Published in final edited form as:

J Orthop Trauma. 2019 June ; 33(6): 301–307. doi:10.1097/BOT.0000000000001445.

Inter-Rater Reliability of The Modified Radiographic Union Score for Diaphyseal Tibial (mRUST) Fractures with Bone Defects

Stuart L. Mitchell, MD[1,2], **William T. Obremskey, MD, MPH, MMHC**[3], **Jason Luly, MS**[2], **Michael J. Bosse, MD**[4], **Katherine P. Frey, PhD, MPH, RN**[2], **Joseph R. Hsu, MD**[4], **Ellen J. MacKenzie, PhD**[2], **Saam Morshed, MD, PhD**[5], **Robert V. O'Toole, MD**[6], **Daniel O. Scharfstein, ScD**[7], **Paul Tornetta III, MD**[8], and **Major Extremity Trauma Rehabilitation Consortium (METRC)**

[1] –Department of Orthopaedic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD

[2] –Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

[3] –Department of Orthopaedic Surgery, Vanderbilt University, Nashville, TN

[4] –Department of Orthopaedic Surgery, Carolinas Medical Center, Charlotte, NC

[5] –Department of Orthopaedic Surgery, University of California San Francisco, San Francisco, CA

[6] –Department of Orthopaedic Surgery, University of Maryland School of Medicine, Baltimore, MD

[7] –Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

[8] –Department of Orthopaedic Surgery, Boston University, Boston, MA

Abstract

Corresponding Author: Paul Tornetta III, MD, Boston Medical Center, 850 Harrison Ave, Dowling 2 north, Boston, MA 02118, ptornetta@gmail.com.

Conflicts of Interest: Paul Tornetta III, MD receives Smith-Nephew and Wolters-Kluwers royalties. None of the other authors have conflicts of interest to disclose.

Presented as a poster at the Annual Meeting of the Orthopaedic Trauma Association, National Harbor, MD, October 2016 and as a poster at the Annual Meeting of MHSRS 2016.

METRC Corporate Authors:
Brooke Army Medical Center: Christina M. Hylden, MD; Daniel J. Stinner, MD; ***Carolinas Medical Center***: Erica Andrews McArthur, MD; Kaitlyn M. Hurst, MD; Katherine Sample, MD; ***Denver Health & Hospital Authority:*** Corey Henderson, MS; ***Hennepin County Medical Center***: Andrew H. Schmidt, MD; Jerald R. Westberg, BA; ***Inova Fairfax Hospital***: Robert A. Hymes, MD; Lolita Ramsey, PhD, RN, CCRC; ***MetroHealth Medical Center***: Heather A. Vallier, MD; Mary A. Breslin, BA; ***University of Miami Ryder Trauma Center***: Gregory A. Zych, DO; Gabriela M. Zych, BS, CCRC; ***St. Louis University Hospital***: Lisa K. Cannada, MD; ***Tampa General Hospital***: Hassan R. Mir, MD, MBA, FACS; Rafael Serrano, MD; Barbara Steverson, MHA; ***University of Maryland R Adams Cowley Shock Trauma Center***: Andrew G. Dubina, MD; Andrea Howe, BS; ***University of Oklahoma***: David Teague, MD; ***University of California San Francisco***: Theodore Miclau, MD; ***Vanderbilt University Medical Center***: Kristin R. Archer, PhD, DPT; Basem Attum, MD, MS; Robert H. Boyce, MD; Vamshi Gajari, MBBS; A. Alex Jahangir, MD; Andres Rodriguez-Buitrago, MD; Manish K. Sethi, MD; ***Wake Forest Baptist University Medical Center***: Eben A. Carroll, MD; Holly Pilson, MD; Robert C. Weinschenk, MD; Martha B. Holden, AAS, AA; ***Walter Reed National Military Medical Center***: Benjamin K. Potter, MD; Xochitl Ceniceros, PhD, RN; Jean-Claude G. D'Alleyrand, MD; Wade T. Gordon, MD; Whitley B. Lucio; Sandra L. Waggoner, BA ***METRC Coordinating Center at the Johns Hopkins Bloomberg School of Public Health***: Renan C. Castillo, PhD; Tara Taylor, MPH

Objectives: To evaluate inter-rater reliability of the modified Radiographic Union Score for Tibial (mRUST) fractures among patients with open, diaphyseal tibia fractures with a bone defect treated with intramedullary nails (IMNs), plates, or definitive external fixation (ex-fix).

Design: Retrospective cohort study.

Setting: 15 level one civilian trauma centers; 2 military treatment facilities.

Patients/Participants: Patients ≥18-years-old with open, diaphyseal tibia fractures with a bone defect ≥1 centimeter surgically treated between 2007 and 2012.

