

Validation of the Radiographic Union Score for HUmeral fractures (RUSHU): A retrospective study in an independent centre

William Fordyce¹ , Grace Kennedy¹, James R Allen^{1,2} ,
Mohamed Abdelmonem¹ , Jonathan Evans³ ,
Jonathan Thomas Evans^{4,5} and Paul Guyver¹

Abstract

Background: Early diagnosis and fixation of fractures unlikely to unite can prevent months of morbidity. The Radiographic Union Score for Humeral fractures (RUSHU) is a summative scoring system developed to aid identification of patients at higher risk of developing humeral shaft non-union. Plain radiographs taken six weeks after injury are given a score between four and 12 based on signs of union. Our aim was to assess the validity of the RUSHU prognostic model in an external population.

Methods: The radiographs of fifty-seven patients were scored independently according to RUSHU methodology by three reviewers (blinded to patient outcome). Interobserver intraclass correlation (ICC) was calculated.

Results: Of the cohort, six (10.5%) progressed to non-union after six months. We observed an interobserver ICC co-efficient of 0.89 (95%CI 0.84, 0.93) in RUSHU score at six weeks. Median score was significantly higher in the union cohort (10 vs 5 $p < 0.001$). Using the score of < 8 to predict non-union gave an area under the ROC curve of 0.87 (95%CI 0.83, 0.90).

Conclusions: In this retrospective single-centre study, we have demonstrated good inter-rater reliability. We would suggest that the RUSHU model be assessed in further external validation studies. RUSHU has the potential to reduce morbidity of delayed treatment of non-union.

Keywords

Humeral shaft fracture, non-union, risk stratification scoring system

Date received: 24th November 2021; revised: 22nd March 2022; accepted: 6th April 2022

Introduction

In the United Kingdom, diaphyseal humeral fractures account for 1% of adult fractures, with an incidence of between 12.9 and 14.5/100,000 per year.^{1,2} The majority of these can be managed non-operatively, with immobilization in a plaster or humeral brace.^{3,4} Of those treated non-operatively, non-union rates have been reported as high as 30%.⁵ Early diagnosis and surgical fixation of those fractures at high risk of non-union has the potential to prevent months of morbidity and loss of independence.

Fixation of diaphyseal humeral fractures is not without risks, for example iatrogenic radial nerve injury has been reported to occur between 4.2% and 6% after operative fixation.^{6,7} Evidence-based scoring systems can aid clinicians in predicting the likelihood of non-union of various fractures. These have been developed and externally validated for both tibial shaft^{8,9} and femoral neck fractures.^{10,11}

The Radiographic Union Score for Humeral fractures (RUSHU) was proposed in 2019.¹² Its aim was to identify diaphyseal humeral fractures which are likely to progress to

¹Trauma and Orthopaedics, Derriford Hospital, University Hospitals Plymouth NHS Trust, Plymouth, UK

²Trauma and Orthopaedics, Huddersfield Royal Infirmary, Huddersfield, HD3 3EA, UK of Great Britain and Northern Ireland

³Health Services and Policy Research, University of Exeter Medical School, Smeall Building St Luke's Campus, Exeter EX2 4SG, UK of Great Britain and Northern Ireland

⁴University of Bristol, Bristol, UK of Great Britain and Northern Ireland

⁵Torbay and South Devon NHS Foundation Trust, Torquay, Torbay, UK of Great Britain and Northern Ireland

Corresponding author:

William Fordyce, Trauma & Orthopaedics, Derriford Hospital, University Hospitals Plymouth NHS Trust, Plymouth, UK.

Email: william.fordyce@nhs.net

non-union using radiographs taken 6 weeks post-injury. Scores are allocated for signs of bridging callous on both the medial and lateral cortices on the anteroposterior (AP) radiograph, and for the anterior and posterior cortices on the lateral radiograph. The sum of these four scores gives a total between 4 and 12; a high total score suggests union likely, whereas a low total score suggests non-union is more likely. Within the index study if all patients with a score < 8 underwent operative intervention, the number of operations needed to avoid one non-union would be 1.5.

External validation of the scoring system is required to assess the predictive value of a test on an external population. This process allows assessment of the prediction model's reproducibility and generalisability to other patients and other populations.¹³

We aimed to assess the validity of the RUSHU prognostic model in our population.

