



Improved reproducibility of high-resolution peripheral quantitative computed tomography for measurement of bone quality

Joshua A. MacNeil, Steven K. Boyd*

*Department of Mechanical and Manufacturing Engineering, Schulich School of Engineering, University of Calgary,
2500 University Drive, N.W., Calgary, Alberta T2N 1N4, Canada*

Received 6 March 2007; received in revised form 2 November 2007; accepted 12 November 2007

Abstract

A human high-resolution peripheral quantitative computed tomography scanner (HR-pQCT) (XtremeCT, Scanco Medical, Switzerland) capable of measuring three important indicators of bone quality (micro-architectural morphology, mineralization and mechanical stiffness) has been developed. The goal of this study was to evaluate the reproducibility of male and female HR-pQCT in vivo measurements, and elucidate the causes of error in these measurements through a comparison with in vitro measurements.

The best possible short-term reproducibility was found using a set of 10 in vitro measurements without repositioning, and a set of 10 with repositioning. Subsequently, in vivo measurements were performed on 15 male and 15 female subjects at baseline and follow-ups of 1 week and 4 months to determine the short- and long-term reproducibility of the system. In addition to the 2D area matching method used in the standard evaluation protocol, a custom developed 3D registration method was used to find the common region between repeated scans.

The best possible reproducibility without movement artifacts and repositioning error was less than 0.5%, while the reproducibility with repositioning error was less than 1.5%. The in vivo reproducibility of density (<1%), morphological (<4.5%) and stiffness (<3.5%) measurements was consistently poorer than the reproducibility of cadaver measurements, presumably due to small movement artifacts and repositioning errors. Using 3D image registration, repositioning error was reduced on average by 23% and 8% for measurements of the radius and tibia sites, respectively.

This study has provided bounds for the reproducibility of HR-pQCT to monitor bone quality longitudinally, and a basis for clinical study design to determine detectable changes.

© 2007 IPEM. Published by Elsevier Ltd. All rights reserved.

Keywords: Osteoporosis; Densitometry; Micro-computed tomography; Reproducibility; Bone

1. Introduction

Areal bone mineral density (BMD) measurements are the traditional method for diagnosing and monitoring osteoporosis (OP) and provide a measure of the average density of a bone sample, when projected onto a 2D surface. It is well known that age related decrease in bone mineral content (BMC) or BMD structurally weakens bone and thus increases the risk of fracture; however, there is an increasing body of evidence suggesting a combination of factors including bone architecture, turnover, damage accumulation and mineraliza-

tion play an important role in the assessment of fracture risk [1]—jointly referred to as *bone quality* [2].

Micro-computed tomography (μ CT) is a non-destructive technique that can measure three important determinants of bone quality: micro-architectural morphology, mineralization and mechanical stiffness on a same sample of bone [3,4]. A human high-resolution peripheral quantitative computed tomography scanner (HR-pQCT) (XtremeCT, Scanco Medical, Switzerland; voxel size 82 μ m) recently developed can perform human in vivo measurements at peripheral sites including the wrist and tibia. The standard HR-pQCT morphological measurement protocol includes a direct measure of trabecular number (Tb.N), and derived measures of bone volume ratio (BV/TV^d), trabecular thickness (Tb.Th), trabecular separation (Tb.Sp) and cortical thickness (Ct.Th) [5,6].

* Corresponding author. Tel.: +1 403 220 4173; fax: +1 403 282 8406.

E-mail addresses: jamacnei@ucalgary.ca (J.A. MacNeil),
sboyd@ucalgary.ca (S.K. Boyd).

From a previous study, it was determined that additional morphometric parameters could be measured by HR-pQCT including direct measurements of BV/TV*, Ct.Th, connectivity density (Conn.D), and degree of anisotropy (DA) [7]. In addition to morphology, mineralization measurements can be performed including total BMD, trabecular (Tb) BMD, and cortical (Ct) BMD. Thus, changes in each of these compartments can be assessed independently from the effects of physical size.

Morphological and density measurements have been shown to have good correlations to mechanical stiffness and strength in the distal radius [8,9]; however, these measures typically provide an average value for a measurement of a whole bone sample. Therefore, it is possible that two bone samples can have similar average values for their bone structure and density, but different overall structures and density distributions and may respond to loading in different manners. Finite element modeling based on high-resolution structural information allows the response of a bone sample to mechanical loading to be examined, and has been previously shown to be effective at predicting Colles' type fractures using full wrist joint models as well as using a section of the radius [8,10–12]. Finite element modeling using in vivo measurements from the human radius and tibia has the potential to provide insight into disease- and treatment-related bone quality changes, and the reproducibility of those measurements are therefore important to determine.