Intervention: Three of six orthopaedic traumatologists reviewed and applied mRUST scoring criteria to radiographs from the last clinical visit within 13 months of injury.

Main Outcome Measurements: Inter-rater reliability was assessed using Krippendorff's Alpha (KA) statistic; Intraclass correlation coefficient (ICC) is presented for comparison to previous publications.

Results: 213 patients met inclusion criteria including 115 IMNs, 24 plates, 29 ex-fixes, and 45 cases that no longer had instrumentation at evaluation. All reviewers agreed on the pattern of scoreable cortices for 90.4% of IMNs, 88.9% of those without instrumentation, 44.8% of rings, and 20.8% of plates. Thirty-one (15%) cases, primarily plates and ex-fixes, did not contribute to KA and ICC estimates because <2 raters scored all cortices. The overall KA for the 85% that could be analyzed was 0.64 (ICC 0.71). For IMNs, plates, ex-fixes, and no instrumentation, KA (ICC) was 0.65 (0.75), 0.88 (0.90), 0.47 (0.62), and 0.48 (0.57), respectively.

Conclusions: In tibia fractures with bone defects, the mRUST appears similarly reliable to prior work in patients treated with IMN but is less reliable in those with plates or ex-fixes, or after removal of instrumentation.

Level of Evidence: Diagnostic, Level I

Keywords

radiographic union; bone defect; mRUST score; inter-rater reliability; tibia fracture; fracture healing

Introduction

Limited consensus exists regarding the exact definition of a "healed fracture."[1–3] Radiographic evaluation of union is used to augment observed clinical outcomes such as presence of pain with bearing weight, or tenderness with palpation of the fracture site. In an attempt to make this assessment more objective, Whelan et al developed a scoring system based on the appearance of each of the cortices on the anteroposterior (AP) and lateral radiographic views of tibial diaphyseal fractures treated with intramedullary nails (IMN), which they referred to as the Radiographic Union Score for Tibial (RUST) fractures.[4] Two orthopaedic residents, two community orthopaedic surgeons, and three orthopaedic traumatologists rated 45 sets of radiographs, and they reported an intraclass correlation coefficient (ICC), a measure of inter-rater reliability, of 0.86. To expand utilization of this scoring system, Litrenta et al analyzed its reliability and applicability to patients with

metadiaphyseal fractures.[5] This application included an important change to the range of the score where the investigators subdivided the category “no fracture line, bridging callus” into two distinct categories: “callus present” and “bridging callus.” The resulting modified RUST (mRUST) has a maximum score of 16 compared to the maximum standard RUST score of 12. Using 27 sets of radiographs of metadiaphyseal distal femur fractures rated by 12 orthopaedic trauma surgeons, the authors found slightly higher inter-rater reliability for the mRUST than the standard RUST (ICC 0.68 vs. 0.63). This finding held for the subgroup of patients treated with nails (mRUST ICC 0.74; RUST ICC 0.67) and for the subgroup treated with plates (mRUST ICC 0.59; RUST ICC 0.53).

Several studies assessing the utility of the RUST and/or mRUST have considered them to be “valid” or “valid and reliable” in adult patients with tibia diaphyseal fractures treated with IMN.[4–10] RUST and/or mRUST criteria have also been applied in other patient populations such as in pediatric Osteogenesis Imperfecta patients,[11] pediatric congenital tibial pseudoarthrosis (Neurofibromatosis Type I) patients,[12] and adult high-tibial osteotomy patients (modification of RUST criteria).[13] However, no published studies have examined the reliability of any cortical scoring system in patients with bone defects, nor in a cohort of patients treated with definitive external fixation. The current study evaluated inter-rater reliability of the mRUST at both the composite and cortex level in a cohort of patients with operatively treated open tibia shaft fractures with associated bone loss.

Methods

There were 739 patients initially identified in a retrospective review of long bone fractures treated with definitive surgical stabilization between June 2007 and May 2012 from 15 level one civilian trauma centers and two military treatment facilities as part of the Major Extremity Trauma and Rehabilitation Consortium (METRC)[14] RETRODefect study. Included were skeletally mature patients (≥18-years-old) with a bone defect of ≥1 centimeter in length and ≥50% cortical loss. For the purpose of segmenting the results, the defect size was defined as the average bone loss measurement across the anterior, posterior, medial, and lateral cortices on the first post-fixation radiograph. Inclusion for this current study was limited to patients with open, diaphyseal tibia shaft fractures (AO/OTA 42)[15] who had final AP and lateral radiographs between 3 and 13 months after definitive fixation and who did not undergo an amputation prior to the final radiograph. Patients with more than one type of definitive fixation instrumentation (e.g., both IMN and plating) seen on the final (rated) radiograph were excluded. Radiographs for all potential cases were reviewed by a senior orthopaedic traumatologist to confirm that selection criteria were met.

