Virtual Mechanical Testing from Low-Dose CT Scans Predicts Tibial Fracture Time to Union and Outperforms Subjective Outcomes Scoring

Dailey, Hannah L., PhD[1]; Schwarzenberg, Peter, MS[1]; Daly, Charles J., BSc[2]; Boran, Sinéad A.M., MB BCh BAO MCh FRCSI[4]; Maher, Michael M., MD FRCSI[3]; Harty, James A., MB BCh BAO MSc FRCSI[4]

Institutions at which work was performed:

Lehigh University, Bethlehem, Pennsylvania, USA
Cork University Hospital, University College Cork, Cork, Ireland

Authors' affiliations:

[1] Department of Mechanical Engineering & Mechanics, Lehigh University, Bethlehem, Pennsylvania, USA
[2] OrthoXel, DAC, Cork, Ireland
[3] Department of Radiology, Cork University Hospital, University College Cork, Wilton, Cork, Ireland
[4] Department of Orthopaedic Surgery, Cork University Hospital, University College Cork, Wilton, Cork, Ireland

Corresponding author:
Hannah Dailey (hannah.dailey@lehigh.edu)

Virtual Mechanical Testing from Low-Dose CT Scans Predicts Tibial Fracture Time to Union and Outperforms Subjective Outcomes Scoring

Abstract (325 Words Max)

Background: Quantitative outcomes assessment remains a persistent challenge in orthopaedic trauma. Although patient-reported outcomes measures (PROMs) and radiographic assessments such as RUST scores are frequently used, very little evidence has been presented to support their validity for measuring structural bone formation or biomechanical integrity.

Methods: A sequential cohort of tibial shaft fracture patients was prospectively recruited for observation following standard reamed intramedullary nailing in a Level I trauma center. Follow-ups at 6, 12, 18, and 24 weeks included X-rays and completion of PROMs (EQ-5D and pain scores). Low-dose computed-tomography (CT) scans were also completed at 12 weeks. Scans were reconstructed in 3D and subjected to virtual mechanical testing via the finite element method to assess fracture limb torsional rigidity relative to intact bone.

Results: Patients reported progressive longitudinal improvement in mobility, self-care, activity, and health over time, but the PROMs were not correlated with structural bone healing. RUST scoring showed moderate intra-rater agreement (ICC = 0.727), but the scores at 12 weeks were not correlated with time to union ($R^2 = 0.103$, $p = 0.193$) and were only moderately correlated with callus structural integrity ($R^2 = 0.346$, $p = 0.010$). In contrast, patient-specific virtual torsional rigidity (VTR) was significantly correlated with time to union ($R^2 = 0.383$, $p = 0.005$) and clearly differentiated one case of delayed union (VTR = 10%, union at 8 months) from the rest of the normally healing cohort (VTR > 60%, median union time 19 weeks) using CT data alone.

Conclusions: PROMs provide insight into the natural history of the patient experience after tibial fracture, but have limited utility as a measure of structural bone healing. RUST scoring, although repeatable, is not a valid longitudinal predictor of time to union. In contrast, virtual mechanical testing from low-dose CT scans provides a quantitative and objective structural callus assessment that reliably predicts time to union and may enable early diagnosis of compromised healing.

Level of Evidence: Diagnostic Level II.

Introduction

Clinical research in orthopaedic trauma requires assessments that can capture the gradual progress of bone healing in the context of the patient's return to normal activity. Patient-reported outcome measures (PROMs) such as pain scores and multi-dimensional health status instruments are becoming increasingly central to robust trial design.[1–3] Unfortunately, interpreting PROMs data can be challenging in trauma research. For example, in a trial of angular-stable nailing of distal tibia fractures, pain scores were highly variable and showed a strong trend toward decreasing pain over time without a detectable difference in pain between groups based on fixator type.[4] Similarly, pilot data for the Warwick Hip Trauma Evaluation (WHiTE) trial indicated that nearly 1,000 patients would be needed detect a clinically significant difference in EuroQol 5-Dimension (EQ-5D) scores between implant groups.[5] PROMs data can also lead to non-inferiority conclusions, such as in a recent trial of immediate versus delayed weight-bearing after tibial nailing.[6]

In contrast to PROMs, radiographic assessments offer the promise of objectivity, but historically have been limited by concerns with reliability.[7] In response to this need, the Radiographic Union Scale for Tibial Fractures (RUST) score was developed as a structured semi-quantitative method for assessing callus. RUST scoring has demonstrable reliability,[8,9] has been adapted for use in distal femur fractures,[10] has been used to diagnose nonunion fractures (RUST > 10 for union),[11–13] and has been adopted in the design of large randomized controlled trials.[14] Despite this wide use, no data currently exists to assess whether RUST scores are a reliable measure of structural bone healing in clinical, not preclinical, application.

Accordingly, the objective of this study was to assess bone healing in a sequential cohort of tibial fracture patients using a comprehensive suite of radiographic and patient-reported outcomes measures and to critically evaluate these instruments with reference to a new objective biomechanical gold standard: the torsional rigidity of the fractured limb relative to the intact bone derived using patient-specific image-based finite element models from low-dose CT scans.

Methods

Study Design

Adult (age 18+ years) tibial shaft fracture patients were prospectively recruited from April 1, 2016 to December 31, 2016 in Cork University Hospital, a Level I trauma centre serving southwest Ireland. All fractures were treated by reamed intramedullary nailing. Patients were followed-up at 6, 12, 18, and 24 weeks or until clinical union, which was defined as radiographic union with pain free ambulation. All cases were also reviewed at least one year after surgery.

Outcome measures included EQ-5D and NRS pain scores, RUST scores, and quantitative CT-derived morphometric and structural measures of callus. To minimize bias, RUST scoring was completed by the senior clinician using blinded radiographs that were randomly shuffled such that films from each follow-up were not presented in chronological order or grouped by patient. RUST scores were repeated on the blinded shuffled films by the same rater after a period of four months to assess intra-rater agreement. None of the CT-derived measures were made available for comparison to the RUST scores or clinical findings until after all evaluations were completed.

The study design was reviewed and approved by the local Institutional Review Board and all patients provided written informed consent.