Bone quality measurements for the assessment of fracture risk will likely consist of a baseline bone quality measurement, and follow-up longitudinal measurements that will be compared to the baseline measurement. This will allow for an assessment of the rate of change of bone quality as well as determination of least significant changes. The reproducibility of in vivo measurements will be affected by sources of error including: inherent error in the system (stochastic error), movement artifacts, physical positioning, and scanning region selection for maximizing repeated measurements. Although movement artifacts can be reduced by providing careful instructions to the subject and immobilization (strapping the extremity into the sample holder, Fig. 1), movement artifacts cannot be completely removed. Repositioning is challenging, particularly because landmarks used for selecting the scanning region may not be clear. It is nearly impossible to select the exact same region in follow-up scans and this error results in reduced reproducibility due to inconsistent overlap. While the overlap error is partially accounted for automatically by matching slice-by-slice bone areas for determination of common regions (Image Processing Language, v4.29, Scanco Medical), the 2D area matching method does not account for 3D repositioning errors. Therefore, a 3D method using image registration [13] may offer an improvement over the 2D area matching method and reduce the effects of repositioning between follow-up scans resulting in improved short- and long-term reproducibility.

The goals of this study are to: (i) determine the affect of movement artifacts and repositioning error by finding the



Fig. 1. Standard forearm fixation (top) and lower leg fixation (middle) in the HR-pQCT forearm scanning cast provided by the manufacturer. After completion of this study, forearm fixation with the addition of foam wedges (bottom) to decrease possible movement artifacts during scanning are routinely used to further reduce movement artifacts.

inherent reproducibility of the system (best case scenario), (ii) determine the short- and long-term reproducibility of morphological, density, and finite element stiffness measurements of the distal radius and tibia (with 2D area matching correction); and (iii) determine if 3D image registration can improve the short- and long-term reproducibility of bone quality measurements by reducing repositioning error. The reproducibility of measurements of males and females will be determined separately to determine if combining these measurements could artificially improve the reproducibility of bone quality measurements of females [6].

2. Methods

2.1. Subjects

A single cadaver forearm was used to determine the best possible reproducibility of the system. The best possible reproducibility can only be assessed using in vitro mea-



Fig. 2. Standard HR-pQCT measurement location for the distal radius (left) and tibia (right) showing the reference landmark location (solid line) and the scanning region (dashed lines).

measurements because in vivo measurements inherently contain movement artifacts. The fresh frozen forearm was disarticulated at the elbow (left forearm, male, 81 years old), thawed and scanned with the soft tissue intact.

In vivo measurements were performed on a total of 30 human subjects: 15 male (ages 20–37 years) and 15 female (ages 20–40 years). No subjects had a history of previous fractures or bone-related ailments (self reported). All subjects (1 male dropped out before the 4 month scan) were scanned at both their dominant and non-dominant radius and tibia (4 sites) at baseline and follow-ups of 1 week and 4 months (Fig. 1). All procedures were approved by the Office of Medical Bioethics at the University of Calgary.

2.2. Scanning by HR-pQCT

Both in vitro and in vivo measurements used the same method to identify the 3D region to be scanned. The identification was based on a ‘scout view’ (2D X-ray) from which landmarks could be identified (Fig. 2). The landmark on the tibia was the distal articular plateau (articulating surface with the talus) and the landmark on the radius was a notch between

the articulating surfaces for the lunate and navicular joint in the wrist. The scanning region was offset 22.5 mm for the tibia and 9.5 mm for the radius as specified by the manufacturer.

The single cadaver forearm was scanned 10 times without repositioning to determine the inherent reproducibility of the scanning system. It was then scanned again 10 times with repositioning (forearm was removed from the scanner and reinserted) at the standard distal location for in vivo patient HR-pQCT measurements to determine the effects of repositioning.