Six experienced, fellowship-trained orthopaedic trauma surgeons participated in the mRUST scoring process. Each case was reviewed by exactly three surgeons, assigned at random, from this pool of six. The raters were blinded to the initial radiographs, defect size, and bone grafting status. Reviewers were provided information on injury date, fixation date, and the time elapsed between injury and radiographs. Raters were instructed to score each case using mRUST criteria. Each of the cortices (four total) on the AP and lateral radiographs were graded as: 1 = No callus; 2 = Callus present without bridging; 3 = Bridging callus, but visible fracture line; 4 = Fracture line not visible (remodeled). The mRUST score is the sum

of the four cortical scores (range, 4–16). Each cortex was scored individually. In some cases, assigned reviewers did not score a cortex because they felt they could not adequately assess the injury due to problems such as an obstructed view of the fracture site. At least two out of the three assigned raters for each case must have scored all four cortices in order for the scoring to be included in sum total mRUST analysis. In the event of multiple, discrete defects, reviewers were instructed to score the area of the largest defect.

Inter-rater reliability was assessed for the sum total mRUST score as well as for the score given to each cortex. Results were stratified by the type of instrumentation seen on the radiographs at the time of evaluation: IMN, plate, definitive external fixation, or none (indicating previous removal of all instrumentation). Reliability of the sum total mRUST score was also assessed based on (a) presence/absence of bone grafting, (b) time from grafting to scored radiograph (30–90 days and >90 days), and (c) defect size: small (<2.5 cm), medium (2.5–5.0 cm) and large (>5.0 cm).

Krippendorff's Alpha (KA) statistic, appropriate for ordinal data, was the primary method used to assess inter-rater reliability.[16,17] ICC, appropriate for continuous data, is presented to enable comparison to previous work.[18] For both KA and ICC, a value of 1 indicates complete agreement among raters. Visual displays of agreement/disagreement of raters within individuals are also presented to aid interpretation. Analyses were conducted in R 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria) and STATA (StataCorp LLC, College Station, TX, USA).

Results

Among the 739 patients identified as eligible for the RETRODefect study, 213 patients (Figure 1) including 183 males and 30 females, with an average age of 34.3 years (range, 18–68) met the selection criteria for this analysis. The average time between initial definitive fixation and the rated radiographs was 294±85 days. The average bone defect size was 3.6±2.9 cm. Initial definitive fixation was accomplished using an IMN in 118 (55%) cases, plate and screws in 24 (11%), or multiplanar (ring) external fixator in 71 (33%). There were 63 cases (30%) in which the final radiograph had either no fixation instrumentation (n=45) or different instrumentation from the initial definitive fixation (n=18). Of the 45 patients who had no instrumentation, 41 were originally treated with definitive external fixation. Bone grafting was performed in 109 (51%) subjects at an average of 188±90 days prior to the scored radiographs.

Of the 213 patients that met selection criteria, all 4 cortices could be scored by all three reviewers for 158 (74%) cases, by two of three reviewers for 24 (11%) cases, by one of three reviewers for 21 (10%), and by zero of three reviewers for 10 (5%) cases. Thus, there were 182 (85%) cases contributing to the estimation of KA and ICC scored by ≥2 raters. There were 31 (15%) cases scored by <2 raters that could not contribute to the estimate of KA and ICC (Table 1). Of these excluded cases, 14 patients had a plate (58% of plated patients), 11 had a ring (38% of rings), 5 had an IMN (4% of IMNs), and 1 patient had no instrumentation (2% of patients with no instrumentation). Thus, the majority of the non-contributing cases had either a plate or ring seen on the rated radiograph (25 out of 31,

81%), suggesting an obstructed view of the fracture callous secondary to the instrumentation. Among the 182 contributing cases, 104 of 110 IMNs (95%), 3 of 10 plates (30%), 11 of 18 rings (61%), and 40 of 44 cases with no instrumentation (91%) could be scored by all three reviewers.

Sum Total mRUST Analyses

Table 2 summarizes inter-rater reliability of the sum total mRUST scores, and Figure 2 provides a visual display of the concordance of mRUST scores given to each individual case. The overall KA was 0.64 (95% CI: 0.54–0.71). The KAs for the IMN, Plate, Ring and No Instrumentation groups were 0.65 (95% CI: 0.53–0.74), 0.88 (95% CI: 0.60–0.94), 0.47 (95% CI: 0.13–0.69) and 0.48 (95% CI: 0.28–0.63), respectively.

