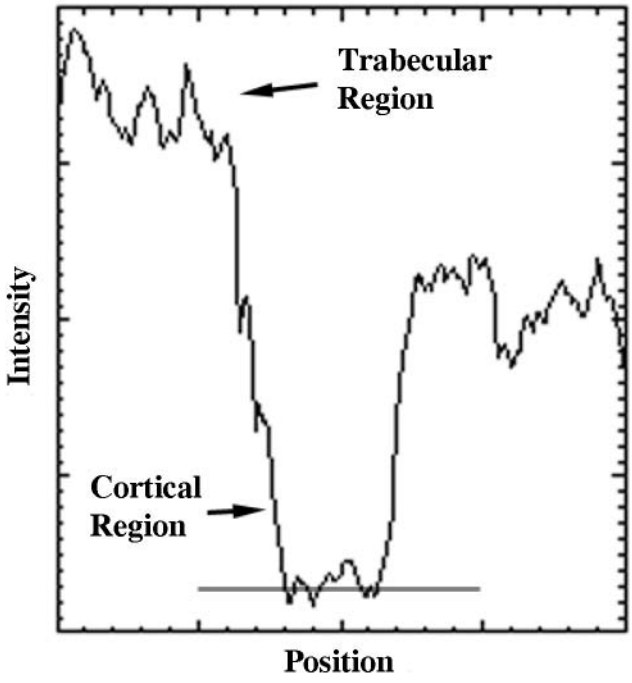
Automatic Bone/Marrow Dual Reference Thresholding. Bone/marrow segmentation requires the determination of two reference intensity levels, I_B and I_M, for bone and marrow respectively. I_B was found by sampling the cortical bone intensity at multiple locations on multiple slices. The samples were taken from intensity profiles measured perpendicular to the trabecular ROIs on slices with a distinct cortical shell (Fig. 4). An averaging width of 7 pixels was used to calculate the profiles to reduce variations from the trabecular structure. The first local intensity minimum outside the trabecular ROI was found for each profile, and the cortical intensity was taken as the mean value of the intensity profile within ± 1 pixel of this location. These values were then averaged for 10 profiles on each slice, and the mean of 20–30 slice values was taken for a global I_B. The upper reference, I_M, was set using the intensity histogram for the trabecular regions of 20 central slices. This histogram consists of a single, non-symmetric peak, due to partial volume effects, noise, and variations in the marrow signal intensity. I_M was taken as the intensity at the peak position plus half the peak width. Using the mean intensity of the trabecular region, I_O, the threshold is set to give a fraction of bone pixels, $f = N_{bone}/N_{total}$, satisfying the equation $fI_B + (1–f)I_M = I_O$ [5]. The resulting binarized image for one slice is shown in Fig. 1c.

Structure Analysis

2D Histomorphometric Analysis. For analysis purposes, ROIs were generated from the full trabecular ROIs by eroding each region by 4 pixels to avoid small inclusions of both the cortical shell and the dense transition region between the cortex and the trabecular bone. Standard 2D algorithms were used to compute the apparent trabecular structural parameters analogous to histomorphometric measures: bone fraction (app.BV/TV), separation (app.Tb.Sp), thickness (app.Tb.Th) and number (app.Tb.N). For statistical analysis, slices were grouped into five slice (2.5 mm thick) groups starting 7 mm from the distal endplate. In addition, the third through fifth groups were further averaged to give a single 7.5 mm thick composite region, labeled ‘R3–5’, for each scan.



a



b

Fig. 4a,b. Typical slice and intensity profiles used for finding the bone intensity reference level. **a** A slice approximately 2 cm from the distal endplate. Radial *white lines* show schematically how the intensity profiles are taken. **b** shows the averaged intensity along one such profile.

