



## ■ TRAUMA

# Sensitivity and specificity of modified RUST score using clinical and radiographic findings as a gold standard

**Y. Plumarom,  
B. G. Wilkinson,  
M. C. Willey,  
Q. An,  
L. Marsh,  
M. D. Karam**

From Department  
of Orthopaedics  
and Rehabilitation,  
University of Iowa  
Hospitals and Clinics,  
Iowa City, Iowa, USA

## Aims

The modified Radiological Union Scale for Tibia (mRUST) fractures score was developed in order to assess progress to union and define a numerical assessment of fracture healing of metadiaphyseal fractures. This score has been shown to be valuable in predicting radiological union; however, there is no information on the sensitivity, specificity, and accuracy of this index for various cut-off scores. The aim of this study is to evaluate sensitivity, specificity, accuracy, and cut-off points of the mRUST score for the diagnosis of metadiaphyseal fractures healing.

## Methods

A cohort of 146 distal femur fractures were retrospectively identified at our institution. After excluding AO/OTA type B fractures, nonunions, follow-up less than 12 weeks, and patients aged less than 16 years, 104 sets of radiographs were included for analysis. Anteroposterior and lateral femur radiographs at six weeks, 12 weeks, 24 weeks, and final follow-up were separately scored by three surgeons using the mRUST score. The sensitivity and specificity of mean mRUST score were calculated using clinical and further radiological findings as a gold standard for ultimate fracture healing. A receiver operating characteristic curve was also performed to determine the cut-off points at each time point.

## Results

The mean mRUST score of ten at 24 weeks revealed a 91.9% sensitivity, 100% specificity, and 92.6% accuracy of predicting ultimate fracture healing. A cut-off point of 13 points revealed 41.9% sensitivity, 100% specificity, and 46.9% accuracy at the same time point.

## Conclusion

The mRUST score of ten points at 24 weeks can be used as a viable screening method with the highest sensitivity, specificity, and accuracy for healing of metadiaphyseal femur fractures. However, the cut-off point of 13 increases the specificity to 100%, but decreases sensitivity. Furthermore, the mRUST score should not be used at six weeks, as results show an inability to accurately predict eventual fracture healing at this time point.

**Cite this article:** *Bone Jt Open* 2021;2-10:796–805.

**Keywords:** Metadiaphyseal fracture, Femur fracture, Fracture healing, Modified RUST score

## Introduction

The ability to accurately determine fracture healing and to predict union is central to patient care and to measure the success of various fracture interventions. Traditional clinical methods for evaluating fracture healing include the absence of pain with weightbearing or with palpation at the fracture site.<sup>1</sup> Radiological criteria of fracture healing include many factors ranging from cortical continuity, visibility of a fracture

line, specific number of bridging cortices or simply surgeon impression.<sup>1-3</sup> Radiological scoring systems have also been described.<sup>1-3</sup> One such scoring system is the Radiological Union Scale for Tibia (RUST) fractures, which numerically evaluates and assesses progression to union after intramedullary (IM) nailing of tibia fractures.<sup>2,4</sup> Subsequently, a modified RUST (mRUST) score was described using four scores for cortices rather than three in attempts to better assess union and bridging

Correspondence should be sent to Yanin Plumarom; email: yaninyo24@gmail.com; yanin.p@pcr.ac.th

doi: 10.1302/2633-1462.210.BJO-2021-0071.R1

*Bone Jt Open* 2021;2-10:796–805.

callus in metadiaphyseal fracture area.<sup>3</sup> The mRUST score is considered a reliable method for evaluation of metadiaphyseal fracture healing.<sup>3</sup> To our knowledge, the accuracy, sensitivity, and specificity of the mRUSTscore in evaluating fracture union have not been determined. The goal of this study was to determine the time dependent sensitivity and specificity of the mRUSTscore in radiologically predicting fracture union of distal femur fractures, and subsequently to determine optimal cut-off values of mRUST at various time points (six weeks, 12 weeks, 24 weeks, and final follow-up).

## Methods

After obtaining Institutional Review Board approval (IRB ID #: 201612761), a cohort of 143 consecutive adult patients with 146 sets of radiographs of distal femur fractures, including periprosthetic, treated surgically from 2011 to 2016 at a single institution (Department of Orthopaedics and Rehabilitation, University of Iowa Hospitals and Clinics, USA) were identified. We excluded the radiographs with partial articular fractures ( $n = 14$ ), referred or old nonunion cases ( $n = 5$ ), incomplete follow-up less than 12 weeks ( $n = 17$ ), follow-up only one time point ( $n = 2$ ), unclear shadow of radiographs ( $n = 2$ ), compression technique was used ( $n = 1$ ), or patients in which bone allograft was used ( $n = 1$ ). All surgeons employed standard locked plating, angle blade, and nailing techniques. All periprosthetic fractures, type A fractures, and type C fractures with a simple articular line were treated with a small distal incision for plate/nail insertion, indirect reduction of the fracture and percutaneous insertion of proximal shaft screws. The remaining type C fractures were treated with a lateral arthrotomy, with or without lag screw (articular portions only), and percutaneous insertion of proximal shaft screws. The metaphyseal portion of the fracture was not directly reduced, and surgical fixation was bridging to provide relative stability. After applying our exclusion criteria, 104 sets of radiographs of distal femur fractures were included in the study. Of this cohort, 93 fractures (89.4%) were treated with locking plate constructs (51 periprosthetic fractures), one (0.96%) with angle blade plate, and ten (9.6%) with intramedullary nails. Operative fixation was performed by one of three fellowship trained orthopaedic traumatologists (JLM, MDK, MCW). Patients routinely returned for follow-up at two weeks, six weeks, 12 weeks, and 24 weeks, or until fracture union was determined by the surgeons. Clinical history and physical examination were performed at each follow-up visit and recorded in the medical record. Orthopaedic Trauma Association/AO Foundation (OTA/AO) fracture classifications were demonstrated in Table I.

