Scanner influence on the mechanical response of QCT-based finite element analysis of long bones

Yekutiel Katz[1]*, Gal Dahan[1]*, Jacob Sosna[2], Ilan Shelef[3], Evgenia Cherniavsky[4], Zohar Yosibash[1]

[1]*School of Mechanical Engineering, Tel Aviv University, Ramat Aviv, Israel*
[2]*Department of Radiology & Medical Imaging, Hadassah Hebrew University Medical Center, Jerusalem, Israel*
[3]*Diagnostic Imaging Institute, Soroka University Medical Center, Beer Sheva, Israel*
[4]*Department of Radiology, Barzilai Medical Center, Ashkelon, Israel*

* Equal contribution

Abstract contains 256 words

Paper contains 3385 words

Corresponding author
Prof Zohar Yosibash
Head, Computational Mechanics and Experimental Biomechanics Labs
School of Mechanical Engineering
POB 6139001
Tel Aviv, Israel
Tel: +972-3-640-8809
Email: yosibash@tauex.tau.ac.il

Abstract

Purpose: Patient-specific QCT-based finite element (QCTFE) analyses enable highly accurate quantification of bone strength. We evaluated CT scanner influence on QCTFE models of long bones.

Methods: A femur, humerus, and proximal femur without the head were scanned with K_2HPO_4 phantoms by seven CT scanners (four models) using typical clinical protocols. QCTFE models were constructed. The geometrical dimensions, as well as the QCT-values expressed in Hounsfield unit (HU) distribution was compared. Principal strains at representative regions of interest (ROIs), and maximum principal strains (associated with fracture risk) were compared. Intraclass correlation coefficients (ICCs) were calculated to evaluate strain prediction reliability for different scanners. Repeatability was examined by scanning the femur twice and comparing resulting QCTFE models.

Results: Maximum difference in geometry was 2.3%. HU histograms before phantom calibration showed wide variation between QCT scans; however, bone density histogram variability was reduced after calibration and algorithmic manipulation. Relative standard deviation (RSD) in principal strains at ROIs was <10.7%. ICC estimates between scanners were >0.9. Fracture-associated strain had 6.7%, 8.1%, and 13.3% maximum RSD for the femur, humerus, and proximal femur, respectively. The difference in maximum strain location was <2 mm. The average difference with repeat scans was 2.7%.

Conclusions: Quantification of strain differences showed mean RSD bounded by ~6% in ROIs. Fracture-associated strains in "regular" bones showed a mean RSD bounded by ~8%. Strains were obtained within a $\pm$10% difference relative to the mean; thus, in a longitudinal study only changes

larger than 20% in the principal strains may be significant. ICCs indicated high reliability of QCTFE models derived from different scanners.

Key words QCT; Femur; Finite element analysis; Humerus; Patient-specific; Personalized medicine

Introduction

Personalized finite element (FE) models of bones based on quantitative computed tomography (QCT) have been used extensively to estimate bone stiffness and strength. Among these, the human femur and humerus attracted significant interest (Dahan et al., 2016; Dall'Ara et al., 2013; Enns-Bray et al., 2016; Hazrati Marangalou et al., 2012; Helgason et al., 2014; Keaveny et al., 2008; Keyak et al., 2005; Keyak et al., 1990; Nishiyama et al., 2013; Pise et al., 2009; Schileo et al., 2008a; Trabelsi et al., 2011; Yosibash et al., 2007b) because of their importance for fracture risk prediction due to osteoporosis, optimal management of bony metastases, and more. QCT-based FE (QCTFE) analysis is considered one of the best current techniques for noninvasive assessment of femoral and humeral strength (Keaveny et al., 2010).

Validation of QCTFE predictions has been performed by comparing predicted strain on bone surface and fracture load to these measured in ex vivo experiments (Dahan et al., 2016; Trabelsi et al., 2011; Yosibash et al., 2007b). Most validation studies were based on QCT scans performed on a single scanner; however, to introduce QCTFE into clinical practice for longitudinal studies in diverse clinical centers, scanner influence on QCTFE predictions must be quantified. QCTFE models are directly affected by QCT data: (a) bone geometry is reconstructed from the set of points sampled in the scan; (b) bone density in each voxel, a proxy for material properties, is derived from CT-values expressed in Hounsfield unit (HU). Additionally, indirect effects are present due to algorithmic manipulations of QCT data: (a) boundary corrections and density averaging are performed at the voxel level to overcome the partial volume effect and to reduce noise; (b) material properties are determined using empirical relationships based on HUs and assigned to the FE model at specific locations (integration points).

The feasibility of introducing QCTFE into clinical use was recently addressed in a study examining QCTFE model dependence on scan protocols (Dragomir-Daescu et al., 2015). The paper concluded that predicted strength and stiffness of QCTFE models generated from 'high' and 'low' resolution scans could be different if CT settings and reconstruction techniques vary. A second study (Giambini et al., 2015) found that material property estimation may be dependent on scan parameters such as voltage and current as well as post-processing techniques, including the reconstruction kernel, with the differences potentially affecting estimates of strength and stiffness. The basic assumption that QCTFE models are independent of CT scanner was questioned based on a comparison of density and strength evaluated with two different scanners (Carpenter et al., 2014). The study concluded that significant inter-scanner differences remain even after using phantom calibration.

This study compares patient-specific QCTFE models using data obtained with multiple CT scanners and hospitals, while considering the steps in the QCTFE model construction pipeline. We analyzed scans of two human femurs and one humerus surrounded by five K_2HPO_4 phantoms. QCT scans were performed using standard clinical protocols. We hypothesized that a carefully designed and implemented pipeline would allow development of reliable QCTFE models of human bones, *i.e.* models that are virtually independent of the CT scanner used to acquire data.

Materials and Methods

A femur from a 70-year-old female (51 kg, 170 cm), a humerus from a 57-year-old male (55 kg, 180 cm) (National Disease Research Institute, Philadelphia, PA), and a proximal femur without the head and neck from a 55-year-old male (73 kg, 175 cm) (ScienceCare Anatomical, Phoenix, AZ), were scanned by seven different CT scanners at four medical centers. In total, four CT

scanner models from three manufacturers were used: Brilliance (64) and iCT (256) by Philips Healthcare (Eindhoven, The Netherlands), Somatom Definition Edge (128) by Siemens Medical Systems (Erlangen, Germany) and Optima CT660 (128) by GE HealthCare (Milwaukee, WI, USA).

Five liquid K_2HPO_4 phantoms (concentrations: 0, 50, 100, 200 and 300 mg/cc), prepared according to (Mindways Software, 2002) were used. The bones and phantoms were immersed in a water bath during all scans (Fig. 1A). Scans were performed according to a standard clinical protocol: exposure 210–500 mAs, X-ray tube voltage 120 kV, and slice thickness and spacing 1mm in all protocols except for scan G, where it was 1.25 mm (see Table 2). Images were reconstructed using a "soft/body" filter/convolution kernel (B, B31f, and body filters for Philips, Siemens, and GE scanners, respectively). Details on CT scanners, medical centers and scanning protocols are presented in Tables 1 and 2.

Brief summary of QCTFE pipeline

FE models were generated using the various QCT scans according to a validated procedure (Yosibash et al., 2007a; Yosibash et al., 2007b). Strains, stiffness, and fracture loads computed by QCTFE models were previously validated in double-blind in vitro studies, where the modeling pipeline was described in detail (Trabelsi et al., 2011; Yosibash et al., 2014; Yosibash et al., 2010). Briefly, bone boundaries were identified on each slice and surfaces were imported as a point cloud to a computer-aided design software program (SolidWorks by Dassault Systems, Waltham, MA, USA). All QCTFE models were aligned in a uniform coordinate system. This ensured that the same boundary conditions were applied and that results were extracted at the same locations in all models. Alignment was performed using an "iterative closest point" algorithm (CloudCompare,

126 http://www.danielgm.net/cc/). Geometrical differences between models were due to differences in
127 pixel size and the set of points from which each model was reconstructed. After alignment, these
128 differences were assessed in 9 to 11 cross sections located along the bones at 12–25 mm intervals.
129 Dimensions in *x* and *y* directions were extracted and the maximum differences at each section were
130 compared to the average values.

131 HU calibration was performed by K_2HPO_4 phantoms: HU values were averaged along each
132 phantom using a 10×10 mm^2 area at the center of 10 slices (3000–5800 pixels), see Fig. 1B. Five
133 average HU values (five known densities in phantoms), determined a linear relationship between
134 K_2HPO_4 density and HU: $\rho_{K2HPO4} = m \times HU + n$ (in all scans $R^2 > 0.998$). Values for *m* and *n* for
135 each scan are given in last two columns of Table 2.

136 Due to the partial volume averaging effect in voxels on bone surface, a boundary correction
137 algorithm was applied. A total width of ~1 mm, corresponding to 2 pixels, was usually affected.
138 HU values in all voxels were corrected for noise effects by a moving average algorithm on a cubic
139 volume of 27 bone voxels. Young's modulus (pointwise "stiffness") was thereafter computed
140 based on ash density (ρ_{ash}), using the empirical relationship described in (Schileo et al., 2008a) and
141 the relationship between hydroxyapatite and K_2HPO_4 densities (Goodsitt, 1992):

142 $$\rho_{ash}\left[\frac{gr}{cc}\right] = 1.061 \times \rho_{K2HPO4} + 0.08 \quad (1)$$

143 CT scanner output data was compared by inspecting histograms of both raw HU and resulting ρ_{ash}
144 obtained after moving average and boundary correction algorithms. An inhomogeneous Young's
145 modulus (*E*), based on ρ_{ash} for cortical (Keller, 1994) and trabecular bone (Keyak et al., 1993) was

assigned to the QCTFE models, considering only values of ρ_{ash} >0 $\left[\frac{gr}{cc}\right]$):

$$E_{cort} = 10200 \times \rho_{ash}^{2.01}\ [MPa], \qquad \rho_{ash} \geq 0.486\ \left[\frac{gr}{cc}\right] \tag{2}$$

$$E_{trab} = 2398\ [MPa], \qquad 0.3 < \rho_{ash} < 0.486\ \left[\frac{gr}{cc}\right] \tag{3}$$

$$E_{trab} = 33900 \times \rho_{ash}^{2.2}\ [MPa], \qquad \rho_{ash} \leq 0.3\left[\frac{gr}{cc}\right] \tag{4}$$

The Poisson ratio was set to $\nu = 0.3$, as commonly used in bone QCTFE models (Schileo et al., 2014; Varga et al., 2018; Wirtz et al., 2000; Yosibash et al., 2007a).