Micro-Finite Element analysis. A regular rectangular volume of interest (VOI) extending over 15 slices (7.5 mm) was determined for each subject for finite element modeling. The segmented image within the VOI was used to generate a μFE model by converting the voxels that represent bone tissue to equally shaped 8-node brick elements. The tissue element properties were taken to be linear elastic and isotropic, with a Young's modulus of 10 GPa and a Poisson's ratio of 0.3 for all models. Using a special-purpose FE-solver, six FE-analyses were performed for each specimen, representing compression and shear tests in three orthogonal spatial directions [16]. A homogenization approach was used to calculate the full stiffness matrix for the specimen as a whole from the results of these analyses. An optimization procedure was then used to find a new coordinate system aligned with the best orthogonal symmetry directions of the specimen. The stiffness matrix was rotated to this new orthogonal coordinate system, and the three Young's moduli, three Poisson's ratios and three shear moduli were calculated in these principal directions. The matrices were sorted such that the Young's modulus in the primary direction (E_1) represents the largest modulus of the specimen and the third Young's modulus (E_3) the smallest one: $E_1 \geqslant E_2 \geqslant E_3$. The shear moduli were denoted as G12, G23 and G13. The advantage of this optimization and rotation procedure is that the actual values found for the elastic parameters are independent of the rotation of the specimen.

Study Population and Reproducibility Scanning Protocol

Images for analysis were taken from a reproducibility study involving 8 subjects: 7 women (age range 24–62 years) and 1 man (age 39 years). After a coronal locator scan, a HR volumetric image was acquired. The subjects were each scanned at least three times, and left the scanner table and were repositioned after each HR scan. Typical total examination time for three independent HR acquisitions with coronal locator scans was 60 min for 60 slice acquisitions.

Results

Robustness and Efficiency

The automatic segmentation worked on all the normal volunteer scans, finding radius regions on between 48 and 60 slices for each 60-slice volume acquisition. It failed to identify any radius region only on slices with significantly degraded image quality. This degradation was due to multiple effects, including proximity of the cortical endplate, coil intensity drop-off from incorrect positioning of the wrist, and aliasing in the slice direction due to inaccurate setting of the scan frequency following magnetic field shimming. For the purpose of generating more complete masks for serial image alignment, trabecular ROIs were added manually on slices where there was a discernible trabecular region and cortical shell.

In all cases the automatically generated ROIs required manual adjustment on some of the most distal slices. On these ultradistal slices of the radius the threshold-based segmentation algorithm used here would fail to include areas with a very high apparent trabecular bone density, and would include areas outside the bone in regions where the apparent cortical thickness was comparable to

the apparent thickness of the larger trabeculae. These errors were corrected by manually adding or subtracting from the ROI, as shown in Fig. 2b. This was required both to avoid a bias toward lower trabecular densities and to avoid analyzing non-trabecular regions.

Operator time for segmentation varies greatly depending on image quality, operator experience, operator fatigue and other factors. However, in timed tests using a high-quality image, a fresh, skilled operator took 5 min to perform the manual adjustment of the automatically generated ROIs. A complete manual segmentation of the same data set took 15 min. This 3-fold time difference actually underestimates the time savings of the automated system, as the operator fatigue factor greatly increases with multiple data sets, and it was found to be necessary to have a second operator check manually defined regions to avoid segmentation errors when multiple patient scans needed to be analyzed.

The image registration succeeded on 16 of 17 short-term follow-up scans for the reproducibility study. Of 26 12-month follow-up scans in the longitudinal study the registration failed initially for 2 scans, but both were successfully registered with manual adjustment of the registration surface-matching parameters. As the registration program was set up initially for whole brain images, this indicates some optimization of parameters may be required for reliable fully automatic registration of bone images. When registering two masks, rough measures of the alignment may be taken as the degree of overlap of the two volumes and the mean distance between the two surfaces. For the reproducibility data the registration resulted in a volume overlap of 95.4 ± 1.4%, with a mean distance between surfaces of 0.22 ± 0.04 mm. The associated values for the unaligned images were 82.2 ± 9.8% and 0.83 ± 0.19 mm. For the 12-month follow-ups the volume matching was 95.5 ± 1.8% versus initial values of 65.3 ± 16%. This significantly poorer initial alignment for the 12-month follow-ups illustrates the difficulty in position matching over extended longitudinal studies.