Short- and long-term reproducibility was determined through in vivo patient measurements of the radii and tibiae and were scanned at baseline and follow-ups of 1 week and 4 months. The default human peripheral in vivo scanning protocol (60 kVp, 1000 μ A, 100 ms integration time) was used for each scan. The total time for each measurement was approximately 3 min and resulted in 110 slices or a 9.02 mm long 3D representation of the radius and tibia (Fig. 3). The effective dose from a single scan was approximately 3 μ Sv, which is well below the recommended annual dose limit (approximately 50 mSv) for general public (International Commission on Radiological Protection). The baseline ‘scout view’ was used for comparison to 1 week

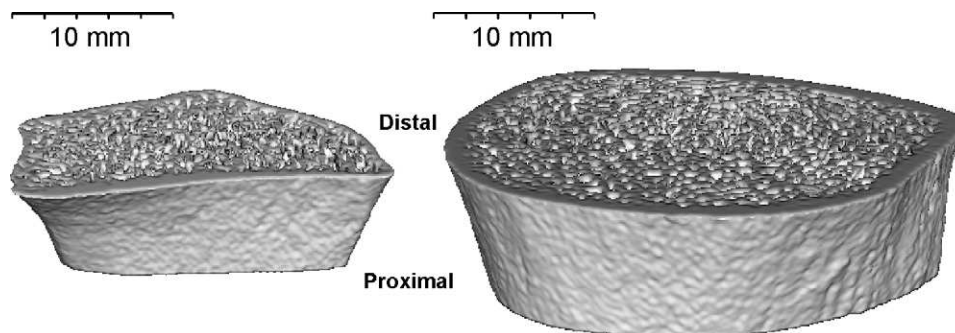


Fig. 3. Representative HR-pQCT distal radius (left) and tibia (right) cross-sections. Morphological, density and, stiffness measurements were performed on all samples. The same scales are used to highlight the typical relative size difference.

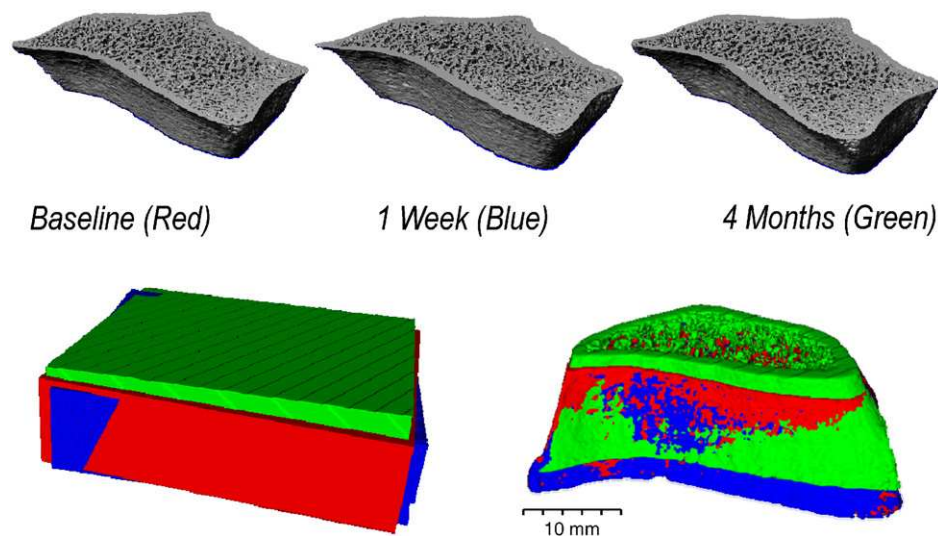


Fig. 4. Example using 3D image registration to find the common region between scans at baseline (top left), 1 week (top center), and 4 months (top right). The overlap between the representative volumes for the scans (bottom left) is used to identify the overlap between the three-segmented images (bottom right). The common region (intersection) is then extracted and analyzed in all three images to improve reproducibility. Note that the misalignment (bottom left) includes both translational and rotational components.

and 4 months follow-up scans as per the standard operational protocol for the HR-pQCT system.

2.3. Density and morphological measurements

Measurements performed on the HR-pQCT images were categorized as either a default patient evaluation or extended evaluation. The first method used the default patient evaluation protocol available on the system. The radius and tibiae were segmented from the image data using a semi-automated contouring method. The evaluation software automatically extracted both the cortical and trabecular mineralized phase using a Laplace Hamming filter and segmented the cortical and trabecular bone both using default values specified by the manufacturer. These values are used consistently for all HR-pQCT scanners. The standard evaluation protocol resulted in measures of total BMD, cortical (Ct.) BMD, trabecular (Tb.) BMD, BV/TV^d, Tb.N, Tb.Sp, Tb.Th, and Ct.Th.

