Virtual Mechanical Tests Out-Perform Morphometric Measures for Assessment of Fracture Healing *In Vivo*

Peter Schwarzenberg[1], Karina Klein[2], Stephen J. Ferguson[3,4], Brigitte von Rechenberg[2,4],

Salim Darwiche[2,4], Hannah L. Dailey[1]

[1]Department of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, PA, USA; [2]Musculoskeletal Research Unit (MSRU), Vetsuisse Faculty, University of Zurich, Switzerland; [3]Institute for Biomechanics, ETH Zurich, Switzerland; [4]Center for Applied Biotechnology and Molecular Medicine (CABMM), University of Zurich, Switzerland.

Corresponding Author: Peter Schwarzenberg
Packard Lab, Room 367
19 Memorial Drive West
Bethlehem, PA 18015
1-304-698-3230
Peter@Lehigh.edu

Running Title: Validation of Image-Based Models

Author Contributions:

PS designed the computational study, performed all computational work, analyzed and interpreted the data, and drafted the manuscript.

KK performed surgeries, radiographic assessment, and provided clinical advice and support.

SJF performed all biomechanical testing and aided in data analysis and interpretation.

BvR performed surgeries, radiographic scoring, and provided clinical advice and support.

SD designed the computational study, analyzed and interpreted the data, and was critical in drafting the manuscript.

HLD designed the computational study, analyzed and interpreted the data, and was critical in drafting the manuscript.

All authors have read and approved the final manuscript

Abstract

Finite element analysis (FEA) with models derived from computed tomography (CT) scans is potentially powerful as a translational research tool because it can achieve what animal studies and cadaver biomechanics cannot – low-risk, non-invasive, objective assessment of outcomes in living humans who have actually experienced the injury or treatment being studied. The purpose of this study was to assess the validity of CT-based virtual mechanical testing with respect to physical biomechanical tests in a large animal model. Three different tibial osteotomy models were performed on 44 sheep. Data from 33 operated limbs and 20 intact limbs was retrospectively analyzed. Radiographic union scoring was performed on the operated limbs and physical torsional tests were performed on all limbs. Morphometric measures and finite element (FE) models were developed from CT scans and virtual torsional tests were performed with four different material assignment techniques. In correlation analysis, morphometric measures and radiographic scores were unreliable predictors of biomechanical rigidity, while the virtual torsion test results were strongly and significantly correlated with measured biomechanical test data, with high absolute agreement. Overall, the results validated the use of virtual mechanical testing as a reliable *in vivo* assessment of structural bone healing. This method is readily translatable to clinical evaluation for noninvasive assessment of the healing progress of fractures with minimal risk. Clinical significance: Virtual mechanical testing can be used to reliably and non-invasively assess the rigidity of a healing fracture using clinical-resolution CT scans and that this measure is superior to morphometric and radiographic measures.

Keywords: Finite element analysis; Computed tomography, Tibial shaft fracture; Ovine osteotomy;

Introduction

Finite element analysis (FEA) with models derived from computed tomography (CT) scans is increasing in prevalence as a research and clinical diagnostic tool in orthopaedics.[1–8] This technique is able to accurately capture both the complex geometries and localized material acquired by the CT scan to build representative animal-specific or patient-specific models.[9,10] FEA can then be used to simulate different loading scenarios and assess the structural biomechanics of the anatomy of interest.

With animal studies and cadaver biomechanics, structural outcomes can be directly measured. However, outcome assessments such as benchtop biomechanical testing and histology cannot be performed in clinical settings with human patients. To that end, image-based models are potentially powerful as a clinically translatable research tool because they can achieve what animal studies and cadaver biomechanics cannot – low-risk, non-invasive, objective assessment of outcomes in living humans who have actually experienced the injury or treatment being studied. For example, we recently developed CT-derived computational methods to calculate the virtual torsional rigidity (VTR) of human tibial fractures after 12 weeks of healing.[9] In a pilot study, we showed that this measure is strongly correlated with clinical healing outcomes such as time to clinical union.[10] Furthermore, integrating image-based models in preclinical in vivo studies may reduce the required number of experimental animals, in accordance with 3R principles, by allowing non-invasive follow-up assessments of structural healing kinetics.

One barrier to widespread clinical use of patient-specific modeling techniques is that unlike in preclinical animal models, the quantitative predictions from these models cannot be easily validated due to a lack of physical biomechanical testing data in humans. Previous investigators have tried to bridge the gap between preclinical and clinical research techniques by using biomechanical testing of small animal osteotomy models to validate clinical assessments such as the radiographic union score (RUS or RUST when applied to the tibia). Recently, two independently conducted murine studies compared radiographic assessments with biomechanical data and CT-derived callus morphometry measures.[11,12] When interpreted together, these studies suggest that radiographic scoring is less reliable (lower inter-observer consistency) and less predictive of biomechanics (weaker correlation) at early timepoints in normal healing condition and with compromised fracture healing conditions compared to later timepoints in normal healing conditions. One previous study examined the repeatability of radiographic scoring when applied in well-healed ovine tibial osteotomies, but did not report correlations with the results of physical biomechanical testing.[13] These gaps in our ability to assess fracture union quantitatively across the spectrum of healing could be resolved by adoption of CT-based virtual biomechanics, but the necessary validation data showing that the virtual tests replicate the physical results have not yet been presented.

Accordingly, the purpose of this investigation was to assess the validity of CT-based virtual mechanical testing with respect to physical biomechanical testing in a large animal model. For this purpose, we used a series of standardized ovine (sheep) tibia osteotomies stabilized with internal fixation for which both CT scans and postmortem biomechanical testing data were available and which represented a wide spectrum of healing responses. The hypothesis of this study was that the results of physical and virtual biomechanical testing would be strongly correlated.

Methods

Specimen Information:

Forty-four adult female Swiss alpine sheep (2-3 years old, weighing 59 – 87 kg) were part of two previously completed research studies with three different tibial osteotomy models (Figure 1) stabilized by medial plating. In total, there were 33 operated limbs and 20 intact control limbs that had both mechanical test data and CT scans available for analysis in this study. Taken together, these animals comprised three experimental datasets across a wide spectrum of healing responses. Dataset 1 consisted of data from seven animals with a 3 mm gap defect stabilized with a 12-hole stainless steel plate (broad straight veterinary 3.5 mm locking compression plate (LCP), 159 mm in length, with 3.5 mm bicortical screws; DePuy Synthes®). Dataset 2 consisted of data from 18 animals with a 3 mm gap defect stabilized with a six-hole titanium plate (broad 4.5/5.0 mm LCP, 115.8 mm in length, with 5 mm bicortical screws; DePuy Synthes®). Dataset 3 consisted of data from eight animals with a 17 mm defect augmented with autografts and stabilized with a 13-hole stainless steel plate (broad straight veterinary 3.5 mm LCP, 172 mm in length, with 3.5 mm bicortical screws; DePuy Synthes®). The 3 mm defect models (Datasets 1 and 2) represented a non-critical size defect capable of spontaneously healing. Sheep in Datasets 1 and 2 were sacrificed at 9 weeks. The 17 mm graft model (Dataset 3) represented a critical size defect that would not spontaneously heal without autograft augmentation. Sheep in Dataset 3 were sacrificed 12 weeks after surgery. All animal experiments were conducted according to the Swiss laws of animal protection and welfare and authorized by the local governmental veterinary authorities (license numbers ZH071/17 and ZH183/17).