Methods

We analysed an existing database of all patients with humeral shaft fractures managed at University Hospitals Plymouth NHS Trust between September 2005 and December 2012.

Of the 144 identified patients, a total of 87 patients were excluded (Figure 1) following the inclusion and exclusion criteria from the original article.¹² The inclusion criteria were patients with a fracture of the humeral diaphysis, aged 16 or older at the time of injury and for whom an adequate AP and lateral radiograph was available between 6 and 9 weeks. Patients who had operative intervention or who were lost to follow-up before 6 months post-injury and pathological fractures were excluded. Patients with non-operatively treated humeral shaft fractures were placed in a plaster of Paris above-elbow backslab or U-slab at presentation were then changed to a functional humeral brace in fracture clinic.

Of the remaining cohort of 57 patients, information was obtained from hospital records to determine patient demographics, mechanism of injury, fracture pattern classification (using the AO classification system) and fracture location within the bone (proximal, middle or distal third).

Patients within the final cohort were categorised into two groups based on their outcome: union (the fracture radiologically or clinically united by six months), or non-union (the fracture had neither clinically or radiologically united by six months). Three reviewers, all orthopaedic registrars (MA, postgraduate year 8, JA postgraduate year 6 and GK postgraduate year 5), who were blinded to patient outcome, were shown the radiographs obtained six weeks after injury and independently gave scores for the four cortices. These scores were allocated according to the methodology of the RUSHU article - (1 suggesting no callus, 2 suggesting some callus, 3 suggesting bridging callus).

Figure 2 demonstrates some examples of how different radiographs would be scored.

Reviewer scores were entered into the database by the primary data collector (WF). The spreadsheet was subsequently anonymised to both reviewer identity and patient identity.

Statistical analysis

Statistical analysis was performed by an independent reviewer. Analysis replicated the original study and the methodology was corroborated following discussion with the RUSHU statistician.

Interobserver and intraobserver reliability was assessed using intraclass correlation coefficients (ICC) using the mixed effects model in Stata 15 (Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC). When conducting a reliability study using ICC, researchers must try to obtain at least 30 heterogeneous samples and involve at least 3 raters.¹⁴ We have fulfilled these conditions and hence, upon interpreting the ICC, values less than 0.5 indicate poor reliability, 0.5–0.75 indicates moderate reliability, 0.75–0.9 is indicative of good reliability, and values greater than 0.9 indicate excellent reliability.¹⁴

Using a K-sample equality-of-medians test, we assessed the null hypothesis that there was no difference between the median scores for those that went on to union compared with non-union.

A receiver operator characteristic curve (ROC) was drawn, with the data dichotomised using the groups of the original study; RUSHU score ≥ 8 and < 8 for predicting a high risk of non-union.

An area under the curve (AUC) was calculated. An AUC of 0.5 suggests no discrimination (between union and non-union based on score), 0.7–0.8 was considered acceptable, 0.8–0.9 was considered excellent and > 0.9 was considered outstanding.¹⁵ The Positive Predictive Value and Negative Predictive Value were calculated using the two-by-two contingency table method.

Results

Patient demographics corresponding to the 57 patients in the final cohort are presented in Table 1. Of this cohort, 51 patients (89.5%) went on to union, and six patients (10.5%) went on to non-union.

Of the six patients with fractures which progressed to non-union after six months, all were scored less than 8 by all three reviewers. Of the 51 patients that went on to union, 18 (35.3%) were scored less than 8 by at least one reviewer. The distribution of RUSHU scores for the overall cohort is demonstrated in Figure 3.

The intraobserver ICC was 0.96 (95% confidence interval (CI) 0.94, 0.98) and interobserver ICC was 0.89