The standard evaluation protocol provided by the manufacturer based on the 2D area matching method was used to determine the common region between baseline, 1 week, and 4 months for the first round of evaluations. To determine if an improvement in reproducibility could be obtained using 3D image registration, the full 3D scans from 1 week and 4 months were registered to the baseline scan with rigid registration using Mattes mutual information with a linear interpolator in a multi-resolution scheme (Fig. 4). For further details about the 3D registration methods see Boyd et al. [13]. The standard evaluation protocol was repeated on the 3D registered data, with the exception of Ct.Th (it required a proprietary file that could not be created). In addition to the standard morphological parameters, additional evaluations including direct measurements of BV/TV*, Ct.Th, Conn.D,

and DA were also completed based on the 2D area matching method and 3D image registration.

2.4. Mechanical stiffness

The HR-pQCT radii and tibia-segmented 3D datasets were automatically converted to finite element meshes using the voxel conversion approach [14]. The boundary conditions applied were uniaxial displacement to 1% strain. The apparent Young's modulus (E) of the cross-sections was calculated based on the reaction force (RF) to induce 1% strain, and average area of the slices of the section. The radius models typically contained 10 million degrees of freedom, while the tibia models typically contained 54 million degrees of freedom and were solved using custom element by element finite element software on a parallel shared memory supercomputing system (64 bit 1.9 GHz processors, 4 GB of RAM per processor) [15].

Table 1
HR-pQCT reproducibility measured in percent root mean squared coefficient of variance

Measurements radius	No repositioning (%)	Repositioning (%)
BV/TV ^d (%)	0.000	0.715
Tb.N (mm ⁻¹)	0.170	0.515
Tb.Th (mm)	0.262	1.07
Tb.Sp (mm)	0.195	0.468
Ct.Th (mm)	0.000	0.820
BMD (g/cm ³)	0.197	0.289
Tb.BMD (g/cm ³)	0.000	0.714
Ct.BMD (g/cm ³)	0.096	0.332

Measurements were performed on a cadaver forearm without and with repositioning. The 2D area matching method used in the standard evaluation protocol was used to select the common region for the follow-up measurements.

2.5. Statistical analysis

Root mean square coefficient of variance was used to determine the absolute error of the system and the effects of repositioning. Root mean square coefficient of variance was used to determine the short-term reproducibility of the bone density, architecture and mechanical stiffness measurements. The long-term reproducibility of the measurements were determined based on the standard error of the estimate of the repeated measurements [16]. The long-term reproducibility was calculated using the standard error of the estimate, which performs a regression analysis, and determines the standard error for each measurement from the line of best fit from the regression analysis. Therefore, a trend of decreasing or increasing morphological, density or strength parameters will have less impact on the long-term reproducibility. To test for statistical differences between two root mean square coefficient of variances, an analysis of variance was used based on a method developed by Sokal and colleagues [17].

3. Results

The best possible reproducibility from the cadaver measurements is given (Table 1) and the absolute error in the scanning system was found to be less than 0.3% for the standard measurement protocol. This represents the upper limit of reproducibility of the system without movement artifacts and without repositioning errors. The system reproducibility without movement artifacts, but with repositioning error was found to be less than 1.2% for all measurements in the standard measurement protocol, and represents the upper limit of the 2D area matching method for finding the common region between multiple scans.

In vivo measurements that were observed to have motion artifacts (based on qualitative observations of artifacts such as blurring or streaks in reconstructed 2D slices) were excluded from the study, which resulted in 22 radius and 22 tibia measurements of males and 22 radius and 22 tibia measurements of females. In vivo reproducibility based on short- and long-term follow-up scans of male and female subjects at the radius and tibia are provided (Table 2). The average common regions using the 2D area matching method from the default patient evaluation for measurements of radii and tibiae were found to be 92% and 91% for males, respectively, and 96% and 95% for females, respectively.

For the radius site, the root mean square coefficient of variance (CV_{rms}) for BMD, Tb.BMD, Ct.BMD, and BVTVD measurements were found to be generally less than 1%. The reproducibility of Tb.N was greater ($CV_{rms} < 3.9\%$), and subsequently the semi-derived measurements had similar reproducibility, including Tb.Th ($CV_{rms} < 4.0\%$) and Tb.Sp ($CV_{rms} < 4.5\%$).

For the tibia site, the CV_{rms} for BMD, Tb.BMD, Ct.BMD, and BVTVD were less than 0.55%, while CV_{rms} of Tb.N, Tb.Th, and Tb.Sp measurements were less than 3.7%. The long-term reproducibility of measurements of the radius and tibia were found to be equal to or slightly better than the short-term reproducibility. The long-term reproducibility precision was based on the standard error of the estimate from a regression analysis and therefore, accounted for trends in the measurements that may have been contributing to differences.