Radiographic Union Scoring:

Plain film radiographs were taken postmortem at the time of sacrifice in three different planes: anteroposterior (AP), mediolateral (ML), and an angled plane between the AP and ML viewed from the cranial aspect of the limb. Radiographs were evaluated by two independent, board certified expert reviewers (Prof. Dr. med. vet. Brigitte von Rechenberg, Dipl. ECVS and Prof. Dr. med. vet. Mark Flückiger, Dipl. ECVDI). Semiquantitative radiographic union scores were assigned following a scoring system based in part on the modified radiographic union score for tibial fractures (mRUST) method.[14–16] Our scoring approach included the callus bridging assessment of clinical mRUST, except that due to medial plate fixation, only the lateral callus was scored from the AP view due to the presence of the medial plate. Half-point scores were allowed for greater granularity on the uniform osteotomies. The resulting unicortical score is hereafter referred to as lateral callus granular mRUST [score range 1 - 4 in 0.5 increments]. The callus bridging scores were performed on the AP radiograph. Additionally, the following scoring criteria we included for callus maturity, which are not typically assessed in clinical mRUST: cortical callus formation (AP projection) [0-4], cis-cortex callus formation (AP projection) [0-4], trans-cortex callus formation (AP projection) [0-4], cranial cortical gap (ML projection) [0-4], caudal cortical gap (angled projection) [0-4], and callus opacity scores (AP projection) [0-3]. All scoring components were summed, resulting in a comprehensive radiographic union score (cRUS) on an interval range [1-27] with higher scores representing more advanced healing. A scoring breakdown can be seen in the supplementary digital content.

Micro-Computed Tomography (µCT) Scanning:

After animal sacrifice, tibiae were excised, stripped of soft tissue, wrapped in saline-soaked gauze, and µCT scanned using an XtremeCT II Micro-CT scanner (Scanco Medical AG, Bruettisellen, Switzerland) with an X-ray voltage of 68 kVp and X-ray current of 1470 µA. The resulting scans had an isotropic resolution of 60.7 µm. A phantom (Scanco KP70 phantom, QRM, Moehrendorf, Germany) was scanned in the same scanner at identical settings allowing for conversion from native Hounsfield Units (HU) to calibrated mineral density (ρ_{QCT}, mgHA/cm^3).

Biomechanical Data

Physical torsion tests were performed on all included samples using a custom-made fixture on an Instron E10000 electro dynamic testing machine (Instron, Massachusetts, US). Axial loads and torques were measured with a calibrated load cell (±10 kN / ±100 N-m). Each tibia was prepared by stripping the periosteum on both ends and fixing the tibia in the test frame using Beracryl embedding medium. Periosteum was not stripped in the region of the callus, only at the embedded ends. To minimize bone movement in the Beracryl, four adjunctive screws were inserted perpendicular to the mechanical axis of the bone in the distal epiphyses. Potting depth was adjusted to ensure a minimal distance of 10 mm between the Beracryl and the nearest screw hole from the fracture plate. The spacing between the proximal and distal Beracryl surfaces after embedding was recorded and ranged between 140 mm and 160 mm. While potting, the diaphysis was kept moist using saline-soaked gauze. Mechanical tests were performed by quasi-statically preloading the limb with 5 N axially, which was held for the entire test, and then applying internal rotation at 5 °/min. Biomechanical torsional rigidity was calculated as a linear regression of the loading curve between 6 and 10 Nm multiplied by the gauge length, or the distance between the surfaces of the two Beracryl pots, of the test specimen during testing.

Scan processing and 3D model construction:

CT scans were processed using the Mimics Innovation Suite (Materialise, Leuven, Belgium) following a similar work flow as Schwarzenberg et al.[9] To better replicate the clinical scans used in the previous study and ensure future translatability, all of the µCT scans were initially down-sampled to an isotropic resolution of 400 µm prior to segmentation. Density threshold values of 400-2500 HU and 2500-4000 HU were chosen to initially segment the callus and cortical bone respectively (Figure 2c-d). Proximal and distal cut-planes were selected just interior to the most proximal and distal screw holes, resulting in two planar cortical surfaces for FE boundary condition (BC) application.

Following preliminary threshold-based segmentation of callus and cortical bone, all models were screened for formation of high-density tissue at the cortical-callus boundary, which can occur in samples with more advanced healing. To accomplish this, each callus mask was reviewed in a slice-by-slice manner throughout the CT-stack. In each slice, voxels of high-density new tissue that fell outside the contours of the original callus mask were assigned to a new tissue mask region. This step ensured that the callus morphometric measures were accurate and that the cortical bone fragments represented exclusively the original bone contour and none of the new tissue growth. After the mask for the new tissue was segmented, it was subtracted from the density-based cortical bone mask to create the true pre-existing cortical bone mask (Figure 2e-f). These regions are referred to as callus and bone within this work. Final geometries were visually inspected by a surgeon to ensure the virtual models were anatomically representative of the biological samples. The final models for all operated limbs can be seen in Figure 3.

After the callus and bone masks were accurately segmented, the masks were united into a single surface model for preparation of volumetric discretization. The united models were wrapped with a gap closing distance of 1 mm and a smallest detail of 0.5 mm to produce cohesive surfaces.

Finite element model creation:

The united surface models were exported to a dedicated toolkit in the Mimics Innovation Suite, 3-Matic, where the surface models were smoothed to further reduce CT scan noise and to fill any small defects. Next a triangular mesh was applied to the surface of each model with maximum edge lengths of 0.4 mm and a linear tetrahedral volumetric mesh was applied to the body of each model with maximum interior edge lengths of 0.875 mm. These parameters are based on a previously published mesh convergence study conducted with human fracture callus models[9].

Elementwise material properties were initially applied to the FE meshes (Figure 2h) by three non-species-specific methods: one linear and two power law material assignments taken from the literature (Figure 4). First, a direct conversion from HU to elastic modulus was achieved using a linear scaling law.[17] For the two power law material assignments, the native HU density measure of the CT scan was first converted to calibrated mineral density (ρ_{QCT}) using the phantom and then to ash and apparent density using equations from Schileo et al.[18] Once voxel density values were in ash and apparent densities, material scaling law equations were used from Keyak et al.[19] and Morgan et al.[20] respectively to calculate elastic modulus. These two power law equations were chosen to represent the lower and higher end of elastic modulus assignment from two literature reviews.[21,22] A species-specific scaling law for ovine tibial cortical bone was not identified from a literature search and was subsequently developed in this study.