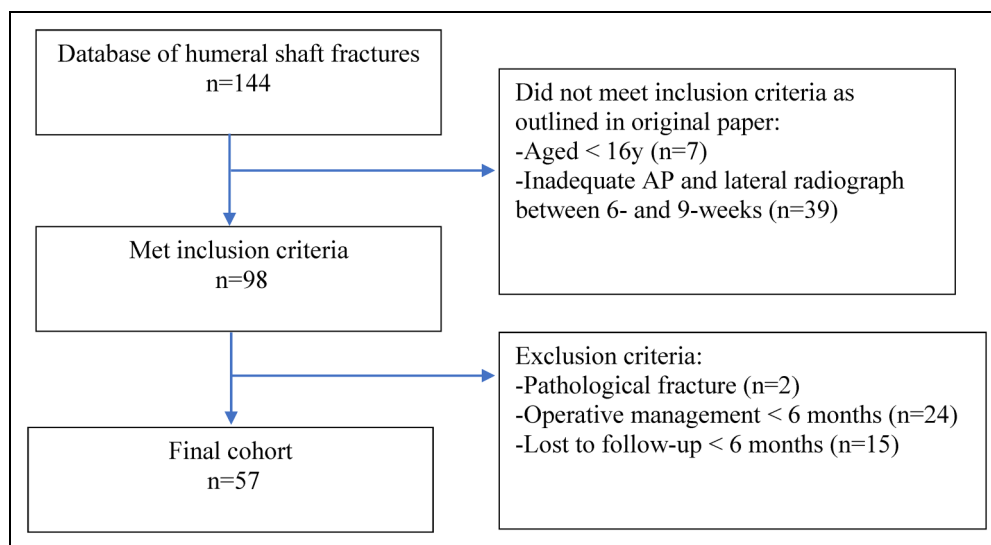


Figure 1. Flowchart of inclusion and exclusion criteria utilised in the present study.

(95% CI 0.84, 0.93), demonstrating excellent reliability of scores by each reviewer, and good agreement between reviewers.¹⁴

The median reviewer score of those fractures progressing union was 10 (interquartile range (IQR) 7-12), and of those fractures that progressed to non-union the median was 5 (IQR 4-6). There was strong evidence against the null hypothesis that there was no difference between the median scores for those that went on to union compared with non-union ($p < 0.001$).

A ROC curve was produced for a score of < 8 predicting non-union, which demonstrated excellent discrimination with an AUC = 0.87 (95% CI 0.83, 0.90) (Figure 4). Sensitivity analyses were performed using cut offs of 7 and 9 to assess whether the cut off of 8 was the best for our study population. The area under the ROC curve was highest using a cut off of 8, supporting the choice of cut off by the original authors (See Supplementary Material).

A score of < 8 had a positive predictive value (PPV) of 31% (95% CI 19.5, 44.5) of progressing to non-union (Table 2). In the index study, the table presented used scores for reviewer one only, this is provided for comparison (Table 3). A score of ≥ 8 had a negative predictive value (NPV) of 100% (95% CI 96.8, 100). Based on this PPV of 31%, three operations would be required to prevent one established non-union if all RUSHU < 8 were offered fixation at six weeks.

Discussion

RUSHU score was introduced by Oliver et al.¹² as a modality for predicting humeral shaft fracture non-union. The aims of this study were to assess the validity of the RUSHU prognostic model in an external population and

to assess reliability of the RUSHU scoring methodology using reviewers independent to the original study team. Similar to the original study, we demonstrated that a score of < 8 was significantly predictive of non-union but found higher inter- and intra-observer reliability (0.98 and 0.96, respectively) than that found in the original study (0.79 and 0.91, respectively). In concordance with the recognized prognostic features of non-union, within this study cohort we found a greater proportion within the subgroups of male gender, proximal third fractures and high energy injuries.

Our findings corroborate the assertion that RUSHU appears to be reliable at early identification of humeral shaft fractures which are likely to progress to non-union. However, based on the PPV of 31%, three operations would be required to prevent one established non-union if all RUSHU < 8 were offered fixation at six weeks, a much higher value than that found in the original study where 1.5 operations would be required to prevent one established non-union. RUSHU < 8 within our cohort would have resulted in twice as many operations to prevent one non-union. This substantial difference in PPV would have a large impact on any cost-analysis regarding the benefit of offering operative intervention based on the RUSHU score < 8 . As well as operating on a proportion of patients that would have united, there would also be patients that would have developed asymptomatic non-unions receiving unnecessary operations. In our study population, if a lower cut off score of 7 is used, one patient who went on to non-union would not have been identified. We therefore support the original score of 8 selected by Oliver et al.

