Improving agreement in assessing subtrochanteric fracture healing among orthopedic surgeons using the Radiographic Union Score for Hip (RUSH)

Tian Jian Zhou[1,2]*, Song Jiang[1], Jin Ke Ren[1], Xuan Zhang[1], Wang Xing Liu[1], Peng Yan[1], Jian Wang Li[1], Tong Zeng[1] and Zhong Shi Xu[1]*

Abstract

Background　Prompt identification of fracture healing is crucial for medical decision-making and clinical research. This study aims to assess the effectiveness of the Radiographic Union Score for Hip (RUSH) in subtrochanteric fractures and determine the optimal scoring threshold for fracture healing.

Methods　We obtained 94 sets of X-ray films from the postoperative follow-up of 35 patients who underwent intramedullary nail fixation for subtrochanteric fractures. Six orthopedic trauma surgeons evaluated the imaging data and determined fracture healing based on subjective judgment. They then scored the X-ray images using the RUSH form. After four weeks, the same observers re-evaluated 47 randomly selected sets of radiographs. Subsequently, the observers and study designer conducted a consensus meeting where they agreed on whether the fractures had healed. We used Fleiss kappa to evaluate agreement among observers based on subjective impressions, and the intraclass correlation coefficient assessed RUSH score consistency. Regression analysis examined the relationship between scores and fracture healing consensus using a receiver operator characteristic (ROC) curve.

Results　Based on overall impressions, the agreement level among orthopedic trauma surgeons for determining the healing status of subtrochanteric fractures was moderate (Kappa = 0.564, 95% CI: 0.511–0.616). However, utilizing the RUSH scoring system can improve the consistency to a substantial level of agreement (ICC = 0.748, 95% CI: 0.663–0.817). Regarding intraobserver reliability, there were significant differences among orthopedic trauma surgeons in judging fracture healing based on overall impressions. However, using the RUSH scoring form, the attending group achieved better intraobserver consistency than the resident group. Regression analysis and ROC curve analysis revealed a strong correlation between the RUSH total score, cortical bone score, cancellous bone score, and consensus on fracture healing. The AUC values are 0.769 (95% CI: 0.670–0.868), 0.779 (95% CI: 0.681–0.876), and 0.771 (95% CI: 0.674–0.867), respectively.

*Correspondence:
Tian Jian Zhou
zhou.tianjian@szhospital.com
Zhong Shi Xu
xuzhongshi@mail.sustech.edu.cn
Full list of author information is available at the end of the article

Conclusions Our study highlights the effectiveness of the RUSH scoring system in enhancing interobserver agreement and intraobserver consistency when evaluating subtrochanteric fracture healing in orthopedic trauma surgeons. We propose setting 21 points as the threshold for the simplified RUSH scoring system to determine fracture healing. This system includes only eight items related to cortical bone, with a total score of 24 points.

Keywords Hip fracture, Intramedullary nail, Fracture healing, Radiological assessment, Reliability

Background

Compared to femoral neck and intertrochanteric fractures, subtrochanteric fractures are particularly susceptible to mechanical complications, including delayed union, nonunion, and hardware failure [1, 2]. These complications lead to more severe functional impairments and an increased disease burden. Therefore, assessing fracture healing is essential for orthopedic trauma surgeons, as it directly influences critical management decisions [3].

The clinical definition of fracture healing typically relies on various factors, including physical examination findings, imaging studies, and patient self-assessment [3]. Plain radiographs remain the primary method for assessing fracture healing, but there is variation in the consistency of different observers [4]. In 2010, Whelan et al. introduced the Radiological Healing Score (RUST) system for evaluating tibial fractures. Among the various criteria assessed, cortical bridging emerged as the most reliable indicator of fracture healing [5]. Bhandari et al. developed the Radiographic Union Score for Hip fractures (RUSH) to evaluate the healing of femoral neck fractures. This system allows for a qualitative assessment of crucial healing parameters, including callus formation, cortical bridging, fracture line resolution, and trabecular integration at the fracture site [6].

Subtrochanteric fractures primarily affect the cortical bone, with the fracture line extending into the cancellous bone within the trochanteric region. Therefore, using the RUSH score is highly suitable for determining the healing status of subtrochanteric fractures. In a study evaluating subtrochanteric femoral fracture healing with the Radiographic Union Score (RUS), an RUS of ≥7 at 12 weeks predicted clinical union with 88% sensitivity and 63% specificity [7]. Due to the cutoff value of 7 used in this study's RUS system, it can be inferred that the RUSH scoring system was not employed. To our knowledge, there have been no reports on using this scoring system to assess the healing of subtrochanteric fractures.