Virtual torsion testing

In preclinical studies, torsion testing is the preferred mode of assessment due to its robustness to variation in setup orientation. Structural FE simulations replicating this torsional test were carried out in ANSYS 17.2 (Canonsburg, Pennsylvania). The boundary conditions were rigid fixation on the distal end and 1 degree of applied rotation on the proximal end, with twisting about the aligned mechanical axis of the tibia. Additionally, using the same multi-point constraint (MPC) contact of the applied axial rotation, the proximal end was fixed in the anterior-posterior and medial-lateral translational directions to best replicate the physical test setup (Figure 2i). For each model, the virtual torsional rigidity (VTR) was calculated as the moment reaction from the applied loading (M) multiplied by the working length of the test segment (L) divided by the applied angle of twist (ϕ): $VTR = ML/\phi$ [Nm^2/°].

Material Optimization

An optimization technique similar to Eberle et al.[23] was performed on the N = 20 intact tibiae to find a material assignment law of the form $E = a\rho_{QCT}^{b}$ that best fit the data. In brief, value ranges for a ($5{,}000 \leq a \leq 20{,}000$) and b ($1 \leq b \leq 3$) were chosen from the same literature review as above[21,22] to represent the current published data across multiple anatomic sites. Sixteen design points were chosen to span the parameter space and the torsional rigidity of each intact model was calculated by assigning all elementwise material properties within all N = 20 models using the coefficients a and b at those points. A root mean square error (RMSE) was calculated comparing each design point (coefficients a and b) to the biomechanical data and a 2nd order response surface was fit to the a, b, and RMSE values. The minimum predicted RMSE of the response surface was calculated by 120,801 sample points in the design space, defined by a resolution of 25 $\left(MPa\frac{cm^3}{gHA}\right)$ in the a dimension and 0.01 (-) in the b dimension. After the initial response surface minimum was calculated, three additional refinement design points were iteratively computed from the surface minimum. Finally, the minimum a and b values for the response surface with 19 design points (16 preliminary and 3 refinement) was used to assign elementwise material properties in all N = 33 operated limb models and N = 20 intact limb models. The resulting virtual torsional rigidities (VTRs) were calculated for further statistical analysis.

Statistical Analysis

Descriptive statistics were generated using Microsoft Excel and MATLAB (R2016a-R2019a, The MathWorks, Inc., Natick, Massachusetts). Additional statistical analyses were conducted in SPSS 25 (IBM Corp., Armonk, NY). Pearson correlations were performed to assess the strength of the linear association between various outcome measures and the measured biomechanical rigidity from physical testing. Correlation coefficients were interpreted as follows[24,25]: weak $R^2 \leq 0.3$, moderate $0.3 < R^2 < 0.6$, or strong $R^2 \geq 0.6$. One-way repeated-measures ANOVAs were computed to determine group differences between the biomechanical rigidity and the simulated VTR data. Additionally, root mean squared error (RMSE) analyses were performed to evaluate the relative performance of the four material property assignment techniques (three non-species-specific, one ovine-optimized). All reported values are medians and interquartile ranges (IQR) unless otherwise reported and statistical significance was at $p < 0.05$.

Results

Structural Data: Biomechanical Testing

The results of physical biomechanical testing for all samples are shown in Figure 5 and Table 1. Dataset 3 (17 mm defect with graft) had the lowest torsional rigidity, as expected in this delayed healing model. Between-groups comparisons were not performed between Datasets 1, 2 and 3 because these experiments were not conducted concurrently with intent to compare outcomes. Instead, these torsional rigidity results were compiled into a single osteotomy dataset, representing a range of healing responses, and were subsequently used as a benchmark to assessing the reliability of various outcomes measures (morphometric, radiographic, and virtual biomechanics) with respect to the measured physical properties of the specimens.

Morphometric Data

Summary morphometric data for the three osteotomy datasets individually and for all samples together is presented in Table 2. Callus volume was weakly-to-moderately and significantly correlated with biomechanical torsional rigidity in Datasets 2 and 3 and for the 33 compiled osteotomies overall. Callus density was weakly and non-significantly associated with biomechanical torsional rigidity in all datasets and for the 33 compiled osteotomies overall.

Radiographic Data

Radiographic scoring data can be seen in Table 3. The lateral callus granular mRUST score was moderately and significantly correlated with biomechanical torsional rigidity in Dataset 1 only, but weakly and non-significantly correlated with the other two datasets and the 33 compiled osteotomies overall. Similarly, the comprehensive radiographic union score (cRUS) was strongly and significantly correlated with biomechanical torsional rigidity in Dataset 1 only, with weak and non-significant correlations for the other two datasets and the 33 compiled osteotomies overall.

Structural Data: Virtual Torsional Rigidity (VTR)

Virtual torsional rigidity (VTR) data from all simulations is shown in Figure 5 and Table 4. As described above, each of the N = 33 osteotomy limbs and N = 20 intact limbs were subjected to virtual testing with four different material assignment approaches: three non-species-specific scaling laws (Synder linear, Morgan power law, and Keyak power law) and one optimized species-specific ovine power law. Across all osteotomy datasets, the compiled osteotomies, and the intact limbs, regardless of material assignment law, the virtual and physical biomechanical tests were moderately-to-strongly and significantly correlated. Unsurprisingly, the VTRs resulting from the optimized ovine-specific material law performed as well or better than the non-optimized non-species-specific material assignment approaches.

Figure 5 compares the predicted VTR values for the compiled osteotomies and the intact limbs to the physical biomechanical torsional rigidity. Significant differences are indicated with respect to physical biomechanical torsional rigidity. The non-species-specific material assignment laws significantly over-predicted rigidity, despite their high correlation with the biomechanical test results. The RMSE analysis showed that in both the fractured and intact models, the Keyak power law material assignment method had the lowest relative error of the non-species-specific material assignment laws (Table 5).

Discussion

The goal in many preclinical orthopaedic studies is to measure the healing progress of the fractured bone with a surrogate measure such as torsional stiffness or rigidity. To accomplish this, animals need to be sacrificed and bones excised for physical testing, which cannot be done in clinical studies or at interim timepoints in an animal model. Currently, morphometric measurement of callus formation (volume and density) is a commonly used technique that can be performed *in vivo* to evaluate fracture healing progress prior to animal sacrifice.[26,27] Furthermore, the clinical gold-standard radiographic scoring systems such as RUST and mRUST are used in both preclinical and clinical settings to track healing progress. In this study, our method of virtual torsion testing outperformed all of these measures when predicting the biomechanical rigidity of the bone. Virtual torsion testing with the species-specific material model also produced strong and statistically significant correlations and low error between the measured and simulated biomechanical properties across a wide range of healing responses.