A lower overall proportion of our cohort did not progress to union in comparison to the reference study. This may be

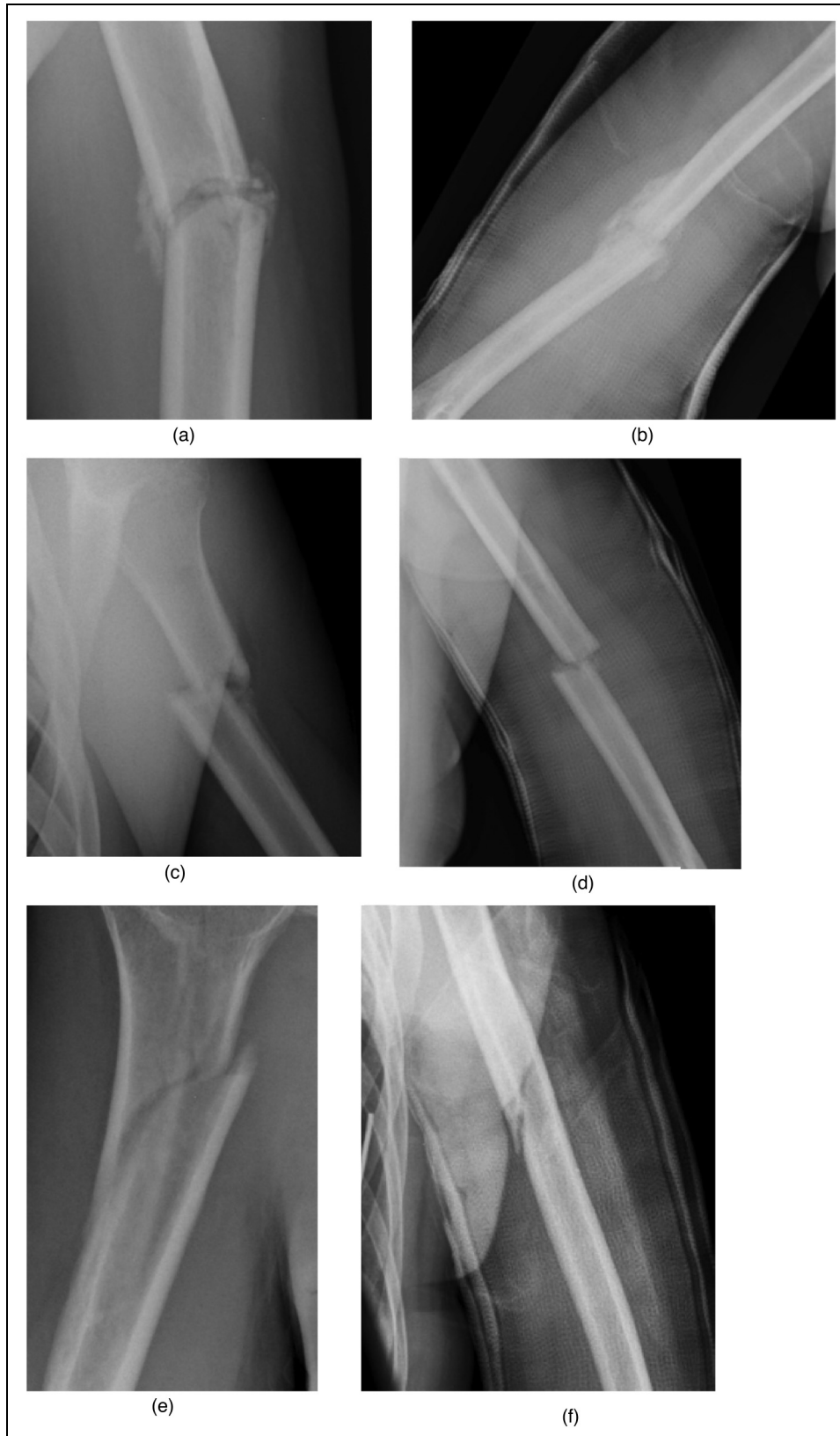


Figure 2. shows examples of radiographs that would be attributed different RUSHU scores. A score of 1 for no evidence of callus, 2 for some callus and 3 for bridging callus is attributed to each cortex. 2a & b suggest bridging callus, each cortex scores 3. 2c & d show some callus, but not obviously bridging, each cortex scores 2. 2e & f do not show convincing evidence of callus, each cortex scores 1.

Table 1. Demographics of the cohort.