**Evaluation of radiological and clinical healing.** AP and lateral radiographs of the knee and femur were independently reviewed by three investigators (YP, MCW,

**Table I.** Characteristics of distal femur fractures.

Characteristic	Total fractures, n (%)
<b>OTA/OA fracture classification</b>	
33A	
33A1	11 (10.6)
33A2	3 (2.9)
33A3	20 (19.2)
33C	
33C1	10 (9.6)
33C2	15 (14.4)
33C3	2 (1.9)
<b>Periprosthetic fractures</b>	
Hip	7 (6.7)
Knee	30 (28.8)
Interprosthesis	6 (5.8)

AO, AO Foundation; OTA, Orthopaedic Trauma Association.

MDK). Callus formation of each cortex was evaluated and scored using the mRUST score.<sup>3</sup> Assessments were performed on radiographs taken at six weeks, 12 weeks, 24 weeks, and final follow-up for each fracture, which was routinely followed up in the clinic. The mRUST score was assessed using both AP and lateral radiographs and all four cortices were scored as demonstrated in Figure 1 and Table II.

In order to decrease the effect of a potential learning curve in applying the mRUST score, prior to any radiological assessment, all three investigators reviewed and came to a consensus on how to apply the mRUST score to each cortex.<sup>3,5-7</sup> However, actual scoring of the 104 sets of radiographs was done by the investigators independently from each other. They were allowed access to radiographs of a single time point only, and were blinded to radiographs from the next time point. Once evaluation and scores for a specific time point were documented, the evaluators were then allowed to view radiographs for the next time point for subsequent scoring at six weeks, 12 weeks, 24 weeks, and final follow-up, respectively. In cases of obstructed visualization of the lateral cortex, our observers were instructed to use consolidation of the fracture line at the lateral cortex to best apply the mRUST score. Scores provided by each investigator were summed and divided to calculate the average mRUST score for each follow-up time point.

The ability of the averaged mRUST score to predict eventual fracture union was determined. The gold standard for union was determined by the operating surgeon (JLM, MDK, MCW) as documented in the medical record, and was based on clinical history and exam findings (patient-reported absence of pain, painless ambulation, absence of limp, no tenderness to palpation of fracture site) and evaluation of radiographs (cortical continuity, bridging cortices, visibility of fracture line) at the time of follow-up. Radiographs of all patients documented as fracture union, that had follow-up less than six months

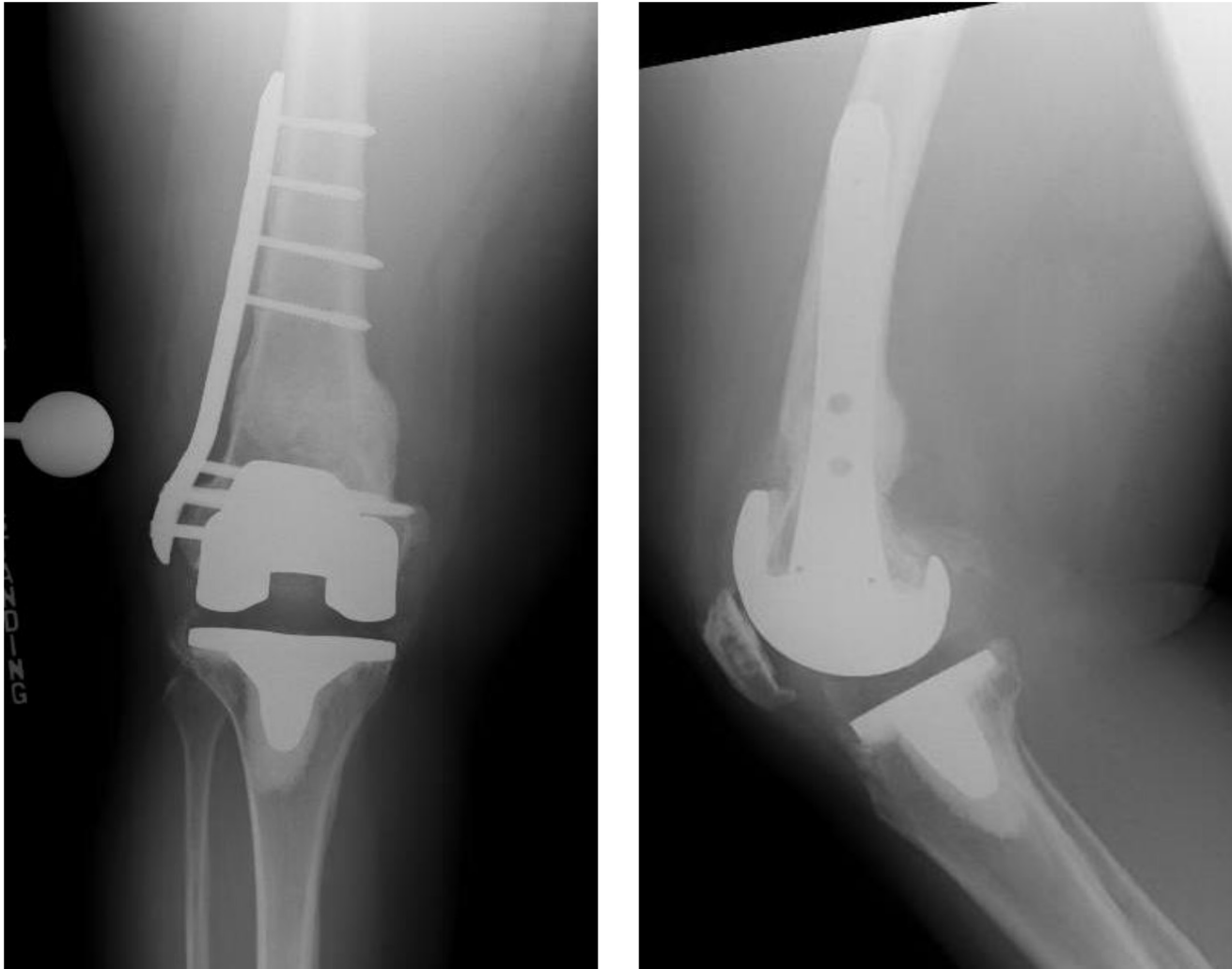


Fig. 1

Radiographs illustrating a modified Radiological Union Scale for Tibia fractures (mRUST) score of 15 at final follow-up. Cortical score: lateral = 4, medial = 4, anterior = 3, and posterior = 4.

were independently reviewed by three authors (YP, BGW, JLM) for confirmation of union. Similarly, nonunion was determined by the operating surgeon as documented in the medical chart based on clinical and radiological findings. All nonunion cases were independently reviewed by three authors (YP, BGW, JLM) to confirm nonunion and determine construct condition, time of nonunion, time of revision, and revision type for nonunion. Mean length of follow-up for both union and nonunion cases was collected from the medical record. Mean follow-up for all patients ( $n = 104$ ) was 12.5 months (3 to 55; median 9). Mean follow-up for patients with fracture union ( $n = 92$ ) was 12.1 months (3 to 55; median 8.5). The mean time of clinical and radiological finding union from the surgeons from medical record was 11.5 weeks (4 to 28; median 12).