The study results also indicate that morphometric measures of callus formation should be interpreted with caution. Callus volume and density are often considered surrogate measures of healing progress or even indirect indicators of biomechanical properties. In this context, callus volume is analogous to the amount of new tissue visible on plain film radiographs, while callus density can help infer tissue maturity and structural integrity. Both are interpreted as signs of healing.[28] In this study, callus volume had a moderate correlation with measured torsional rigidity in Dataset 1 and Dataset 3 and weak correlations with torsional rigidity in Dataset 2. Dataset 3 was designed to represent a critical size defect which is more indicative of a delayed healing case which may have more dependency on callus volume than a non-critical size defect model. Furthermore, this dependency on callus volume may be attributed to Dataset 3 being a delayed healing model in which new bone is still expected to significantly remodel and condense after the analyzed endpoint of 12 weeks. However, within all datasets, callus density showed little association with biomechanics, which could be because all callus present was of similar maturity. In a preclinical model with more widely varying timepoints, callus density could be a more important indicator of healing progress. In our controlled osteotomy models the coefficients of variation for callus volume (36%, 38% and 24%) were much higher than for callus density (10%, 8% and 7%). This is consistent with callus volume and density patterns observed in humans at 12 weeks[10]. Notably, the plates used for fixation were not the same across datasets, which could explain some of the differences in healing response, although as previously stated, between-groups comparison of the fixators was not the objective of this investigation. In fact, by design we aimed to show the validity of our hypothesis across various fixation techniques and defect sizes. The observed low correlation of morphometric measures with outcome parameters seen in this study are consistent with findings from our clinical pilot studies, which used the same virtual mechanical testing technique and showed that morphometric parameters are not strongly correlated with simulated VTR or with clinical outcomes such as time to union.[10]

Additionally, the radiographic scoring results do not provide a consistent picture. The radiographic scores in Dataset 1 had much stronger correlations with biomechanical rigidity than the other two models. The callus in Dataset 1 also tended to be the largest (Table 2). This could suggest that visual radiographic scoring is more reliable when there is more callus present. If true, this would be consistent with murine data showing that radiographic scoring is less reliable and less predictive of biomechanics at early timepoints and with compromised fracture healing compared with later timepoints in normal healing conditions[29].

Overall, this study validated the robustness of virtual mechanical testing as an *in vivo* method for assessing structural bone healing. With the optimized species-specific material assignment law for ovine tibial cortical bone, we found a strong and statistically significant correlation between virtual and physical torsional rigidity across all sample subgroups, which was superior to both radiographic and morphometric assessments from CT. Even with non-species-specific material models, the virtual mechanical test results were still moderately-to-strongly and significantly correlated with physical biomechanical testing. This finding is particularly notable because it indicates that virtual mechanical testing may be a powerful *in vivo* assessment tool, even when a robust species- and site-specific material assignment law is not available.

This study had a few additional minor limitations that may be restricting the predictive power and correlations between measures. First, as in all physical biomechanical torsion tests, there were small potting artifacts where the bone was able to slightly deform within the PMMA pot. This could reduce the measured stiffness and rigidity of the biomechanical data, raising the relative error between physical and simulated data. Additionally, while higher resolution µCT scans were acquired, the scans were down sampled to replicate the resolution of clinical scanners. Starting with a less resolute data set eliminates the possibility of more sophisticated structural material modeling, such as the use of fabric tensors, and could have also introduced mask boundary definition errors in the virtual models. However, µCT scan data is not available in a clinical setting or at interim *in vivo* timepoints in large animal models, so the clinical resolution chosen here has clear translational relevance.

Finally, it is important to note that the optimized scaling law we developed for the ovine tibial diaphysis using the intact limbs may not accurately model the properties of callus tissue at all stages of maturity. At the timepoint our scans were taken, fracture callus consisted of fibrous tissue, some cartilage, and woven bone, which is not as organized or mineralized as cortical bone. During coupled remodeling, callus microstructure, composition, and mechanics change adaptively[30], and the modeling approach we are using my not be adequate to capture all of those processes. Optimization of new material property scaling laws for combined cortical-callus structures should be a target for continuing research and would undoubtedly reduce the RMSE of the virtual prediction.

Conclusions

Image-based structural mechanics modeling from CT scans is an objective, quantitative method for assessing fracture healing that can closely replicate physical biomechanical testing. Using the methods described herein, we validated a technique for virtual torsional testing using ovine tibial osteotomy fracture models representing a wide range of healing scenarios. The virtual mechanical testing results were better predictors of trends in measured biomechanical properties than radiographic scoring or morphometric analysis of callus from CT scans in these models. This method is readily translatable to preclinical or clinical settings for noninvasive assessment of the structural healing progress of fractured long bones with minimal risk.

Acknowledgements

The preclinical studies from which this data was obtained were funded by the Johnson & Johnson Family of Companies and Biomech Innovations. The authors wish to thank Beat Lechmann (Johnson & Johnson Family of Companies) and Stefano Brianza (Biomech Innovations AG) for agreeing to grant access to the ovine studies data for our analyses. The authors also wish to thank Beat Lechmann (Johnson & Johnson Family of Companies) for Figure 1 graphics and Prof. Mark Flückiger for performing radiographic assessments together with BvR. The authors want to thank the National Science Foundation for supporting Peter Schwarzenberg with an IIE-GIRE travel grant. BvR is a paid consultant and receives research support as a principal investigator from Johnson & Johnson Family of Companies. HLD has stock or stock options, receives research or institutional support as a principal investigator, and receives financial or material support from OrthoXel. Additionally, HLD serves on a committee in the Orthopaedic Research Society (ORS).

References

1. Jaecques SVN, Van Oosterwyck H, Muraru L, et al. 2004. Individualised, micro CT-based finite element modelling as a tool for biomechanical analysis related to tissue engineering of bone. Biomaterials 25(9):1683–1696.

2. Shefelbine SJ, Simon U, Claes L, et al. 2005. Prediction of fracture callus mechanical properties using micro-CT images and voxel-based finite element analysis. Bone

36(3):480–488.

3. Trabelsi N, Yosibash Z, Wutte C, et al. 2011. Patient-specific finite element analysis of the human femur-A double-blinded biomechanical validation. J. Biomech. 44(9):1666–1672 Available from: http://dx.doi.org/10.1016/j.jbiomech.2011.03.024.

4. Taddei F, Falcinelli C, Balistreri L, et al. 2016. Left–right differences in the proximal femur's strength of post-menopausal women: a multicentric finite element study. Osteoporos. Int. 27(4):1519–1528.

5. Sternheim A, Giladi O, Gortzak Y, et al. 2018. Pathological fracture risk assessment in patients with femoral metastases using CT-based finite element methods. A retrospective clinical study. Bone 110:215–220 Available from: https://doi.org/10.1016/j.bone.2018.02.011.

6. Falcinelli C, Di Martino A, Gizzi A, et al. 2019. Mechanical behavior of metastatic femurs through patient-specific computational models accounting for bone-metastasis interaction. J. Mech. Behav. Biomed. Mater. 93(January):9–22 Available from: https://doi.org/10.1016/j.jmbbm.2019.01.014.

7. Petfield JL, Hayeck GT, Kopperdahl DL, et al. 2017. Virtual stress testing of fracture stability in soldiers with severely comminuted tibial fractures. J. Orthop. Res. 35(4):805–811.

8. De Jong JJA, Christen P, Plett RM, et al. 2017. Feasibility of rigid 3D image registration of high-resolution peripheral quantitative computed tomography images of healing distal radius fractures. PLoS One 12(7):1–12.

9. Schwarzenberg P, Maher MM, Harty JA, Dailey HL. 2019. Virtual structural analysis of tibial fracture healing from low-dose clinical CT scans. J. Biomech. 83:49–56 Available from: https://doi.org/10.1016/j.jbiomech.2018.11.020.

