Inter-Rater Reliability of Modified RUST Scoring for Diaphyseal Tibia Fractures with Bone Defects

Paul Tornetta III, MD[1]*; Ellen MacKenzie, PhD*[2]*; William Obremskey, MD, MPH*[3]*;*
CAPT (ret) Michael J. Bosse, MD[4]*; Joseph Hsu, MD*[4]*; Saam Morshed, MD*[5]*;*
Jason Luly, MS[6]*; Robert V. O'Toole, MD*[7]
[1]*Boston Medical Center, Boston, Massachusetts, USA;*
[2]*Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA;*
[3]*Vanderbilt Orthopaedic Institute, Nashville, Tennessee, USA;*
[4]*Carolinas Medical Center, Charlotte, North Carolina, USA;*
[5]*UCSF/SFGH Orthopaedic Trauma Institute, San Francisco, California, USA*
[6]*Johns Hopkins University, Baltimore, Maryland, USA;*
[7]*University of Maryland, Shock Trauma, Baltimore, Maryland, USA*

Background/Purpose: Cortical scoring systems, such as the Radiographic Union Score for Tibia Fractures (RUST) were developed to improve the grading of callus formation and radiographic progression toward union in tibial diaphyseal fractures. The modified RUST (mRUST) added an additional scoring level and has demonstrated better agreement in diaphyseal and metadiaphyseal fractures. These methods have shown excellent agreement that has led to adoption in many fracture trials. However, cortical scoring has not been evaluated in the face of fractures with bone defects, where assigning a score may be more difficult. The current study seeks to evaluate the agreement of the mRUST in patients with operatively treated open tibial fractures associated with bone loss as it may relate to reporting in clinical trials.

Methods: All skeletally mature patients (≥18 years) with open diaphyseal tibia shaft fractures and a bone defect >1 cm treated operatively over a 5-year period at 17 centers were included. Patients with amputations were excluded. Defects were divided by their largest gap as <2.5 cm, 2.5-5.0 cm, or >5.0 cm. Radiographs between 11 and 13 months postinjury were selected for scoring. If no radiographs were available during this time fame the final available radiograph was used. Three experienced orthopaedic surgeons from a pool of 6 were randomly assigned to apply the mRUST to each case. Each of the cortices on the AP and lateral radiographs were graded as: 1 = no callus, 2 = callus present without bridging, 3 = bridging callus, and 4 = fracture line not visible (remodeled). The mRUST score is the sum of the 4 cortical scores (4-16). If any cortex could not be assessed due to an implant, the score was not calculated. Raters were blinded to the original films, defect size, and whether the patient received a bone graft. Inter-rater reliability of mRUST was assessed using two measures: (1) the intraclass correlation coefficient (ICC), and (2) Krippendorf's alpha. For both measures, a score close to 1 indicates agreement and a score close to 0 indicates non-agreement. Krippendorf's alpha is preferred for ordinal data such as the modified RUST. Absolute ICC was computed to allow comparison with other research on the mRUST.

Results: 234 patients (202 M, 32 F; average age 34 [range, 18-68]), met inclusion criteria. The average time between definitive fixation and the selected radiograph was 278 ± 103 days. All raters were able to score all four cortices on 171 (73%) subjects; two raters scored all four cortices on 28 (12%), and no raters were able to score all four cortices on 12 (5%) subjects. 160

(68%) subjects were internally fixed and 74 (32%) were treated with a ring external fixator. 112 (48%) had bone graft an average of 187 ± 91 days prior to the scored radiograph. The agreement based on Krippendorf's alpha was 0.67 (CI: 0.59-0.74) and the ICC was 0.69 (CI: 0.62-0.74). Higher agreement was seen for subjects treated with internal fixation as compared with external fixation (0.72 and 0.70 vs 0.54 and 0.51 for the Krippendorf's alpha and ICC, respectively). Bone grafting did not affect the level of agreement, but intermediate defects yielded slightly better agreement than did smaller or larger defects (ICC: 0.8 vs 0.65 and 0.64; Krippendorf's alpha: 0.79 vs 0.65 and 0.57).

Conclusion: Cortical scoring systems have become common tools in reporting radiographic progression toward union in lower extremity fracture trials. The purpose of this study was to assess one such cortical scoring method in the environment of open tibia fractures with bone loss. Agreement was found to be lower than prior trials of metaphyseal and diaphyseal fractures fixed with nails and plates. The agreement in the face of external fixation was the lowest reported in the literature. This may limit the usefulness of cortical scoring in determining the progression to union in patients with open fractures and bone loss, particularly if external fixation is used.

See pages 49 - 106 for financial disclosure information.