Cortex-Level Analyses

Among the 213 cases in the initial cohort, all three reviewers agreed on the pattern of scoreable cortices (i.e., which cortices can or cannot be scored) for 90.4% of cases with IMNs, 88.9% of cases with no instrumentation, 44.8% of cases still in rings, and 20.8% of cases with plates. A Table, Supplemental Digital Content 1 displays inter-rater reliability estimates by cortex, overall and stratified by treatment instrumentation. Overall, the inter-rater reliability was lower for the lateral and posterior cortices compared to medial and anterior cortices. When stratified by instrumentation type, inter-rater reliability was lowest for the lateral cortex in the Ring (KA 0.04) and the “None” groups (KA 0.36), for the anterior cortex in the Plate group (KA 0.37) and for the posterior cortex in the IMN group (KA 0.59).

Other Subgroup Analyses

Inter-rater reliability did not differ substantially by presence or absence of bone grafting (KA 0.66 vs. 0.61) or the time from grafting to scored radiograph (KA 0.60 for 30–90 days vs. 0.66 for grafting >90 days). Inter-rater reliability for cases with small (<2.5 cm) and large (>5.0 cm) defects was comparable (KA 0.60 vs. 0.57); agreement for medium (2.5–5.0 cm) defects was higher (KA 0.77).

Discussion

Cortical scoring systems have become common tools in reporting radiographic progression towards union in lower extremity fracture trials. Since the initial description by Whelan et al, [4] the RUST and/or mRUST has been applied to, and reliability assessed in, multiple patient populations[5–7,10–12] and in animal studies.[19–21] Among adult patients, the RUST has been judged to be most reliable in the assessment of diaphyseal tibia fractures that are treated with an IMN, with reported ICCs ranging from 0.67 to 0.87.[4,6,7,10] The inter-rater reliability of mRUST in humans was first evaluated by Litrenta et al,[5] who reported ICCs of 0.74 (95% CI: 0.68–0.81) and 0.59 (95% CI: 0.51–0.67) for metadiaphyseal distal femur fractures treated with IMNs and plates, respectively.[5] In contrast, the current study focused on a population of patients with open, diaphyseal tibia fractures with a bone defect of at least 1 centimeter with at least 50% cortical loss treated operatively with IMN, plating, or definitive external fixation, with or without bone grafting. Relative to the Litrenta study,[5] inter-rater

reliability, as measured by ICCs, was nearly identical for IMNs (ICC 0.75), but higher for plates (ICC 0.90).[5] However, the inter-rater reliability results for plates and rings must be viewed with great caution as 58% (n=14) of plate and 38% (n=11) of ring cases could not contribute to the estimation of KA and ICC because they could not be scored by ≥2 raters. Additionally, only 13% of plates and only 38% of rings had all 4 cortices scored by all three reviewers, substantially reducing the utility of this measure in tibial fractures with a bone defect and treated with plates or rings.

The current study introduces two new subgroups to analysis of the inter-rater reliability of the mRUST: patients treated with definitive multiplanar external fixation, and cases with no instrumentation present due to removal prior to the radiographic evaluation. Metallic instrumentation can block a raters' view(s) of the cortices; inter-rater reliability for cases with external fixators in place might be expected to be similar to the Litrenta study of plating (ICC 0.59), although they did not report difficulty with scoring radiographs.[5] In fact, the ICC for external fixators was 0.62, but this is based on a subset of only 62% cases that could be scored by ≥2 raters. Conversely, inter-rater reliability for cases with no instrumentation was expected to be at least as high as IMNs (ICC 0.74)[5] since there were no implants blocking the view of the cortices. However, the inter-rater reliability was lower than anticipated with an ICC of 0.57. Notably, the majority of the cases with no instrumentation (n=40/45) were initially treated with definitive external fixation. We hypothesize that the relatively low level of reliability may be due to the severity of the original injury in patients treated with definitive external fixation as these were frequently grafted and can heal with an abnormal appearance to the tibia with the potential for less discrete cortices (Figure 3). Further, the mRUST scores can be influenced by the amount of fracture callus present, so differences in agreement may also be due to the different biomechanical environments seen in patients treated with IMN vs. plates vs. ring fixators.[5]

Nearly all prior RUST and mRUST studies examined narrowly defined patient populations.[4–7,10–12] In the current study there was relatively greater heterogeneity in treatment. Specifically, inclusion was not limited to a specific fixation device at the time of initial definitive fixation surgery. Some subjects in this study underwent revision fixation surgery and either had additional, or entirely different, instrumentation seen on the final rated radiograph. In nearly one-third of cases (n=63/213) the instrumentation present on the rated radiographs was different than the initial definitive fixation (e.g., patient initially treated with plating was later converted to IMN, patient initially treated with a ring fixator had it removed, etc.). Although patient selection criteria was limited to only three types of definitive fixation, there were four different groups based on the type(s) of, or lack of, instrumentation present on the rated radiographs. Thus, the results of the present study may be more generalizable to real-world orthopaedic trauma populations.