10. Dailey HL, Schwarzenberg P, Daly CJ, et al. 2019. Virtual Mechanical Testing Based on Low-Dose. J. Bone Jt. Surg. 101(13):1193–1202 Available from: https://journals.lww.com/jbjsjournal/Fulltext/2019/07030/Virtual_Mechanical_Testing_Based_on_Low_Dose.9.aspx#pdf-link.

11. Fiset S, Godbout C, Crookshank MC, et al. 2018. Experimental Validation of the Radiographic.:1871–1878.

12. Cooke ME, Hussein AI, Lybrand KE, et al. 2018. Correlation between RUST assessments of fracture healing to structural and biomechanical properties. J. Orthop. Res. 36(3):945–953.

13. Litrenta J, Tornetta P, Ricci W, et al. 2017. In Vivo Correlation of Radiographic Scoring (Radiographic Union Scale for Tibia Fractures) and Biomechanical Data in a Sheep Osteotomy Model: Can We Define Union Radiographically? J. Orthop. Trauma 31(3):127–130.

14. Whelan DB, Bhandari M, Stephen D, et al. 2010. Development of the radiographic union score for tibial fractures for the assessment of tibial fracture healing after intramedullary fixation. J. Trauma - Inj. Infect. Crit. Care 68(3):629–632.

15. Leow JM, Clement ND, Tawonsawatruk T, et al. 2016. The radiographic union scale in tibial (RUST) fractures: Reliability of the outcome measure at an independent centre. Bone Jt. Res. 5(4):116–121 Available from: http://www.bjr.boneandjoint.org.uk/cgi/doi/10.1302/2046-3758.54.2000628.

16. Litrenta J, Tornetta P, Mehta S, et al. 2015. Determination of Radiographic Healing: An Assessment of Consistency Using RUST and Modified RUST in Metadiaphyseal Fractures. J. Orthop. Trauma 29(11):516–520.

17. Snyder SM, Schneider E. 1991. Estimation of mechanical properties of cortical bone by computed tomography. J. Orthop. Res. 9(3):422–431.

18. Schileo E, Dall'Ara E, Taddei F, et al. 2008. An accurate estimation of bone density improves the accuracy of subject-specific finite element models. J. Biomech. 41(11):2483–2491.

19. Keyak JH. 1994. Correlations between orthogonal mechanical properties and density of trabecular bone: Use of different densitometric measures. Biomed. Mater. 28:1329–1336.

20. Morgan EF, Bayraktar HH, Keaveny TM. 2003. Trabecular bone modulus-density relationships depend on anatomic site. J. Biomech. 36(7):897–904.

21. Helgason B, Perilli E, Schileo E, et al. 2008. Mathematical relationships between bone density and mechanical properties: A literature review. Clin. Biomech. 23(2):135–146.

22. Knowles NK, Reeves JM, Ferreira LM. 2016. Quantitative Computed Tomography (QCT) derived Bone Mineral Density (BMD) in finite element studies: a review of the literature. J. Exp. Orthop. 3(1):36 Available from: http://jeo-esska.springeropen.com/articles/10.1186/s40634-016-0072-2.

23. Eberle S, Göttlinger M, Augat P. 2013. Individual density-elasticity relationships improve accuracy of subject-specific finite element models of human femurs. J. Biomech. 46(13):2152–2157 Available from: http://dx.doi.org/10.1016/j.jbiomech.2013.06.035.

24. Mukaka MM. 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. 24(3):69–71.

25. Akoglu H. 2018. User's guide to correlation coefficients. Turkish J. Emerg. Med. 18(3):91–93.

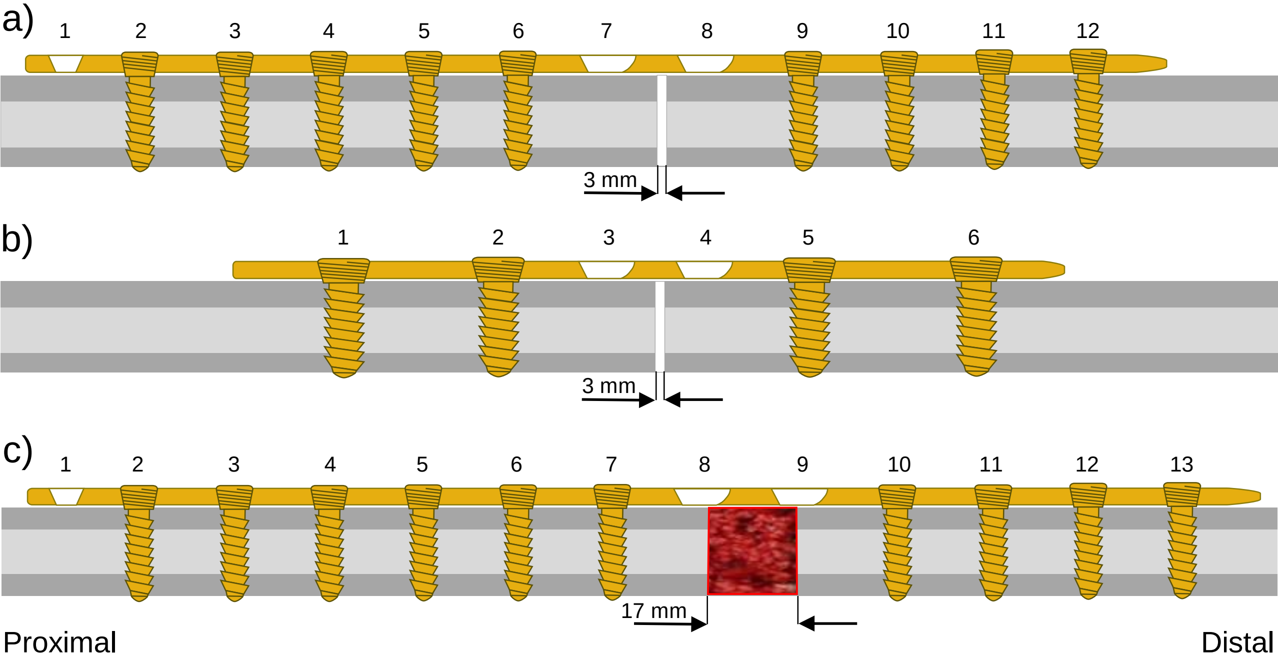
26. Bottlang M, Tsai S, Bliven EK, et al. 2016. Dynamic stabilization with active locking plates delivers faster, stronger, and more symmetric fracture-healing. J. Bone Jt. Surg. - Am. Vol. 98(6):466–474.