The reproducibility of the stiffness measurements were comparable to the morphological measurements. The reproducibility for measurements of males and females were significantly different ($p < 0.05$) for the estimation of reaction force (RF) under 1% strain. Short- and long-term measure-

Table 2

Radius and tibia short- and long-term reproducibility of the HR-pQCT system for males and females measured in percent root mean squared coefficient of variance

Avg. com. region	Radius male		Radius female		Tibia male		Tibia female	
	92%		91%		96%		95%	
Measurements	Short term	Long term	Short term	Long term	Short term	Long term	Short term	Long term
BV/TV ^d (%)	0.613	0.682	0.522	0.487	0.525	0.437	0.790	0.840
Tb.N (mm ⁻¹)	3.60	3.55	3.36	3.20	3.59	3.15	3.89	3.58
Tb.Th (mm)	3.41	3.37	3.10	2.74	3.62	3.39	3.85	3.47
Tb.Sp (mm)	4.40	3.94	3.28	2.95	3.51*	3.31	4.00*	3.76
Ct.Th (mm)	1.35	1.43	0.663	0.772	0.859	0.887	2.24	2.04
BMD (g/cm ³)	0.624	0.634	0.386	0.458	0.481	0.529	1.03	0.997
Tb.BMD (g/cm ³)	0.662	0.688	0.489	0.495	0.535	0.490	0.912	0.851
Ct.BMD (g/cm ³)	0.339	0.327	0.404	0.290	0.247	0.275	0.582	0.548
E	2.36	2.27	1.44	1.40	1.60*	1.60*	3.55*	3.47*
RF (N)	2.47	2.36	1.44†	1.40†	1.86*	1.83*	4.28†*	4.18†*
von Mises stress	2.48	2.40	1.38	1.36	1.87	1.87	3.19	3.14
Strain energy den.	3.41	3.30	1.89	1.86	2.56*	2.55*	4.22*	4.17*

The common region was determined using the standard 2D area matching method from the standard evaluation protocol. The reproducibility of morphological, density (gray shading), and stiffness measurements was determined. A † symbol indicates a significant difference ($p < 0.05$) between males and females. A * symbol indicates a significant difference between the tibia and radius.

Table 3

Radius and tibia short- and long-term reproducibility of the HR-pQCT system for males and females using 3D image registration to find the common region between longitudinal scans

Avg. com. region	Radius male		Radius female		Tibia male		Tibia female	
	86%		82%		88%		88%	
Measurements	Short term	Long term	Short term	Long term	Short term	Long term	Short term	Long term
BV/TV ^d (%)	0.587	0.590	0.474	0.475	0.484	0.471	0.800	0.800
Tb.N (mm ⁻¹)	3.52	3.50	3.22	3.14	3.26	3.24	3.33	3.30
Tb.Th (mm)	3.33	3.30	2.97	2.88	3.35	3.40	3.28	3.28
Tb.Sp (mm)	3.90	3.92	3.00	2.94	3.43	3.49	3.47	3.51
Ct.Th (mm)	1.30	1.34	1.32	1.32	2.20	2.23	1.55	1.57
BV/TV* (%)	1.05	1.04	0.936	0.923	1.03	1.04	0.898	0.932
Conn.D (mm ⁻³)	6.39	6.31	9.15†	8.83†	6.47*	6.43*	4.93†*	4.97†*
DA	2.07	2.04	2.16	2.14	2.17	2.22	2.84	2.83
BMD (g/cm ³)	0.462	0.472	0.393	0.388	0.333	0.327	0.691	0.688
Tb.BMD (g/cm ³)	0.587	0.590	0.473	0.475	0.484	0.471	0.800	0.800
Ct.BMD (g/cm ³)	0.227	0.229	0.278	0.272	0.281	0.278	0.271	0.256

Reproducibility is measured in percent root mean squared coefficient of variance and standard error of the estimate. The common region was found using 3D image registration. The precision of morphological, density, and stiffness measurements was determined.