Cortical scoring has been shown to correlate both with clinical markers of fracture healing[8] and with biomechanical healing.[20,21] Previous analyses[5,21] suggest an mRUST threshold of 13 to define radiographic union. Out of the 182 included patients, 39 (18%) had ≥1 score above and another score strictly below this threshold (IMN: 15%, None: 38%, Plate: 0%, Ring: 17%) (Figure 2). These results show marked variability in mRUST scoring around the union threshold based on one set of radiographs per patient, measured generally late in the

healing process (mean/median 294/321 days after initial definitive fixation). Given that mRUST is used for assessing progress toward union, our study would have greater generalizability for the assessment of fracture union in this population if the reviewed radiographs were more uniformly distributed along the healing continuum.

The limitations in this study include those inherent to any uncontrolled, retrospective study: heterogeneity of the patient population, variability in length of follow-up, and challenges with collection of data from multiple centers by multiple research staff. There were 31 (15%) subjects who were excluded from our estimates of inter-rater reliability analysis because their radiographs could not be scored by ≥2 raters. We were unable to account for this disagreement in reviewers' ability to score radiographs because typical consensus models do not account for raters that abstain from rating. If accounted for, this disagreement would certainly have lowered inter-rater reliability estimates. Thus, the reported KA and ICC from our study are not fully representative of the true inter-rater reliability, particularly for patients with plates or multiplanar external fixators. The high exclusion rates in these subgroups also limited interpretation and strength of conclusions because of the wide confidence intervals observed. In addition, cases that underwent bone grafting were not excluded from analysis: 17 (16% of grafted cases) of these cases were within 90 days of grafting surgery, 14 of which were included in the inter-rater reliability analysis. Although we did not find a difference in the inter-rater reliability of grafted cases between 30–90 and >90 days after grafting, some surgeons may find it difficult to accurately score radiographs close to the time of graft surgery. Lastly, the raters in our study were highly-experienced orthopaedic traumatologists, suggesting that our estimates of inter-rater reliability could be higher than might be expected when applied to the broader community of surgeons.

In this retrospective study, the mRUST appears similarly reliable to prior studies in the evaluation of patients with open, diaphyseal tibia fractures with associated bone defects treated with intramedullary nails, but the use of this cortical scoring either in patients treated with plates and with ring fixators or in cases with removal of all instrumentation is less reliable. Since the assessment of fracture union continues to be a challenging task for orthopaedic surgeons, especially among patients with open, diaphyseal tibia fractures with associated bone defects, a prospective, controlled study is needed to definitively assess the utility of this tool in patients with bone defects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Source of Funding: Grant supported by Department of Defense USAMRMC, Contract Number: W81XWH-09–20108, and by T32 AR067708 (S.M.) from the National Institutes of Health.

References

1. Cook GE, Bates BD, Tornetta P, et al. Assessment of fracture repair. J Orthop Trauma. 2015;29 Suppl 12:57 Accessed Nov 27, 2017. doi: 10.1097/BOT.0000000000000470.