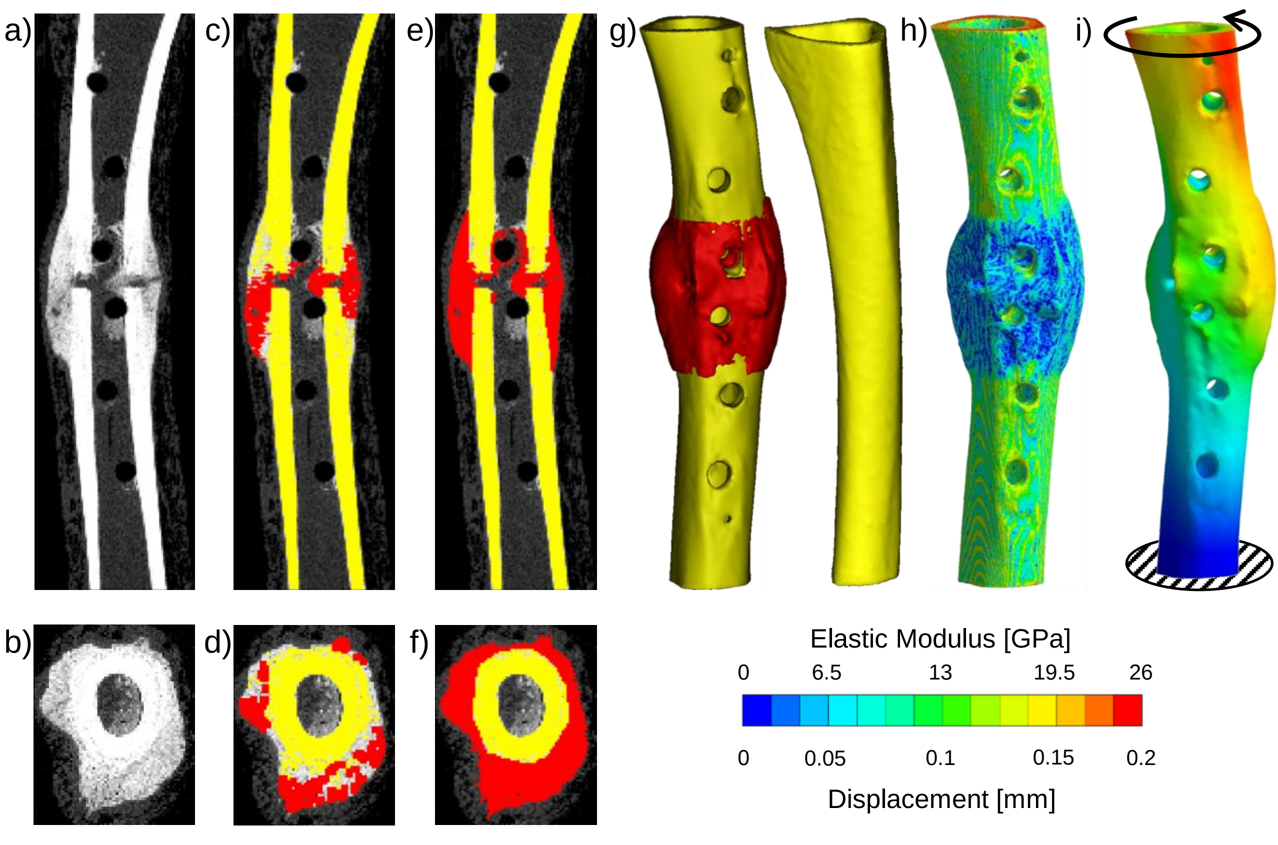
27. Richter H, Plecko M, Andermatt D, et al. 2015. Dynamization at the Near Cortex in Locking Plate. J Bone Jt. Surg Am. 97(3):208–215.

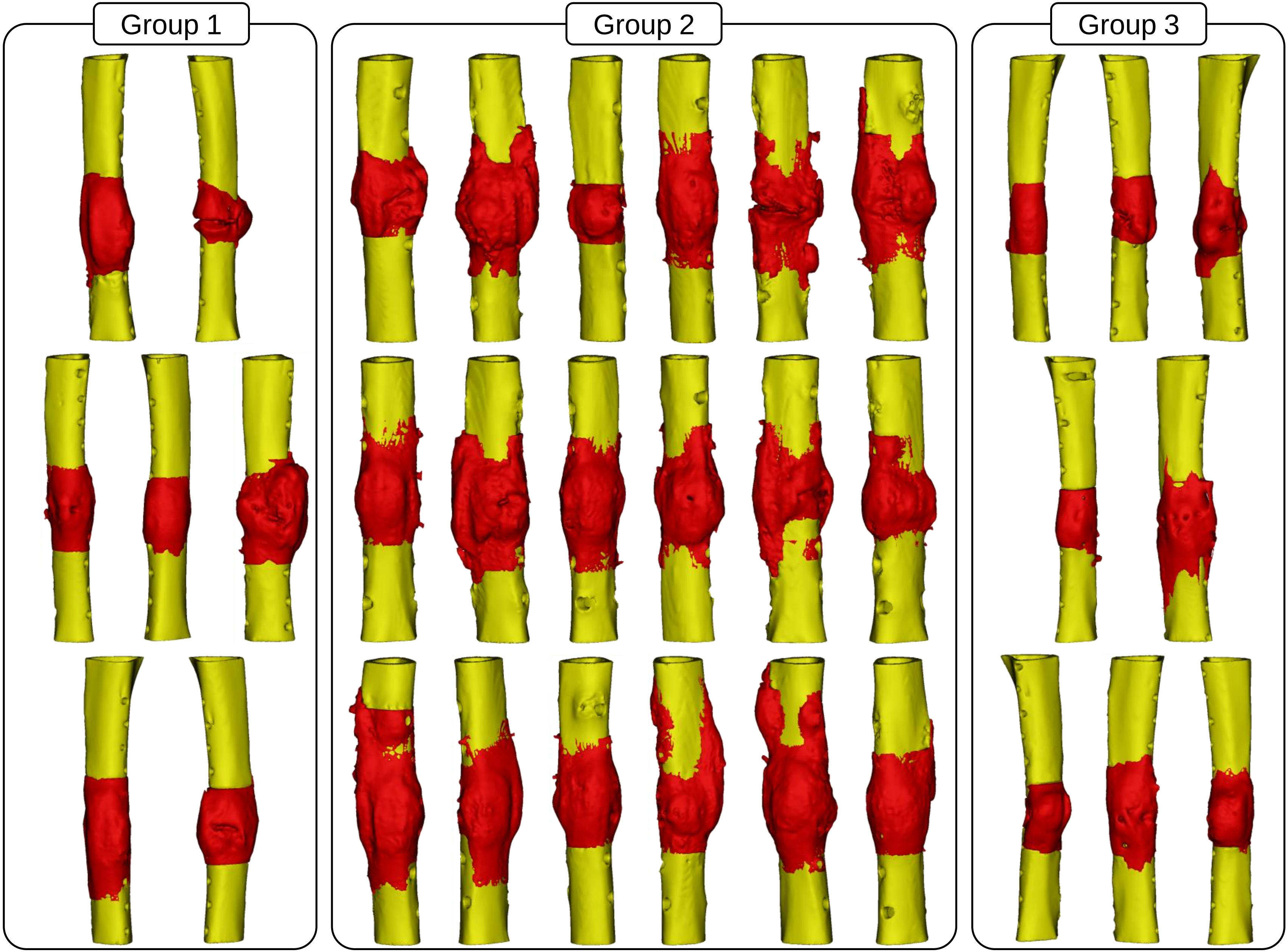
28. Augat P, Merk J, Genant HK, Claes L. 1997. Quantitative assessment of experimental fracture repair by peripheral computed tomography. Calcif. Tissue Int. 60(2):194–199.