On the femur head, a hip contact force representing a stance position of 1250N was applied parallel to the shaft axis on a 335 mm^2 area. On the humerus head, a 230N$\hat{x}$ + 650N$\hat{z}$ load was applied on a 320 mm^2 area with the z axis rotated at $20^{\circ}$to the shaft axis and the x axis pointing laterally. On the proximal femur, a normal displacement u_n = 0.75 mm was applied on an area of 810 mm^2 inclined by $35^{\circ}$to the shaft axis (as illustrated in Fig. 2). All three bones were clamped at the distal surface ($\vec{u} = 0$). A mesh of $p-$tetrahedral elements with a maximum element volume of ~150mm^3 was generated. Linear elastic QCTFE analyses were performed by increasing the polynomial degree over elements from p=1 to p=8, to allow monitoring of numerical errors in energy norm and strains (Szabó and Babüska, 1991). We emphasize the advantages of p-FE methods (Trabelsi et al., 2011; Yosibash et al., 2007b) over conventional FE methods: they accurately represent bone surfaces and polynomial degree (p) of shape functions is increased systematically to achieve convergence. p-FEs also allow a smooth variation of material properties within the model, elements are much larger, may be far more distorted, and produce considerably faster convergence rates.

To isolate the effect of the QCT reconstruction algorithm, we also performed FE analyses with a

homogeneous Young's modulus, E=10 GPa.

Comparison between QCTFE models was performed by means of principal strains (compression/tension), as they may serve as a criterion to predict femoral/humeral fracture (Bayraktar et al., 2004; Dahan et al., 2016; Schileo et al., 2014; Schileo et al., 2008b; Yosibash et al., 2010). Regions of interest (ROIs) on the bone surface were arbitrarily selected on the shaft, neck, and intertrochanteric/intertuberosity zones (13, 12, and 8 ROIs on the femur, humerus and proximal femur surfaces respectively, Fig. 2). ROIs were 3 mm lines resembling the commonly used strain gauges (SG) in validation experiments (Dahan et al., 2016; Yosibash et al., 2007b). The average strain value along each line was compared relative to the mean strain value at a given ROI across all QCTFE models from the different scanners. Standard deviation of the strains was calculated for each ROI. Relative standard deviation (RSD [%], also known as coefficient of variation (CV)), defined by the ratio of the standard deviation to the mean strain in the ROI is reported. The mean RSD (averaged across all ROIs) for each bone is also presented. Based on the ROI strainsm, linear correlations were plotted for each pair of the QCTFE models.

To evaluate the reproducibility of results obtained from different scanners, Intraclass Correlation Coefficient (ICC) estimates (two-way random effects, absolute agreement, single rater (Koo and Li, 2016; McGraw and Wong, 1996)) and their 95% confidence intervals (CI) were calculated using SPSS (IBM, Chicago, IL, USA). ICCs were calculated based on strain values in the different ROIs.

In addition to ROIs, indicating a global agreement between QCTFE models, the maximum tension and compression principal strains and their locations (surrogate of load to fracture and fracture location) were compared (averaged over a circular surface with a radius of r=2 mm). To quantify

and compare the obtained locations, a reference point representing the 'average' location was computed using least-mean-squares. We report mean, maximum, and RSD of the distance from the reference point.

Within-scanner repeatability of the QCTFE was examined by performing two consecutive scans of the femur on each CT scanner. Maximum tension and compression principal strains were computed and compared.

Results

The relative difference in dimension at each cross section is presented in Figure 3. The maximum relative difference in dimensions between all models was less than 2.3%. All QCTFE models had 1,170,000–1,390,000 degrees of freedom and converged to <6.5% relative error in energy norm at p=8. Strains converged to <1% relative error between consecutive solutions (increased polynomial degree). The influence of the geometrical reconstruction, as examined by the FE models with constant E=10 GPa, is presented in Table 3. The maximum RSD in ROI principal strain that can be attributed to the geometry reconstruction algorithm (segmentation process and pixel size in the various QCT scans in the three considered bones) is 5.5%; however, the average RSD is between 2.1% and 3.1%. Detailed strain values are provided in the Supplemental Material (Appendix B).

Histograms of scans' output data are shown in Figure 4. Raw HU histograms can be seen in Figure 4A-C; manipulated raw HU resulting in ash density (ρ_{ash}), which is used to determine inhomogeneous material properties is shown in Figure 4D-F. Since one proximal femur was cut and cleaned in a previous study, some of the fat and marrow tissue was lost and replaced by air (HU=-1000), accounting for the many negative values in the proximal femur's histogram.

The mean and the RSD% of the absolute maximum principal strains at each ROI for all QCTFE models are presented in Figure 5. These represent all differences in QCTFE models including geometry, material property distribution, and algorithmic influence due to HU manipulation. Maximum RSD in the femur is 7.3%, in the humerus 7.8%, and in the proximal femur 11%; average RSDs are 4.97%, 4.72%, and 6.1%, respectively. Detailed graphs showing the specific values obtained in each FE model are presented in Appendix A. Correlations between each pair of models yielded linear regression equations with slopes varying between 0.93 to 1.05 ($r^2 > 0.96$ for all pairs), the complete linear correlation matrix is provided in Appendix C. ICC estimates and their corresponding 95% CIs are 0.959 [0.901, 0.986], 0.988 [0.971, 0.996], and 0.981 [0.952, 0.995] for the femur, humerus, and proximal femur, respectively.

The location and magnitude of maximum principal strain is the most important information to the medical community because it is associated with risk of fracture. Maximum tensile and compressive principal strains in the femur and proximal femur, and compressive principal strain in the humerus (no tensile significant strains at most loading conditions) are presented in Figure 6. In the femur, a 6.7% RSD was obtained in the maximum tensile principal strain, with the largest maximum difference about 10% from the mean value for scanner A. In the humerus, an 8.1% RSD was obtained in the maximum compressive principal strain, with maximum differences for scanners F and G of about -10% and 10% from the mean, respectively. In the proximal femur, a 13.3% RSD was obtained in the maximum compressive principal strain with maximum differences for scanners D and G of about -15% and 15% from the mean value, respectively. The differences in the location of the obtained principal strains compared to the reference point are summarized in Table 4; the maximum difference was <2 mm.

Within-scanner differences in the maximum tensile and compressive principal strains was less

than 4.1% for six scanners, but was 7.7% for scanner G. The average difference across all seven scanners was 2.7% (standard deviation 2%).

Discussion

This study investigated whether a rigorous QCT protocol combined with a unique pipeline for the generation of a QCTFE model enabled reliable values of principal strain in three human bones, regardless of the CT scanner used for image acquisition. Principal strains are quantities used to estimate risk of fracture and indirectly assess bone stiffness, so are of special interest.

Bone dimensions from the various QCT scans showed very small differences (maximum relative difference <2.3%). These differences introduced a maximum RSD of 5.5% in the ROI principal strains for the three bones when homogeneous material properties were used. The average differences were much smaller, usually half of the maximum RSD.

Considering the material properties, the cortex (high HU value) dominates the mechanical behavior of bones. Discrepancy in low HUs has a negligible influence and is deemed unimportant. For example, a tissue having HU <200 (ρ_{ash} < 0.15 gr/cc), corresponds to E < 0.5 GPa, considerably smaller than typical values of E = 10–20 GPa of the cortex. Focusing on high HU values in the histograms, discrepancies between raw HU values obtained from the different QCT scans are clearly visible, suggesting that the HU without callibration is scanner-dependent (Fig. 4A-C). Boundary correction and moving average algorithms only influenced the likelihood of HU values already present in the bone (the y axis in the histograms). In contrast, HU values (the x axis) are affected by the phantom calibration alone; therefore, phantom calibration reduces scanner dependence, although some variation is still present (Fig. 4D-F).

The maximum RSD was <8% in principal strains along the bones in a femur and humerus, and <11% in a proximal femur without the head. The mean RSD was about half of the maximum RSD, ≤5% in "regular" bones and <6.1% in the proximal femur. However, the total maximum difference observed between two scanners in all three bones and all ROIs was 20%. Linear correlations between each pair of models showed very good agreement.

ICC values >0.9 indicate excellent reliability (Koo and Li, 2016) among the different scanners. However, these findings should be interpreted with caution. The ICC highly depends on the variance between the examined subjects, i.e. the ROI strains. These were taken from different regions along different bones and thus characterized by high variability. Additionally, the ICC provides only a general measure of the agreement level, with no quantitative estimate of the variability between the scanners.

The maximum tensile/compressive principal strains, the quantities of interest when risk of fracture is considered, were obtained at similar locations in all models (Fig. 6). The high variability in the maximum principal strains values obtained in the proximal femur (RSD <13.3%) can be related to the highly complex geometry at the unusual location of the maximum strains (Fig. 6-C) and thus may not be representative. Using a clinical “soft/body” filter/convolution kernel at 120 kV and calibration with liquid K2HPO4 phantoms, the maximum RSD between QCTFE maximum principal strains in "regular" bones (such as the femur and humerus), is <8%. Nonetheless, there may be differences up to 20% between maximum principal strains estimated using two scanners with the most extreme differences; thus, in a longitudinal study only changes larger than 20% in principal strains may be significant. Such analyses are important, for example, to determine prognosis in a patient with osteoporosis or metastases. It is noteworthy that previous experience suggests that in pathological situations such as benign or metastatic tumors, changes in the strain

fields for the pathological femur are of tens of percent compared to the healthy femur, so the 20% difference from different scanners would be of smaller significance in these cases (Sternheim et al., 2018).

Repeated scans on same CT systems resulted in an average difference of 2.7% in the maximum principal strains. These findings suggest that one should expect similar or higher variability between CT scanners.

Exposure (mAs) has a negligible influence on differences in QCTFE results between CT systems, as demonstrated in previous studies (Giambini et al., 2015; Nazarian et al., 2008). We conducted two identical scans where a change in the exposure level was the only difference in protocol and obtained almost no change in the HU histograms.