CT Scan Protocol

Patients underwent low-dose CT scanning at 12 weeks after surgery. CT scans were performed on a GE (General Electric Healthcare; Waukesha, Wisconsin, USA) Discovery CT750 HD with X-ray tube voltage of 80kV, current-time product of 10mAs, gantry rotation speed of 0.4

seconds, and a pitch of 0.51. A pure iterative reconstruction algorithm, Model-Based Iterative Reconstruction (MBIR), was used for image reconstruction with a resolution improvement filter kernel, RP05.[15,16] The resulting scan resolution had a slice thickness of 0.625 mm and median pixel spacing of 0.373 mm.

CT Scan Processing and Injured Limb Reconstruction

Image sets were processed using the Mimics Innovation Suite v20 (Materialise; Leuven, Belgium). A segmentation workflow was developed to create 3D surface models by applying density-based threshold rules to identify different tissues. Callus included tissue within threshold values of 400-1400 Hounsfield units (HU) and cortical bone included tissue within 1400-2700 HU. Automated surface optimization sequences were applied to refine the segmented models and then the surfaces were enclosed and volumetrically discretized to create a unified tetrahedral mesh for finite element (FE) analysis. The mesh was refined and elementwise mechanical properties were interpolated from voxel radiodensities using a previously published elastic modulus scaling law based on local Hounsfield units of the original CT scan: $E = 0.00704 \times HU$ [GPa].[17] Each scan was also digitally reconstructed to recreate an intact bone model representative of the pre-injury anatomy. The complete workflow for generating the injured and intact models is summarized in Figure 1 and discussed in detail elsewhere.[18]

A virtual torsion test was chosen as the summary measure of mechanical integrity in each model. Torsional stiffness relative to intact paired controls is often used as a summary indicator of healing in preclinical models,[19–23] because it is direction independent.[24] Torsion testing yields virtual torsional rigidity: $VTR = ML/\phi$ [N-m^2/°], where M is the twisting moment, L is the length of the test segment, and ϕ is the angle of twist. All structural simulations were carried out in ANSYS 17.2 (ANSYS; Canonsburg, PA, USA).

Statistical Analysis

Descriptive statistics were generated using Microsoft Excel (2016) and MATLAB (R2016a, The MathWorks; Natick, Massachusetts, USA). All other statistical analyses were carried out in IBM SPSS Statistics 24 (IBM Corp; Armonk, NY, USA). Repeated measures testing for differences over time was carried out using Friedman non-parametric tests as is appropriate for ordinal and non-normally distributed interval data. One-way repeated measures ANOVA was also used for testing differences in normally-distributed interval variables. Relationships between outcomes measures were assessed using Pearson's correlations. Normality was assessed by Shapiro-Wilk and Kolmogorov-Smirnov testing.

Sources of Funding

Funding to support study management was provided by a grant from OrthoXel, DAC.

Results

Patient and Injury Characteristics

Twenty-four patients met the inclusion criteria and 19 completed the follow-up protocol including low-dose CT scan (see Figure 2). Demographic and injury characteristics for these patients are shown in Table 1. The median age was 39 years. Two patients (11%) were female and three (16%) had open injuries. The predominant fracture morphology was OTA/AO type A (15 cases, 79%). Injury mechanisms included: sports (5), simple fall (4), blow (3), fall from height (2), crush (2), pedestrian struck by vehicle (2), and other (1). The median time to clinical union was 18.7 weeks. There was one delayed union, indicated by progressive autodynamization and eventual union at 36 weeks. There were no nonunions or reoperations to induce union.

Patient Reported Outcome Measures (PROMs)

Complete descriptive statistics for all EQ-5D component scores and the NRS pain scores are shown in Table 2 and selected statistically or clinically meaningful temporal trends are illustrated in Figure 3. Patients tended to report significantly improved mobility, capacity for self-care, engagement in usual activities, and general health over time. Post-hoc testing showed that this trend was generally only statistically significant between the first and the last time points. One exception was the usual activities score, where patients indicated the greatest difficulty at the six-week follow-up, which was significantly different from the scores reported at all later time points. Pain scores were very low for most patients and were steady over time, both in the EQ-5D pain component score and on the NRS pain scale.

RUST Scores

RUST scores at the four follow-ups are also shown in Figure 3, together with time to union. The intra-class correlation coefficient, ICC(3,1) for test-retest reliability of RUST scoring with consistency effects (two-way random single measures procedure in SPSS) was 0.727 (95% CI 0.597 – 0.820). RUST scores significantly increased over time ($p = 0.001$), and after six weeks, became non-normally distributed with a notable ceiling effect, with at least 75% of patients achieving RUST $\geq$ 10 from 18 weeks onward.

CT-Derived Morphometry and Virtual Mechanical Testing

Complete morphometric data for each patient-specific model is provided in Table 3, with selected values illustrated in Figure 4. For each patient-specific model, the fractured limb VTR was normalized by that patient's own reconstructed intact tibia to produce a dimensionless indicator of healing relative to the pre-injury state: normalized VTR. Across all patients, median normalized VTR was 99% (IQR 86% – 113%) at 12 weeks.

Correlations between Outcomes Measures

Relationships between PROMs, RUST scores, and CT-derived properties at 12 weeks were assessed by Pearson's correlations. RUST scores were not significantly correlated with CT-derived VTR [N-m^2/°] of the fractured or intact tibia (all $p > 0.217$), but RUST scores had a moderate and significant positive correlation with normalized (fractured/intact) VTR ($p = 0.010$, see Figure 5). Callus volume had no observable correlation with normalized VTR ($R^2 = 0.012$, $p = 0.651$). Callus density showed a weak and non-significant correlation with normalized VTR ($R^2 = 0.167$, $p = 0.082$). None of the EQ-5D component scores or pain measures were correlated with normalized VTR (all $R^2 < 0.051$, all non-significant). RUST scores at 12 weeks were not significantly correlated with time to union, but normalized VTR from 12-week CT was moderately and significantly correlated with the final clinical outcome (see Figure 5).

Discussion

When evaluating a new clinical assessment method, we must consider whether the technique produces data that is objective (free from bias), precise (repeatable and reproducible), systematically accurate (in agreement with true values, when known), and clinically useful.[25] For this discussion, we will focus on the virtual structural analysis technique and its resulting summary parameter, normalized VTR, which indicates the extent of structural healing proportional to the individual's own intact bone.