**Statistical analysis.** The study was designed in conjunction with a statistician (QA). All data are reported as means with associated standard deviation. All data

**Table II.** Assessment tool for the modified Radiological Union Scale for Tibia fractures (mRUST) scores.

Score per cortex*	Radiological callus criteria
1	No callus
2	Callus present
3	Bridging callus
4	Remodelled, fracture not visible

\*The individual cortical scores (anterior, posterior, medial, and lateral) are added to provide a mRUST. The mRUST score was totalled for each fracture to equal a minimum score of 4 or a maximum score of 16. Low scores indicate poor fracture healing and callus formation, while high scores correlate with fracture healing and remodelling.

analysis, including receiver operating characteristic (ROC) curves generation by logistic regression, was performed using SAS software, version 9.3 (SAS Institute, USA). The sensitivity, specificity, and accuracy for each cut-off point which is the least of mRUST score that define union radiologically was determined for each separate time point.

**Table III.** Intraclass correlation coefficients (ICCs) among investigators using the modified Radiological Union Scale (mRUST) for tibia fractures score.

Timing	ICC (95% confident interval)
Six weeks	0.58 (0.47 to 0.67)
12 weeks	0.54 (0.43 to 0.64)
24 weeks	0.50 (0.37 to 0.62)
Final	0.57 (0.45 to 0.69)

The appropriate cut-off points were selected. The interpretation of ROC curve is similar to a single point the ROC space; the closer the point on the ROC to the ideal coordinate, the more accurate the test is. The closer the point on the ROC curve to the diagonal, the less accurate the test is.<sup>8</sup> The better the diagnostic test, the more quickly the true positive rate nears 1 (or 100%). A near-perfect diagnostic test would have a ROC curve that is almost vertical from (0,0) to (0,1) and then horizontal to (1,1).<sup>9</sup>

The AUC is a popular measure of the accuracy of a diagnostic test. In general, higher AUC values indicate better test performance. The possible values of AUC range from 0.5 (no diagnostic ability) to 1.0 (perfect diagnostic ability).<sup>10</sup> The accuracy of a diagnostic test, classified by the AUC, is summarized as follows: 0.9 < AUC < 1.0 equals excellent, 0.8 < AUC < 0.9 equals good, 0.7 < AUC < 0.8 equals acceptable, and 0.6 < AUC < 0.7 equals not good classification, respectively.<sup>8,11</sup>

There are several criteria for determination of the most appropriate cut-off value in a diagnostic test with continuous results. Mostly based on receiver operating characteristic (ROC) analysis, there are various methods to determine the test cut-off value. The most common criteria are the point on ROC curve where the sensitivity and specificity of the test are equal; the point on the curve with minimum distance from the left-upper corner of the unit square; and the point where the Youden's index is maximum.<sup>9,12</sup>

While the ROC curve and corresponding AUC give an overall picture of the behaviour of a diagnostic test across all cut-off values, there remains a practical need to determine the specific cutoff value that should be used for individuals requiring diagnosis. If the cost of each diagnostic decision is known, as well as the positive condition prevalence, the optimal cutoff value is the one that minimizes cost. However, cost and prevalence values are typically unknown and unattainable. In this case, a recommended approach is to find the cutoff with highest Youden Index, or equivalently, the highest sensitivity plus specificity.<sup>9</sup> Therefore, in this study, we selected the cut-off points using this method to determine the most appropriate sensitivity and specificity.

Intraclass correlation coefficients (ICCs) with 95% confidence intervals (CIs) were used to measure agreement in the observer's mRUST scores. The ICC, used to quantify agreement for a continuous variable, is equivalent to the

**Table IV.** Results of mean modified Radiological Union Scale for Tibia fractures (mRUST) score from three investigators at each time point.

Time point	n	Mean (SD)	Minimum to maximum
Six weeks	104	6.71 (1.81)	4.00 to 11.33
12 weeks	102	9.73 (1.97)	5.00 to 13.67
24 weeks	82	11.94 (2.32)	5.00 to 16.00
Final	82	13.83 (2.04)	6.67 to 16.00

SD, standard deviation.

quadratically weighted kappa ( $\kappa$ ) for categorical data. The weighted kappa, as described by Fleiss,<sup>4</sup> adjusts the observed proportion of agreement by correction for the proportion of agreement that could have occurred by chance alone. As they are numerically equivalent, the same guidelines for interpretation of kappa values can be applied to the ICC. Landis and Koch<sup>10</sup> suggest kappa of 0 to 0.2 represents "slight agreement," 0.21 to 0.40 "fair agreement," 0.41 to 0.60 "moderate agreement," and 0.61 to 0.80 "substantial agreement." A value above 0.80 is considered almost "perfect agreement." The value of the ICC ranges from +1, in which case there is "perfect agreement, to -1, which corresponds to "absolute disagreement."

## Results

**Demographic data.** A total of 104 distal femur fractures (61.7% female; mean age 65.4 years) were included for radiological evaluation via the mRUST score. In all, 12 (11.5%) nonunions were identified through a combination of radiological and clinical assessment with an average follow-up of 12.5 months.

**Radiological assessment of callus using mRUST score.** The reliability of the mRUST score was assessed by comparing the four sub-scores provided by each independent observer for each follow-up time point using ICCs. Overall, there was moderate agreement with ICC values of 0.57, 0.52, 0.56, and 0.61 at six weeks, 12 weeks, 24 weeks, and final follow-up, respectively (Table III).