Methods

Study design

The study designer (TJZ)selected 94 sets of anteroposterior and lateral X-ray films from the medical records database. These films were part of the postoperative follow-ups for 38 patients undergoing intramedullary nail fixation for subtrochanteric fractures. Six observers, including three orthopedic attendings and three orthopedic residents, independently evaluated the imaging data and determined fracture healing based on subjective judgment. They then scored the X-ray images using the Radiographic Union Score for Hip (RUSH) form. After four weeks, the same observers re-evaluated 47 sets of radiographs selected and rerandomized by the research designer from the original 94 sets. Finally, a consensus meeting was held involving all observers and the study designer, during which they agreed on the healing status of the fractures. The flowchart of the study design is shown in Fig. 1.

Participants

Our review panel comprised three orthopedic residents who had completed residency training (with an average orthopedic clinical practice experience of 3.3 years, ranging from 3 to 4 years) and three orthopedic attending physicians with expertise in the treatment of hip fractures (with an average experience of 10.0 years, ranging from 9 to 12 years). We recruited observers from two distinct seniority groups to investigate potential variations in scoring based on work experience and to evaluate the reliability of the RUSH scoring system.

Imaging criteria

The ethics committee of our hospital approved the research protocol. We thoroughly reviewed the medical records database from January 1, 2014, to December 31, 2021. In this period, we identified 443 patients diagnosed with "intertrochanteric fracture of the femur," "fracture of the greater trochanter," or "subtrochanteric fracture," all of whom had undergone internal fixation surgery.

For inclusion in this study, the following criteria were strictly enforced: (1) the fracture must have been a subtrochanteric type treated with intramedullary nail fixation, (2) X-ray imaging must have been performed no earlier than three weeks postoperatively, and (3) there must be at least a 30-day interval between successive X-ray examinations for the same patient. The exclusion criteria were equally precise: (1) fractures other than subtrochanteric were excluded, (2) fixation was achieved using a plate or dynamic hip screw was not considered, and (3) incomplete preoperative or postoperative follow-up imaging data led to exclusion.

Ultimately, the study comprised 38 cases and 94 follow-up imaging datasets meeting the criteria. Each dataset
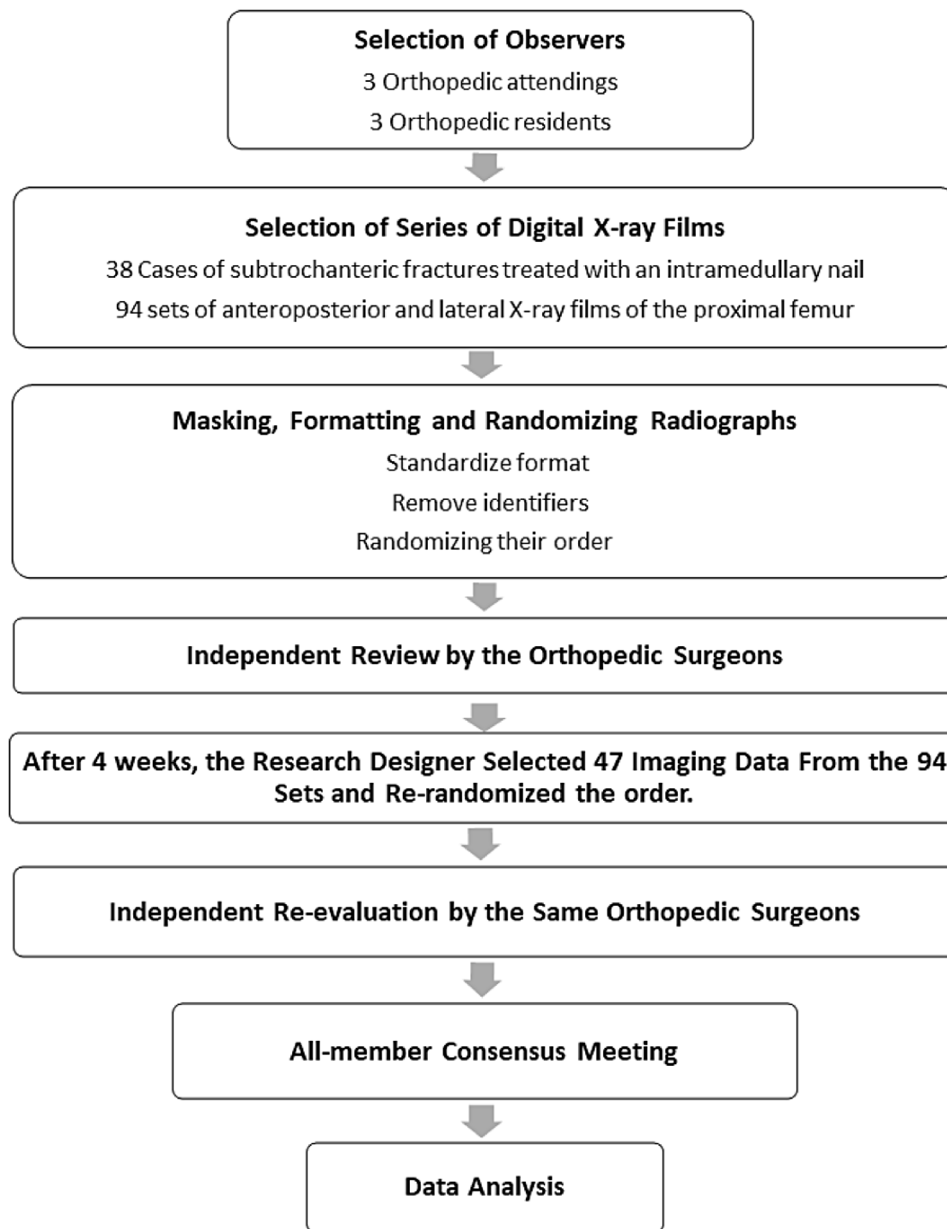
Selection of Observers