Several previous studies have emphasized the use of phantoms to quantify HU variability between CT scanners. A study by (Suzuki et al., 1991) investigated the variance between QCT scans of cadaveric vertebrae together with K_2HPO_4 phantoms obtained from 16 CT scanners, including 10 different models. They reported "good correlations ($r > 0.97$) … observed for QCT values obtained by various CT scanners." An excellent review (Cann, 1988) showed good reproducibility of QCT values obtained from several scanners, although CT systems used in the study are by now outdated. These publications however did not evaluate differences in strains computed by QCTFE. To the best of our knowledge, the only study addressing the latter issue examined the femoral fracture load predicted by QCTFE analyses, as assessed with two different CT scanners (Carpenter et al., 2014).

It is important to note that most publications using QCTFE were based on the classical h-version

of the FE method, which does not include numerical convergence studies. Numerical errors inherent in h-version techniques may contribute substantially to the differences in predictions. As an example, a difference in strains of ~3% was found at the neck between the FE analysis performed at p=4 compared to the converged strains at p=8.

Limitations and Future Studies

This study did not consider body size effects that may be taken into consideration using anthropomorphic standardization phantoms (ASPs) to assess inter-scanner differences. Carpenter, et al. (Carpenter et al., 2014) found that differences in body size may result in varying estimates of femur strength between scanners. In addition, tube voltage (kV) has an influence on HU distribution histograms. We performed scans using 80–140 kV (data not presented here) and will address the influence of kV on estimation of bone's mechanical response in a future publication.

We studied three manufacturers and four CT scanners at four different medical centers. While these are believed to represent the majority of CT scanners in the Western world, future studies may include a larger number of CT systems and manufacturers.

The conclusions of our study are based only on descriptive statistics. Future studies should include more subjects, i.e. more bones of the same kind.

Finally, although past investigations suggest the use of phantoms in the range of 0–300 mg/cc (Bessho et al., 2007; Miura et al., 2017; Yosibash et al., 2014), cortical tissue has an equivalent mineral density of about 1200 mg/cc, about four times higher. The slope parameter m in Table 2 is sensitive to inaccuracies in the extreme concentration phantoms (0 and 300 mg/cc), thus it

316 strongly influences predicted strains. To improve the calibration, two alternatives will be
317 considered in a future investigation: Using phantom concentrations covering the entire range of
318 HUs, or preferably, developing phantom-less methods as suggested in (Lee et al., 2017; Suzuki
319 et al., 1991).

320

321 ***Conclusions***

322 QCTFE analyses of long bones performed by seven different CT scanners showed a maximum
323 20% differences in principal strains. In a longitudinal study, only changes larger than 20% in
324 principal strains may be significant.

325

326 **Conflicts of Interest**

327 Prof. Yosibash has a financial interest in PerSimiO, a technology company specializing in
328 personalized simulations in orthopedics. Prof. Sosna is a member of the PerSimiO Board of
329 Advisors. Neither Prof. Yosibash or Prof. Sosna derived any direct financial support, research
330 support, or other benefit from the company that could jeopardize their independence in this
331 research. Primary data was jointly controlled by Yekutiel Katz and Gal Dahan. They and the
332 remaining authors have no conflict of interest to declare that could bias the presented work.

333 **Disclosures and Acknowledgements**

334 The study was funded in part by a grant from the Milgrom Foundation for Science grant. The
335 authors thank Dr. Gabriel Bartal and Mr. Isaac Yunes at Meir Medical Center, Mr. David

Shushan at the Soroka Medical Center, and Mrs. Nathalie Greenberg from the Hadassah-Hebrew University Medical Center for their help and collaboration on the acquisition of the QCT scans. The authors thank Prof. Joyce Keyak from the University of California at Irvine in the USA and Dr. Nir Trabelsi from the Shamoon College of Engineering in Israel for interesting discussions. The authors thank Dr. Lena Novack from Ben Gurion University of the Negev and Prof. Malka Gurfine from Tel-Aviv University for their assistance on the statistical analysis and Mrs. Shifra Fraifeld from the Hadassah-Hebrew University Medical Center for helpful editorial assistance.

References

Bayraktar, H.H., Morgan, E.F., Niebur, G.L., Morris, G.E., Wong, E.K., Keaveny, T.M., 2004. Comparison of the elastic and yield properties of human femoral trabecular and cortical bone tissue. Journal of biomechanics 37, 27-35.

Bessho, M., Ohnishi, I., Matsuyama, J., Matsumoto, T., Imai, K., Nakamura, K., 2007. Prediction of strength and strain of the proximal femur by a CT-based finite element method. Journal of biomechanics 40, 1745-1753.

Cann, C.E., 1988. Quantitative CT for determination of bone mineral density: a review. Radiology 166, 509-522.

Carpenter, R.D., Saeed, I., Bonaretti, S., Schreck, C., Keyak, J.H., Streeper, T., Harris, T.B., Lang, T.F., 2014. Inter-scanner differences in in vivo QCT measurements of the density and strength of the proximal femur remain after correction with anthropomorphic standardization phantoms. Medical engineering & physics 36, 1225-1232.

Dahan, G., Trabelsi, N., Safran, O., Yosibash, Z., 2016. Verified and validated finite element analyses of humeri. Journal of biomechanics 49, 1094-1102.

Dall'Ara, E., Luisier, B., Schmidt, R., Kainberger, F., Zysset, P., Pahr, D., 2013. A nonlinear QCT-based finite element model validation study for the human femur tested in two configurations in vitro. Bone 52, 27-38.

Dragomir-Daescu, D., Salas, C., Uthamaraj, S., Rossman, T., 2015. Quantitative computed tomography-based finite element analysis predictions of femoral strength and stiffness depend on computed tomography settings. Journal of biomechanics 48, 153-161.

Enns-Bray, W.S., Ariza, O., Gilchrist, S., Widmer Soyka, R.P., Vogt, P.J., Palsson, H., Boyd, S.K., Guy, P., Cripton, P.A., Ferguson, S.J., Helgason, B., 2016. Morphology based anisotropic finite element models of the proximal femur validated with experimental data. Medical engineering & physics 38, 1339-1347.

Giambini, H., Dragomir-Daescu, D., Huddleston, P.M., Camp, J.J., An, K.N., Nassr, A., 2015. The Effect of Quantitative Computed Tomography Acquisition Protocols on Bone Mineral Density Estimation. Journal of biomechanical engineering 137, 114502.

Goodsitt, M.M., 1992. Conversion relations for quantitative CT bone mineral densities measured with solid and liquid calibration standards. Bone and mineral 19, 145-158.

Hazrati Marangalou, J., Ito, K., van Rietbergen, B., 2012. A new approach to determine the accuracy of morphology-elasticity relationships in continuum FE analyses of human proximal femur. Journal of biomechanics 45, 2884-2892.

Helgason, B., Gilchrist, S., Ariza, O., Chak, J.D., Zheng, G., Widmer, R.P., Ferguson, S.J., Guy, P., Cripton, P.A., 2014. Development of a balanced experimental-computational approach to understanding the mechanics of proximal femur fractures. Medical engineering & physics 36, 793-799.

Keaveny, T.M., Hoffmann, P.F., Singh, M., Palermo, L., Bilezikian, J.P., Greenspan, S.L., Black, D.M., 2008. Femoral bone strength and its relation to cortical and trabecular changes after treatment with PTH, alendronate, and their combination as assessed by finite element analysis of quantitative CT scans. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 23, 1974-1982.

Keaveny, T.M., Kopperdahl, D.L., Melton, L.J., 3rd, Hoffmann, P.F., Amin, S., Riggs, B.L., Khosla, S., 2010. Age-dependence of femoral strength in white women and men. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 25, 994-1001.

Keller, T.S., 1994. Predicting the compressive mechanical behavior of bone. Journal of biomechanics 27, 1159-1168.

Keyak, J.H., Fourkas, M.G., Meagher, J.M., Skinner, H.B., 1993. Validation of an automated method of three-dimensional finite element modelling of bone. Journal of biomedical engineering 15, 505-509.

Keyak, J.H., Kaneko, T.S., Tehranzadeh, J., Skinner, H.B., 2005. Predicting proximal femoral strength using structural engineering models. Clinical orthopaedics and related research, 219-228.

Keyak, J.H., Meagher, J.M., Skinner, H.B., Mote, C.D., Jr., 1990. Automated three-dimensional finite element modelling of bone: a new method. Journal of biomedical engineering 12, 389-397.

Koo, T.K., Li, M.Y., 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine 15, 155-163.

Lee, D.C., Hoffmann, P.F., Kopperdahl, D.L., Keaveny, T.M., 2017. Phantomless calibration of CT scans for measurement of BMD and bone strength-Inter-operator reanalysis precision. Bone 103, 325-333.

McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol Meth 1, 30-46.

Mindways Software, I., 2002. CT calibration phantom user's guide. Mindways Software, Inc., Austin, TX, USA.

Miura, M., Nakamura, J., Matsuura, Y., Wako, Y., Suzuki, T., Hagiwara, S., Orita, S., Inage, K., Kawarai, Y., Sugano, M., Nawata, K., Ohtori, S., 2017. Prediction of fracture load and stiffness of the proximal femur by CT-based specimen specific finite element analysis: cadaveric validation study. BMC Musculoskelet Disord 18, 536.

Nazarian, A., Snyder, B.D., Zurakowski, D., Muller, R., 2008. Quantitative micro-computed tomography: a non-invasive method to assess equivalent bone mineral density. Bone 43, 302-311.

Nishiyama, K.K., Gilchrist, S., Guy, P., Cripton, P., Boyd, S.K., 2013. Proximal femur bone strength estimated by a computationally fast finite element analysis in a sideways fall configuration. Journal of biomechanics 46, 1231-1236.

Pise, U.V., Bhatt, A.D., Srivastava, R.K., Warkedkar, R., 2009. A B-spline based heterogeneous modeling and analysis of proximal femur with graded element. Journal of biomechanics 42, 1981-1988.

Schileo, E., Balistreri, L., Grassi, L., Cristofolini, L., Taddei, F., 2014. To what extent can linear finite element models of human femora predict failure under stance and fall loading configurations? Journal of biomechanics 47, 3531-3538.