There were some patients without follow-up radiographs at some time points, leaving 104, 102, 82, and 82 sets of radiographs to evaluate mRUST at six weeks, 12 weeks, 24 weeks, and at final follow-up, respectively. The mRUST score from each investigator was summed and averaged. Mean mRUST scores were 7, 10, 12, and 14 at six weeks, 12 weeks, 24 weeks, and at final follow-up, respectively (Table IV).

**Selection of appropriate cut point.** The most appropriate sensitivity and specificity is summarized in Table V.

When the mRUST score was assessed as predictive of healing at the six-week and 12-week time points, the optimal cut points were six and nine, respectively. At these earlier time points, there was the sensitivity and specificity of 85.2% and 41.3%, and 78.8% and 81.3% at six weeks and 12 weeks, respectively. The accuracy of six weeks time point was 52.9% with positive predictive

**Table V.** Results of sensitivity, specificity of each cut point of mean modified Radiological Union Scale for Tibia fractures (mRUST) scores in each period time point.

Time point	Cut point	Sensitivity, %	Specificity, %	1-specificity, %	
Six weeks	4	100.0	0.0	100.0	
	5	92.60	21.3	78.7	
	<b>6</b>	<b>85.20</b>	<b>41.3</b>	<b>58.7</b>	
	7	51.90	58.7	41.3	
	8	44.40	78.7	21.3	
	9	22.20	92.0	8.0	
	10	14.8	96.0	4.0	
	11	7.4	100.0	0.0	
	12 weeks	5	100.0	0.0	100.0
		7	98.8	37.5	62.5
		8	89.4	62.5	37.5
<b>9</b>		<b>78.8</b>	<b>81.3</b>	<b>18.8</b>	
10		58.8	93.8	6.3	
11		37.6	100.0	0.0	
12		20.0	100.0	0.0	
24 weeks	13	5.9	100.0	0.0	
	5	100.0	0.0	100.0	
	6	98.6	0.0	100.0	
	8	98.6	71.4	28.6	
	<b>10</b>	<b>91.9</b>	<b>100.0</b>	<b>0.0</b>	
	11	81.1	100.0	0.0	
	12	62.2	100.0	0.0	
	13	41.9	100.0	0.0	
	14	20.3	100.0	0.0	
	15	9.5	100.0	0.0	
Final	16	4.1	100.0	0.0	
	10	100.0	50.0	25.0	
	11	97.3	50.0	25.0	
	<b>12</b>	<b>89.6</b>	<b>100.0</b>	<b>0.0</b>	
	13	74.7	100.0	0.0	
	14	62.7	100.0	0.0	
	15	41.3	100.0	0.0	

Bold shows the most appropriate cut points, sensitivity, and specificity selected.

**Table VI.** Summary of sensitivity, specificity, accuracy, PPV, NPV, and AUC of each cut point at each time point.

Time point	Cut point	Sensitivity, %	Specificity, %	Accuracy, %	PPV, %	NPV, %	AUC
Six weeks	6	85.2	41.3	52.9	34.3	88.6	0.6593
12 weeks	9	78.8	81.3	79.2	95.7	41.9	0.8864
24 weeks	10	91.9	100.0	92.6	100.0	53.8	0.9826
Final	12	89.6	100.0	90.1	100.0	33.3	0.9805

AUC, area under the ROC curve; NPV, negative predictive value; PPV, positive predictive value.

value (PPV) was 34.3% and negative predictive value (NPV) was 88.6%. However, the AUC was 0.6593 which is classified as “not good”, according to commonly used classification systems for AUC as a diagnostic test.<sup>10</sup> The accuracy of the 12-week time point was 79.2% with positive predictive value (PPV), and was 97.5% and negative predictive value (NPV) was 41.9%. The AUC at the same cut point was 0.8864.

When the mRUST score was assessed as predictive of healing at the 24-week time point, the optimal cut points were ten. Compared to gold standard for union at this time point, the most appropriate sensitivity and specificity of the mRUST of ten was 91.9% and 100%, respectively. Using the same cut-off point, positive predictive value (PPV) was 100%, while negative predictive value (NPV) was 53.8% with the highest accuracy of 92.6% (summarized in Table VI).

Increasing the cut-off point to 13 at the 24-week time point yielded decreased sensitivity (41.9%), but specificity remained 100%. The AUC was 0.9826 which was classified as “excellent”, according to commonly used classification using AUC for a diagnostic test.<sup>8</sup> (Table V).

The ROC curve for the mean mRUST score compared to the gold standard at each time point is demonstrated in Figures 2 to 5.

## Discussion

The ability to accurately determine fracture healing and to predict union is central to patient care, and to measure the success of various fracture interventions. Traditional clinical methods for evaluating fracture healing include the absence of pain with weightbearing or with palpation at the fracture site.<sup>1</sup> Radiological criteria of fracture healing includes many factors ranging from cortical continuity, visibility of a fracture line, specific number of bridging cortices, or simply surgeon impression.<sup>1-3</sup> Therefore, it is important to have a method to evaluate radiological healing precisely. However, radiological assessment of fracture healing remains difficult as no clear consensus has been reached on assessing or determining bony union. RUST and mRUST were developed in order to assign a numerical value to tibial shaft fracture and metadiaphyseal healing after operative fixation.<sup>3,6</sup>

Kooistra et al<sup>5</sup> showed the RUST score had better reli-

ability when compared with a surgeon’s general impression or the number of cortices bridged by callus in the follow-up of tibial fractures. The mRUST score has been developed in order to evaluate cortical scoring systems in metadiaphyseal fractures.<sup>3</sup> Litrenta et al<sup>3</sup> showed substantial agreement and increased ICCs for the mRUST compared to the RUST score in the assessment