2. Bhandari M, Guyatt GH, Swiontkowski MF, Tornetta P, 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. Accessed Nov 21, 2017. [PubMed: 12352564]
3. Kooistra BW, Sprague S, Bhandari M, Schemitsch EH. Outcomes assessment in fracture healing trials: A primer. J Orthop Trauma. 2010;24 Suppl 1:71 Accessed Nov 21, 2017. doi: 10.1097/BOT.0b013e3181ca3fbd.
4. 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. https://www.ncbi.nlm.nih.gov/pubmed/19996801. Accessed Nov 21, 2017. doi: 10.1097/TA.0b013e3181a7c16d. [PubMed: 19996801]
5. Litrenta J, Tornetta 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. Accessed Nov 21, 2017. doi: 10.1097/BOT.0000000000000390. [PubMed: 26165265]
6. Filho Azevedo, Silva de Fernando Antonio, Cotias RB, Azi ML, Teixeira, de Almeida Armando Augusto. Reliability of the radiographic union scale in tibial fractures (RUST). Rev Bras Ortop. 2017;52(1):35–39. Accessed Nov 21, 2017. doi: 10.1016/j.rboe.2016.05.006. [PubMed: 28194379]
7. 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 Joint Res. 2016;5(4):116–121. Accessed Nov 21, 2017. doi: 10.1302/2046-3758.54.2000628. [PubMed: 27073210]
8. Cekiç E, Alıcı E, Yeşil M. Reliability of the radiographic union score for tibial fractures. Acta Orthop Traumatol Turc. 2014;48(5):533–540. Accessed Nov 21, 2017. [PubMed: 25429579]
9. 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 1:81 Accessed Nov 21, 2017. doi: 10.1097/BOT.0b013e3181ca3fd1.
10. Bhandari M, Kooistra BW, Busse J, Walter SD, Tornetta P, Schemitsch EH. Radiographic union scale for tibial (r.u.s.t.) fracture healing assessment: Preliminary validation. Orthopaedic Proceedings. 2011;93-B(SUPP IV):575 http://bjjprocs.boneandjoint.org.uk/content/93-B/SUPP_IV/575.2.
11. Franzone JM, Finkelstein MS, Rogers KJ, Kruse RW. Evaluation of fracture and osteotomy union in the setting of osteogenesis imperfecta: Reliability of the modified radiographic union score for tibial fractures (RUST). J Pediatr Orthop. 2017 Accessed Nov 21, 2017. doi: 10.1097/BPO.0000000000001068.
12. Richards BS, Wilkes D, Dempsey M, Nurenberg P. A radiographic scoring system to assess healing in congenital pseudarthrosis of the tibia. J Pediatr Orthop B. 2015;24(2):118–122. Accessed Nov 21, 2017. doi: 10.1097/BPB.0000000000000141. [PubMed: 25588045]
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. Accessed Nov 21, 2017. doi: 10.1007/s11999-013-3383-y. [PubMed: 24249537]
14. Building a clinical research network in trauma orthopaedics: The major extremity trauma research consortium (METRC). J Orthop Trauma. 2016;30(7):353–361. Accessed Jan 4, 2018. doi: 10.1097/BOT.0000000000000549. [PubMed: 27333458]
15. Marsh JL, Slongo TF, Agel J, et al. Fracture and dislocation classification compendium - 2007: Orthopaedic trauma association classification, database and outcomes committee. J Orthop Trauma. 2007;21(10 Suppl):1 Accessed Jan 12, 2018. [PubMed: 17211261]
16. Krippendorff K Computing krippendorff's alpha-reliability. 2011 https://repository.upenn.edu/asc_papers/43. Accessed Feb 13, 2018.
17. Krippendorff K Content analysis: An introduction to its methodology. 2nd Edition ed. Thousand Oaks, CA: SAGE; 2004 Accessed Feb 13, 2018.
18. Tawonsawatruk T, Hamilton DF, Simpson A Hamish RW. Validation of the use of radiographic fracture-healing scores in a small animal model. J Orthop Res. 2014;32(9):1117–1119. Accessed Feb 8, 2018. doi: 10.1002/jor.22665. [PubMed: 24895294]

19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psych ol Bull. 1979;86(2): 420–428.

20. 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. Accessed Nov 21, 2017. doi: 10.1097/BOT.0000000000000753. [PubMed: 28072652]

21. Cooke ME, Hussein AI, Lybrand KE, et al. Correlation between RUST assessments of fracture healing to structural and biomechanical properties. J Orthop Res. 2017 Accessed Nov 21, 2017. doi: 10.1002/jor.23710.

