Injury 51 (2020) S55–S62

Contents lists available at ScienceDirect

Injury

journal homepage: www.elsevier.com/locate/injury

Validation of a long bone fracture non-union healing score after treatment with mesenchymal stromal cells combined to biomaterials ☆,☆☆,★,★★,✞,✞✞,†,††,✧,✧✧,✦,✦✦,❍❍

Enrique Gómez-Barrena [a,*], Norma G. Padilla-Eguiluz [b], Eduardo García-Rey [a], Pablo Hernández-Esteban [c], José Cordero-Ampuero [d], Juan C. Rubio-Suárez [c], On behalf of the REBORNE and ORTHOUNION Research Consortia

[a] Servicio de Cirugía Ortopédica y Traumatología, Hospital Universitario La Paz-IdiPAZ and Facultad de Medicina, Universidad Autónoma de Madrid, Madrid, Spain
[b] Facultad de Medicina, Universidad Autónoma de Madrid -IdiPAZ, Madrid, Spain
[c] Servicio de Cirugía Ortopédica y Traumatología, Hospital Universitario La Paz-IdiPAZ, Madrid, Spain
[d] Servicio de Cirugía Ortopédica y Traumatología, Hospital Universitario La Princesa, Madrid, Spain

ARTICLE INFO

Article history:
Accepted 8 February 2020

Keywords:
Non-union
Treatment
MSC
Bioceramics
Radiographic
Score
Validation
Reliability
Internal consistency

ABSTRACT

The available scores to clinically evaluate fracture consolidation encounter difficulties to interpret progression towards consolidation in long-bone non-union, particularly when incorporating biomaterials in the surgical treatment. The aims of this study were to validate the REBORNE bone healing scale in tibia, humerus and femur non-unions treated by a combination of mesenchymal stromal cells (MSCs) and biomaterials, through the interclass correlation (ICC) among raters, and to define reliability and concordance in anteroposterior and lateral radiographs, compared to computed tomography (CT).

Methods: Twenty-six cases from the EudraCT 2011-005441-13 clinical trial underwent bone healing evaluation, if at least 3 out of 4 cortical views clearly identified. Three senior orthopaedic surgeons evaluated radiographs and CTs at 3 and 6 months FU. All cases included preoperative imaging and radiographs at 12 months. The 4-stage scale score was obtained from each cortical view in orthogonal radiographs or CTs. A score of 0.6875 (11/16) was set as a threshold for bone healing. Statistically, ICC evaluated agreement among raters. Cronbach's alpha coefficient tested reliability. Lin's concordance correlation coefficients (CCC) were estimated between mean CT scores and mean radiographic scores. Bland and Altman graphs provided the limits of agreement between both imaging techniques. Sensitivity and specificity were assessed in radiographs (against CT), and the Area Under the Receiver Operating Characteristics

☆ Researchers from the FP7-REBORNE and H2020-ORTHOUNION Research Consortia who contributed to this work:
☆☆ Rosset, P. (Service de Chirurgie Orthopédique et Traumatologique 2, Hôpital Trousseau, Université François-Rabelais de Tours, CHU de Tours, Tours, France).
★ Gebhard, F.; Ehrnthaller, C.; Kalbitz, M. (Department of Traumatology, Hand-, Plastic-, and Reconstructive Surgery, Center of Surgery, University of Ulm, Ulm, Germany).
★★ Rojewski, M; Lotfi, R.; Schrezenmeier, H. (Institut for Transfusion Medicine, Ulm University, and Institute for Clinical Transfusion Medicine and Immunogenetics Ulm, German Red Cross Blood Transfusion Service and University Hospital Ulm, Ulm, Germany).
✞ Hernigou, P. (Orthopaedic Department, Hôpital Henri Mondor, InsermU955 and UPEC (University Paris-Est, Créteil), Créteil, France).
✞✞ Baldini, N.; Ciapetti, G.; Donatti, D.M.; Spazzolini, B. (Laboratory for Orthopaedic Pathophysiology and Regenerative Medicine, Istituto Ortopedico Rizzoli, and Dept. of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy).
† Gonzalo-Daganzo, R.M.; Fernandez, M.N.; Avendaño-Sola, C. (Servicios de Hematología y Farmacología, Hospital Universitario Puerta de Hierro-Majadahonda, and Universidad Autónoma de Madrid, Madrid, Spain).
†† Bunu-Painatescu, C. (Victor Babes University of Medicine and Pharmacy, Timisoara, Romania).
✧ Dominici, M. (Laboratory of Cellular Therapies, Department of Medical and Surgical Sciences for Children & Adults, University - Hospital of Modena and Reggio Emilia, Modena, Italy).
✧✧ Montemurro, T.; Giordano, R. (Cell Factory, Center for Cellular Therapy and Cryobiology, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milano, Italy).
✦ Rouard, H.; Sensebé, L. (Établissement Français du Sang, Paris, France).
✦✦ Layrolle, P. (INSERM U957, Lab. Pathophysiology of bone resorption, Faculty of Medicine, University of Nantes, Nantes, France).
❍❍ This paper is part of a Supplement supported by The Orthopaedic Surgery and Traumatology Spanish Society (SECOT).