In terms of objectivity, the workflow for processing CT scans and carrying out structural mechanics simulations follows a pre-defined sequence of actions applied uniformly for all models, with extremely limited opportunity for introduction of researcher bias. The process has been documented in extensive detail elsewhere.[18] The strict workflow control produces data that is highly repeatable and any two individuals following the proscribed process would produce nearly identical summary measures.

The question of systematic accuracy with patient-specific virtual structural analysis is harder to resolve because the torsional rigidity of partially healed human fractures fixed by IM nailing cannot be directly measured. The closest set of available reference data arises from published values for the torsional rigidity of intact cadaver tibiae, which have an expected mean torsional rigidity of 2.42 ± 0.80 N-m^2/° in physical mechanical testing.[26] For comparison, the mean intact virtual torsional rigidity of the tibiae in our case series was 2.32 ± 0.68 N-m^2/°, indicating that the modeling technique produces global structural parameters that are clearly within the physiological range and very near the accepted reference value based on the limited data available. Systematic model bias may be present, for example in the choice of a mathematical scaling law for defining mechanical properties from radiodensity, but this source of error would be consistent across all models and is likely to be small in light of the close agreement between our results and published cadaver test data.

Finally, we considered whether virtual structural analysis is clinically useful via an *a priori* power analysis based on all data collected at 12 weeks. For each outcome measure, we calculated the required sample size to detect clinically meaningful 20% differences between group means, assuming 80% power and a significance level of 0.05. Based on this data, significant differences between means would be detectable for the following measures with the group sample sizes shown: NRS pain score (N = 273), EQ-5D Mobility Component score (N = 55), objective callus volume from CT (N = 151), RUST score (N = 30), and normalized VTR (N = 30). Notably, unlike in preclinical research, callus volume is not a useful measure of healing progress because the variety of clinical fracture morphologies leads to large variability in the amount of callus. Both

RUST score and normalized VTR had similar expected statistical power, but the utility of RUST scoring should be considered carefully due to ceiling effects and known challenges with interpretation and repeatability.[9] Furthermore, in this analysis, virtual torsional rigidity at 12 weeks successfully predicted time to union, whereas RUST score did not.

Aside from questions of statistical power, this study also illustrates some concerning limitations of some assessment instruments that are commonly used to assess fracture healing and outcomes. First, our data indicated that PROMs such as EQ-5D may be likely to show a strong regression to the mean effect, in which patients report feeling better over time, independent of their structural healing progress. Published normative population data from the EuroQol Group for individuals in the UK shows that the median self-reported health score across all age groups is 90,[27] which is precisely equal to the median health score reported by our participants at 24 weeks after surgery. The fact that none of the of the PROMs were significantly correlated with our objective measure of healing—normalized virtual torsional rigidity—suggests that PROMs provide insight into the psychological response to injury, but are not be a reliable indicator of the physiological progress of fracture union for nailed tibial fractures.

This study also suggests some potential difficulties with the widely used RUST score. Our RUST scores intra-class correlation coefficient (ICC = 0.727) was comparable to the RUST repeatability recently reported by many other investigators,[9,10,22,23,28,29] but was not as good as the original reports.[8] Notably, none of these reports show that RUST scoring meets the strictest definition of ICC>0.9 for an assessment to be deemed “reasonable for clinical measurements”,[30] even in studies using simple transverse osteotomies in animal models.[22,23] Furthermore, our data indicates that RUST scores may not have prognostic validity for predicting longitudinal outcomes, an issue that was raised early in the history of the RUST scoring,[31] but has been neglected in favor of repeated examination of inter- and intra-rater reliability. Interpreted cautiously, our data supports the general conclusion that RUST is useful for differentiating between healed and unhealed fractures, but may be lacking in sensitivity and specificity as a surrogate measure of callus structural integrity, particularly in early healing (non-united RUST < 10).

This study has a few noteworthy limitations. First, we have not included any objective activity monitoring, which if available, would provide a valuable comparison to the EQ-5D Mobility and Usual Activities component scores. We also did not collect data on pain or radiographic progress at very early time points (< 6 weeks). By 6 weeks, patients were reporting low levels of pain with limited need for analgesia, so it is possible that more meaningful early pain data may have been missed. The balance between pain and self-limited activity may also dictate the level of mechanical stimulation that drives callus formation, but this was not quantified in our investigation and may be important for future study. Finally, although the virtual mechanical test method we have employed shows promise, the sample size was small and stronger inferences about diagnostic or prognostic validity, particularly for early nonunion detection, will require further study.

In conclusion, we have shown that normalized virtual torsional rigidity (VTR) from CT-derived patient-specific structural models is predictive of final time to union, offers greater statistical power than PROMs, and avoids difficulties associated with RUST scoring such as repeatability and ceiling effects. Patient-specific modeling does incur increased investigational burden compared to using only patient-reported or standard radiographic measures, but in exchange, it offers precision and accuracy that cannot be achieved by those routine assessments. In the future, virtual mechanical testing may be appropriate or even preferable as a main outcome measure in clinical trials designed to compare healing between groups when sensitivity and objectivity are of paramount importance, such as research focused on implant type or design, surgical approach, injury characteristics, patient risk factors, or behavioral interventions.

Acknowledgements

The authors gratefully acknowledge supporting contributions from Karl James MB, FFR (RCSI), Richard Kavanagh MB, FFR (RCSI), and Owen J. O'Connor MB, FFR (RCSI).