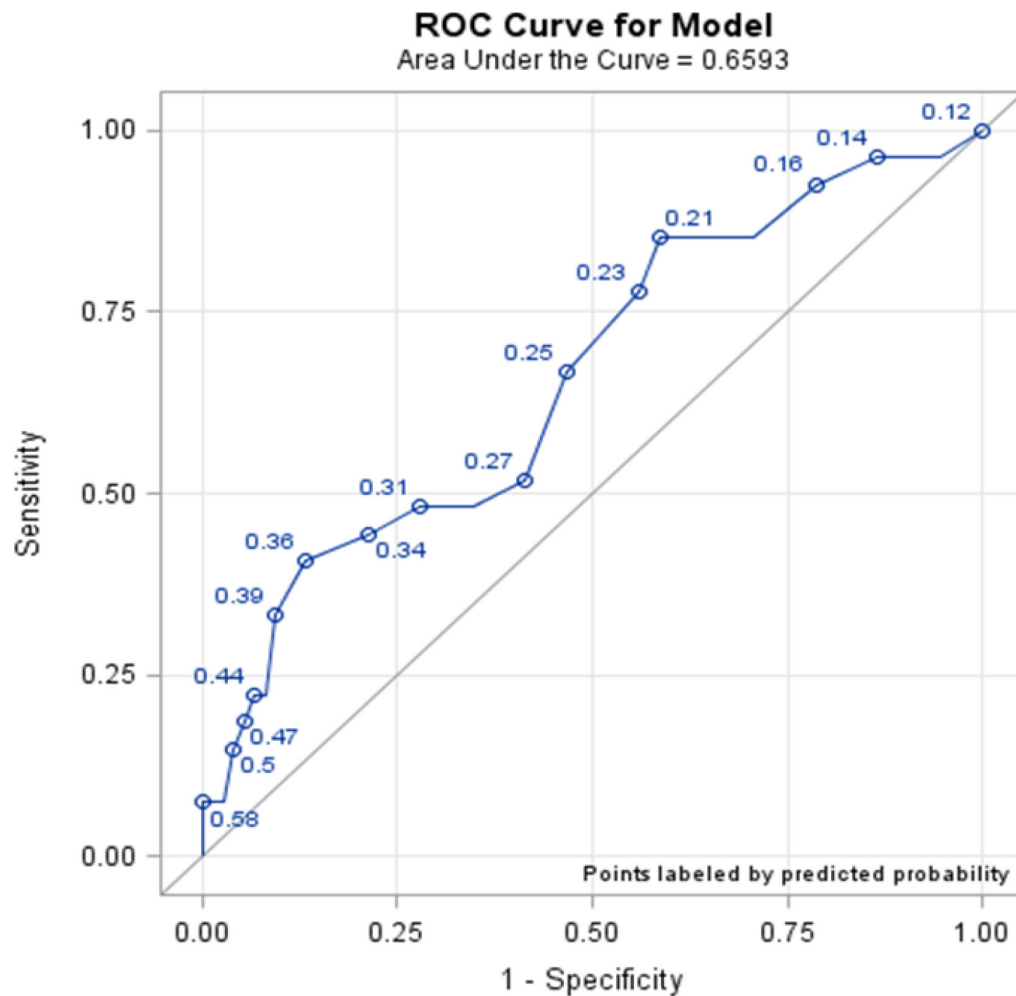


Fig. 2

Receiver operating characteristic (ROC) curve of modified Radiological Union Scale for Tibia fractures (mRUST) score at six weeks.

of metadiaphyseal fracture healing in a series of radiographs of distal femur treated with either plate or retrograde nail. They demonstrated that scores of ten and 13 using the RUST and mRUST scores, respectively, resulted in over 90% of reviewers assigning union on the RUST and mRUST, respectively.<sup>3</sup> Furthermore, they found that a standard RUST score of ten and a mRUST score of 14 provide an excellent definition of union based on surgeons' opinion and biomechanical testing in a sheep osteotomy model.<sup>13</sup> Cooke et al<sup>14</sup> demonstrated that radiologically healed fractures had a mRUST  $\geq$  13 and a RUST  $\geq$  10 and had excellent relationship to structural and biomechanical metrics in an animal study. This is clinically relevant, as scores showed high correlation to physical properties of healing and generally distinguished healed versus non-healed fractures.<sup>14</sup> However, there has been no investigation of sensitivity or specificity of the mRUST score in evaluating eventual fracture union. This study demonstrated the highest sensitivity of 91.9%, specificity of 100% and accuracy of 92.6% of mRUST score of ten at 24-week follow-up.

Cut-off points of the mRUST score in the previous studies did not rely on clinical healing as a gold standard. They solely implemented the agreement of reviewer to determine the score that define union radiologically. Some of those were also investigated from animal studies, which may have different results from clinical situation. Additionally, sensitivity and specificity have not been investigated in the literatures.

The current study supplements the literatures by being the first to assess and determine the sensitivity, specificity, and accuracy of the mRUST scoring system. Thus, the current study allows us to quantify how good and reliable mRUST is at predicting bony union. The ROC curve is a graphic representation of the relationship between both sensitivity and specificity, and it helps select the optimal model by determining the best threshold for the diagnostic test. Recognizing which cut-off points (i.e. mRUST scores) provide high sensitivity or specificity (or both) would prove extremely useful in predicting ultimate fracture union. Furthermore, identified cut-off points could further clinical communication between