Reproducibility

2D Analysis. Eight baseline scans with 17 associated follow-up scans (2 each for 7 subjects, 3 for 1 subject) were processed for reproducibility data, using both the LPF-based coil correction and the phantom-based correction. The phantom corrected follow-up scans were analyzed both independently, using the endplate location for a longitudinal reference, and by registering the images to the subject's baseline scan. For registered images the analysis was done using the baseline ROIs and longitudinal references. The first and last regions (most distal and most proximal respectively) were discarded as they often show large errors in the registered follow-up scans. This left from 5 to 8 slabs 2.5 mm thick for analysis from each scan. Statistical analysis was done both for all slabs ($n = 111$) and for the composite region consisting of slabs 3 through 5 (12–19.5 mm from the distal endplate) for each scan ($n = 16$ for method comparisons, $n = 8$ for reproducibility).

Table 1 gives the results of a paired *t*-test comparing registered and original images. From the results for all regions we see that the registration transformation has a small but significant effect on the app.Tb.Th (2.4% decrease, $p<0.001$) and app.Tb.N (1.9% increase, $p<0.001$). App.Tb.Sp and app. BV/TV show no significant change. Results for the composite region R3–5 are similar, except that the change in Tb.N is no longer significant.

Table 2 gives the results of a paired *t*-test comparing baseline and aligned images corrected with the LPF-based and phantom-based correction algorithms. From the results for all regions ($n = 179$) and for the composite R3–5 region ($n = 24$) we see that there is little or no average difference between the two methods. Only app.Tb.N (0.64%, 0.010 mm^{-1}, $p<0.04$ and Tb.Sp (-2.1%, 0.010 mm^{-1}, $p<0.05$) showed possibly significant changes.

Reproducibility was calculated for both correction methods for all individual 2.5 mm thick sections ($n = 59$) and for the composite R3–5 region ($n = 8$). Seven subjects had 3 scans, while one had 4. However, the one scan that failed to register properly with the baseline was eliminated, reducing to 2 the number of scans for that subject. Means and standard deviations for each region and for R3–5 were calculated for each subject. The population means and standard deviations (SD), repeated measurement percent coefficients of variation (%CV) and standardized CV (%Std.CV, given by 100 × the root mean square SD of the repeated measurements for each subject divided by the SD of the means), for app.BV/TV, app.Tb.N, app.Tb.Sp, and app.Tb.Th are given in Table 3 for the LPF-corrected images. The %CVs varied from 3.4% to 8.3% for all regions. For the larger R3–5 region they range from 2.5% to 4.0%. The %CVs for the phantom-based method were 0–2% higher. As the scans from one subject (subject 8, female, age 37 years) had noticeably worse

Table 1. Mean percent change in 2D trabecular parameters from non-interpolated registration transformation

	app.BV/TV	app.Tb.N	app.Tb.Sp	app.TbTh
All regions	–0.52	1.9**	–0.8	–2.4**
R3–5	–1.2	1.8	–0.4	–3.0**

**$p<0.001$; all others $p>0.05$.

Table 2. Mean percent change in 2D trabecular structure parameters between different correction algorithms

	app.BV/TV	app.Tb.N	app.Tb.Sp	app.TbTh
All regions	0.78	0.64†	–0.47	0.07
R3–5	1.9	0.91	–2.1†	0.95

†$p<0.05$.

signal-to-noise ratio (SNR) than average, %CVs were also calculated for the R3–5 region excluding this subject. This selection (n = 7 group) shows significantly less variation, with %CVs ranging from 2.2% to 3.2%. Percent Std.CVs were lowest for app.BV/TV, followed by app.Tb.Sp and app.Tb.N. App.Tb.Th has a significantly higher %Std.CV, reflecting the small range of values which results from the quantization error of the relatively large pixel size. Only app.BV/TV and App.Tb.Th showed an appreciable improvement in %Std.CV with elimination of the noisier images.