References

1. Gagnier JJ. Patient reported outcomes in orthopaedics. *J Orthop Res*. 2017;35(10):2098-2108.
2. Christensen DL, Dickens JF, Freedman B, et al. Patient-Reported Outcomes in Orthopaedics. *J Bone Joint Surg*. 2018;100(5):436-442.
3. Hamilton DF, Giesinger JM, Giesinger K. It is merely subjective opinion that patient-reported outcome measures are not objective tools. *Bone Jt Res*. 2017;6(12):665-666.
4. Höntzsch D, Schaser K-D, Hofmann G, et al. Evaluation of the effectiveness of the angular stable locking system in patients with distal tibial fractures treated with intramedullary nailing: a multicenter randomized controlled trial. *J Bone Joint Surg Am*. 2014;96(22):1889-1897.
5. Griffin XL, Parsons N, McArthur J, Achten J, Costa ML. The Warwick Hip Trauma Evaluation One: a randomised pilot trial comparing the X-Bolt Dynamic Hip Plating System with sliding hip screw fixation in complex extracapsular hip fractures: WHiTE (One). *Bone Joint J*. 2016;98-B(5):686-689.
6. Gross SC, Galos DK, Taormina DP, Crespo A, Egol KA, Tejwani NC. Can Tibial Shaft Fractures Bear Weight After Intramedullary Nailing? A Randomized Controlled Trial. *J Orthop Trauma*. 2016;30(7):370-375.
7. Bhandari M, Guyatt GH, Swiontkowski MF, Tornetta PI, Sprague S, Schemitsch EH. A Lack of Consensus in the Assessment of Fracture Healing Among Orthopaedic Surgeons. *J Orthop Trauma*. 2002;16(8):562-566.
8. Whelan DB, Bhandari M, Stephen D, et al. Development of the radiographic union score for tibial fractures for the assessment of tibial fracture healing after intramedullary fixation. *J Trauma*. 2010;68(3):629-632.

9. Leow JM, Clement ND, Tawonsawatruk T, Simpson CJ, Simpson AHRW. The radiographic union scale in tibial (RUST) fractures: Reliability of the outcome measure at an independent centre. *Bone Jt Res*. 2016;5(4):116-121.

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

11. Bishop JA, Palanca AA, Bellino MJ, Lowenberg DW. Assessment of compromised fracture healing. *J Am Acad Orthop Surg*. 2012;20(5):273-282.

12. O'Halloran K, Coale M, Costales T, et al. Will My Tibial Fracture Heal? Predicting Nonunion at the Time of Definitive Fixation Based on Commonly Available Variables. *Clin Orthop Relat Res*. 2016:1-11.

13. Van Houten AH, Heesterbeek PJC, Van Heerwaarden RJ, Van Tienen TG, Wymenga AB. Medial open wedge high tibial osteotomy: Can delayed or nonunion be predicted? *Clin Orthop Relat Res*. 2014;472(4):1217-1223.

14. Castillo R, Raja S, METRC Consortium. Improving Pain Management and Longer-Term Outcomes Following High-Energy Orthopaedic Truama. https://metrc.org/54-infopages/crsd/292-painstudydescription. Accessed June 21, 2017.

15. Li G, Liu X, Dodge CT, Jensen CT, Rong XJ. A noise power spectrum study of a new model-based iterative reconstruction system : Veo 3.0. *J Appl Clin Med Phys*. 2016;17(5):428-439.

16. Li K, Tang J, Chen G-H. Statistical model based iterative reconstruction (MBIR) in clinical CT systems: Experimental assessment of noise performance. *Med Phys*. 2014;41(4):041906.

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

18. Schwarzenberg P. (Thesis) Virtual Structural Analysis of Bone Fracture Healing from Low-Dose Clinical CT Scans. Lehigh University 2018;(May).

19. Epari DR, Kassi J-PP, Schell H, et al. Timely Fracture-Healing Requres Optimization of Axial Fixation Stability. *J Bone Joint Surg*. 2007;89(7):1575-1585.

20. Kaspar K, Schell H, Seebeck P, et al. Angle Stable Locking Reduces Interfragmentary Movements and Promotes Healing After Unreamed Nailing. Study of a Displaced Osteotomy Model in Sheep Tibiae. *J Bone Joint Surg Am*. 2005;87(9):2028-2037.

21. Bottlang M, Tsai S, Bliven EK, et al. Dynamic Stabilization with Active Lockign Plates Delivers Faster, Stronger, and More Symmetric Fracture-Healing. *J Bone Joint Surg*. 2016;98:466-474.

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

23. Litrenta J, Tornetta P, Ricci W, et al. 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*. 2017;31(3):127-130.

24. Augat P, Penzkofer R, Nolte A, et al. Interfragmentary movement in diaphyseal tibia fractures fixed with locked intramedullary nails. *J Orthop Trauma*. 2008;22(1):30-36.

25. Langlois J, Hamadouche M. Current recommendations for assessing the reliability of a measurement tool: A survival guide for orthopaedic surgeons. *Bone Jt J*. 2016;98B(2):166-172.

26. Heiner AD. Structural properties of fourth-generation composite femurs and tibias. *J Biomech*. 2008;41(15):3282-3284.

27. Szende A, Janssen B, Cabasés J, eds. *Self-Reported Population Health: An International Perspective Based on EQ-5D*. Vol 16. First. Springer, Dordrecht; 2014.

28. Azevedo Filho FAS de, Cotias RB, Azi ML, Teixeira AA de A. Reliability of the radiographic union scale in tibial fractures (RUST). *Rev Bras Ortop (English Ed*. 2017;52(1):35-39.

29. Perlepe V, Cerato A, Putineanu D, et al. Value of a radiographic score for the assessment of healing of nailed femoral and tibial shaft fractures: A retrospective preliminary study. *Eur J Radiol*. 2018;98(October 2017):36-40.

30. Trevethan R. Intraclass correlation coefficients: clearing the air, extending some cautions, and making some requests. *Heal Serv Outcomes Res Methodol*. 2017;17(2):127-143.

31. Kooistra BW, Dijkman BG, Busse JW, Sprague S, Schemitsch EH, Bhandari M. The radiographic union scale in tibial fractures: Reliability and validity. *J Orthop Trauma*. 2010;24(SUPPL. 1).

Figure Legends

Fig. 1 – Workflow for virtual mechanical testing: A) sagittal slice view of low-dose CT scan without masks applied, B) density-based threshold masks applied for cortical bone (yellow, 1400-2700 HU) and callus (cyan, 400-1400 HU), C) 3D reconstruction after surface optimization, D) 3D tetrahedral volumetric model discretization with elementwise mechanical properties derived from voxel radiodensity, E) virtual torsion test with distal end clamped and proximal end rotated about the anatomic axis of the tibia, and F) reconstructed intact tibia for comparison to fractured bone rigidity.

Fig. 2 – Reconstructed models of all 19 cases. Colors indicate radiodensity, with lower density (blues) in callus zone and higher density (yellows and reds) in the cortex.