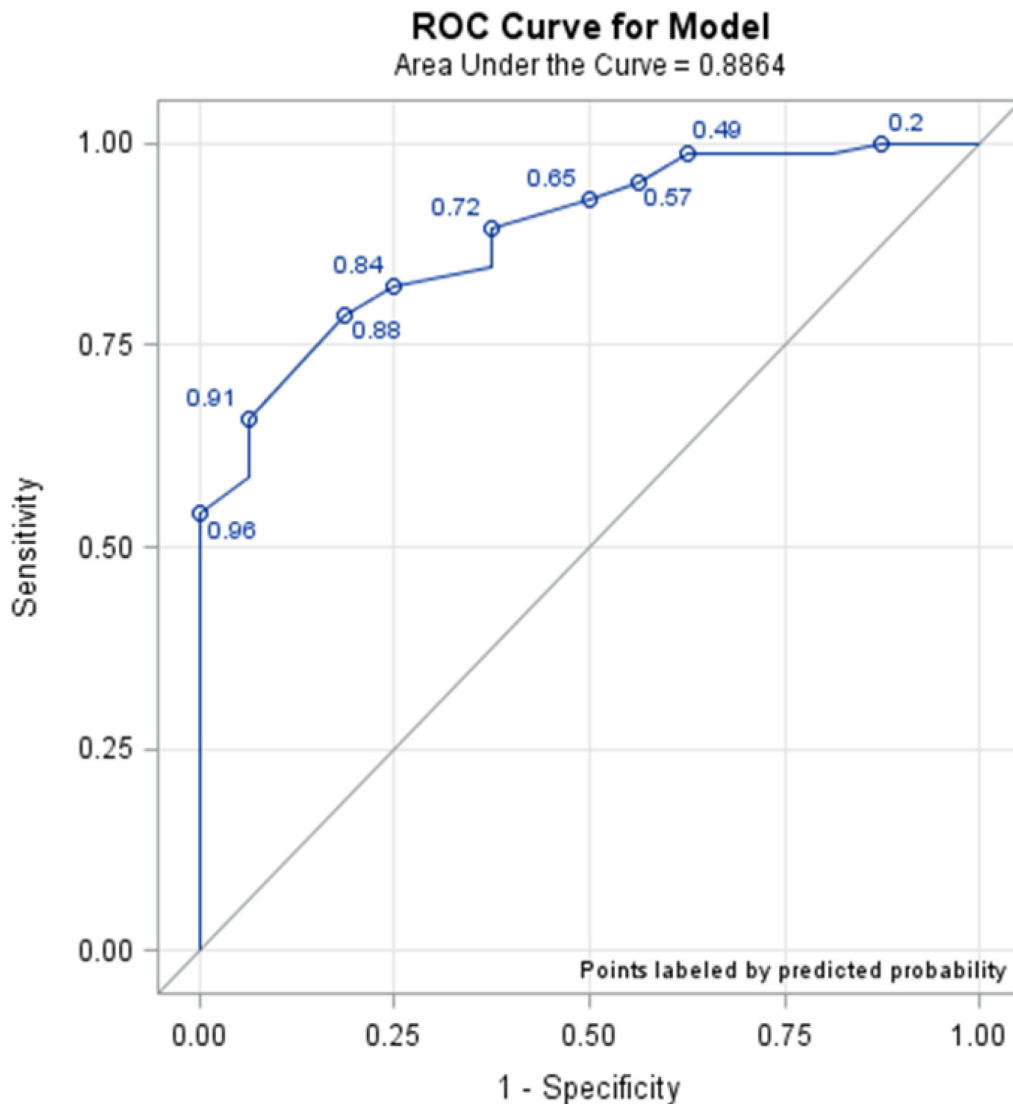


Fig. 3

Receiver operating characteristic (ROC) curve of modified Radiological Union Scale for Tibia fractures (mRUST) score at 12 weeks.

providers, as well as being used for continued research on fracture healing.

Interpretation of the ROC curve is similar to a single point in the ROC space: the closer the point on the ROC curve to the ideal coordinate, the more accurate the test is. The closer the points on the ROC curve are to the diagonal, the less accurate the test is. The faster the curve approaches the ideal point, the more useful the test results are. The AUC provides a way to measure the accuracy of a diagnostic test. The larger the area, the more accurate the diagnostic test is. There are several criteria for determining the most appropriate cut-off value in a diagnostic test with continuous results, based on ROC analysis.

According to the ROC curves in the current study, mRUST performs poorly and should not be used to predict eventual union when used within six weeks

postoperatively (AUC = 0.6593). This can be attributed to the timing of callus formation, and mRUST performs poorly in the early postoperative period as callus formation is likely not robust enough to be radiologically scored.

Conversely, the cut-off point of nine is appropriate for 12 weeks follow-up, and demonstrates a mRUST sensitivity and specificity of 78.8% and 81.3%, respectively. If we used the cut-off point at 13, as done in previous studies, the sensitivity significantly decreased to 5.9% and the specificity increased up to 100%. Both examples showing drastic changes in sensitivity and specificity with different mRUST scores, and highlights the need for validated cut-off points showing how minimal differences in scoring can produce drastic changes in sensitivity and specificity. Nonetheless, results show a cut-off point > ten yields 100% specificity and can be reliably used to make union likely possibility.