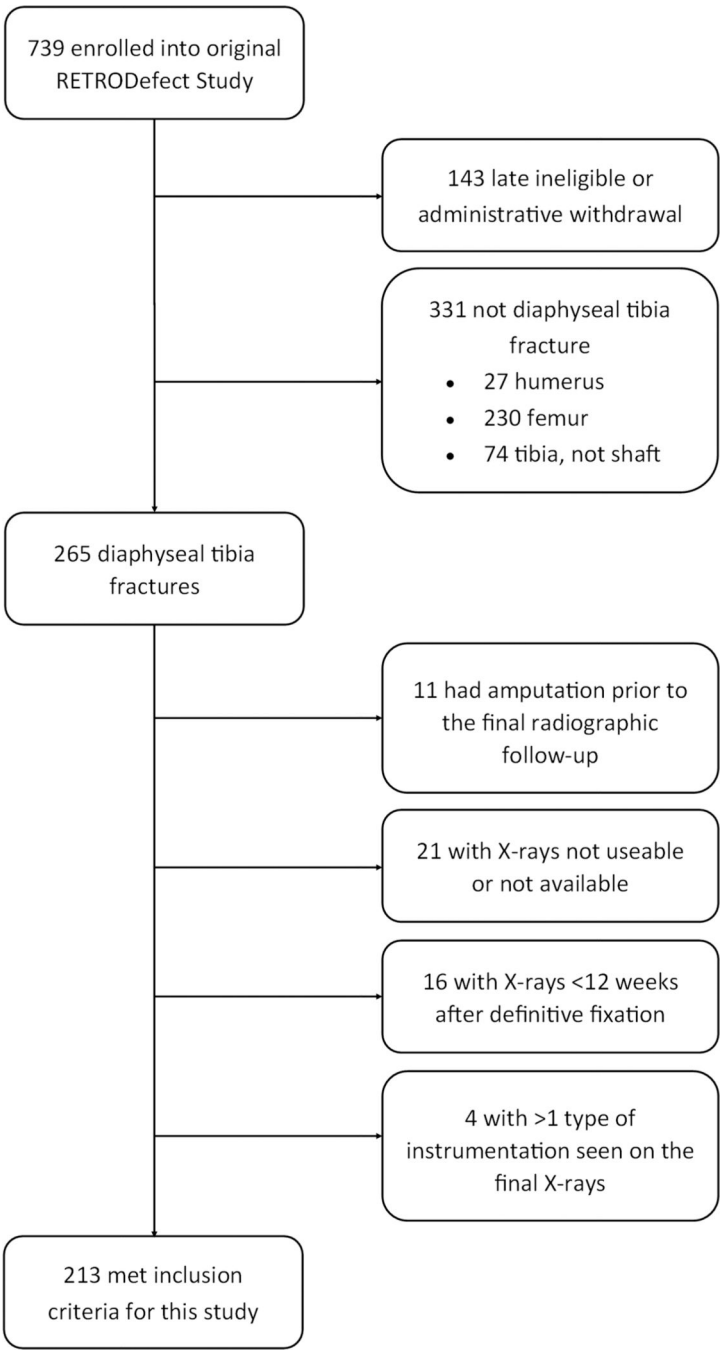

Figure 1. Flow Diagram of Patient Selection

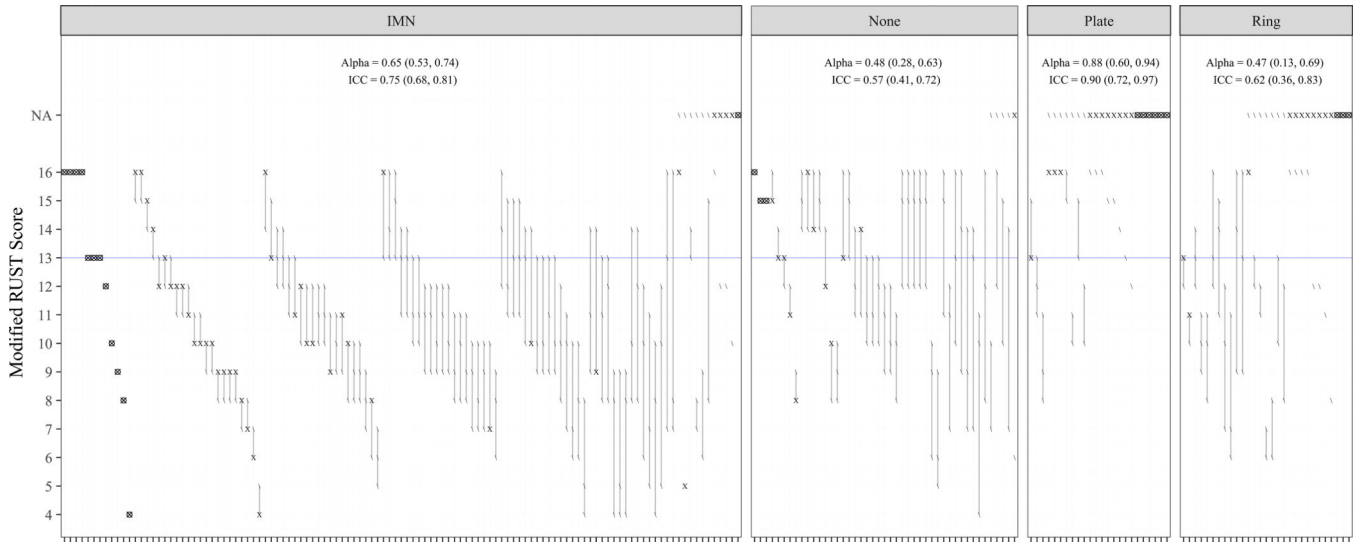

Figure 2. Visualizing disagreement in mRUST by type, or lack of, instrumentation visible on final radiograph applied to 213 cases.

Each rater is represented by a symbol with "/", "X", and "⊗" representing agreement among the three raters. Presence of the "X" symbol indicates that two (of three) raters were in perfect agreement for a given case, and presence of the "⊗" symbol indicates that all three raters were in perfect agreement. The "/" symbol indicates that a single rater provided a different assessment than the other raters. For included cases, the vertical line segment connects the minimum and maximum scores for a single patient. The figure is stratified into four panels by instrumentation group. Within each panel, scoring disagreement (as measured by the length of the line segments) increases from left to right; non-contributing cases (≥2 N/As) are placed furthest to the right. Multiple cases straddle a previously defined score indicating union (mRUST=13 (blue line)). Thirty-nine (18%) cases had ≥1 score above and another score below the threshold (IMN: 15%, None: 38%, Plate: 0%, Ring: 17%).
"IMN" indicates intramedullary nail; "None" indicates cases with no instrumentation seen on rated radiographs; "N/A" indicates that a score could not be computed; Alpha indicates Krippendorff's Alpha; ICC indicates Intraclass Correlation Coefficient

Figure 3. Radiographs from a patient with a clinically healed tibia fracture showing persistent abnormal appearance of the bone

This is an example of a case that may be unreliably scored under the mRUST protocol.