3D Micro-Finite Element Analysis. MicroFE analysis was done on a rectangular VOI defined on the region from 14.5 mm to 22 mm from the distal endplate, and completely enclosed within the 2D trabecular analysis ROIs. Table 4 gives the results for the trabecular bone fraction of this VOI and the μFE-derived mechanical parameters. The mean and %CV for app.BV/TV are slightly lower than those for the 2D analysis, most likely due to the different regions analyzed, as the rectangular VOI excludes much of the trabecular region nearest the cortex which tends visually to appear denser than the central region. The %CVs for Young's moduli vary from 10% to 22% for all subjects, but are significantly reduced, to 3.9% to 8.7%, in the n = 7 group. Similarly, the shear moduli showed a large reduction in %CV, from 14–21% to 6.3–9.3%, for this reduced sample group. The

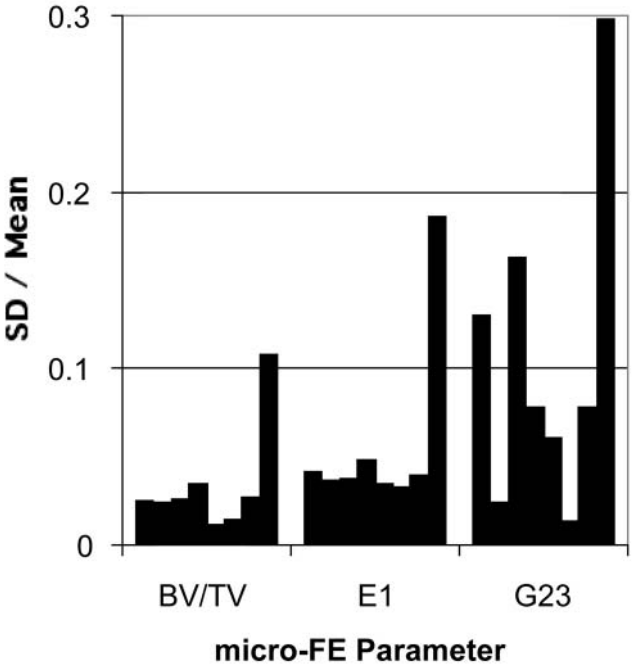


Fig. 5. Coefficients of variation for repeated measurements on each subject for trabecular bone fraction and representative *FE-derived parameters. The far *right bar* for each parameter corresponds to the subject with degraded images.

Table 3. Overall statistical reproducibility measures for 2D apparent trabecular structure parameters of images processed with the low-pass filter (LPF)-based correction

		app.BV/TV	app.Tb.N (1/mm)	app.Tb.Sp (mm)	app.TbTh (mm)
All regions	Mean (SD)	0.358 (0.081)	1.62 (0.25)	0.416 (0.123)	0.219 (0.021)
	%CV	5.2	4.1	8.3	3.4
	%Std.CV	23.1	26.5	28.1	36.0
R3–5	Mean (SD)	0.384 (0.063)	1.71 (0.178)	0.365 0.072)	0.222 (0.015)
	%CV	3.7	2.5	4.0	2.9
	%Std.CV	22.3	24.3	20.2	44.0
n = 7 group[a]					
R3–5	Mean (SD)	0.370 (0.054)	1.67 (0.144)	0.380 (0.063)	0.220 (0.014)
	%CV	2.2	2.2	3.2	2.2
	%Std.CV	14.9	26.0	19.2	33.5

[a]n = 7 group values were calculated excluding one subject with significantly poorer SNR of the images.

Table 4. Statistical reproducibility measures for the rectangular VOI bone fraction and μFE derived mechanical properties

	app.BV/TV	E1	E2	E3	G23	G13	G12
Mean (SD)	0.357 (0.05)	2.3 (0.70)	1.3 (0.50)	0.84 (0.42)	0.39 (0.19)	0.48 (0.21)	0.65 (0.22)
%CV	5.5	10.6	18.1	21.6	20.6	16.8	14.1
%Std.CV	36.8	41.2	55.8	50.6	51.4	45.8	49.7
n = 7 group[a]							
Mean (SD)	0.346 (0.04)	2.2 (0.43)	1.2 (0.26)	0.76 (0.24)	0.35 (0.11)	0.44 (0.13)	0.60 (0.12)
%CV	2.4	3.9	8.3	8.7	9.3	6.7	6.3
%Std.CV	20.7	19.8	37.8	27.0	30.4	23.5	30.7

Young's moduli (E1,2,3) and shear moduli (G23,13,12) are measured in Gpa.
[a]n = 7 group values were calculated excluding one subject with significantly poorer SNR of the images.