Schileo, E., Dall'ara, E., Taddei, F., Malandrino, A., Schotkamp, T., Baleani, M., Viceconti, M., 2008a. An accurate estimation of bone density improves the accuracy of subject-specific finite element models. Journal of biomechanics 41, 2483-2491.

Schileo, E., Taddei, F., Cristofolini, L., Viceconti, M., 2008b. Subject-specific finite element models implementing a maximum principal strain criterion are able to estimate failure risk and fracture location on human femurs tested in vitro. Journal of biomechanics 41, 356-367.

Sternheim, A., Giladi, O., Gortzak, Y., Drexler, M., Salai, M., Trabelski, N., Milgrom, C., Yosibash, Z., 2018. Pathological fracture risk assessment in patients with femoral metastases using CT-based finite element methods. A retrospective clinical study. Bone 110, 215-220.

Suzuki, S., Yamamuro, T., Okumura, H., Yamamoto, I., 1991. Quantitative computed tomography: comparative study using different scanners with two calibration phantoms. The British journal of radiology 64, 1001-1006.

Szabó, B., Babüska, I., 1991. Finite element analysis. John Wiley & Sons, New York.

Trabelsi, N., Yosibash, Z., Wutte, C., Augat, P., Eberle, S., 2011. Patient-specific finite element analysis of the human femur--a double-blinded biomechanical validation. Journal of biomechanics 44, 1666-1672.

Varga, P., Inzana, J.A., Gueorguiev, B., Sudkamp, N.P., Windolf, M., 2018. Validated computational framework for efficient systematic evaluation of osteoporotic fracture fixation in the proximal humerus. Medical engineering & physics 57, 29-39.

Wirtz, D.C., Schiffers, N., Forst, R., Pandorf, T., Weichert, D., Radermacher, K., 2000. Critical evaluation of known bone material properties to realize anisotropic FE-simulation of the proximal femur. Journal of biomechanics 33, 1325--1330.

Yosibash, Z., Padan, R., Joskowicz, L., Milgrom, C., 2007a. A CT-based high-order finite element analysis of the human proximal femur compared to in-vitro experiments. Journal of biomechanical engineering 129, 297-309.

Yosibash, Z., Plitman Mayo, R., Dahan, G., Trabelsi, N., Amir, G., Milgrom, C., 2014. Predicting the stiffness and strength of human femurs with real metastatic tumors. Bone 69, 180-190.

Yosibash, Z., Tal, D., Trabelsi, N., 2010. Predicting the yield of the proximal femur using high-order finite-element analysis with inhomogeneous orthotropic material properties. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences 368, 2707-2723.

Yosibash, Z., Trabelsi, N., Milgrom, C., 2007b. Reliable simulations of the human proximal femur by high-order finite element analysis validated by experimental observations. Journal of biomechanics 40, 3688-3699.

453 **Table 1: CT scanners and hospitals in this study.**

Scan Abbreviation	Medical Center	CT scanner model (# detector rows)
A	MC 1	Brilliance Philips Healthcare (64)
B	MC 2	
C	MC 4	
D	MC 1	iCT Philips Healthcare (256)
E	MC 2	
F	MC 3	Somatom Definition Edge Siemens Medical Systems (128)
G	MC 1	Optima CT660 GE HealthCare (128)

454

455

Table 2: Detailed information on the different QCT scans. All scans performed at 120 kV.

Scan	Scanner (slices)	Date of Scanning	Exposure [mAs]*	Slice spacing & thickness (mm)**	Pixel size [mm]	$\rho_{K_2HPO_4} = m \times HU + n$	
						m	n
A	Philips Brilliance (64)	May 17, 2015	374	1	0.496	0.816	1.82
		Jan 18, 2017	83-283	1	0.601	0.814	-3.20
B	Philips Brilliance (64)	Dec 12, 2015	232	1	0.46	0.795	3.08
		Jan 18, 2017	223-370	1	0.523	0.818	-3.82
C	Philips Brilliance (64)	July 19, 2016	301	1	0.492	0.793	3.85
		Jan 18, 2017	164	1	0.518	0.826	-4.50
D	Philips iCT (256)	May 17, 2015	374	1	0.473	0.794	0.60
		Jan 18, 2017	81-106	1	0.557	0.811	-2.28
E	Philips iCT (256)	Dec 12, 2015	406	1	0.574	0.806	-2.96
		Jan 18, 2017	321-1104	1	0.502	0.813	-3.49
F	Siemens Somatom Definition (128)	July 21, 2015	210	1	0.440	0.869	0.00
		Jan 18, 2017	96-215	1	0.621	0.857	0.89
G	GE Optima CT660 (128)	May 17, 2015	500	1.25	0.523	0.823	1.00
		Jan 18, 2017	3-6	1.25	0.594	0.831	-6.83

* A range of mAs values indicates scans with adaptive current algorithm.

** The slice thickness and slice spacing are equal for each scan.

Table 3: Mean RSD% and maximum RSD% for each of the three bones with homogeneous constant Young modulus.

	Femur	Humerus	Femur w/o head
Mean RSD%	2.08%	3.08%	2.51%
Max RSD%	3.45%	5.53%	4.90%
Location of Max RSD	N2	S6	N3

Table 4: Mean and maximum distances [mm] and standard deviation of the location of the maximum principal strains (QCTFE based on the 7 scanners).

	Femur		Humerus	Proximal Femur	
	ϵ_1	ϵ_3	ϵ_3	ϵ_1	ϵ_3
Mean distance [mm]	0.48	0.91	0.66	0.73	1.00
SD [mm]	0.32	0.40	0.23	0.23	0.52
Max distance [mm]	1.10	1.57	1.05	1.01	1.97

Figures

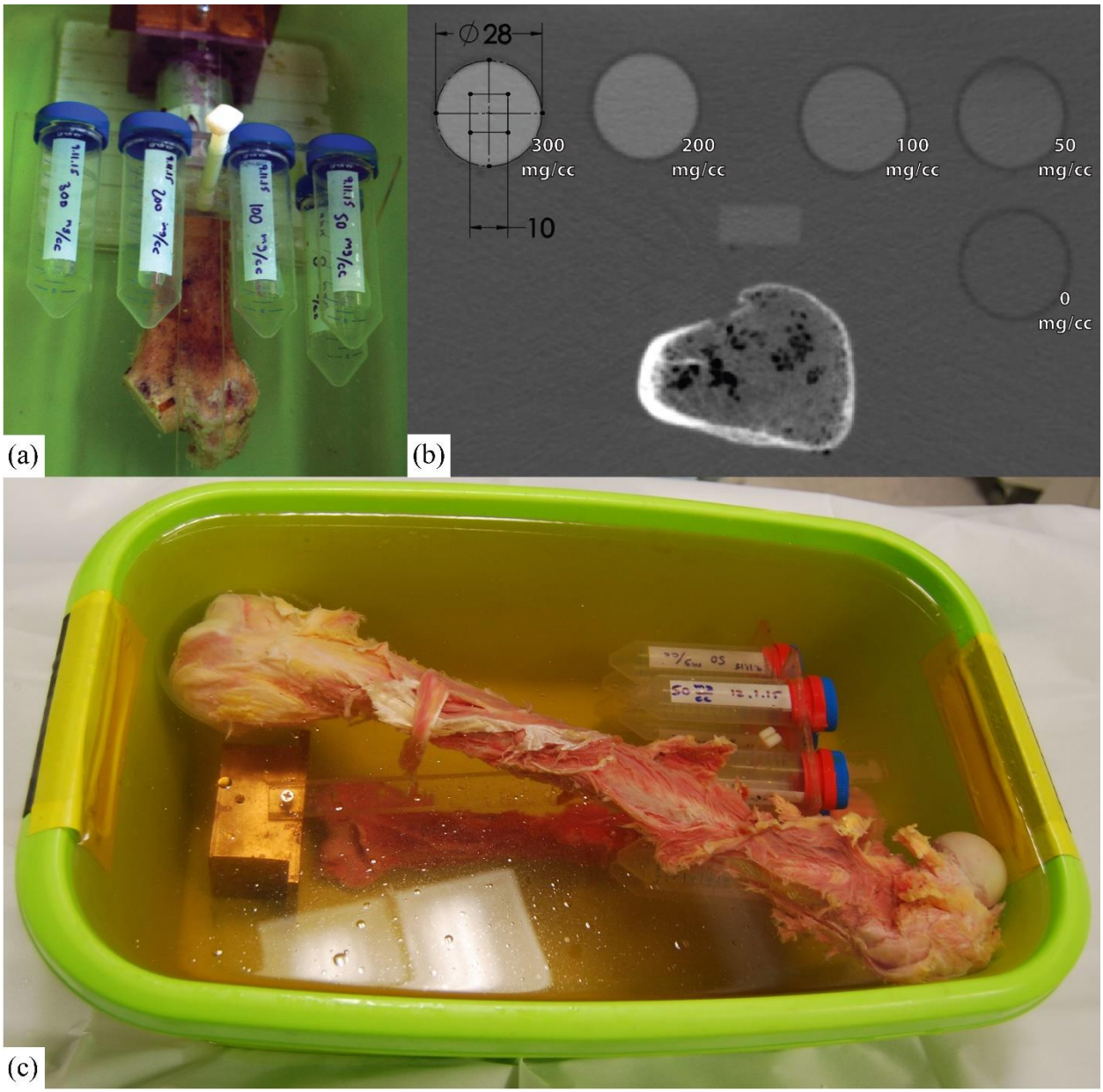


Figure 1 (a) Proximal femur in the water bath surrounded by five liquid potassium phosphate (K_2HPO_4) phantoms. (b) A representative CT slice of the femur and phantoms. (c) Whole femur and humerus in water bath with the liquid phantoms.