29. Schwarzenberg P, Darwiche S, Yoon RS, et al. 2020. Imaging Modalities to Assess Fracture Healing. Curr. Osteoporos. Rep. .

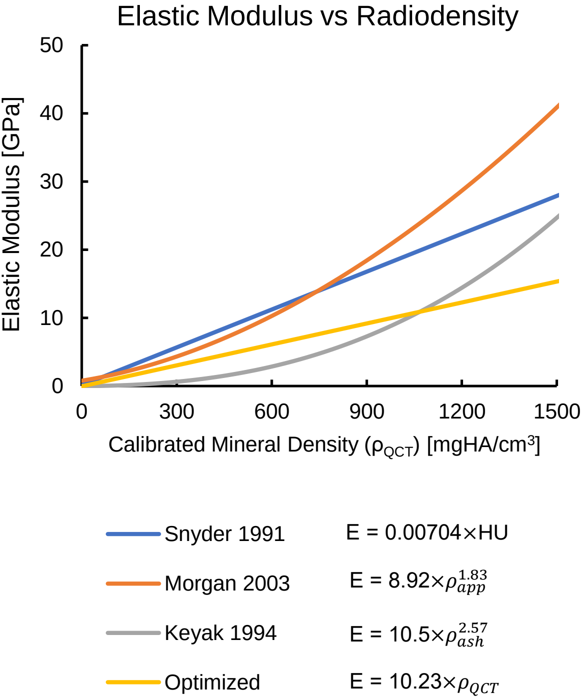
30. Casanova M, Schindeler A, Little D, et al. 2014. Quantitative phenotyping of bone fracture repair: a review. Bonekey Rep. 3(July):1–8.

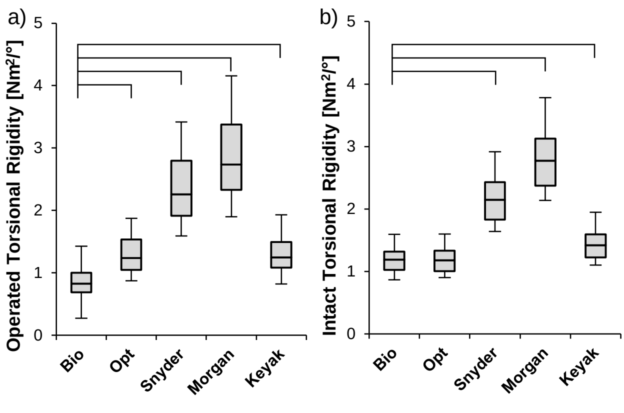

a)
1
2
3
4
5
6
7
8
9
10
11
12
3 mm
b)
1
2
3
4
5
6
3 mm
c)
1
2
3
4
5
6
7
8
9
10
11
12
13
17 mm
Proximal
Distal


a)
b)
c)
d)
e)
f)
g)
h)
i)
Elastic Modulus [GPa]
0
6.5
13
19.5
26
0
0.05
0.1
0.15
0.2
Displacement [mm]



Group 1
Group 2
Group 3


Elastic Modulus vs Radiodensity
Elastic Modulus [GPa]
50
40
30
20
10
0
0
300
600
900
1200
1500
Calibrated Mineral Density (ρ_{QCT}) [mgHA/cm^3]
Snyder 1991
E = 0.00704×HU
Morgan 2003
$E = 8.92\times\rho_{app}^{1.83}$
Keyak 1994
$E = 10.5\times\rho_{ash}^{2.57}$
Optimized
$E = 10.23\times\rho_{QCT}$



a)
b)
Operated Torsional Rigidity [Nm²/°]
Intact Torsional Rigidity [Nm²/°]
0
1
2
3
4
5
Bio
Opt
Snyder
Morgan
Keyak

Table 1: Biomechanical torsional rigidities for each osteotomy group, all osteotomies combined, and the intact tibiae.

Dataset	Sample Size	Measured Torsional Rigidity [$Nm^2/°$] Median	IQR
Dataset 1	7	1.02	(0.79 - 1.16)
Dataset 2	18	0.82	(0.71 - 0.89)
Dataset 3	8	0.72	(0.60 - 1.04)
Compiled Osteotomies	33	0.83	(0.69 - 1.00)
All Intact Tibia	20	1.19	(1.03 - 1.32)

Table 2: Callus morphometric data and correlations to biomechanical torsional rigidity for each group.

	Measure	Median	IQR	R^2 with Biomech data	*p* with Biomech data
Dataset 1	Callus Volume [cm^3]	12.54	(7.30 - 13.63)	0.56	0.055
	Callus Density [$mgHA/cm^3$]	736	(688 - 813)	0.09	0.516
Dataset 2	Callus Volume [cm^3]	10.73	(8.75 - 14.69)	0.23	**0.044**
	Callus Density [$mgHA/cm^3$]	745	(709 - 793)	0.03	0.475
Dataset 3	Callus Volume [cm^3]	10.46	(9.36 - 13.63)	0.53	**0.040**
	Callus Density [$mgHA/cm^3$]	774	(717 - 809)	0.06	0.568
Compiled Osteotomies	Callus Volume [cm^3]	10.79	(8.54 - 14.16)	0.26	**0.002**
	Callus Density [$mgHA/cm^3$]	748	(705 - 797)	0.01	0.636

Table 3: Radiographic scores and correlations to biomechanical torsional rigidity for each group.

	Measure	Median	IQR	R^2 with Biomech data	*p* with Biomech data
Dataset 1	Lateral Callus Granular mRUST	3.0	(3.0 - 3.0)	0.58	**0.048**
	Total Radiographic Score (cRUS)	20.0	(17.5 - 20.0)	0.72	**0.016**
Dataset 2	Lateral Callus Granular mRUST	2.5	(2.13 - 3.0)	0.18	0.075
	Total Radiographic Score (cRUS)	18.5	(14.6 - 19.4)	0.01	0.714
Dataset 3	Lateral Callus Granular mRUST	2.0	(2.0 - 3.0)	0.00	0.888
	Total Radiographic Score (cRUS)	16.3	(15.0 - 18.6)	0.06	0.566
All Samples Together	Lateral Callus Granular mRUST	3.0	(2.0 - 3.0)	0.01	0.650
	Total Radiographic Score (cRUS)	18.5	(15.5 - 19.5)	0.07	0.131

Table 4: Virtual Torsional Rigidities (VTRs) and correlations to biomechanical torsional rigidity for each group.