Table 5. Statistical reproducibility measures for anisotropy of the Young's moduli derived from μFE analysis

	E1/E2	E1/E3	E2/E3
mean (SD)	1.80 (0.10)	2.90 (0.44)	1.61 (0.20)
%CV	5.9	8.0	5.6
%Std.CV	112	52.6	46.2
n = 7 group[a]			
Mean (SD)	1.83 (0.06)	2.99 (0.39)	1.63 (0.20)
%CV	5.3	7.5	5.6
%Std.CV	177	57.2	46.8

[a]n = 7 group values were calculated excluding one subject with significantly poorer SNR of the images.

coefficients of variation across repeated measurements for each subject (Fig. 5) show the much larger spread of measurements for subject 8 with degraded SNR scans. Percent Std.CVs for the μFE measurements range from 42% to 56% for all 8 subjects, but drop by an average of 21% when the n = 7 group is considered, to 20%–38%.

Anisotropy of the trabecular mechanical properties was also investigated. Table 5 shows the reproducibility statistics for the three Young's modulus ratios. The %CVs for the ratios are generally lower than those for the moduli themselves, and show little change when subject 8 is excluded from the analysis. However, the %Std.CVs for these measures are quite large, ranging from 46% to 177%, and are also unaffected by going to the n = 7 group of subjects.

Discussion

The automatic trabecular region segmentation presented here has been found to save significant trained operator time. It also reduces segmentation errors in the analysis of HR-MRI images compared with manual segmentation. The robustness of the algorithm is found to depend a great deal on image quality. In a larger continuing study involving multiple MRI centers with a greater variation in image quality, the algorithm failed initially for approximately 5% of the scans. Upon inspection, these were all found to be scans with poor positioning, or with significantly degraded apparent SNR, as by patient motion effects. In the cases of positioning errors the automatic segmentation can proceed normally if the operator chooses the initial regions on a suitable slice. Further optimization of the initial region-picking algorithm is expected to reduce or eliminate this operator interaction. The failures due to degraded image quality were all found to involve scans that were of insufficient quality to provide useful trabecular structure information, and thus were excluded from the study. Since the segmentation depends on a clear cortical shell, it is possible that images from a spin echo based sequence, such as the FLASE sequence described by Ma et al. [17], which can provide a sharper bone/marrow boundary, may be more amenable to segmentation with this algorithm. Attempts to apply the algorithm to the calcaneus, another area of trabecular bone, have been unsuccessful due to larger trabecular density variations and a thinner cortical shell. Applications in the femur and tibia are being investigated.

This study does not demonstrate a clear advantage or disadvantage to the use of image registration and alignment for serial studies. The use of registration has clear benefits for some comparisons, such as visual qualitative comparisons of baseline scans with follow-ups. It also allows easy analysis of multiple ROIs, as the regions defined on the baseline scans may be applied directly to the registered follow-up images. We have shown that an analysis using baseline ROIs with non-interpolated registration and one using internally referenced ROIs with unaligned images differed only slightly in their results. Only the app.Tb.Th variation is of a size comparable to the reproducibility, and thus might be of concern. However, as app.Tb.Th is the 2D parameter most affected by the quantization effect of the relatively large voxel size of the MR images, we do not currently consider it as a very important parameter. For the system presented here the two methods of analysis are of comparable efficiency. A significant improvement would result from the implementation of an alignment procedure based on the gray-level images, such as that by Woods et al. [18], rather than the binary masks generated from the trabecular ROIs, as this would eliminate the need for segmenting and finding the longitudinal reference location (endplate location) for all follow-up scans. We are currently researching this possibility.

For images at this resolution the intensity histogram is always monomodal, rather than having distinct bone and marrow peaks, due to partial volume effects, noise, and marrow signal intensity variations. Setting a binarization threshold is therefore difficult. The dual-reference system described in this paper has several advantages over others investigated by the authors. Being completely automated it is completely reproducible, and adds a negligible amount of time to the total analysis process. Other automated methods, such as using the histogram directly to set a threshold, were found to greatly reduce the dynamic ranges of the structure measures. Using a region mean to define the marrow reference level, instead of the histogram method, was found to be problematic in the radius. In the ultradistal region there is often no area of pure marrow large enough to obtain a reproducible value. The distance from the endplate required to provide a relatively 'clean' marrow region differs greatly between subjects, and in some cases requires going to positions where there is significant decease in SNR due to the limited field of view of the coil. Further, it is much more subjective, depending on the operator's choice of region.