		Total cohort (57)	Union (51)	Non-union (6)	Original RUSHU population (60)
Sex, n (%)	Male	24 (42.1)	20 (39.2)	4 (66.7)	38 (63.3)
	Female	33 (57.9)	31 (60.8)	2 (33.3)	22 (36.7)
Age, mean (range, SD)		55.7 (16 to 94, 21.7)	55.6 (16 to 94, 22.5)	56.5 (42 to 79, 14.5)	Not reported (18 to 97)
Mechanism, n (% of known)	Fall < 2m	33 (75.0)	32 (80.0)	1 (25.0)	39 (67.2)
	Fall > 2m	7 (15.9)	5 (12.5)	2 (50.0)	6 (10.3)
	Sport	4 (9.1)	3 (7.5)	1 (25.0)	2 (3.4)
	Other				11 (19.0)
	Unknown	13	11	2	2
Side, n (%)	Right	22 (38.6)	20 (39.2)	4 (66.7)	26 (43.3)
	Left	35 (61.4)	31 (60.8)	2 (33.3)	34 (56.7)
Location, n (%)	Proximal	23 (40.4)	19 (37.3)	4 (66.7)	15 (25.0)
	Middle	26 (45.6)	24 (47.1)	2 (33.3)	35 (58.3)
	Distal	8 (14.0)	8 (15.7)	0	10 (16.7)
AO Classification, n (%)	A1	21 (36.8)	18 (35.3)	3 (50.0)	26 (43.3)
	A2	5 (8.8)	5 (9.8)	0	7 (11.7)
	A3	15 (26.3)	15 (29.4)	0	7 (11.7)
	B2	13 (22.8)	11 (21.6)	2 (33.3)	17 (28.3)
	B3	3 (5.3)	2 (3.9)	1 (16.7)	3 (5.0)

because of a lower proportion of our cohort that progressed onto non-union. In the initial study they selected their population based on outcome, giving them a predetermined proportion of 33% non-unions. We included all patients that fit the inclusion criteria, this difference resulted in a lower proportion (10%) of the population who had non-unions. This may have contributed to our lower PPV. This certainly demonstrates the need for further validation regarding this scoring system before it could be used with confidence in clinical practice.

Our NPV of 100% is a result of all non-unions having a score of less than eight. External validation of the scoring system is required to assess the reproducibility and generalisability of a test on an external population.¹³

In this study, we observed excellent inter-reliability. This suggests that the tool is easy to understand and is reliable between users. In fact, our results for reliability were higher than in the original study, showing that the methodology is easy to follow. This tool is convenient as can be performed

anywhere likely to be managing this injury. It does not require any additional investigations above those that would be done as part of routine practice, so is unlikely to increase cost.

Our study has a number of limitations. We had a relatively small cohort, which was collected retrospectively from a departmental database of injuries that occurred between eight and fifteen years before this retrospective review. As a result of this, we were unable to identify the mechanism of injury for some patients due to this time lapse between injury and data collection. Also, we were unable to collect data on other important predictors of non-union, such as smoking, comorbidities and hand dominance given the retrospective nature of the study. Many patients were excluded due to lack of adequate radiographs from 6–8 weeks post injury, which removed a significant proportion of the cohort.

With all of these considerations, it is likely our cohort is not representative of all humeral shaft fractures and patient demographics. Expanding the inclusion criteria in a