Table 1.

Treatment and Injury Characteristics among Patients With and Without at Least Two Valid mRUST Scores

	Entire cohort	<2 Valid mRUST Scores	≥2 Valid mRUST Scores
Total N	213	31 (15%)[≠]	182 (85%)[≠]
Instrumentation on rated X-rays			
IMN	115	5 (4%)	110 (96%)
Defect size			
<2.5 cm	73	3 (4%)	70 (96%)
2.5–5.0 cm	23	2 (9%)	21 (91%)
>5.0 cm	19	0 (0%)	19 (100%)
ORIF w/ plate	24	14 (58%)	10 (42%)
Defect size			
<2.5 cm	12	7 (58%)	5 (42%)
2.5–5.0 cm	9	5 (56%)	4 (44%)
>5.0 cm	3	2 (67%)	1 (33%)
External Fixator	29	11 (38%)	18 (62%)
Defect size			
<2.5 cm	9	2 (22%)	7 (78%)
2.5 – 5.0 cm	6	4 (67%)	2 (33%)
>5.0 cm	14	5 (36%)	9 (64%)
None	45	1 (2%)	44 (98%)
Defect size			
<2.5 cm	15	1 (7%)	14 (93%)
2.5–5.0 cm	14	0 (0%)	14 (100%)
>5.0 cm	16	0 (0%)	16 (100%)
Time to X-rays	294 ± 85	299 ± 88	293 ± 85
Bone defect size			
<2.5 cm	109	13 (12%)	96 (88%)
2.5–5.0 cm	52	11 (21%)	41 (79%)
>5.0 cm	52	7 (13%)	45 (87%)
Bone grafting status			
No	104	13 (12%)	91 (88%)
Yes	109	18 (17%)	91 (83%)

[≠]= row %

Table 2.

Inter-Rater Reliability of the mRUST Score, Overall and Stratified by Final Instrumentation, Defect Size, and Bone Grafting Status

	Not Used in KA and ICC		Used in KA and ICC			
	0/3 scores*	1/3 scores*	2/3 scores*	3/3 scores*	KA (95% CI)	ICC (95% CI)
Overall	10 (5%)	21 (10%)	24 (11%)	158 (74%)	0.64 (0.54, 0.71)	0.71 (0.65, 0.77)
By Instrumentation						
IMN	1 (1%)	4 (3%)	6 (5%)	104 (90%)	0.65 (0.53, 0.74)	0.75 (0.68, 0.81)
Plate	6 (25%)	8 (33%)	7 (29%)	3 (13%)	0.88 (0.60, 0.94)	0.90 (0.72, 0.97)
Ring	3 (10%)	8 (28%)	7 (24%)	11 (38%)	0.47 (0.13, 0.69)	0.62 (0.36, 0.83)
None	0 (0%)	1 (2%)	4 (9%)	40 (89%)	0.48 (0.28, 0.63)	0.57 (0.41, 0.72)
Defect Size (cm)						
<2.5	5 (5%)	8 (7%)	9 (8%)	87 (80%)	0.60 (0.46, 0.71)	0.70 (0.61, 0.78)
2.5–5.0	3 (6%)	8 (15%)	6 (12%)	35 (67%)	0.77 (0.60, 0.85)	0.85 (0.77, 0.91)
>5.0	2 (4%)	5 (10%)	9 (17%)	36 (69%)	0.57 (0.36, 0.71)	0.67 (0.53, 0.79)
Grafted prior to X-ray?						
No	3 (3%)	10 (10%)	10 (10%)	81 (78%)	0.61 (0.47, 0.72)	0.70 (0.61, 0.78)
Yes	7 (6%)	11 (10%)	14 (13%)	77 (71%)	0.66 (0.54, 0.75)	0.73 (0.64, 0.80)
Yes, days prior:						
<30	0 (0%)	0 (0%)	1 (33%)	2 (67%)	—	—
30–90	0 (0%)	3 (21%)	1 (7%)	10 (71%)	0.60 (0.19, 0.79)	0.67 (0.31, 0.90)
>90	7 (8%)	8 (9%)	12 (13%)	65 (71%)	0.66 (0.52, 0.76)	0.74 (0.65, 0.82)

* - the number of raters (0, 1, 2, or 3) out of 3 who provided an mRUST score for all 4 cortices of a given case.

KA indicates Krippendorff's Alpha; ICC indicates Intraclass Correlation Coefficient.