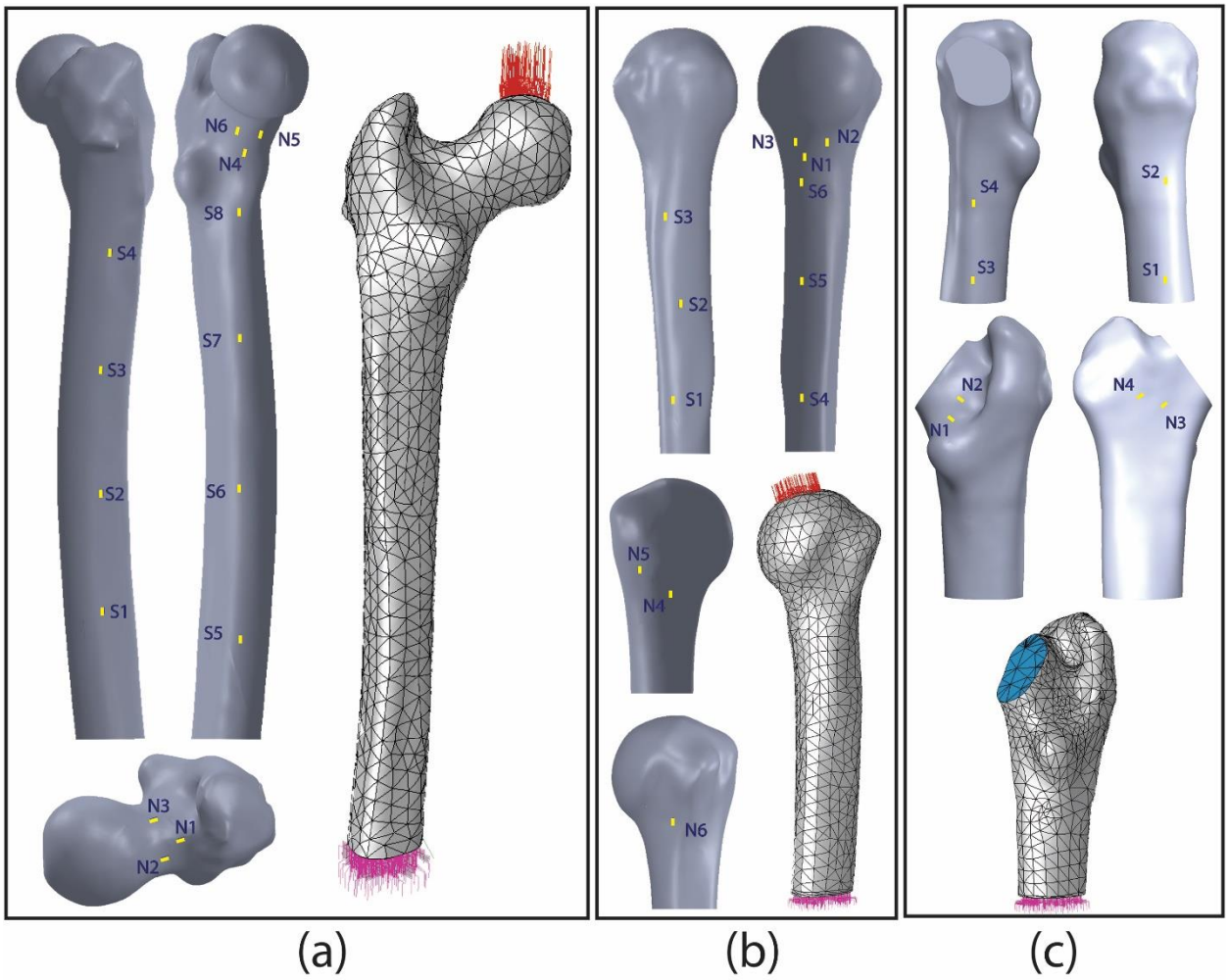


Figure 2 (a) The femur and regions of interest (ROI) at which strains were extracted with the QCT-based finite element analysis (CTFEA) model and boundary conditions. (b) The humerus and ROI with CTFEA model and boundary conditions. (c) The proximal femur with ROI and CTFEA model with displacement boundary conditions.

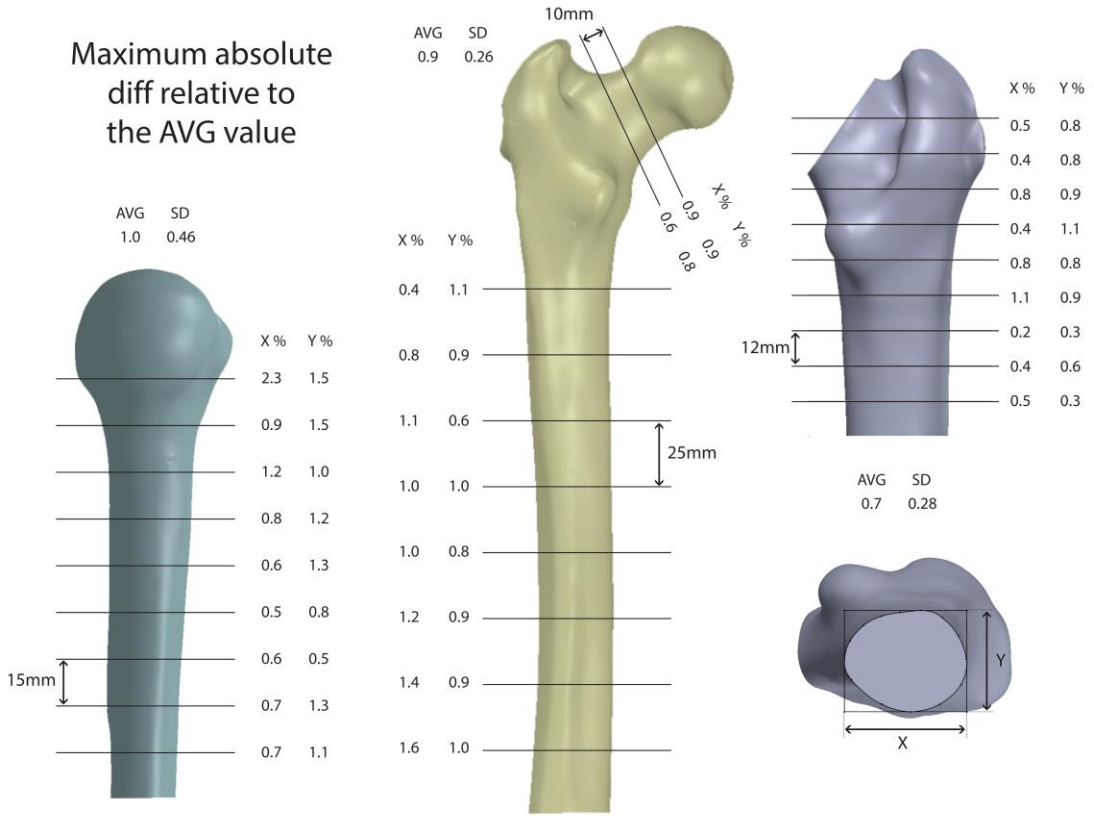
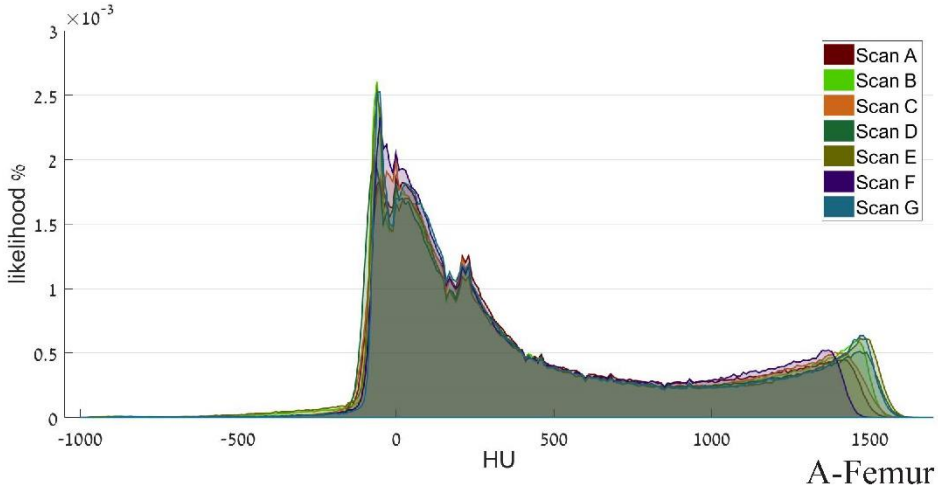
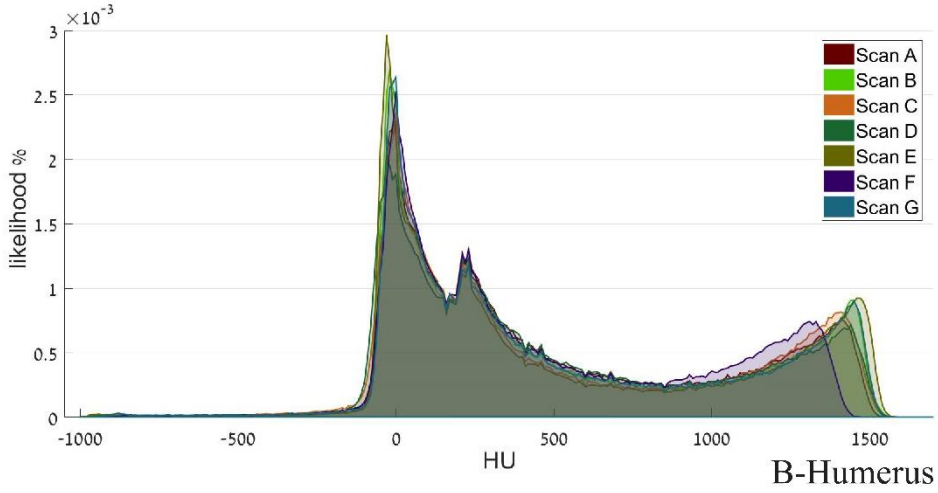
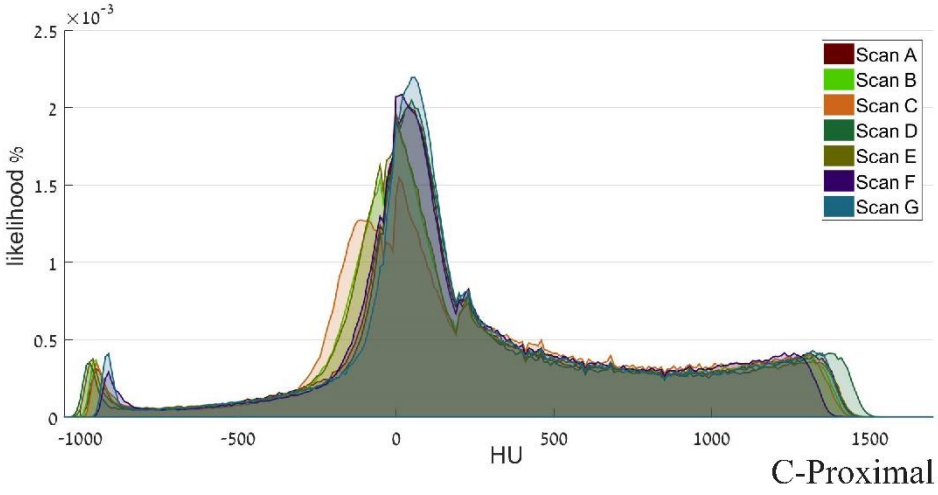