Fig. 3 – Selected clinical findings at four follow-up timepoints: 6, 12, 18, and 24 weeks. The EQ-5D Mobility and Usual Activities components were recorded on an interval scale (1-5) with 1 representing "no difficulty". The EQ-5D Health VAS score was recorded on an interval scale (0-100) with 100 being "best imaginable health state". The NRS Pain Score was also recorded no an interval scale (0-10) with 10 being "worst pain imaginable". RUST scores and time to union were assessed by the senior clinician.

Fig. 4 – From left to right, distributions of CT-derived patient-specific morphometric and structural parameters including: cortical bone and callus density, callus volume, torsional rigidity of fractured and virtually reconstructed intact limbs, and normalized torsional rigidity of the fractured tibia relative to the reconstructed intact bone. At least 75% of patients (Q1 and above) achieved at least 85% of the torsional rigidity of their intact bone by 12 weeks after surgery.

Fig. 5 – Left to right: RUST score was moderately and significantly correlated with CT-derived normalized torsional rigidity at 12 weeks. RUST score was not correlated with time to clinical union. CT-derived normalized torsional rigidity was moderately and significantly correlated with time to clinical union.

Table Captions

Table 1 – Patient and injury characteristics.

Table 2 – Detailed temporal data for all PROMs

Table 3 – Complete morphometric and structural data for all patients.

Table 1 – Patient and injury characteristics.

CT Case Number	Gender	Age [Years]	Injury Type	OTA/AO Classification	Gustilo-Anderson Classification	Tscherne Classification
CT01	Female	51	Closed	42A2		0
CT02	Male	52	Closed	42A1		0
CT03	Male	32	Open	42B3	I	
CT04	Male	32	Closed	42A3		1
CT05	Male	55	Closed	42A1		0
CT06	Male	58	Closed	42A3		3
CT07	Male	33	Closed	42A2		0
CT08	Female	39	Open	42B2	I	
CT09	Male	65	Closed	42C2		1
CT10	Male	33	Closed	42A3		0
CT11	Male	20	Closed	42A3		0
CT12	Male	45	Open	42A2	II	
CT13	Male	50	Closed	42A1		0
CT14	Male	24	Closed	42A1		0
CT15	Male	39	Closed	42A2		0
CT16	Male	29	Closed	42A3		1
CT17	Male	57	Closed	42B3		0
CT18	Male	53	Closed	42A1		1
CT19	Male	18	Closed	42A3		0

Table 2 – Detailed Temporal Data for all PROMs

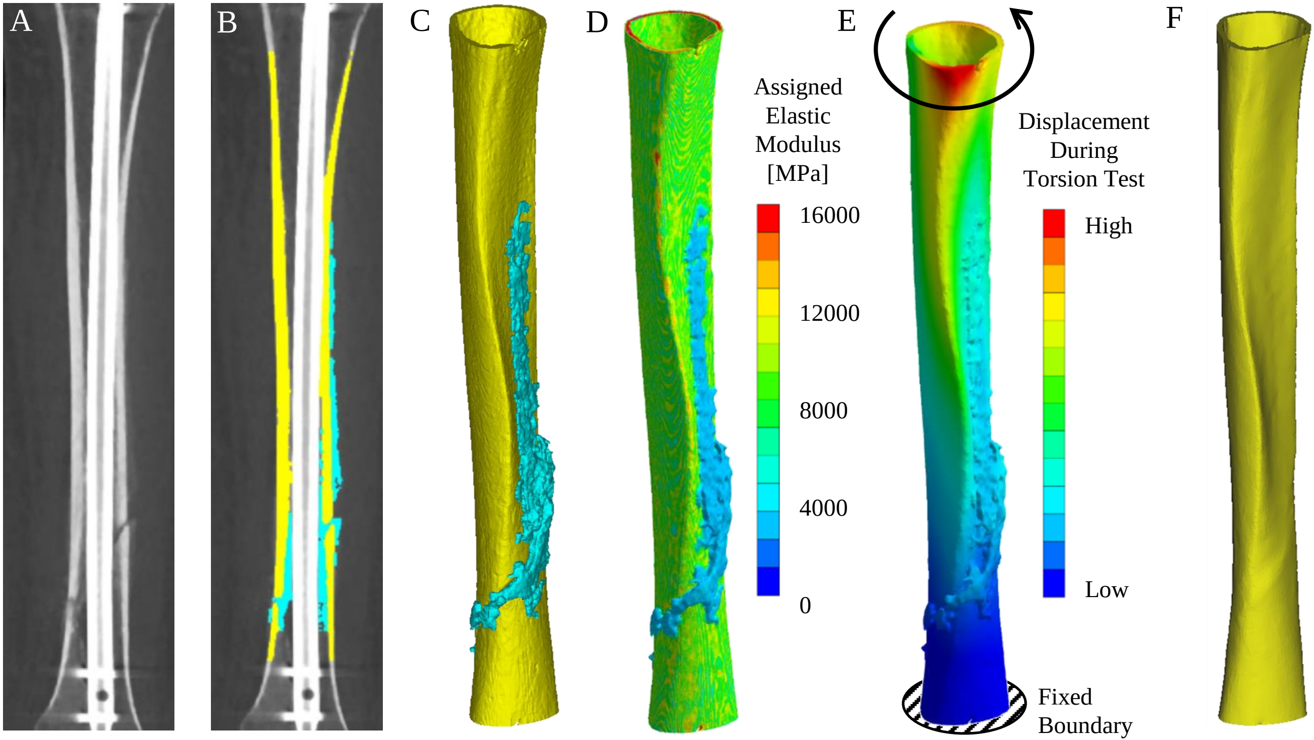
Component	Follow-Up†	Median	Interquartile Range	Mean ± Standard Deviation	Friedman Test *p*-value†
EQ-5D Mobility Component Score[a]	1*	3.0	2.0 – 4.3	3.1 ± 1.1	0.007*
	2	3.0	2.0 – 3.0	2.8 ± 1.0	
	3	2.0	2.0 – 3.0	2.5 ± 1.2	
	4*	2.0	1.0 – 2.5	2.2 ± 1.2	
EQ-5D Self-Care Component Score[a]	1*	2.0	2.0 – 2.3	2.1 ± 0.8	0.002*
	2	1.0	1.0 – 2.0	1.7 ± 0.9	
	3	1.0	1.0 – 1.0	1.3 ± 0.8	
	4*	1.0	1.0 – 1.0	1.3 ± 0.6	
EQ-5D Usual Activities Component Score[a]	1*,**,***	4.0	3.0 – 5.0	3.8 ± 1.1	< 0.001*
	2*	3.0	2.0 – 4.0	2.9 ± 1.2	
	3**	2.3	2.0 – 3.0	2.7 ± 1.2	
	4***	2.0	2.0 – 3.0	2.3 ± 0.9	
EQ-5D Pain/Discomfort Component Score[a]	1	2.0	2.0 – 3.0	2.4 ± 0.7	0.666
	2	3.0	2.0 – 3.0	2.4 ± 0.7	
	3	2.0	2.0 – 3.0	2.2 ± 0.7	
	4	2.0	2.0 – 3.0	2.3 ± 0.7	
EQ-5D Anxiety/Depression Component Score[a]	1	1.0	1.0 – 2.0	1.5 ± 0.8	0.252
	2	1.0	1.0 – 1.0	1.5 ± 1.1	
	3	1.0	1.0 – 1.8	1.4 ± 0.7	
	4	1.0	1.0 – 1.3	1.3 ± 0.7	
EQ-5D Health Component Score[b]	1*	80	65 – 80	74 ± 12	0.040*
	2	80	75 – 90	79 ± 14	
	3	88	80 – 90	81 ± 15	
	4*	90	89 – 92	86 ± 15	
NRS Pain Score[c]	1	2.0	1.0 – 3.5	2.4 ± 2.1	0.324
	2	2.0	1.0 – 4.0	2.7 ± 2.3	
	3	2.0	0.3 – 3.8	2.5 ± 2.4	
	4	2.5	0.8 – 4.0	2.4 ± 1.8	