	Resultant Virtual Torsional Rigidity (VTR) with Each Material Assignment Law	Median	IQR	Units	R with Biomech data	R^2 with Biomech data	*p* with Biomech data
Dataset 1	VTR - Optimized	0.94	(0.70 - 1.20)	Nm2/°	0.89	0.79	**0.008**
	VTR - Snyder	2.34	(1.78 - 2.99)	$Nm^2/°$	0.89	0.78	**0.008**
	VTR - Morgan	2.83	(2.14 - 3.67)	$Nm^2/°$	0.89	0.79	**0.008**
	VTR - Keyak	1.29	(0.87 - 1.56)	$Nm^2/°$	0.92	0.85	**0.003**
Dataset 2	VTR - Optimized	0.86	(0.81 - 0.96)	Nm2/°	0.71	0.50	**0.001**
	VTR - Snyder	2.14	(2.04 - 2.43)	$Nm^2/°$	0.71	0.50	**0.001**
	VTR - Morgan	2.61	(2.50 - 2.94)	$Nm^2/°$	0.74	0.54	**0.001**
	VTR - Keyak	1.24	(1.14 - 1.35)	$Nm^2/°$	0.79	0.63	**< 0.0005**
Dataset 3	VTR - Optimized	0.83	(0.75 - 1.18)	Nm2/°	0.82	0.67	**0.013**
	VTR - Snyder	2.09	(1.88 - 2.94)	$Nm^2/°$	0.82	0.67	**0.013**
	VTR - Morgan	2.5	(2.26-3.55)	$Nm^2/°$	0.82	0.67	**0.012**
	VTR - Keyak	1.08	(0.95 - 1.57)	$Nm^2/°$	0.84	0.71	**0.008**
Compiled Osteotomies	VTR - Optimized	0.87	(0.78 - 1.14)	Nm2/°	0.80	0.63	**< 0.0005**
	VTR - Snyder	2.15	(1.95 - 2.84)	$Nm^2/°$	0.80	0.63	**< 0.0005**
	VTR - Morgan	2.65	(2.35 - 3.44)	$Nm^2/°$	0.81	0.66	**< 0.0005**
	VTR - Keyak	1.22	(1.05 - 1.54)	$Nm^2/°$	0.86	0.73	**< 0.0005**
Intact Limbs	VTR - Optimized	0.81	(0.70 - 0.93)	Nm2/°	0.84	0.70	**< 0.0005**
	VTR - Snyder	2.00	(1.73 - 2.29)	$Nm^2/°$	0.84	0.70	**< 0.0005**
	VTR - Morgan	2.60	(2.23 - 2.97)	$Nm^2/°$	0.84	0.71	**< 0.0005**
	VTR - Keyak	1.34	(1.14 - 1.53)	$Nm^2/°$	0.85	0.73	**< 0.0005**

Table 5: Root mean squared error with biomechanical torsional rigidity for each group.

	Structural Model	RMSE with Biomechanical Rigidity
Dataset 1	Optimized	0.96
	Snyder	2.38
	Morgan	3.06
	Keyak	0.83
Dataset 2	Optimized	0.62
	Snyder	1.90
	Morgan	2.52
	Keyak	0.63
Dataset 3	Optimized	0.86
	Snyder	2.30
	Morgan	3.96
	Keyak	0.77
Compiled Osteotomies	Optimized	0.77
	Snyder	2.11
	Morgan	2.75
	Keyak	0.71
All Intact Limbs	Optimized	0.09
	Snyder	1.16
	Morgan	1.77
	Keyak	0.47

Supplementary Table 1: Scoring breakdown for radiographic assessment. Includes scores for both lateral callus granular mRUST and comprehensive RUS scores.

• Granular mRUST Score (0°, Anteroposterior Plane, Half-Points Allowed)	
1	Fracture with fracture line and no callus formation
2	Fracture with callus formation and a fracture line
3	Fracture with bridging callus, but the fracture line is still visible across both corticies
4	Complete bridging of the callus with no evidence of fracture line
• Cortical Callus Formation (0°, anteroposterior plane)	
0	no callus noted
1	callus not reaching into the defect
2	callus bridging the defect < 50%
3	callus bridging the defect >50% but <100 %
4	callus bridging the defect completely
• Cis-Cortex Callus Formation (0°, Anteroposterior Plane)	
0	no callus noted
1	callus not reaching into the defect
2	callus reaching into the defect, fracture line in callus visible
3	callus reaching into the defect, fracture line in callus less visible
4	callus reaching into the defect, fracture line in callus not visible
• Trans-Cortex Callus Formation (0°, anteroposterior plane)	
0	no callus noted
1	callus not reaching into the defect
2	callus reaching into the defect, fracture line in callus visible
3	callus reaching into the defect, fracture line in callus less visible
4	callus reaching into the defect, fracture line in callus not visible
• Cranial Cortical Gap (275° angled plane)	
0	no callus noted
1	callus not reaching into the defect
2	callus reaching into the defect, fracture line in gap visible
3	callus reaching into the defect, fracture line in gap less visible
4	callus reaching into the defect, fracture line in gap not visible
• Caudal Cortical Gap (265° angled plane)	
0	no callus noted
1	callus not reaching into the defect
2	callus reaching into the defect, fracture line in gap visible
3	callus reaching into the defect, fracture line in gap less visible
4	callus reaching into the defect, fracture line in gap not visible
• Callus Opacity (0°, anteroposterior plane)	
0	soft tissue opacity
1	< 50% of normal
2	> 50-100% of normal

3	> 100% (superimposition)
	Total Compiled Radiographic Union Score (cRUS): 1-27

Figure 1: Illustrations of the plate constructs used in the three osteotomy models. a) Group 1: 3 mm gap defect stabilized with a 12-hole stainless steel plate (broad straight veterinary 3.5 mm locking compression plate (LCP, 159 mm in length, and 3.5 mm bicortical screws). b) Group 2: 3 mm gap defect stabilized with a six-hole titanium plate (broad 4.5/5.0 mm LCP, 115.8 mm in length, and 5 mm bicortical screws). c) Group 3: 17 mm defect augmented with autografts and stabilized with a 13-hole stainless steel plate (broad straight veterinary 3.5 LCP, 172 mm in length, and 3.5 mm bicortical screws).

Figure 2: Virtual model segmentation workflow: a-b) Coronal and axial slice views of raw CT image without masks applied. c-d) Coronal and axial slice views with density thresholds of 400-2500 HU for callus and 2500-4000 HU for bone. e-f) Coronal and axial slice views after contour-based segmentation. g) 3D surface models of operated and intact limbs. h) Contour plot of element-wise elastic modulus finite element model. i) Simulated torsion loading with displacement contours.

Figure 3: Segmented 3D surface models of all operated limbs with callus tissue shown in red and cortical bone in yellow.

Figure 4: Material assignment relationship curves used for virtual mechanical testing included three non-species-specific relations and one optimized material law developed for ovine cortical bone using data collected in this study.

Figure 5: Box plots comparing torsional rigidity in physical biomechanical testing (Bio) to predicted virtual torsional rigidity (VTR) in each group: Opt (optimized ovine cortical bone scaling law) and the non-species-specific laws (Snyder, Morgan, Keyak). Bars show statistically

significant group differences compared to biomechanical testing for all 33 operated limbs (panel a) and 20 intact limbs (panel b).

Table 1: Biomechanical torsional rigidities for each osteotomy group, all osteotomies combined, and the intact tibiae.

Table 2: Callus morphometric data and correlations to biomechanical torsional rigidity for each group.

Table 3: Radiographic scores and correlations to biomechanical torsional rigidity for each group.

Table 4: Virtual Torsional Rigidities (VTRs) and correlations to biomechanical torsional rigidity for each group.

Table 5: Root mean squared error with biomechanical torsional rigidity for each group.

Supplementary Table 1: Scoring breakdown for radiographic assessment. Includes scores for both lateral callus granular mRUST and comprehensive RUS scores.