Table 4

Absolute difference in percent root mean squared coefficient of variance between standard 2D area matching method and 3D image registration for radius and tibia short- and long-term precision of the HR-pQCT system for males and females

Measurements	Radius male		Radius female		Tibia male		Tibia female	
	Short term	Long term	Short term	Long term	Short term	Long term	Short term	Long term
BV/TV ^d (%)	0.026	0.092	0.050	0.010	0.040	−0.030	−0.010	0.040
Tb.N (mm ⁻¹)	0.08	0.05	0.14	0.06	0.33	−0.09	0.56	0.28
Tb.Th (mm)	0.08	0.07	0.13	−0.14	0.27	−0.10	0.57	0.19
Tb.Sp (mm)	0.50	0.02	0.28	0.01	0.08	−0.18	0.53	0.25
BMD (g/cm ³)	0.160	0.160	−0.010	0.070	0.150	0.200	0.340	0.310
Tb.BMD (g/cm ³)	0.075	0.098	0.020	0.020	0.050	0.020	0.110	0.051
Ct.BMD (g/cm ³)	0.110	0.098	0.130	0.020	−0.030	0.000	0.310	0.290

The precision of morphological, density, and stiffness measurements was determined. A positive value is an increase in precision with 3D registration.

ments at the tibia site were found to be significantly more precise than measurements at the radius for measurements of strain energy density (SED) for males, and in females for measurements of *E*, RF, and SED.

3D image registration was used to find the common region for measurements of male and female radii and tibiae, and the calculated overlap was decreased to averages of 86% and 82% for male radii and tibiae sites, respectively, and

88% and 88%, for female radii and tibiae sites, respectively. Using 3D image registration there was an improvement in root mean squared coefficient of variance (Table 3) of over 0.5% for the short-term morphological measurements (Tb.N, Tb.Th, Tb.Sp) of the female radius. Image registration improved the root mean squared coefficient of variance (Table 4) by an average of 0.2% and an average relative increase (Table 5) of 23% for the short- and long-term

Table 5

Relative difference in percent root mean squared coefficient of variance between standard 2D area matching method and 3D image registration for radius and tibia short- and long-term precision of the HR-pQCT system for males and females

Measurements	Radius male		Radius female		Tibia male		Tibia female	
	Short term	Long term	Short term	Long term	Short term	Long term	Short term	Long term
BV/TV ^d (%)	4.43	15.6	10.1	2.53	8.47	−7.22	−1.25	5.00
Tb.N (mm ⁻¹)	2.27	1.43	4.35	1.91	10.1	−2.78	16.8	8.48
Tb.Th (mm)	2.40	2.12	4.38	−4.86	8.06	−0.29	17.4	5.79
Tb.Sp (mm)	12.8	0.51	9.33	0.34	2.33	−5.16	15.3	7.12
BMD (g/cm ³)	35.1	34.3	−1.78	18.0	44.4	61.8	49.1	44.9
Tb.BMD (g/cm ³)	12.8	16.6	3.38	4.21	10.5	4.03	14.0	6.38
Ct.BMD (g/cm ³)	49.3	42.9	45.3	6.62	−12.1	−1.08	114	114

The precision of morphological, density, and stiffness measurements was determined. A positive value is an increase in precision with 3D registration.

reproducibility of the male and female radius when compared to the standard 2D overlap method. However, none of these improvements were statistically significant, and a small decrease in the female short-term reproducibility of BV/TV was found. There was an average improvement in root mean squared coefficient of variance (Table 4) of 0.04% and an 8% average relative improvement (Table 5) for the short- and long-term reproducibility of the male and female tibia measurements. A small decrease in the long-term reproducibility for the structural measurements in females was found.

4. Discussion

This study evaluated the reproducibility of HR-pQCT to determine the bounds of this new technology for longitudinal bone morphological, density, and stiffness measurements. The contribution of movement artifact and repositioning error was evaluated, and by their elimination in the cadaver measurements, the inherent machine reproducibility was determined. The short- and long-term reproducibility of morphological, density and finite element stiffness measurements of the distal radius and tibia using the standard 2D area matching method was found, and it was shown that 3D image registration generally improved those results. The reproducibility for measurements of males and females were statistically different for reaction force to induce 1% strain, but not other parameters measured.