Previous work [13] has shown the need for coil correction for binarized analysis of HR trabecular bone images. In the current work we have compared an image-based LPF-algorithm with a phantom-based algorithm. For this data set the results are similar. However, it must be noted that the phantom correction is

dependent on an accurate landmark setting by the technologist taking the scans, and the reproducibility scans analyzed here were taken before this algorithm was implemented. Thus the landmark accuracy issue was not emphasized in the scanning protocol, and cannot be evaluated post hoc. This may account for some or all of the increased variability of these results over the LPF-based results. In theory, either the phantom correction or a purely analytical correction should be superior to the image-based technique, since the latter will tend to reduce actual low spatial frequency intensity variations. Such variations might be caused, for example, by gradual changes in the bone density along the longitudinal direction.

The structural parameters presented in this paper, which are analogs of standard 2D histomorphometric parameters, may not be the ideal measures of trabecular structure when one has 3D volume images of the resolution attainable with MRI. Other measures proposed include tubularity and transverse contiguity [19] and direct measures of 3D structural parameters [20]. The reproducibility of these parameters, and their dependence on image quality, must be and are being established under clinically realistic scanning conditions. Reproducibility of the parameters discussed here, as well as new parameters, must also be established in a subject group with osteoporosis, to determine whether the loss of the trabecular bone component affects the technique reproducibility. These studies are planned, but become increasingly expensive and also difficult from the patient compliance angle. In our longitudinal studies with some form of therapeutic intervention, we have found that immobilization in elderly osteoporotic subjects is possible and enables us to acquire high-quality images, hence we speculate that the technique reproducibility may be stable across different groups.

In summary, we find that with this analysis system we can measure apparent trabecular structure parameters in a group of normal volunteers to a reproducibility of 2–4%, and apparent mechanical properties to approximately 5%, with %Std.CVs of 15–34% and 20–38% respectively. This reproducibility is only possible with high-quality images without significant motion degradation. When lower-quality images are included the CVs increase significantly, especially for the bone fraction measurement. The effect of image quality is seen even more strongly in the large changes in %CV and %Std.CV for the μFE parameters. The statistics for the anisotropy measures are relatively unaffected by the inclusion of the noisy data set, but the %Std.CVs for these parameters are also larger. This particular case of image degradation would appear to give mainly errors in the app.BV/TV, possibly due to thresholding errors, which then are reflected in the μFE-derived moduli which are highly dependent on the bone fraction. Determining whether this is a general or common effect of this type of image degradation will require further study with images of varying qualities. The relevance of the standardized coefficients of variation will be demonstrated when comparing MR with other modalities that emerge that can provide the same type of information. Longitudinal studies are also under way to determine whether actual bone changes with age and/or drug therapy are large enough to be measured with these techniques. This clearly depends on the magnitude of change we expect to measure, the standard deviation of subject response as well as the technique reproducibility. However, if a mean difference between a treatment and non-treatment group of 2% is to be detected with 95% confidence (with a standard deviation between subjects of 2%), our reproducibility results for app.BV/TV and app.Tb.N indicate that we need a minimum number of 26 subjects per group. For a more diverse response to treatment, with a standard deviation of 4%, we would require a minimum of 65 subjects per group.

The relevance of in vivo bone quality assessment in osteoporosis, using MRI, needs to be established. The initial cross-sectional results, the current technique evaluation in vivo, and continuing longitudinal studies are small steps in that direction. With emerging technologies, assessed by larger multicenter studies in the next several years, these methods may play a valuable part in explaining skeletal changes with aging, osteoporosis and therapeutic response.

Acknowledgements. The authors wish to thank Dr Sarah Nelson for assistance with image registration, and Drs Brian MacDonald (Smith Kline Beecham Pharmaceuticals), Charles Chesnut (UW), and Harry Genant and Steve Harris (UCSF), for assisting in the initiation of a longitudinal study, for which this validation study was essential. We also thank them for many helpful discussions.