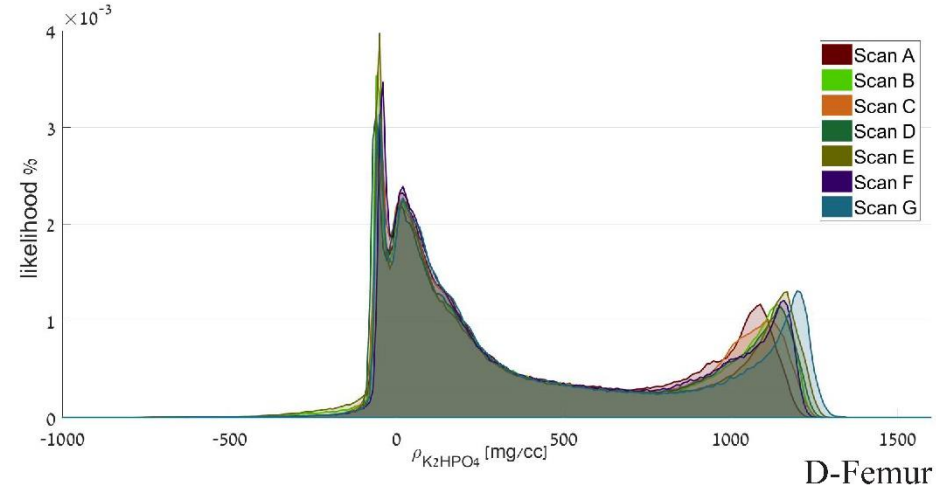
Figure 3 Maximum difference (relative to scan's average) between all scans at several slice locations along the segmented femurs and humerus. (AVG, average; SD, standard deviation; X, x-axis; Y, y-axis)

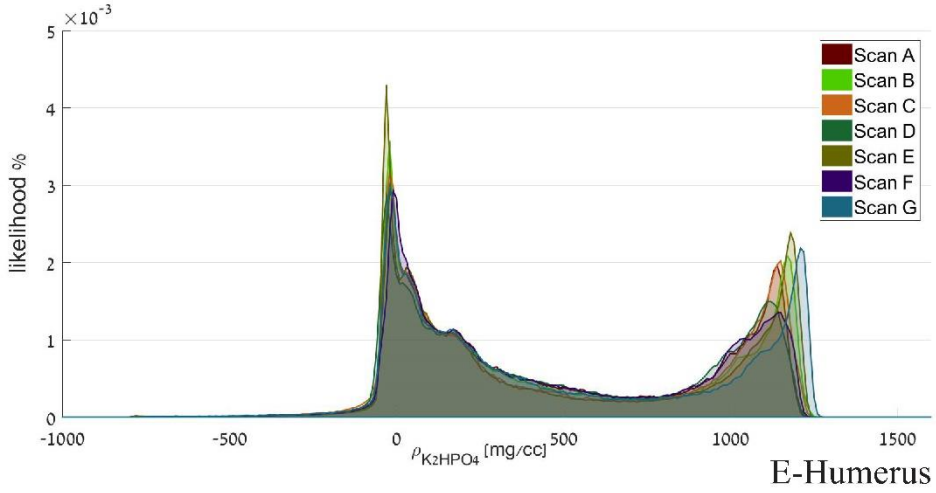

×10^-3
likelihood %
HU
Scan A
Scan B
Scan C
Scan D
Scan E
Scan F
Scan G
A-Femur

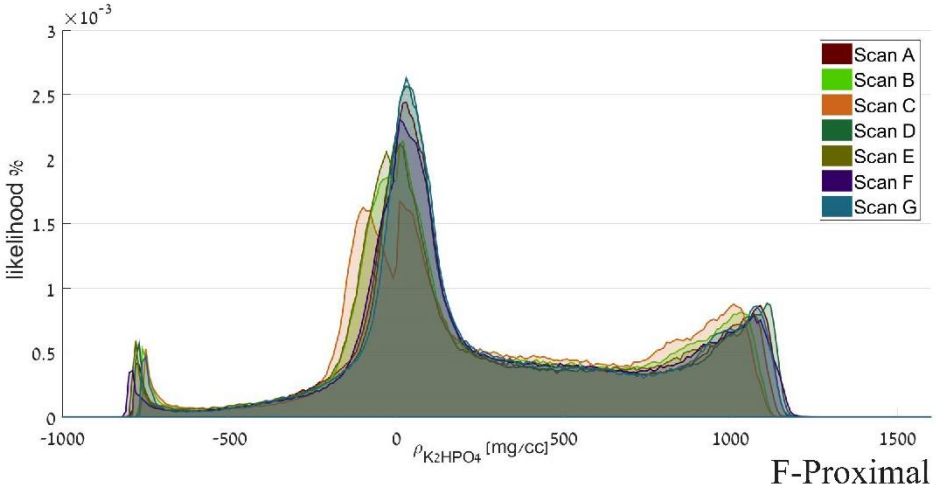

×10^-3
likelihood %
HU
Scan A
Scan B
Scan C
Scan D
Scan E
Scan F
Scan G
B-Humerus


×10^-3
likelihood %
HU
Scan A
Scan B
Scan C
Scan D
Scan E
Scan F
Scan G
C-Proximal

491


×10^-3
likelihood %
Scan A
Scan B
Scan C
Scan D
Scan E
Scan F
Scan G
ρK2HPO4 [mg/cc]
D-Femur


×10^-3
likelihood %
Scan A
Scan B
Scan C
Scan D
Scan E
Scan F
Scan G
ρK2HPO4 [mg/cc]
E-Humerus


×10^-3
likelihood %
Scan A
Scan B
Scan C
Scan D
Scan E
Scan F
Scan G
ρK2HPO4 [mg/cc]
F-Proximal

492

Figure 4 Normalized histograms of Hounsfield unit (HU) values inside the femur, humerus, and proximal femur in scans A through G. The area under the histogram equals 1. (A–C) Histogram of raw HU values. (D–F) Histogram of ρ_{ash} after phantom calibration, moving average, and boundary corrections.

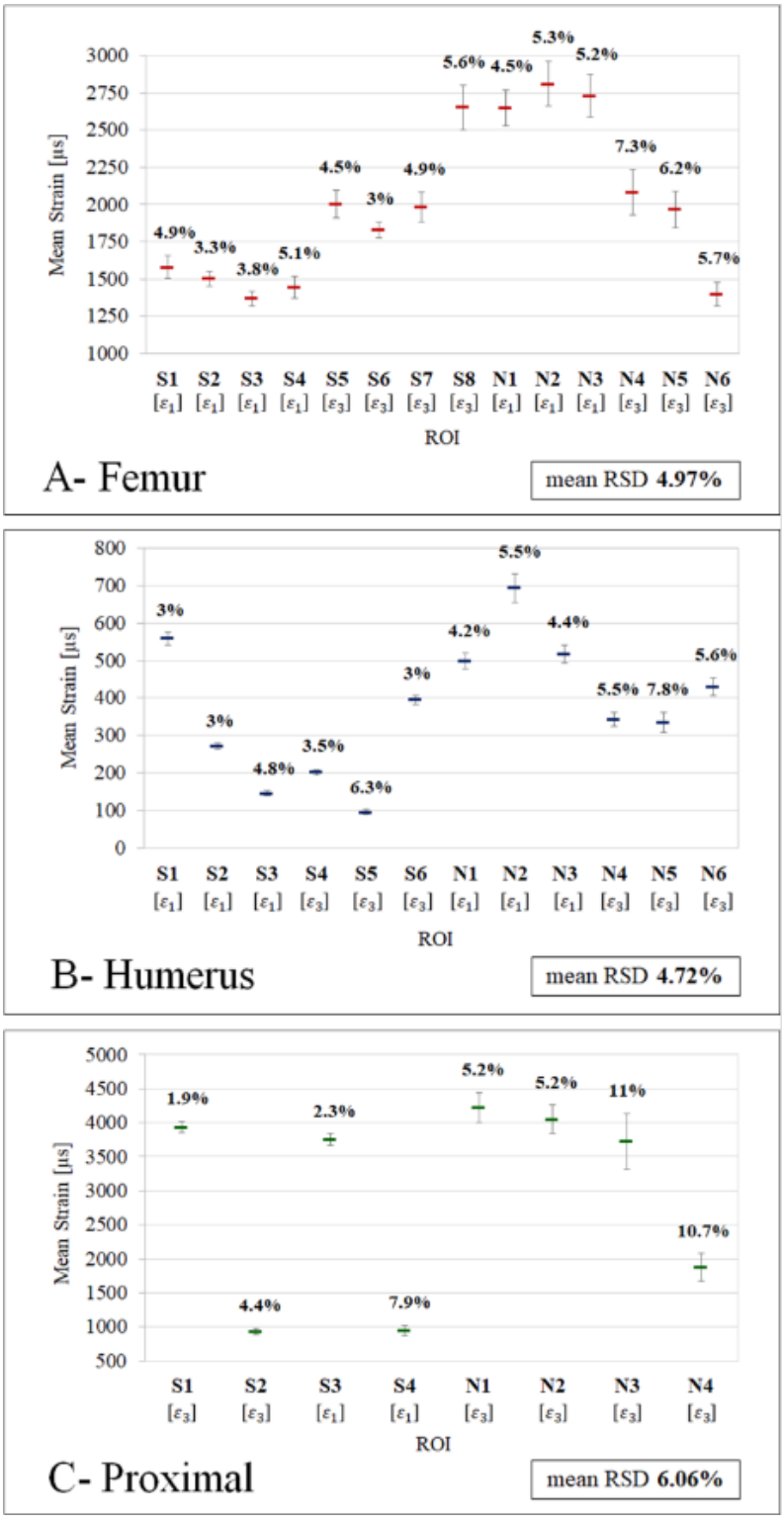


Figure 5 Mean and RSD% values of the maximum absolute principal strains at each ROI across the different scans for the femur (A), humerus (B**)** and proximal femur (C)