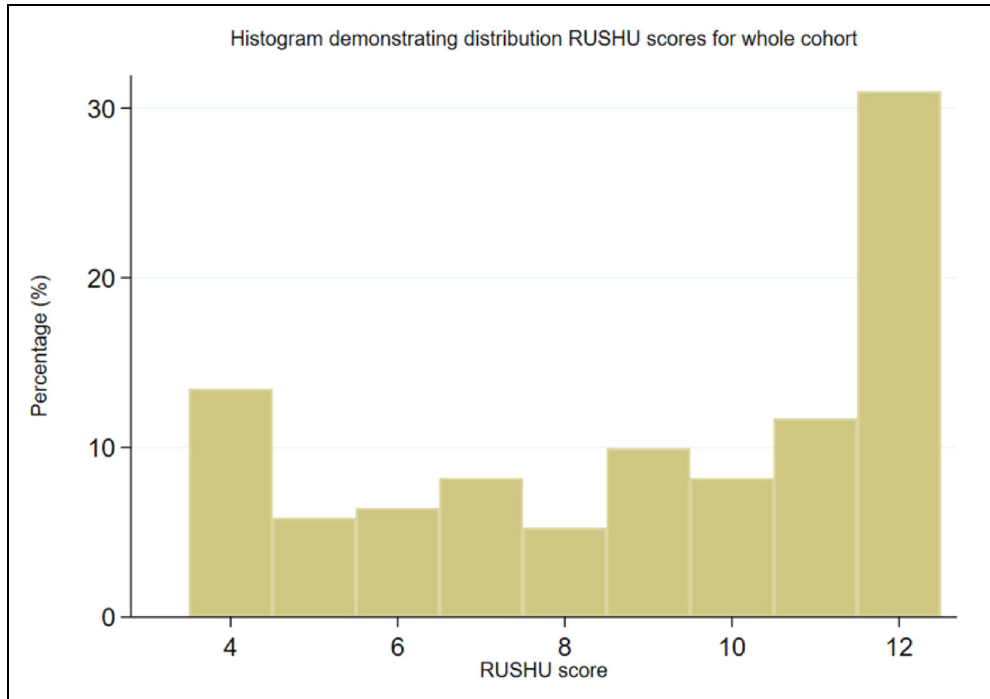


Figure 3. Distribution of RUSHU score for overall cohort.

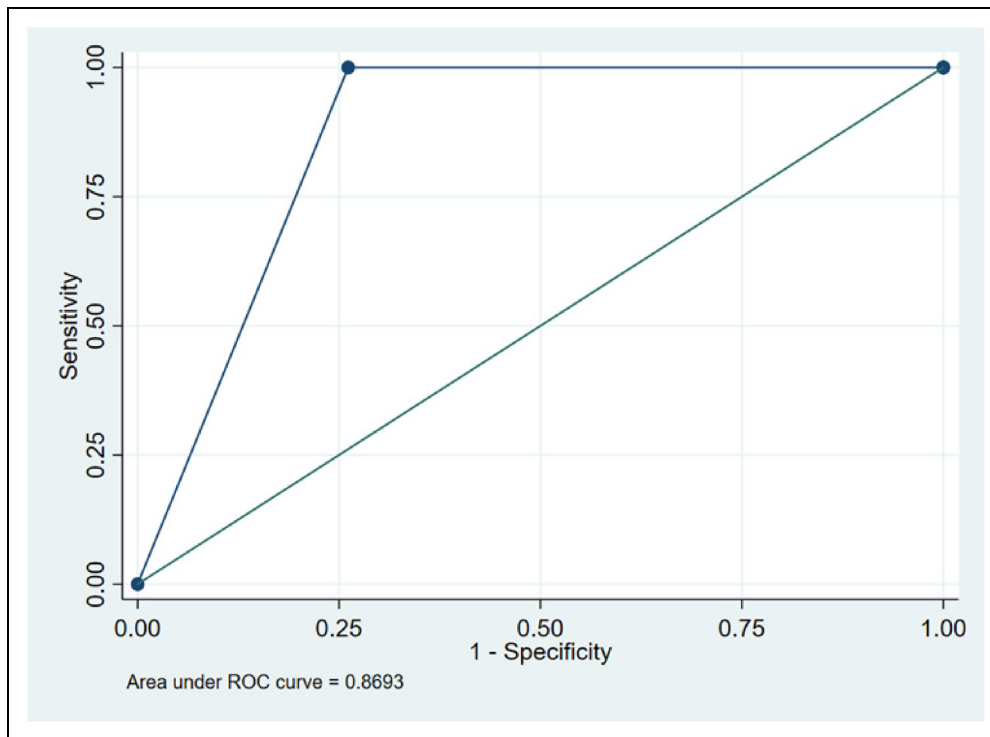


Figure 4. Receiver operating characteristic curve for a score of < 8 predicting non-union.

Table 2. Clinical relevance of Radiographic Union Score for HUmeral fractures (RUSHU) < 8 as a predictor of non-union (using all reviewers' scores).

	Union	Non-union	Total
RUSHU ≥ 8	113	0	113
RUSHU < 8	40	18	58
Total	153	18	171

Chi squared test $p < 0.001$.

Sensitivity 100% (95%CI 81.5, 100), Specificity 73.9% (95%CI 66.1, 80.6), PPV 31.0% (95%CI 19.5, 44.5), NPV 100% (95%CI 96.8, 100), ROC AUC 0.87 (95%CI 0.83, 0.90).

Table 3. Clinical relevance of Radiographic Union Score for HUmeral fractures (RUSHU) < 8 as a predictor of non-union (using first reviewer's scores only).