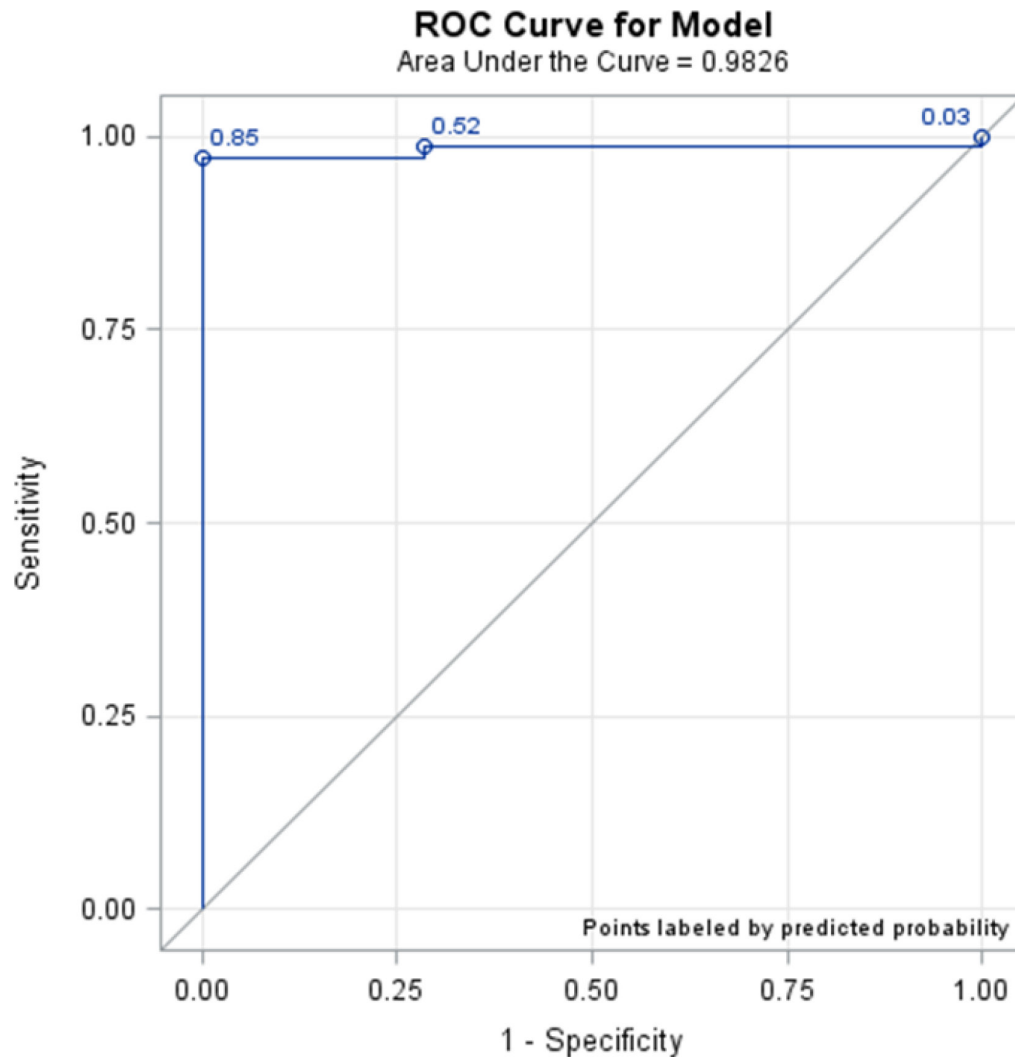


Fig. 4

Receiver operating characteristic (ROC) curve of modified Radiological Union Scale for Tibia fractures (mRUST) score at 24 weeks.

The cut-off point of ten is appropriate for 24 weeks follow-up yielding a mRUST sensitivity and specificity of 91.9% and 100%, respectively with the highest sensitivity plus specificity. The AUC of 0.9826 indicates this is an excellent diagnostic test at this cut-off point.<sup>10</sup> The accuracy at this cut off point was 92.6%, which was the highest value. Once again increasing to a cut-off point of 13, as used in previous studies, the sensitivity remarkably decreased to 41.9%. Similarly, the specificity increased to 100%. However, these findings are helpful, as a mRUST score of 13, with its corresponding 100% specificity at 24 weeks follow-up, will definitively confirm fracture union.

Finally, when we used a cut-off point of 12, the sensitivity and specificity of the mRUST score was 89.6% and 100%, respectively, at final follow-up, which was an appropriate cut-off point as the AUC was 0.9805, classified as “excellent” according to commonly used classification using AUC for a diagnostic test.<sup>10</sup> The accuracy at this cut off point was 90.1%.

There are several limitations to the current study. The accurate calculation of the mRUST score requires evaluation of all four bony cortices. Evaluation of the lateral cortex in fractures treated with lateral plate constructs can be very difficult as the cortex can be obscured by the implant. This difficulty was noted by each investigator. Knowing this, it is not surprising that ICC for plate constructs showed only moderate agreement. This limitation is clearly emphasized by the fact that mRUST scores for intramedullary nail constructs consistently demonstrated higher agreement than plate constructs. This limitation is not unique to our study as previous authors evaluating nails and plates separately showed ICC scores of 0.74 and 0.59, respectively.<sup>3</sup> The authors also showed the lowest agreement dealt with the lateral cortex as full visualization is often difficult.<sup>3</sup> In addition, there were noticeable patients losing to follow-up over time, leaving smaller groups of patient at subsequent follow-up. Furthermore, this is a retrospective chart