A limitation of this study was that movement artifacts could not be explicitly controlled for during in vivo measurements, and were difficult to identify based on the single reconstructed slice that is provided at the time of patient measurement (remaining data is reconstructed online after patient departure). The effect of motion artifacts on the measurement parameters are not fully understood; however, in some cases where patients were re-scanned immediately (for other studies) anecdotal results indicate that density measures are more stable than structural parameters. This is not surprising since the blurring of image data should have little influence the density measurements compared to the structural elements and this is reflected in this study in the better in vivo reproducibility measured for density parameters compared to structural parameters. The reproducibility estimated in this study is representative of the current methodologies used to acquire HR-pQCT scans, and is therefore relevant to the clinical usage of this device. However, it is possible that improved reproducibility could be achieved via reduction of movement artifacts. In the future, a full characterization of the effects of motion artifacts needs to be determined as well as methods to detect them quantitatively. Another clear limitation of this study is that the use of younger subjects may have contributed to an overestimation of reproducibility (i.e., not representative of elderly or osteoporotic patients). Performing a reproducibility study using elderly patients is problematic as their bone quality may be unstable, and elderly

subjects may have more difficulty remaining still (i.e., tremor) thus increasing the risk of movement artifacts. Therefore, the results presented here, even for our in vivo measurements, represent a best case scenario for maximizing reproducibility.

The best possible system reproducibility is less than 0.26% if movement artifact and repositioning errors are removed, and increases to 1.07% if repositioning errors are included. The in vivo reproducibility estimates are higher, and were of similar magnitude to previously reported results for density and morphological measurements [5] when the standard 2D overlap correction method was employed. The application of 3D registration generally improved reproducibility, and the improvement was most notable with respect to the density measurements which ranged from 0.23% to 0.80%. Putting the HR-pQCT reproducibility estimates into perspective, the annual rate of bone loss for healthy women is reported to be 0.5–0.9% and 0.9–1.1% for trabecular and total BMD, respectively [18]. Therefore, taking into account the least significant change that is detectable based on the measured reproducibility of HR-pQCT density measurements [19], it should be theoretically possible to detect BMD changes within the timeframe of approximately a year in the normal aging population, and even sooner for osteoporotic patients with higher bone loss rates.

This system reproducibility can be improved through the use of 3D image registration to reduce repositioning artifacts. The 2D area matching currently used in the standard evaluation protocol adjusts the common region on the basis of removing whole slices from the 3D dataset, and cannot account for out-of-plane (i.e., 3D) differences. A 3D method to select the common region helps to remove the effects of repositioning while maximizing the volume retained for analysis (Fig. 4). Generally, this led to an improvement in reproducibility, most markedly in the measurements of total BMD and Ct.BMD. The variation of the cortex volume along the length of the bone, particularly at the radius site, may have resulted in a higher sensitivity to 3D registration (i.e., more weight when 3D position is corrected) thus explaining the large improvements. In contrast, the effect of a 3D method to select the common region on the reproducibility of the trabecular morphological parameters (and Tb.BMD) was not as strong. This is possibly because the cancellous structure is relatively uniform along the length of the bone in the measurement region. The stiffness measurements take into account the whole 3D bone structure measured, including the cortical and cancellous region, and it was encouraging that the reproducibility of these parameters were similar to the morphological parameters because they may provide a significant addition to bone quality measurements in the future, particularly to distinguish changes in osteoporotic fracture risk.

There were a few cases where 3D registration resulted in a modest reduction in reproducibility (negative signs in Table 4); however, these results predominantly occur in the long-term evaluations which uses the statistical method of the standard error of the estimate of the repeated measurements to

determine reproducibility, and may have been due to a trend artificially detected by the 2D area matching method. Despite improvements in reproducibility through the use of 3D registration, in vivo architectural measurements all had greater than 3% variation, compared to less than 1.2% for measurements containing repositioning error and no movement artifacts. Therefore, this indicates that movement artifacts are a significant source of error the reduction of that type of error will greatly improve reproducibility. Complete immobilization of the limb (Fig. 1) being scanned will likely offer an improvement, and investigations into cast supports to reduce movement may be warranted.

In general, measurements at the radius were found to be less precise than measurements at the tibia and that may be due to differences in the bone size. The same amplitude of error may have less effect on a larger bone. Motion artifacts were also found to occur more frequently in the radius than in the tibia, and this may be due to the radius being a location in the body that is constantly in motion during daily activity, as well to patient positioning difficulties for radius measurements (it is more awkward than positioning the patient for the tibia measurements). Improved methods to immobilize the wrist may help prevent movement artifacts. Another factor affecting radius precision is the difficulty of selecting a consistent landmark, whereas the landmark of the distal tibia is well defined. This difficulty is reflected in the relative improvement of reproducibility provided by the use of 3D registration in the radius (23%) compared to a modest improvement at the tibia (8%).