3 Orthopedic attendings

3 Orthopedic residents

Selection of Series of Digital X-ray Films

38 Cases of subtrochanteric fractures treated with an intramedullary nail

94 sets of anteroposterior and lateral X-ray films of the proximal femur

Masking, Formatting and Randomizing Radiographs

Standardize format

Remove identifiers

Randomizing their order

Independent Review by the Orthopedic Surgeons

After 4 weeks, the Research Designer Selected 47 Imaging Data From the 94 Sets and Re-randomized the order.

Independent Re-evaluation by the Same Orthopedic Surgeons

All-member Consensus Meeting

Data Analysis

Fig. 1 Flowchart illustrating the methodology employed in our study

included anteroposterior and lateral X-rays of the proximal femur on the affected side. The imaging data reflected various recovery stages, with 18.0% of images taken between 3 and 6 weeks post-surgery, 22.3% between 7 and 12 weeks, 22.3% between 13 and 24 weeks, 17.0% between 25 and 52 weeks, and 20.2% beyond 52 weeks. All personal identifiers, such as names, gender, hospital admission numbers, and timestamps, were anonymized. The reviewing physicians were blinded to the follow-up times; no markers identified them, such as suture pins. Importantly, reviewers were not involved in the selection of cases or the processing of imaging data.

Training and evaluation process

Our pre-assessment training involved studying three critical reference articles on the RUSH scoring system, including explanations and images. Chiavaras et al. explained the indicators and scoring criteria of the RUSH scoring system in their literature. The RUSH score incorporates four component scores: cortical bridging, cortical fracture line disappearance, cancellous bone calcification, and cancellous bone fracture line disappearance, and the total RUSH score ranges from a minimum of 10 to a maximum of 30 [8]. Additionally, we selected X-rays from 10 cases of intertrochanteric fractures for independent scoring. We discussed cases with significant scoring

discrepancies to ensure all evaluators reached a consensus on the scoring.

The study designer (TJZ) uploaded 94 sets of digital radiographs to a secure, password-protected platform for online review. Reviewers trained in using the system had 14 days to independently assess each radiograph for fracture healing, classifying each as "healed" or "not healed" based on their subjective judgment. Although fracture healing is gradual and not strictly binary, only fully healed fractures by orthopedic trauma surgeons were classified as "healed." Reviewers also completed the RUSH checklist, evaluating cortex and trabecular fracture lines, cortex bridging, and trabecular consolidation. After four weeks, 47 original images were rerandomized and re-uploaded for an additional two-week review period. Figure 2 displays X-ray images taken during two follow-up periods of the same patient and their corresponding RUSH scores.

The results from both assessment rounds were summarized, and a consensus meeting was held with the study designer and all reviewing surgeons. During this meeting, we discussed the timing of follow-up and serial radiographs for each case. We evaluated whether the X-ray images met the criteria for fracture healing, using the Sanders Traumatic Hip Rating Scale scores from patient follow-ups as a reference. Although clinical function scores are crucial for a comprehensive assessment of fracture healing, they were not the focus of this study. Ultimately, consensus on fracture healing was reached based exclusively on the provided imaging data for each case.

Data analysis

The formula $[2 \times$ (number of reviewers)$^2]$ is usually used for sample size calculation in consistency studies [9]. We had six observers; the sample size requirement was 72 cases (2×6^2). According to the above sample size requirements, the sample size of 94 sets of image data is sufficient to achieve good statistical accuracy. We used the Fleiss-kappa (κ) coefficient to evaluate the consistency of multiple reviewers' subjective judgments on fracture healing. As the RUSH score is a continuous variable ranging from 10 to 30, the intraclass correlation coefficient (ICC) was employed to assess the consistency among observers. Both kappa (κ) and ICC values range from −1.00 (indicating absolute disagreement) to 1.00 (indicating total agreement). According to the guidelines proposed by Landis and Koch, kappa values between 0.00 and 0.20 are categorized as poor agreement, 0.21 to 0.40 as fair agreement, 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as substantial agreement, and 0.81 to 1.00 as almost perfect agreement [10]. Finally, we conducted regression analysis and ROC analysis to explore the possibility of optimizing the RUSH scoring criteria,

specifically the overall score, the overall score for cortical bone items, and the overall score for cancellous bone items.