	Union	Non-union	Total
RUSHU ≥ 8	40	0	40
RUSHU < 8	11	6	17
Total	51	6	57

Chi squared test $p < 0.001$.

Sensitivity 100% (95%CI 54.1, 100), Specificity 78.4% (95%CI 64.7, 88.7), PPV 35.3% (95%CI 14.2, 61.7), NPV 100% (95%CI 91.2, 100), ROC AUC 0.89 (95%CI 0.84, 0.95).

prospective study of all sequential humeral shaft fractures would help address this issue.

Conclusion

In this retrospective single-centre study, we have demonstrated good inter-rater reliability. We would suggest that the RUSHU model be assessed in further external validation studies on different socioeconomic and ethnic populations. We do however agree with the statement of the initial article, that RUSHU could potentially reduce morbidity of delayed treatment of non-union.

Further validation of this tool should be trialed, using different socioeconomic and ethnic populations, in independent individual centres. Another area of additional research would be secondary analysis of RCT data or using a multi-centre collaborative study of sequential patients to provide a more diverse cohort representative of patients who have conservatively managed humeral shaft fractures.

Take home message

- RUSHU is a scoring system for predicting humeral shaft fractures that are unlikely to unite

- We have shown that in our population, this has potential to be a valuable scoring system
- More external validation studies would be valuable to explain some differences between our statistical results and those presented in the initial study.


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

William Fordyce  <https://orcid.org/0000-0002-6837-3589>

James R Allen  <https://orcid.org/0000-0003-3781-1133>

Mohamed Abdelmonem  <https://orcid.org/0000-0002-3627-6434>

Jonathan Evans  <https://orcid.org/0000-0003-0090-1012>

Supplemental Material

Supplemental material for this article is available online.

References

1. Court-Brown CM and Caesar B. Epidemiology of adult fractures: a review. *Injury* 2006; 37: 691–697.
2. Ekholm R, Adami J, Tidermark J, et al. Fractures of the shaft of the humerus: an epidemiological study of 401 fractures. *Bone Joint J* 2019; 101–B: 1300–1306.
3. Sarmiento A, Zagorski JB, Zych GA, et al. Functional bracing for the treatment of fractures of the humeral diaphysis. *J Bone Joint Surg Am* 2000; 82: 478–486.
4. Rutgers M and Ring D. Treatment of diaphyseal fractures of the humerus using a functional brace. *J Orthop Trauma* 2006; 20: 597–601.
5. Harkin E and Large JE. Humeral shaft fractures: union outcomes in a large cohort. *J Shoulder Elbow Surg* 2017; 26: 1881–1888.
6. Wang J-P, Shen W-J, Chen W-M, et al. Iatrogenic radial nerve palsy after operative management of humeral shaft fractures. *J Trauma* 2009; 66: 800–803.
7. Schwab T, Stillhard P, Schibli S, et al. Radial nerve palsy in humeral shaft fractures with internal fixation: analysis of management and outcome. *Eur J Trauma Emerg Surg* 2018; 44: 235–243.
8. Leow JM, Clement ND, Tawonsawatruk T, et al. The radiographic union scale in tibial (RUST) fractures: reliability of outcome measure at an independent centre. *Bone Joint Res* 2016; 5: 116–121.
9. Ross KA, O'Halloran K, Castillo RC, et al. Prediction of tibial non-union at the 6-week time point. *Injury* 2018; 49: 2075–2082.
10. Frank T, Osterhoff G, Sprague S, et al. FAITH investigators. The Radiographic Union Score for Hip (RUSH) identifies

- radiographic non-union of femoral neck fractures. *Clin Orthop Relat Res* 2016; 474: 1396–1404.
11. Chiavaras MM, Bains S, Choudur H, et al. The Radiographic Union Score for Hip (RUSH): the use of a checklist to evaluate hip fracture healing improves agreement between radiologists and orthopedic surgeons. *Skeletal Radiol* 2013; 42: 1079–1088.
 12. Oliver WM, Smith TJ, Nicholson JA, et al. The Radiographic Union Score for HUmeral fractures (RUSHU) predicts humeral shaft nonunion. *Bone Joint J* 2019; 101–B: 1300–1306.
 13. Ramspek CL, Jager KJ, Dekker FW, et al. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2020; 14: 49–58.
 14. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155–163.
 15. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol* 2010; 5: 1315–1316.