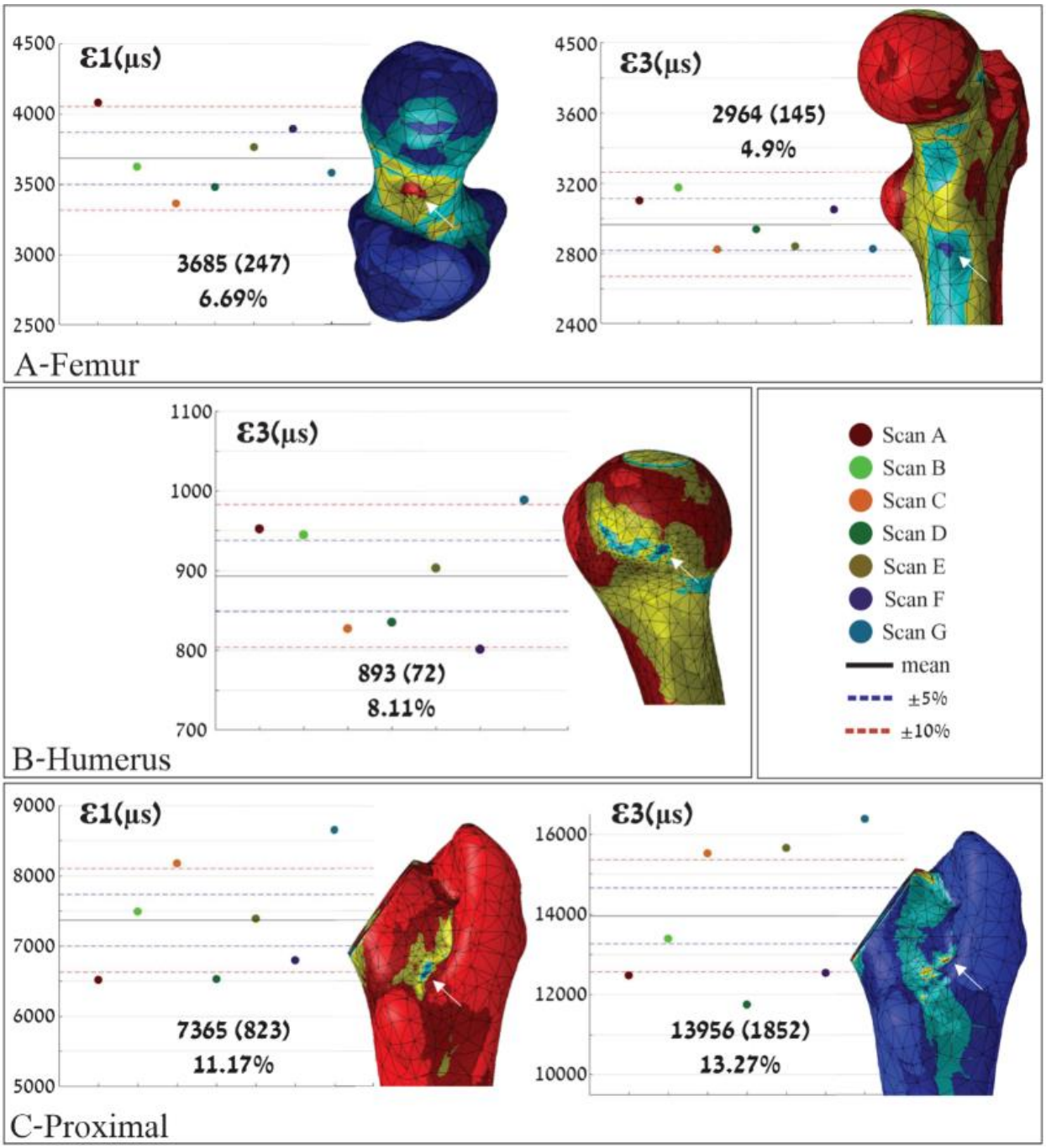


Figure 6 Maximum tensile/compressive (ε_1/ ε_3) principal strains, in the femur (A), humerus (B), and proximal femur (C). In each panel the mean (SD) and RSD% of the maximal strain is presented. The humerus is under compressive state; thus, no tensile graph is presented.

Supplementary Material

Appendix A – Full results of the maximum absolute principal strains at ROIs in the different FE models generated by the data in the different scans for the femur (A.1), humerus (A.2) and proximal femur (A.3). In each panel, the mean (SD) and RSD% for the ROI strain is presented, including the mean RSD% for the entire bone.

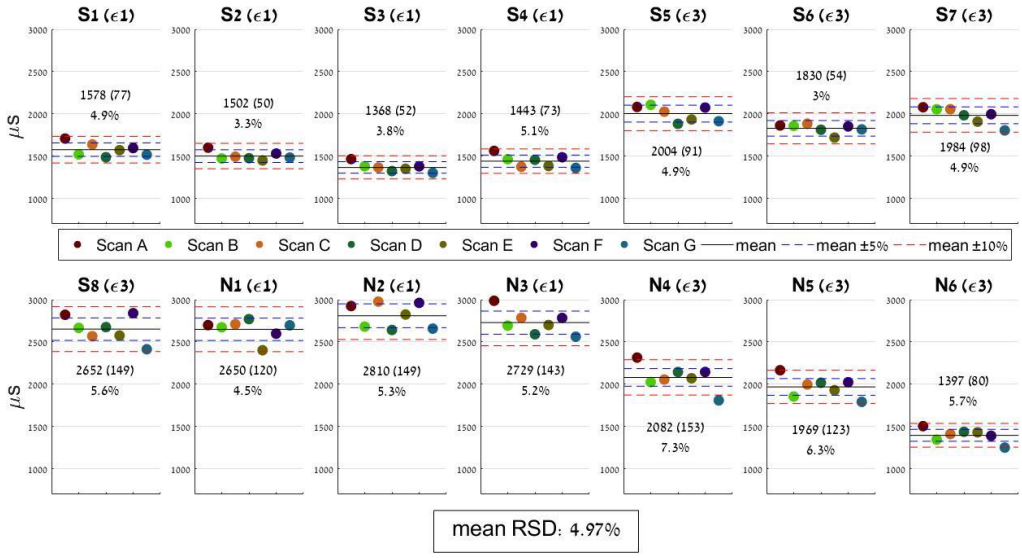


Figure A.1 - Maximum absolute principal strains at ROIs in the different scans for the femur.

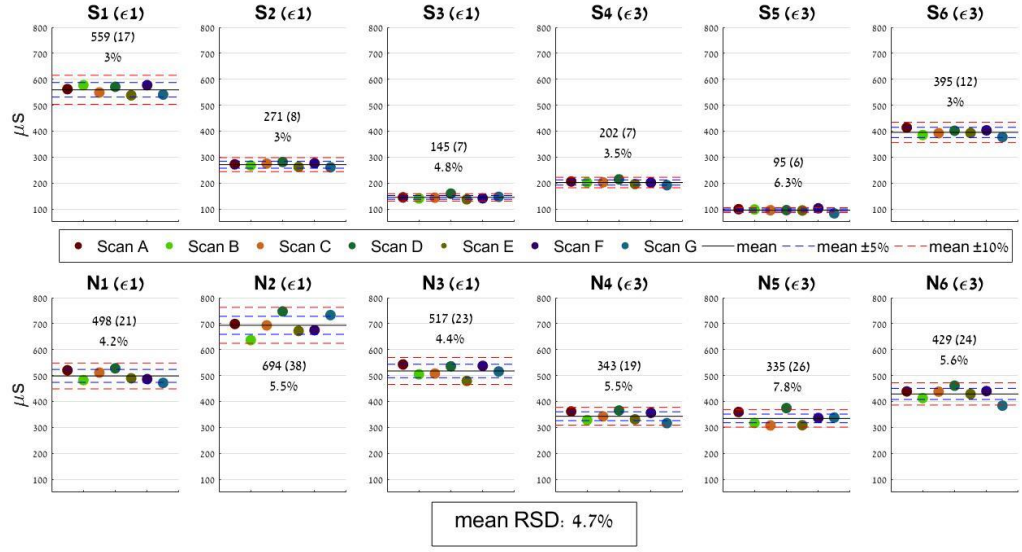
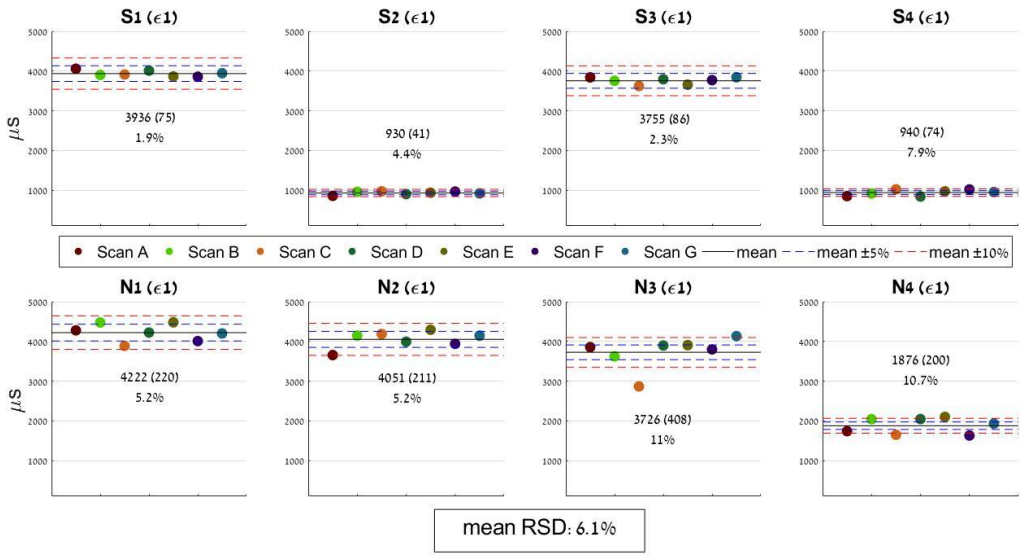


Figure A.2 - Maximum absolute principal strains at ROIs in the different scans for the humerus.