Result

Overall, the agreement among orthopedic doctors in their subjective assessments of fracture healing was moderate, with a kappa value of 0.564 (95% CI: 0.511–0.616). Subgroup analysis revealed that resident doctors had slightly higher consistency (kappa=0.658, 95% CI: 0.542–0.775) than senior attending doctors (kappa=0.486, 95% CI: 0.370–0.603). However, after a four-week interval, the overall consistency in fracture healing judgments decreased to a fair level (kappa=0.324, 95% CI: 0.250–0.398). Among senior attending doctors, consistency remained higher (kappa=0.450, 95% CI: 0.294–0.624) than resident doctors (kappa=0.133, 95% CI: 0.032–0.298). There was considerable variation in intraobserver reliability, with one observer showing almost perfect agreement, one demonstrating high consistency, three exhibiting moderate consistency, and one showing poor consistency. Table 1 details the specific kappa values for intraobserver reliability among the observers.

Regarding the subjective assessment and consensus on fracture healing, the initial analysis demonstrated a moderate correlation between each observer's subjective evaluation and the final consensus, as indicated by the Kendall tau-b correlation coefficient. The correlation coefficients for each reviewer are presented in detail in Table 2. In the second evaluation, the Kendall tau-b correlation coefficients between the observers' subjective assessments of fracture healing and the final consensus on union demonstrated significant variability, ranging from low to high correlation. These specific correlation coefficients are detailed in Table 3.

In the assessment of interobserver and intraobserver agreement using the RUSH score, among the 94 cases of imaging data in the initial evaluation, there was a high level of agreement among orthopedic trauma surgeons regarding the total RUSH score (ICC=0.748, 95% CI: 0.663–0.817). The agreement was higher when considering the cortical-only total score, which includes cortical bridging and the cortical fracture line (ICC=0.787, 95% CI: 0.706–0.849). The cancellous-only total score, reflecting trabecular calcification and the cancellous fracture line, had a slightly lower agreement (ICC=0.668, 95% CI: 0.589–0.744). Group analysis showed that senior attendings had higher consistency in RUSH total scores than residents (senior attendings: ICC=0.780, 95% CI: 0.643–0.862; residents: ICC=0.674, 95% CI: 0.527–0.779).

After four weeks, the orthopedic surgeon re-evaluated the RUSH scores for the 47 imaging data sets. The interobserver agreement for the total RUSH score decreased (ICC=0.574, 95% CI: 0.397–0.721). Despite

Zhou *et al. BMC Musculoskeletal Disorders* (2024) 25:798

Page 5 of 9

Table C:

Cortex Index	Location	Score
Bridging (No Cortical Bridging, Score=1; Some Cortical Bridging, Score=2; Complete Cortical Bridging, Score=3)	Medial	2
	Lateral	2
	Anterior	2
	Posterior	1
Fracture Line (Fully Visible, Score=1; Some Evidence, Score=2; No Evidence, Score=3)	Medial	2
	Lateral	2
	Anterior	1
	Posterior	1
Trabecular Consolidation (No Consolidation, Score=1; Some Consolidation, Score=2; Complete Consolidation, Score=3)		2
Trabecular Fracture Line (Fully Visible, Score=1; Some Evidence, Score=2; No Evidence, Score=3)		2
Total RUSH Score		17

Table F:

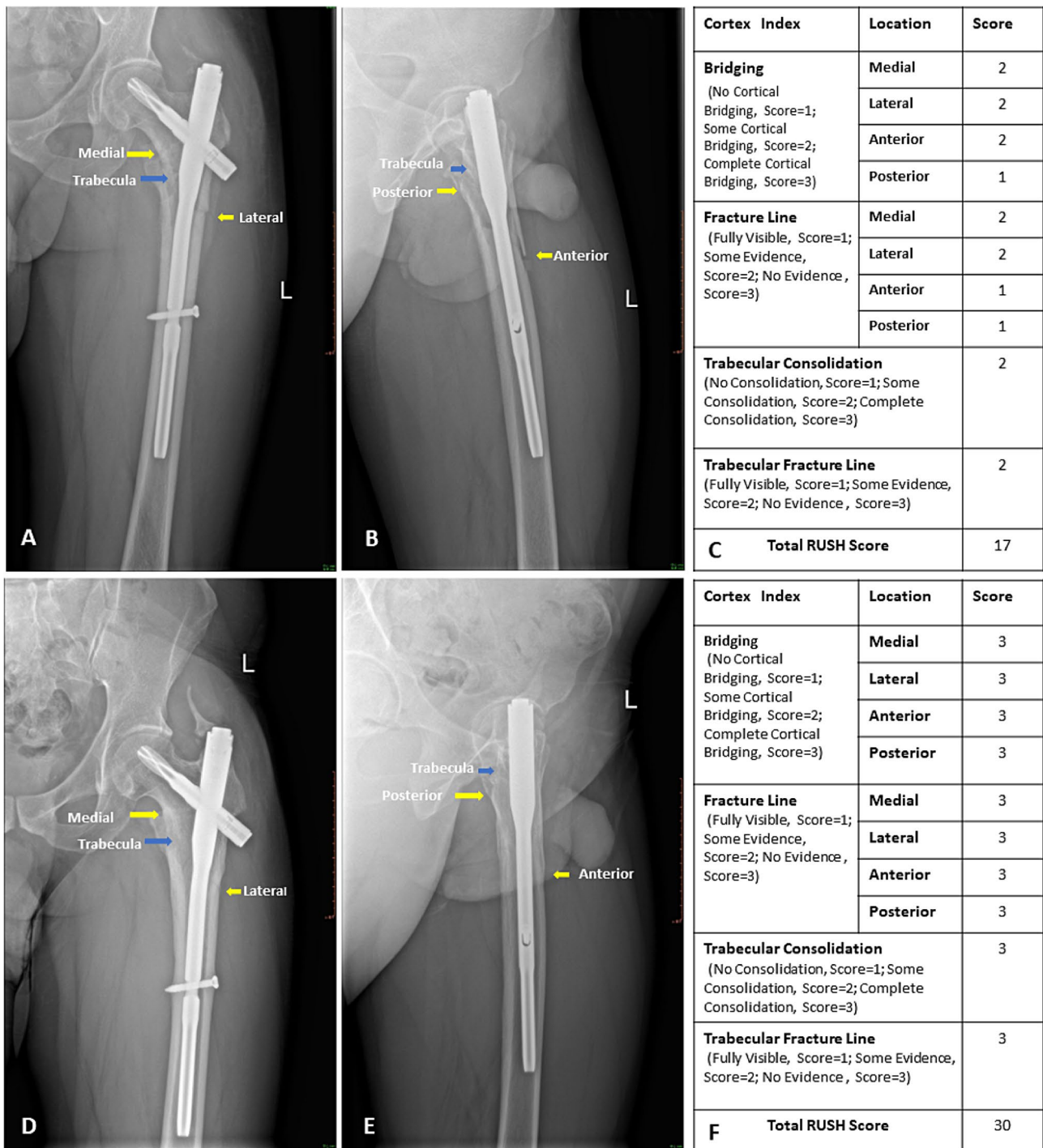| Cortex Index	Location	Score
Bridging (No Cortical Bridging, Score=1; Some Cortical Bridging, Score=2; Complete Cortical Bridging, Score=3)	Medial	3
	Lateral	3
	Anterior	3
	Posterior	3
Fracture Line (Fully Visible, Score=1; Some Evidence, Score=2; No Evidence, Score=3)	Medial	3
	Lateral	3
	Anterior	3
	Posterior	3
Trabecular Consolidation (No Consolidation, Score=1; Some Consolidation, Score=2; Complete Consolidation, Score=3)		3
Trabecular Fracture Line (Fully Visible, Score=1; Some Evidence, Score=2; No Evidence, Score=3)		3
Total RUSH Score		30

Fig. 2 X-ray images with their corresponding RUSH scores at different periods of the same patient. Figures **A** and **B** depict early postoperative X-rays of a patient, while Figure **C** represents the RUSH scores for **A** and **B**. Figures **D** and **E** show X-rays from later follow-up stages of the same case, with Figure **F** displaying the RUSH scores for **D** and **E**

this decline, the senior attending physician group maintained higher consistency (ICC=0.698, 95% CI: 0.550–0.810) than the resident physician group (ICC=0.554, 95% CI: 0.246–0.748).

Statistical analysis of intraobserver consistency revealed that, after four weeks, two observers from the senior attending group demonstrated nearly perfect agreement, with ICC values of 0.859 and 0.825, while one observer showed a high degree of consistency

Table 1 Intraobserver reliability of fracture healing assessment based on orthopedic trauma surgeons' subjective impressions

Residents			Attendings		
Number	Kappa	Significance(*P*)	Number	Kappa	Significance(*P*)
Reviewer1	0.823	0.000	Reviewer 3	0.782	0.000
Reviewer2	0.081	0.480	Reviewer 4	0.496	0.001
Reviewer6	0.443	0.001	Reviewer 5	0.394	0.003

Table 2 The correlation between subjective impressions-based judgments and the consensus on fracture healing (first round)

Residents			Attendings		
Number	Kendall tau-b	95%CI	Number	Kendall tau-b	95%CI
Reviewer1	0.462	0.348–0.563	Reviewer 3	0.564	0.464–0.650
Reviewer2	0.658	0.574–0.729	Reviewer 4	0.666	0.583–0.736
Reviewer6	0.577	0.479–0.661	Reviewer 5	0.508	0.399–0.602