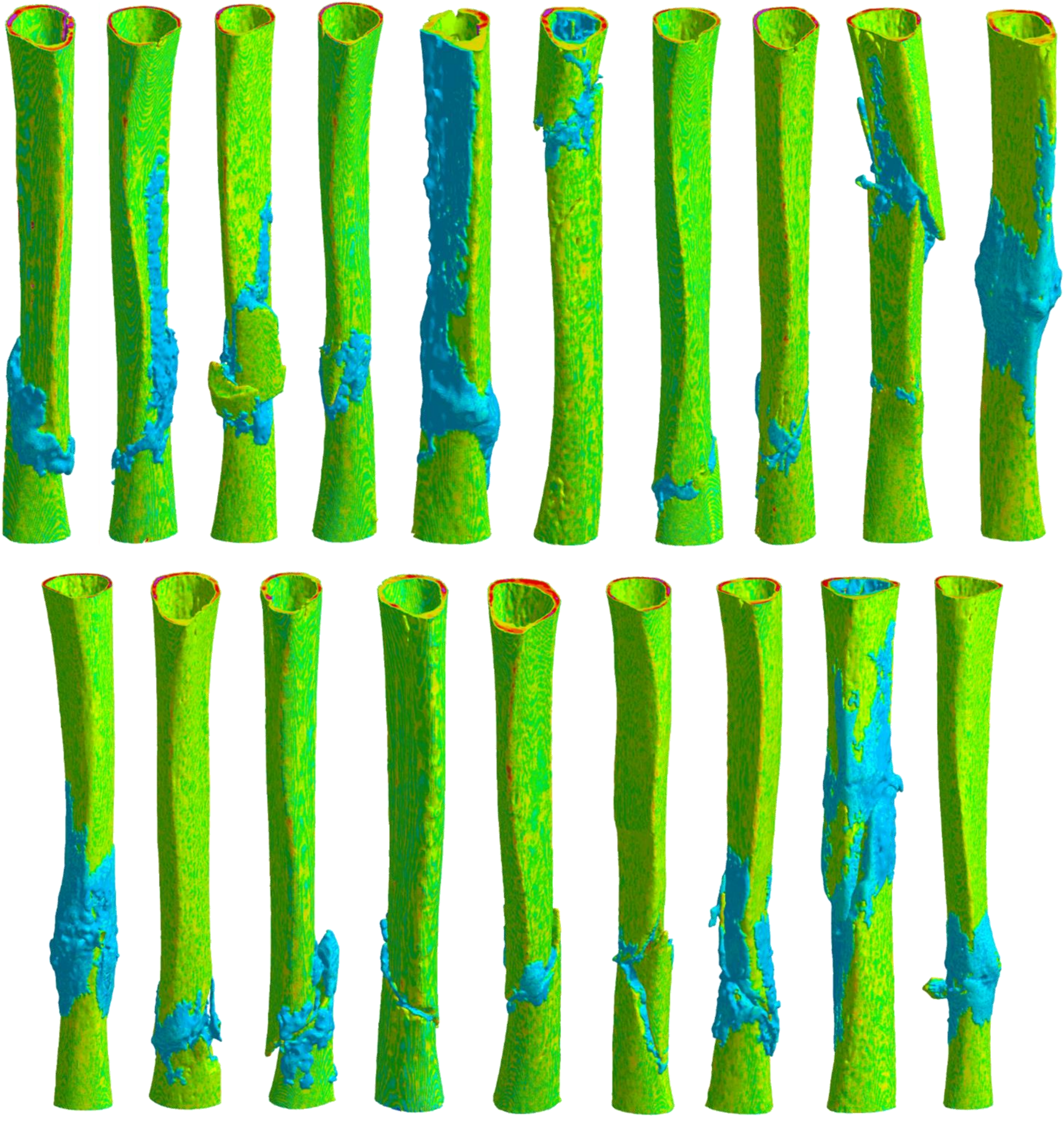
[a] 1 = No Problems, 2 = Slight Problems, 3 = Moderate Problems, 4 = Severe Problems, 5 = Unable
[b] 0 = worst health you can imagine, 100 = best health you can imagine
[c] Current pain level rating: 0 = No Pain, 1-3 = Mild Pain, 4-6 = Moderate Pain, 7-10 = Severe Pain
† Significant differences between time points (LSD post-hoc) at $p < 0.05$ level are indicated if Friedman test was significant overall