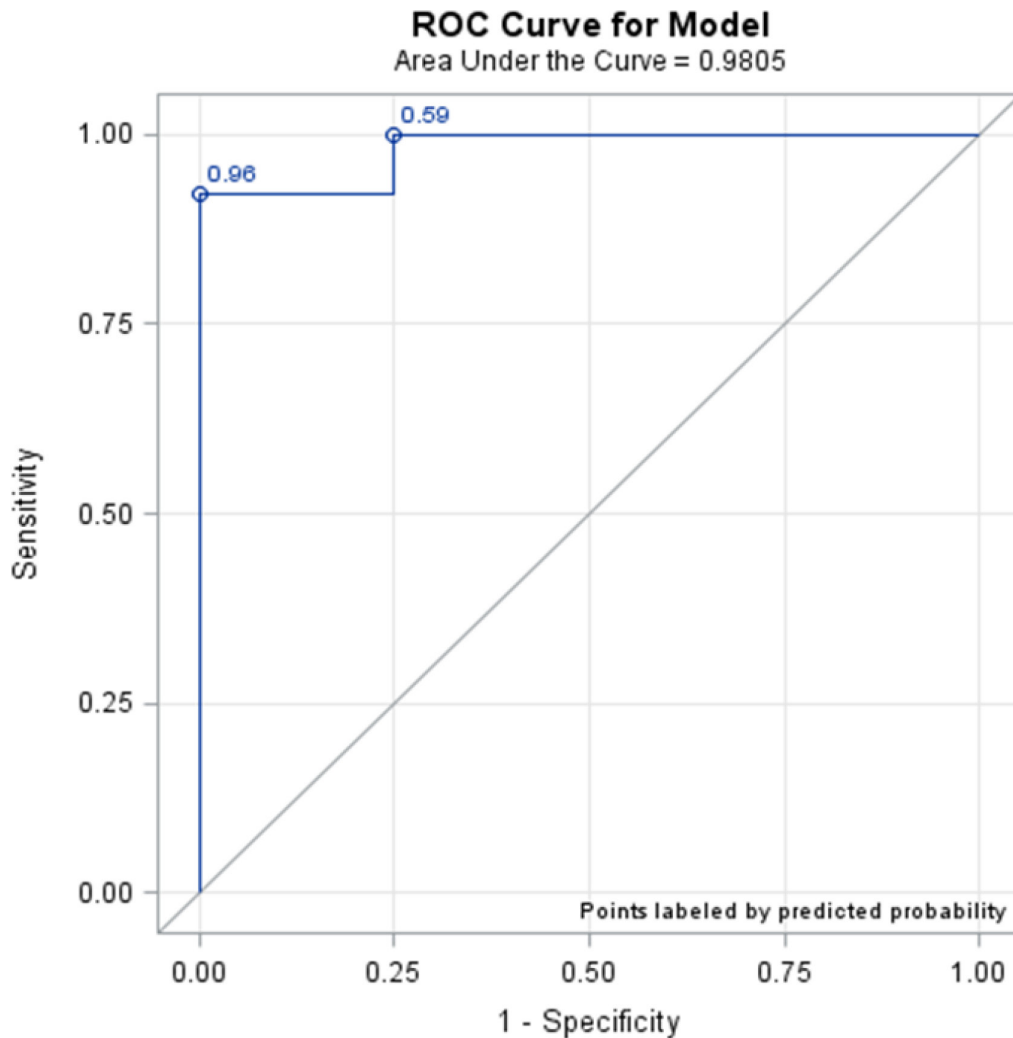


Fig. 5

Receiver operating characteristic (ROC) curve of modified Radiological Union Scale for Tibia fractures (mRUST) score at final follow-up.

review and subject to imperfections in the medical record. Most notable is the potential variation in subjective evaluation of healing between physicians at follow-up. This, however, in our opinion reflects accurately the practice occurring in routine clinic visits and therefore stands as a satisfactory bank of information. Lastly, with a low prevalence of nonunion (11.5%), the diagnosis of nonunion in the population may result in low accuracy, as accuracy is influenced by the prevalence of the disease in a selected population.

To our knowledge, this is the first study to evaluate the sensitivity, specificity, accuracy, and cut-off points of the mRUST score. Our results show good to excellent accuracy ranges in assessing metadiaphyseal femur fracture healing. Most notably, a mRUST score of ten at 24 weeks follow-up yields the highest sensitivity and specificity of 91.9% and 100%, respectively, with the highest accuracy of the mRUST score was 92.6%. For 100% specificity at 24-week follow-up, a mRUST score of at least ten must be

used to assure union; however, this significantly reduces sensitivity.

In conclusion, the AUCs ranged from good to excellent according to commonly used classifications using AUC as a diagnostic test at 12 weeks, 24 weeks, and final follow-up. Importantly, the mRUST performs poorly in the early postoperative period ( $\leq$  six weeks) and should not be used to reliably assess or predict ultimate bony union in this time period.



#### Take home message

- The modified Radiological Union Scale for Tibia (mRUST) score of ten points at 24 weeks can be used as a viable screening method with the highest sensitivity, specificity, and accuracy (91.9%, 100%, and 92.6%, respectively) for healing of metadiaphyseal femur fractures.
- The cut-off point of 13 increases the specificity to 100%, but decreases sensitivity.
- The mRUST score should not be used at six weeks, as results show an inability to accurately predict eventual fracture healing at this time point.

## References

- Bhandari M, Guyatt GH, Swiontkowski MF, Tornetta P 3rd, 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.
- Whelan DB, Bhandari M, McKee MD, et al.** Interobserver and intraobserver variation in the assessment of the healing of tibial fractures after intramedullary fixation. *J Bone Joint Surg Br.* 2002;84-B(1):15–18.
- Litrenta J, Tornetta P 3rd, 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.
- Fleiss JL.** *Statistical Methods for Rates and Proportions.* 2nd edition. New York: John Wiley & Sons. 1981.
- 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(Supplement 1):S81–S86.
- 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.
- Leow JM, Clement ND, Tawonsawatruk T, Simpson CJ, Simpson A, et al.** The radiographic union scale in tibial (rust) fractures. *Bone Joint Res.* 2016;5(4):116–121.
- Zhu W, Zeng N, Wang N.** Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *Health Care and Life Sciences NESUG.* 2010.
- One ROC curve and cutoff analysis In: Chapter 546, NCSS Statistical Software. Ncss. Com
- Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174.
- Mandrekar JN.** Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5(9):1315–1316.
- Habibzadeh F, Habibzadeh P, Yadollahie M.** On determining the most appropriate test cut-off value: The case of tests with continuous results. *Biochem Med (Zagreb).* 2016;26(3):297–307.
- Litrenta J, Tornetta P 3rd, 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.
- Cooke ME, Hussein AI, Lybrand KE, et al.** Correlation between rust assessments of fracture healing to structural and biomechanical properties. *J Orthop Res.* 2017;9999:1–9.

### Author information:

- Y. Plumarom, MD, Orthopaedic Surgeon, Department of Orthopaedics and Rehabilitation, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA; Department of Orthopaedic Surgery, Phramongkutklao Hospital and College of Medicine, Bangkok, Thailand.
- B. G. Wilkinson, MD, Orthopaedic Surgeon
- M. C. Willey, MD, Orthopaedic Surgeon
- Q. An, MPH, Statistician
- L. Marsh, MD, Orthopaedic Surgeon
- M. D. Karam, MD, Orthopaedic Surgeon  
Department of Orthopaedics and Rehabilitation, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA.

### Author contributions:

- Y. Plumarom: Data curation, Visualization, Writing – original draft.
- B. G. Wilkinson: Visualization, Writing – original draft.
- M. C. Willey: Writing – review & editing.
- Q. An: Validation.
- L. Marsh: Writing – review & editing.
- M. D. Karam: Writing – review & editing.

### Funding statement:

- No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article

### ICMJE COI statement:

- J. L. Marsh reports employment by the University of Iowa Hospitals and Clinics, grants/grants pending from DoD and Arthritis Foundation, and royalties from Biomet and Oxford Press, all of which is unrelated to this work. M. C. Willey declares grants/grants pending from the Orthopaedic Trauma Association, the US Department of Defense, and the Orthopaedic Research and Education Foundation, all of which is also unrelated.

### Open access funding:

- The authors report that they received open access funding for this manuscript from
- Department of Orthopaedic Surgery, Phramongkutklao Hospital and College of Medicine, Bangkok, Thailand.

### Ethical review statement:

- IRB ID #: 201612761

© 2021 Author(s) et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc-nd/4.0/>