Note 95%CI*(95% Confidence intervals)

Table 3 The correlation between subjective impressions-based judgments and the consensus on fracture healing (second round)

Residents			Attendings		
Number	Kendall tau-b	95%CI	Number	Kendall tau-b	95%CI
Reviewer1	0.282	0.092–0.452	Reviewer 3	0.498	0.336–0.632
Reviewer2	0.393	0.215–0.547	Reviewer 4	0.207	0.012–0.386
Reviewer6	0.440	0.268–0.585	Reviewer 5	0.737	0.633–0.815

Note 95%CI*(95% Confidence intervals)

Table 4 Intraobserver consistency of the RUSH score

Residents			Attendings		
Number	ICC*	95%CI*	Number	ICC*	95%CI*
Reviewer1	0.828	0.711-0.900	Reviewer 3	0.859	0.717–0.926
Reviewer2	0.523	-0.062-0.794	Reviewer 4	0.642	0.438–0.784
Reviewer6	0.642	0.391–0.795	Reviewer 5	0.825	0.700–0.900

Note ICC* (Intraclass correlation coefficient), 95%CI*(95% Confidence intervals)

(ICC=0.642). Among the resident physicians, the intraobserver agreement ranged from almost perfect to moderate, with ICC values of 0.828, 0.523, and 0.642, respectively (Table 4).

We conducted a regression analysis on the average total RUSH score, the average cortical bone item score, and the average cancellous bone item score, followed by plotting ROC curves (Fig. 3). The results demonstrated a strong correlation between all three metrics and fracture healing consensus, with AUC values of 0.769 (95% CI: 0.670–0.868), 0.779 (95% CI: 0.681–0.876), and 0.771 (95% CI: 0.674–0.867), respectively. The optimal cutoff values, as determined by the Youden index, were 25.917 for the total score, 21.415 for the cortical score, and 4.915 for the cancellous score. The cortical bone item score had the highest AUC. In practice, a cortical bone item score greater than 21 can be used as a threshold for determining fracture healing, corresponding to a sensitivity of 0.660 and a specificity of 0.864.

Discussion

Although the fracture healing process typically progresses through four stages—hematoma and inflammation, soft callus formation, hard callus formation, and remodeling—clinical decision-making and research often simplify these outcomes into two categories: healed or non-healed [11]. However, there is no universally accepted standard for assessing fracture healing in clinical practice and research [12]. The routine use of biopsies or advanced imaging techniques is often impractical due to cost and invasiveness, leading clinicians to rely on more accessible but less precise methods like conventional radiography [3]. Traditional X-ray imaging offers several advantages, including wide availability, low cost, and a low radiation dose. Nevertheless, there is ongoing uncertainty regarding the consensus among orthopedic surgeons regarding assessing fracture healing [12].

Whelan et al. have emphasized that the reliable assessment of cortices bridging suggests its utility in evaluating tibial fracture healing, particularly when an intramedullary nail is present [13]. This approach is formalized through the Radiographic Union Score for Tibial (RUST), which aims to standardize fracture healing assessments based on plain radiographs [5]. Similarly, Bhandari et al. developed a Radiographic Union Score for Hip (RUSH) checklist to improve the femoral neck fracture healing assessment. This scoring system enhances radiographic evaluations' consistency and boosts interobserver and intraobserver reliability [6]. RUSH was constructed by integrating various definitions and criteria for fracture healing found in the literature, such as bone bridging and
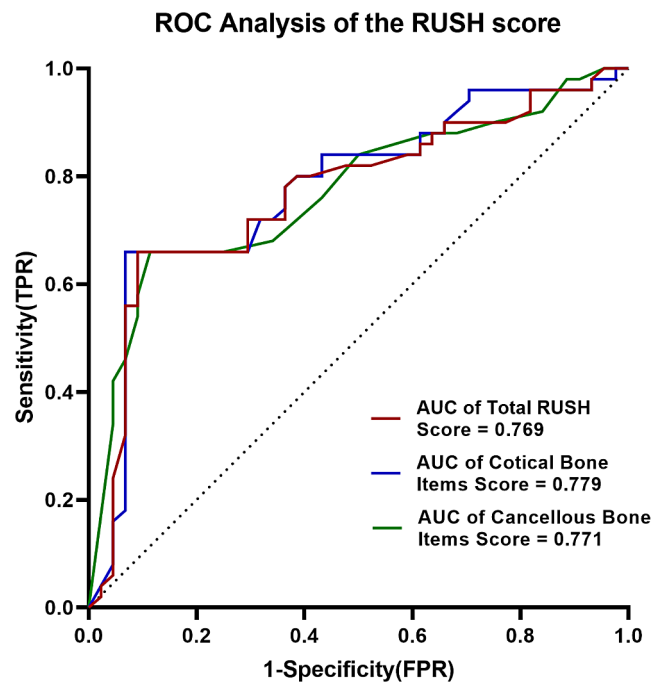
ROC Analysis of the RUSH score

Fig. 3 Illustrates the receiver operating characteristic (ROC) analysis of the RUSH score

the disappearance of the fracture line. Furthermore, in a study focused on standardizing intertrochanteric fracture healing assessment, Chiavaras et al. utilized the RUSH methodology, which significantly enhanced consensus among radiologists and orthopedic surgeons [8].