References

1. Wehrli FW, Hwang SN, Song HK. New architectural parameters derived from micro-MRI for the prediction of trabecular bone strength. Technol Health Care 1998;6:307–320.
2. Majumdar S, Genant HK. Assessment of trabecular structure using high resolution magnetic resonance imaging. Stud Health Technol Inform 1997;40:81–96.
3. Chung HW, Wehrli FW, Williams JL, Wehrli SL. Three-dimensional nuclear magnetic resonance microimaging of trabecular bone. J Bone Miner Res 1995;10:1452–61.
4. Majumdar S, Kothari M, Augat P, et al. High-resolution magnetic resonance imaging: three-dimensional trabecular bone architecture and biomechanical properties. Bone 1998;22:445–54.
5. Majumdar S, Genant HK, Grampp S, et al. Correlation of trabecular bone structure with age, bone mineral density, and osteoporotic status: in vivo studies in the distal radius using high resolution magnetic resonance imaging. J Bone Miner Res 1997;12:111–8.
6. Majumdar S. A review of magnetic resonance (MR) imaging of trabecular bone micro-architecture: contribution to the prediction of biomechanical properties and fracture prevalence. Technol Health Care 1998;6:321–7.
7. Wehrli FW, Ma J, Hopkins JA, Song HK. Measurement of R'2 in the presence of multiple spectral components using reference spectrum deconvolution. J Magn Reson 1998;131:61–8.
8. Wehrli FW, Hwang SN, Ford JC, Jara H. Method for image-based T2* measurement in bone marrow. In: Proceedings of the second annual meeting of the ISMRM, San Francisco, 1994. p 201.
9. Newitt DC, Majumdar S, Jergas MD, Genant HK. Decay characteristics of bone marrow in the presence of a trabecular bone network: in vitro and in vivo studies showing a departure from monoexponential behavior. Magn Reson Med 1996;35:921–7.

10. Yablonskiy D, Reinus W, Stark H, Haacke E. Quantitation of T2′ anisotropic effects on magnetic resonance bone mineral density measurement. Magn Reson Med 1997;37:214–21.
11. Jergas M, Majumdar S, Keyak J, et al. Relationships between Young's modulus of elasticity, ash density and magnetic resonance imaging (MRI) derived effective transverse relaxation time T2* in human tibial specimens. J Comput Aided Tomog 1995;19:472–9.
12. van Rietbergen B, Weinans H, Huiskes R, Odgaard A. A new method to determine trabecular bone elastic properties and loading using micromechanical finite-element models. J Biomech 1995;28:69–81.
13. Newitt DC, Majumdar S. Correction for receiver coil inhomogeneity profiles for quantitative analysis of trabecular bone structure from high resolution MRI. In: Proceedings of the seventh annual meeting of the ISMRM, Philadelphia, PA, 1999, p 1046.
14. Wald LL, Carvajal L, Moyher SE, et al. Phased array detectors and an automated intensity-correction algorithm for high-resolution MR imaging of the human brain. Magn Reson Med 1995;34:433–9.
15. Nelson SJ, Majumdar S. Automated alignment of MR images of the radius. In: Twelfth annual meeting of the Society of Magnetic Resonance in Medicine, New York, 1993, p 392.
16. van Rietbergen B, Odgaard A, Kabel J, Huiskes R. Direct mechanics assessment of elastic symmetries and properties of trabecular bone architecture. J Biomech 1996;29:1653–7.
17. Ma J, Wehrli FW, Song HK. Fast 3D large-angle spin-echo imaging 3D FLASE. Magn Reson Med 1996;35:903–10.
18. Woods RP, Cherry SR, Mazziotta JC. Rapid automated algorithm for aligning and reslicing PET images. J Comput Assist Tomogr 1992;16:620–33.
19. Hwang S, Wehrli F, Williams J. Probability-based structural parameters from three-dimensional nuclear magnetic resonance images as predictors of trabecular bone strength. Med Phys 1997;24:1255–61.
20. Laib A, Rüegsegger P. Comparison of structure extraction methods for in vivo trabecular bone measurements. Comput Med Imaging Graphics 1999;23:69–74.

Received for publication 11 June 2001
Accepted in revised form 6 November 2001