562
563 **Figure A.3** - Maximum absolute principal strains at ROIs in the different scans for the proximal femur.
564

565

Appendix B - Maximum absolute principal strains at ROIs in the different FE models generated by the data in the different scans for the constant material properties E = 10 GPa. This is to investigate the influence of the geometry difference on the principal strains. In each panel, the mean (SD) and RSD% for the ROI strain is presented, including the mean RSD% for the entire bone.

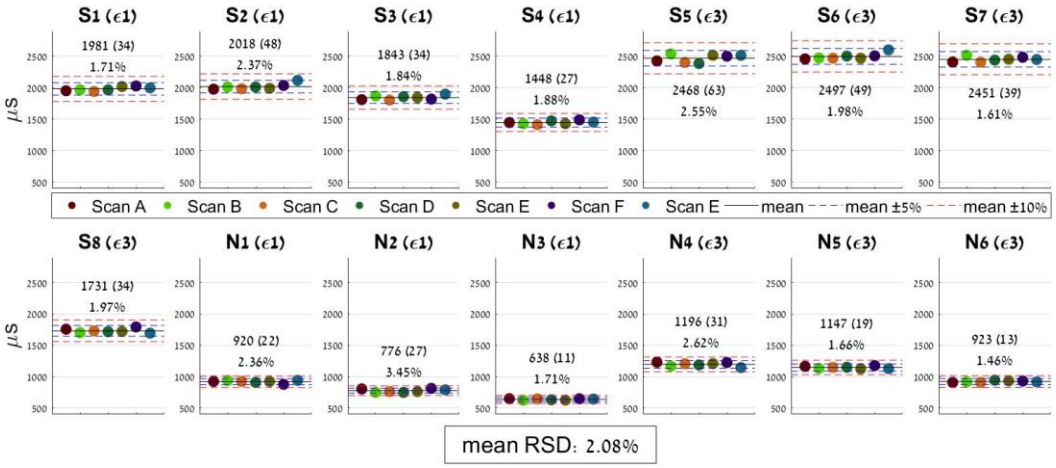


Figure B.1 Maximum absolute principal strains at ROIs in the different scans for the homogeneous femur with constant material properties E = 10 GPa.

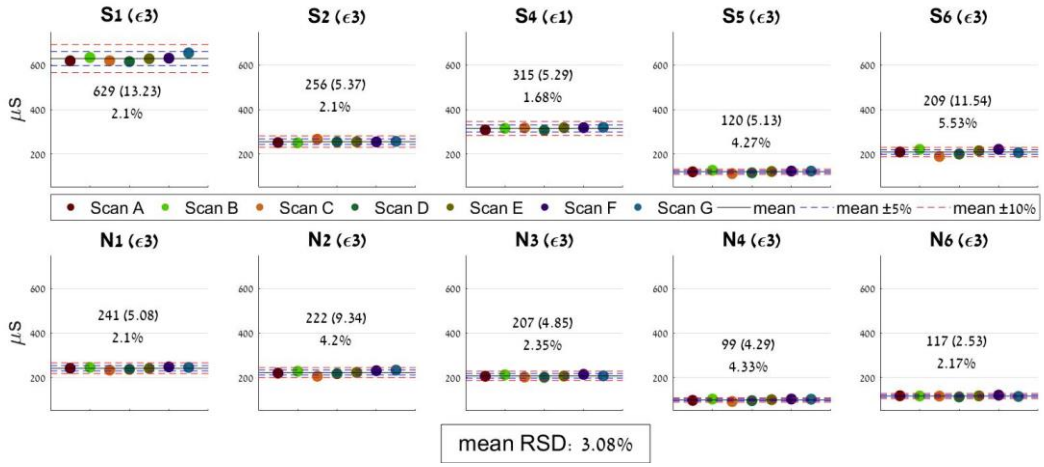


Figure B.2 Maximum absolute principal strains at ROIs in the scans for the homogeneous humerus with constant material properties E = 10 GPa. S3 and N5 are not shown since the strains there are close to zero.

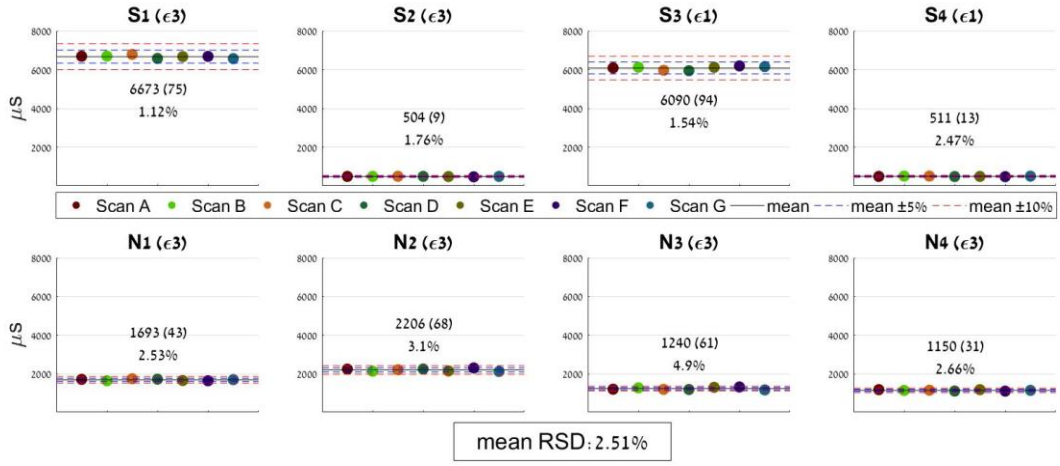


Figure B.3 Maximum absolute principal strains at ROIs in the scans for the homogeneous proximal femur with constant material properties E = 10 GPa.

Appendix C – Linear correlation matrix

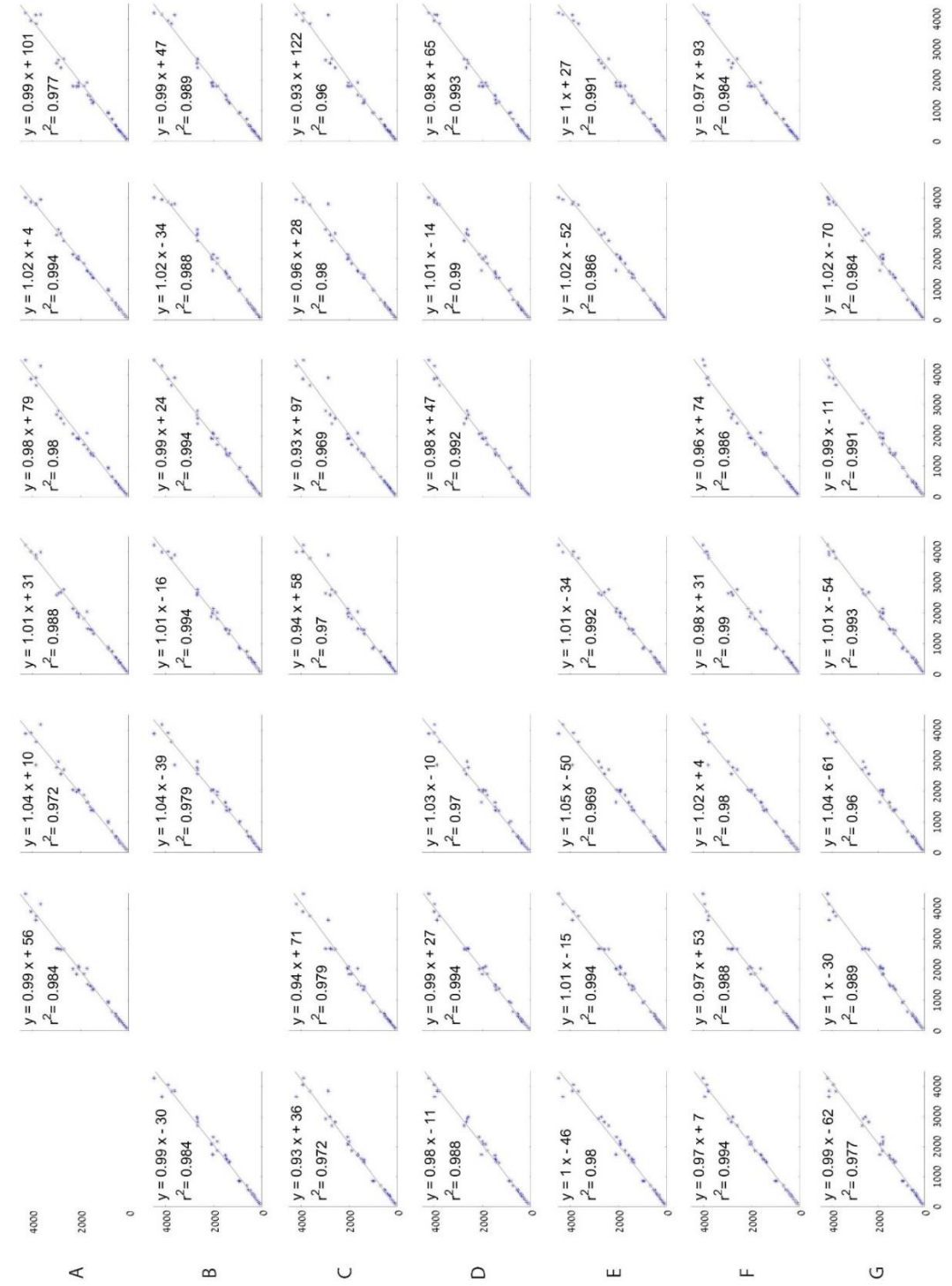


Figure C Linear correlation matrix showing agreement in strains between all pairs of QCTFE models obtained from the different CT scanners.