Our study assessed both interobserver and intraobserver reliability in evaluating subtrochanteric fracture healing. The results indicate that the consistency of orthopedic trauma surgeons' subjective impressions of healing was only moderate (Kappa=0.564, 95% CI: 0.511–0.616). This level of agreement is higher than that reported by Chiavaras et al. for overall impression-based fracture healing assessments by radiologists and orthopedic trauma surgeons for intertrochanteric fractures (ICC=0.34, 95% CI: 0.11–0.52) [8], and exceeds the consensus level reported by Bhandari et al. for intertrochanteric fractures assessed by surgeons and radiologists (ICC=0.50, 95% CI: 0.33–0.62) [14]. It states that the agreement among surgeons in subjective assessments of hip fracture healing on X-rays is moderate, regardless of whether the fractures are femoral neck, intertrochanteric, or subtrochanteric.

Using the RUSH scoring system, the six observers achieved a high level of consistency with an ICC of 0.748 (95% CI: 0.663–0.817). This ICC value is between those reported by Chiavaras (0.66) [8] and Bhandari (0.88) [14]. When comparing different experience levels, senior attendings at our center demonstrated greater consistency in overall RUSH scores than orthopedic residents. However, after a four-week interval, the general

agreement among observers slightly decreased to an ICC of 0.574 (95% CI: 0.397–0.721), although senior attendings still showed higher consistency than residents. These results suggest that clinical experience may improve the reliability of orthopedic surgeons in using the RUSH scoring system for evaluating fracture healing.

Analysis of intraobserver agreement revealed that after four weeks, two senior attendings achieved nearly perfect agreement, while one attained a high level of consistency. In contrast, orthopedic residents exhibited a range of internal consistency, including nearly perfect, moderate, and high levels. These results suggest that the intraobserver agreement among our orthopedic trauma surgeons using the RUSH scoring system for subtrochanteric fractures is not as robust as reported in other studies. However, the greater experience of senior orthopedic trauma surgeons in interpreting X-ray images contributes to their higher consistency in scoring.

Chiavaras et al. performed separate statistical analyses for the components of fracture line disappearance, including four cortical areas and one cancellous area, as well as the items related to calcification, which included cortical bridging and cancellous trabecular consolidation. Their ICC values for these components were 0.68 (95% CI: 0.56–0.77) for cortical bridging and 0.57 (95% CI: 0.43–0.69) for cancellous consolidation [8]. However, we performed separate analyses on 8 cortical and 2 cancellous bone-related items. The ICC for cortical bone-related items was 0.787 (95% CI: 0.706–0.849) in the initial evaluation of 94 imaging data sets, whereas the

Zhou *et al. BMC Musculoskeletal Disorders*   (2024) 25:798

Page 8 of 9

ICC for cancellous bone-related items was 0.668 (95% CI: 0.589–0.744). This data shows that our segmentation method yielded higher reliability than Chiavaras' study due to subtrochanteric fractures predominantly involving cortical bone, making identifying cortical fracture lines more straightforward than assessing cancellous bone fracture lines. On the contrary, accurately distinguishing the disappearance of fracture lines and evaluating the degree of trabecular calcification in the cancellous bone region of subtrochanteric fractures poses more significant challenges for orthopedic trauma surgeons.

Frank et al. evaluated postoperative X-rays of 250 femoral neck fractures at six months, which were treated using cancellous bone screws or sliding hip screws. They employed the RUSH scoring system for their assessment. With just two observers involved, the study demonstrated high interobserver consistency, achieving an ICC of 0.81 (95% CI: 0.76–0.85). Furthermore, ROC analysis revealed that a RUSH score of less than 18 exhibited a perfect specificity and positive predictive value of 100% for identifying radiographic nonunion [9]. Through regression analysis and ROC curve plotting, we identified strong correlations between the average total RUSH score, average cortical bone item score, average cancellous bone item score, and the consensus on fracture healing. The optimal cutoff thresholds were determined to be 25.917 for the total score, 21.415 for cortical bone items, and 4.915 for cancellous bone items. For practical application, we chose a cutoff value of 21 points for the total score of cortical bone-related items. This threshold enhanced sensitivity (0.660) and specificity (0.864) in evaluating fracture healing. Compared to Frank's investigation, orthopedic surgeons at our center use more stringent criteria for assessing the healing of subtrochanteric fractures, allowing only minimal visibility of cortical bone fracture lines or absence of callus in one direction. We did not perform ROC analysis for nonunion, so no threshold for assessing nonunion has been established.