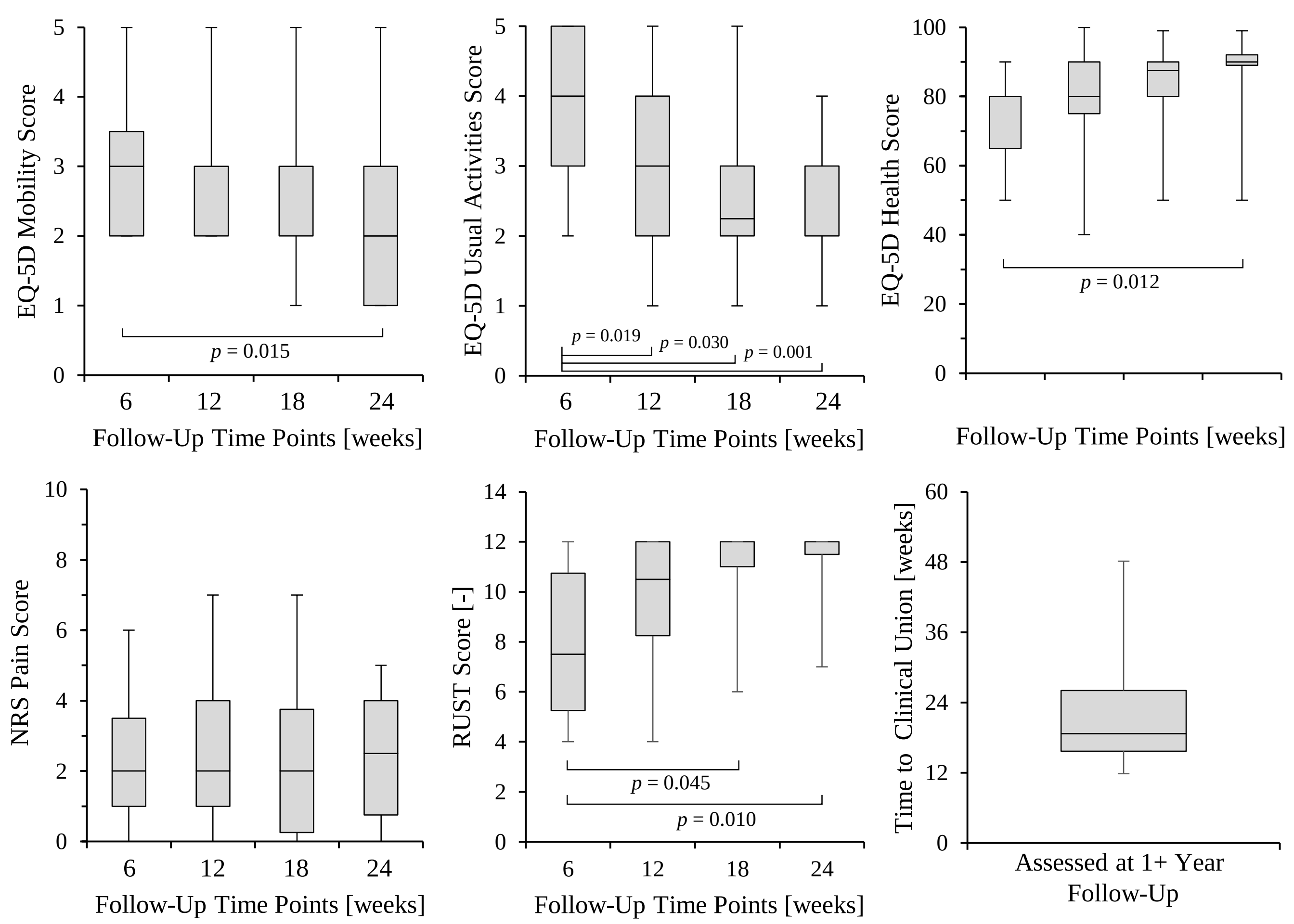
Table 3 – Complete morphometric and structural data for all patients

CT Case	Cortical Segment Length [mm]	Median Cortical Bone Density [HU]	Median Callus Density [HU]	Callus Volume [cm^3]	Virtual Torsional Rigidity (VTR) - Fractured [N-m^2/°]	Virtual Torsional Rigidity (VTR) - Intact [N-m^2/°]	Normalized VTR (Fractured/ Intact) [-]	12-Week RUST Score from X-Ray [-]
CT01	202	2121	806	7.4	1.48	1.28	1.16	12
CT02	269	2059	694	14.2	3.66	3.21	1.14	8
CT03	263	2034	743	9.0	1.63	2.68	0.61	8
CT04	275	2062	700	5.3	2.41	2.43	0.99	11
CT05	189	1737	839	15.3	1.94	1.48	1.31	12
CT06	230	1926	738	8.2	1.37	1.68	0.81	11
CT07	291	1940	782	9.1	4.06	3.70	1.10	12
CT08	265	2053	772	8.1	1.73	1.73	1.00	5
CT09	261	1984	767	32.2	2.82	3.31	0.85	11
CT10	201	1916	925	14.8	1.99	1.73	1.15	12
CT11	251	2017	784	23.6	2.61	2.09	1.25	12
CT12	256	1954	706	11.8	2.71	2.86	0.95	9
CT13	265	1961	681	8.8	0.19	1.85	0.10	4
CT14	231	1868	781	2.3	2.09	1.88	1.11	9
CT15	264	2069	793	4.0	2.34	2.59	0.90	5
CT16	268	2103	763	8.7	2.60	3.20	0.81	10
CT17	259	1929	692	13.0	1.68	1.91	0.88	12
CT18	213	1732	948	29.1	1.61	1.84	0.88	--*
CT19	276	1922	833	13.7	2.83	2.56	1.10	10
All (Median)	**261**	**1961**	**772**	**9.1**	**2.09**	**2.09**	**0.99**	**10.5**