This study has determined the reproducibility of HR-pQCT for monitoring parameters related to bone quality at peripheral sites over time. The reproducibility of density measurements was better than the reproducibility of morphological and stiffness measurements, likely due to the reduced sensitivity of BMD measurements to motion artifact. The application of 3D registration had the largest impact on density measurements, and although 3D registration is not currently a standard option for HR-pQCT systems, it is relatively straightforward to apply and not excessively computationally demanding [13]. It could be incorporated as standard protocol in the future, and the maximization of reproducibility in HR-pQCT is important for the longitudinal monitoring of bone quality measurements for improved diagnosis and treatment of diseases such as osteoporosis.

Conflicts of interest

None of the authors have any conflicts of interest to disclose.

References

- [1] Müller R. The Zurich experience: one decade of three-dimensional high-resolution computed tomography. *Top Magn Reson Imaging* 2002;13:307–22.

- [2] Osteoporosis prevention, diagnosis, and therapy. *Jama* 2001;285:785–95.
- [3] Rüeggsegger P, Koller B, Müller R. A microtomographic system for the nondestructive evaluation of bone architecture. *Calcif Tissue Int* 1996;58:24–9.
- [4] Feldkamp LA, Goldstein SA, Parfitt AM, Jesion G, Kleerekoper M. The direct examination of three-dimensional bone architecture in vitro by computed tomography. *J Bone Miner Res* 1989;4:3–11.
- [5] Boutroy S, Bouxsein ML, Munoz F, Delmas PD. In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography. *J Clin Endocrinol Metab* 2005;90:6508–15.
- [6] Khosla S, Riggs BL, Atkinson EJ, Oberg AL, McDaniel LJ, Holets M, et al. Effects of sex and age on bone microstructure at the ultradistal radius: a population-based noninvasive in vivo assessment. *J Bone Miner Res* 2006;21:124–31.
- [7] MacNeil JA, Boyd SK. Accuracy of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Med Eng Phys* 2007;29:1096–105.
- [8] Pistoia W, van Rietbergen B, Lochmuller EM, Lill CA, Eckstein F, Rueggsegger P. Estimation of distal radius failure load with micro-finite element analysis models based on three-dimensional peripheral quantitative computed tomography images. *Bone* 2002;30:842–8.
- [9] Newitt DC, Majumdar S, van Rietbergen B, von Ingersleben G, Harris ST, Genant HK, et al. In vivo assessment of architecture and micro-finite element analysis derived indices of mechanical properties of trabecular bone in the radius. *Osteoporos Int* 2002;13:6–17.
- [10] Pistoia W, van Rietbergen B, Lochmuller EM, Lill CA, Eckstein F, Rueggsegger P. Image-based micro-finite-element modeling for improved distal radius strength diagnosis: moving from bench to bedside. *J Clin Densitom* 2004;7:153–60.
- [11] Ulrich D, van Rietbergen B, Laib A, Rueggsegger P. Load transfer analysis of the distal radius from in-vivo high-resolution CT-imaging. *J Biomech* 1999;32:821–8.
- [12] Ulrich D, Rietbergen B, Laib A, Rueggsegger P. Mechanical analysis of bone and its microarchitecture based on in vivo voxel images. *Technol Health Care* 1998;6:421–7.
- [13] Boyd SK, Moser S, Kuhn M, Klinck RJ, Krauze PL, Muller R, et al. Evaluation of three-dimensional image registration methodologies for in vivo micro-computed tomography. *Ann Biomed Eng* 2006;34:1587–99.
- [14] 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.
- [15] Su R, Campbell GM, Boyd SK. Establishment of an architecture-specific experimental validation approach for finite element modeling of bone by rapid prototyping and high resolution computed tomography. *Med Eng Phys* 2006.
- [16] Langton CM, Njeh CF. The physical measurement of bone. Series in medical physics and biomedical engineering. Bristol, Philadelphia: Institute of Physics Pub.; 2004.
- [17] Sokal RR, Braumann CA. Significance tests for coefficients of variation and variability profiles. *System Zool* 1980;29:50–66.
- [18] Butz S, Wuster C, Scheidt-Nave C, Gotz M, Ziegler R. Forearm BMD as measured by peripheral quantitative computed tomography (pQCT) in a German reference population. *Osteoporos Int* 1994;4:179–84.
- [19] Bonnick SL. Bone densitometry in clinical practice: application and interpretation. Current clinical practice. Totowa, NJ: Humana Press; 1998.