Our study has several limitations. Firstly, our orthopedic trauma surgeons were new to using the RUSH scoring system. Despite receiving training, variations in their understanding and interpretation of the system could have affected the results. Furthermore, due to the infrequent occurrence of subtrochanteric fractures, we had to use multiple sets of X-ray images from some patients at different follow-up times. Although we randomized the image order twice, some residual bias may still exist. Additionally, during the X-ray imaging process, variations in operator technique and patient cooperation sometimes prevented us from obtaining ideal orthogonal views for both frontal and lateral images. Some images were obtained at oblique angles, which may have impacted the scoring accuracy. Nonetheless, these conditions reflect real-world clinical practice and contribute to the generalizability of our findings.

Conclusions

To summarize, our study demonstrates that utilizing the RUSH scale enhances the consistency of evaluating subtrochanteric fracture healing among orthopedic trauma surgeons, regardless of their experience level. A simplified RUSH scoring system, focusing on eight cortical bone-related items, can be utilized in subtrochanteric fractures. It is essential to acknowledge that the ideal method for assessing fracture healing should incorporate clinical and radiographic assessments. Future research should investigate incorporating clinically relevant data into this scale to improve its reliability.

Data availability
The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The Medical Ethics Committee of Shenzhen People's Hospital approved the study, with the committee's reference number LL-KY-2021070. We confirm that appropriate informed consent was obtained from all subjects and their legal guardian(s). Our study did not involve any participants who were minors or illiterates.

Consent for publication
We obtained consent from all enrolled patients or authorized representatives of the participating patients to anonymize their imaging data for publication purposes.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Orthopedic Surgery, Shenzhen Key Laboratory of Musculoskeletal Tissue Reconstruction and Function Restoration, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University, The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, Guangdong 518020, P.R. China
[2]The First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong 510405, P.R. China

Zhou *et al. BMC Musculoskeletal Disorders*   (2024) 25:798

Page 9 of 9

References

1. Krappinger D, Wolf B, Dammerer D, Thaler M, Schwendinger P, Lindtner RA. Risk factors for nonunion after intramedullary nailing of subtrochanteric femoral fractures. Arch Orthop Trauma Surg. 2019;139(6):769–77.
2. Aguado HJ, Castillon-Bernal P, Ventura-Wichner PS, Cervera-Diaz MC, Abarca-Vegas J, Garcia-Florez L, et al. Impact of subtrochanteric fractures in the geriatric population: better pre-fracture condition but poorer outcome than pertrochanteric fractures: evidence from the Spanish Hip Fracture Registry. J Orthop Traumatol. 2022;23(1):17.
3. Fisher JS, Kazam JJ, Fufa D, Bartolotta RJ. Radiologic evaluation of fracture healing. Skeletal Radiol. 2019;48(3):349–61.
4. Atwan Y, Schemitsch EH. Radiographic evaluations: which are most effective to follow fracture healing? Injury. 2020;51(Suppl 2):S18–22.
5. Whelan DB, Bhandari M, Stephen D, Kreder H, McKee MD, Zdero R, 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):629–32.
6. Bhandari M, Chiavaras M, Ayeni O, Chakraverrty R, Parasu N, Choudur H, et al. Assessment of radiographic fracture healing in patients with operatively treated femoral neck fractures. J Orthop Trauma. 2013;27(9):e213–9.
7. Thamyongkit SS-n, Kulachote P, Sirisreetreerux N, Chulsomlee N, Pengrung K, Chitrapazt N. Jaovisidha, Suphaneewan. Reliability and Usefulness of Radiographic Union Score (RUS) for fracture Healing Assessment in Subtrochanteric femoral fractures. J Med Assoc Thai. 2020;103(11):1148–54.
8. Chiavaras MM, Bains S, Choudur H, Parasu N, Jacobson J, Ayeni O, 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(8):1079–88.
9. Frank T, Osterhoff G, Sprague S, Garibaldi A, Bhandari M, Slobogean GP, et al. The Radiographic Union score for hip (RUSH) identifies Radiographic Nonunion of femoral Neck fractures. Clin Orthop Relat Res. 2016;474(6):1396–404.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
11. Morshed S, Corrales L, Genant H, Miclau T 3. Outcome assessment in clinical trials of fracture-healing. J Bone Joint Surg Am. 2008;90(Suppl 1):62–7.
12. 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):562–6.
13. Whelan DB, Bhandari M, McKee MD, Guyatt GH, Kreder HJ, Stephen D, 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(1):15–8.
14. Bhandari M, Chiavaras MM, Parasu N, Choudur H, Ayeni O, Chakravertty R, et al. Radiographic union score for hip substantially improves agreement between surgeons and radiologists. BMC Musculoskelet Disord. 2013;14:70.