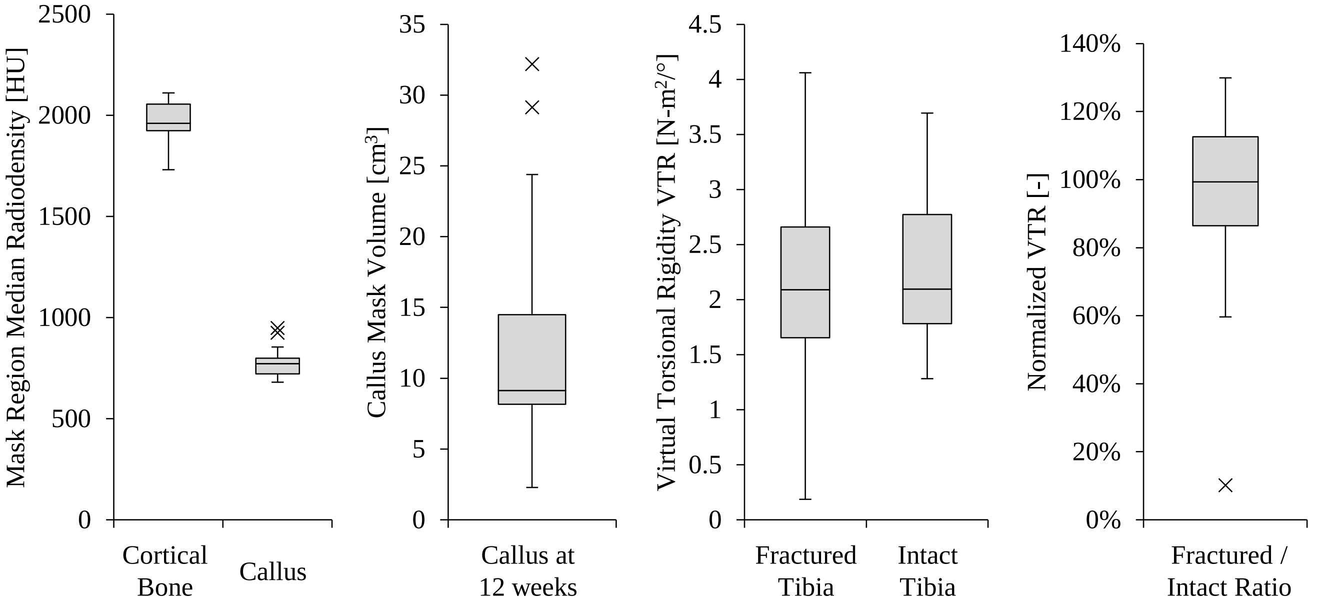
* Patient completed CT scan and PROMs at 12 weeks, but had no 12-week X-rays; for reference, this individual's RUST score at 18 weeks was 12.


A
B
C
D
Assigned Elastic Modulus [MPa]
16000
12000
8000
4000
0
E
Displacement During Torsion Test
High
Low
Fixed Boundary
F

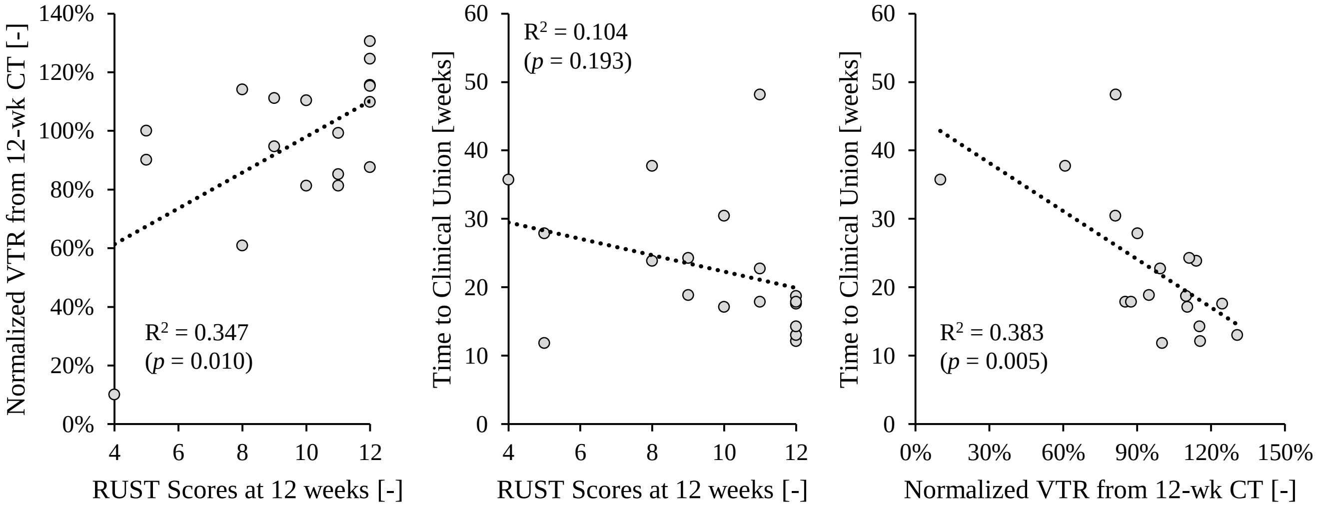




EQ-5D Mobility Score
p = 0.015
Follow-Up Time Points [weeks]
EQ-5D Usual Activities Score
p = 0.019
p = 0.030
p = 0.001
Follow-Up Time Points [weeks]
EQ-5D Health Score
p = 0.012
Follow-Up Time Points [weeks]
NRS Pain Score
Follow-Up Time Points [weeks]
RUST Score [-]
p = 0.045
p = 0.010
Follow-Up Time Points [weeks]
Time to Clinical Union [weeks]
Assessed at 1+ Year Follow-Up



Mask Region Median Radiodensity [HU]
2500
2000
1500
1000
500
0
Cortical Bone
Callus
Callus Mask Volume [cm³]
35
30
25
20
15
10
5
0
Callus at 12 weeks
Virtual Torsional Rigidity VTR [N-m²/°]
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Fractured Tibia
Intact Tibia
Normalized VTR [-]
140%
120%
100%
80%
60%
40%
20%
0%
Fractured / Intact Ratio



Normalized VTR from 12-wk CT [-]
140%
120%
100%
80%
60%
40%
20%
0%
R² = 0.347
(p = 0.010)
4
6
8
10
12
RUST Scores at 12 weeks [-]
Time to Clinical Union [weeks]
60
50
40
30
20
10
0
R² = 0.104
(p = 0.193)
4
6
8
10
12
RUST Scores at 12 weeks [-]
Time to Clinical Union [weeks]
60
50
40
30
20
10
0
R² = 0.383
(p = 0.005)
0%
30%
60%
90%
120%
150%
Normalized VTR from 12-wk CT [-]