* Corresponding author.
E-mail addresses: egomezb@salud.madrid.org, enrique.gomezbarrena@uam.es (E. Gómez-Barrena).

https://doi.org/10.1016/j.injury.2020.02.030

(ROC) Curve was estimated. The probability to predict bone consolidation with REBORNE scores obtained from radiographs was modelled.
Results: An ICC of 0.88 and 0.91 (CT and radiographs) confirmed agreement in the REBORNE score for non-union bone healing, with an inter-rater reliability of 0.92 and 0.95. Scores through the radiographic evaluation were found equivalent to the CTs at 6 months FU. A CCC of 0.79 was detected against CT. The radiographic scores in the REBORNE bone healing scale correctly classified bone consolidation in 77%, with an accuracy of 83% based on ROC curves.
Conclusions: The REBORNE score measured with CT or radiographic images was reliable among raters at a follow-up time above 6 months for long bone non-union fractures. The REBORNE scale measured with radiographs proved valid to assess consolidation against CT measurements.

Introduction

The evaluation of bone healing after treating fracture non-unions is still debatable at the clinical setting. Imaging, mechanical testing, serologic markers and clinical assessment have been advocated to interpret bone healing [1,2]. When the clinician needs to conclude on the status of bone healing, a compound analysis is performed, incorporating radiographic and clinical data, including pain at rest and at weight-bearing, as seen in a recent international survey [3]. Orthogonal radiographic views of the fractured segment allow evaluation of callus size, cortical continuity and progressive loss of fracture line, considered the basic radiographic data of bone consolidation [4]. However, these signs are more difficult to evaluate in case of complex fracture non-unions, particularly after treatment. Computed tomography (CT) scanning is considered mandatory to confirm bone healing in non-unions, with the help of 3D reconstruction [5]. Eventually, other methods may be required to clarify bone healing and fluorodeoxyglucose positron emission computer tomography (FDG-PET-CT) is becoming a useful option [5]. The final decision about healing corresponds to the treating surgeon, who will take into account all the available information. Time usually helps to confirm bone healing through callus observation in the follow-up. Cortical remodelling, particularly if absence of pain with weight-bearing occurs, frequently supports the impression of definite bone healing.

When evaluation of clinical trials comes into question, precise timing to define bone healing is required, and robust criteria are needed. This has fostered bone healing quantification, with the need to set thresholds of bone healing versus no healing. With the precedent of Hammer et al. [6], who established a radiographic scoring method for tibial fractures, the RUST (Radiographic Union Scale for Tibial fractures) score was developed based on cortical bridging and fracture line visibility [7].

The RUST score was designed to evaluate diaphyseal tibial fractures treated by intramedullary nailing, based in 3 simplified categories (callus absent and visible fracture line, callus present with visible fracture line, and callus present with invisible fracture line) which may not interpret more complex bone healing scenarios. The application to other bones, fractures in the metaphysodiaphyseal location, and other forms of fixation besides the intramedullary nail, have not been clarified yet. Besides, a clear threshold of bone healing for the RUST score has not been established, which limits its potential generalization.

Furthermore, in the non-union setting, the progression of callus and fracture line may not be so definite. When the surgical treatment incorporates radiopaque material, such as bioceramics, bone healing assessment becomes more complex and definition of bone to material separation may not be evident. Simplistic approaches may lose the healing progression, particularly in early evaluations. How to establish differences between a repaired non-union which may progress to healing versus the one that is not healing, at a specific time and evaluation, poses significant problems to evaluate the treatment efficacy in non-unions.

In this context, the aim of this study was to validate a 4-stage scale imaging of bone healing by estimating the ICC with the ratings of 3 qualified orthopaedic surgeons, and by defining the reliability and concordance of the scores in anteroposterior and lateral radiographs, compared to CT scans, in tibia, humerus and femur non-unions treated by a combination of mesenchymal stromal cells (MSCs) and biomaterials.

Material and methods

The REBORNE score

The *REBORNE bone healing score* was defined to perform a detailed evaluation of long bone non-union consolidation in radiographs and CTs. REBORNE was the FP7-EU project of basic and clinical bone regeneration research [8] that initiated this and other clinical trials on bone regenerative medicine.

Three senior orthopaedic surgeons, with more than 15-year experience in dealing with long bone fractures, participated as raters in this evaluation. Each rater assessed four postoperative cortical views for each bone (medial and lateral cortices in AP radiographic views and in coronal and/or transverse CT sections; also, anterior and posterior cortices in lateral radiographic views and in sagittal and/or transverse CT sections).

In each of these cortices, four stages were assessed at each imaging exam (Table 1, Fig. 1), including fracture unchanged from preoperative views (stage 1, 1 point), presence of initial, non-continuous callus (stage 2, 2 points), presence of continuous callus with fracture still apparent (stage 3, 3 points), and presence of callus with same density as cortical (stage 4, 4 points). Fig. 1 (a,b,c,d)

Table 1
Overview of the REBORNE scale.

Cortical score	Stage
1	Fracture unchanged*
2	Callus but non-continuous
3	Callus continuous but fracture still apparent
4	Callus with same density as cortical
0	Non interpretable/non visible

REBORNE scale is calculated by the sum of the cortical scores (internal, external, anterior, and posterior), divided by maximum punctuation expected (which is calculated multiplying the number of evaluated cortices per 4).
Non-union radiological consolidation is set from a Reborne scale score of 0.6875 (11/16).
With radiographs, the Reborne-scale is only valid if at least 3 cortices can be evaluated.
*Unchanged from preoperative image test.

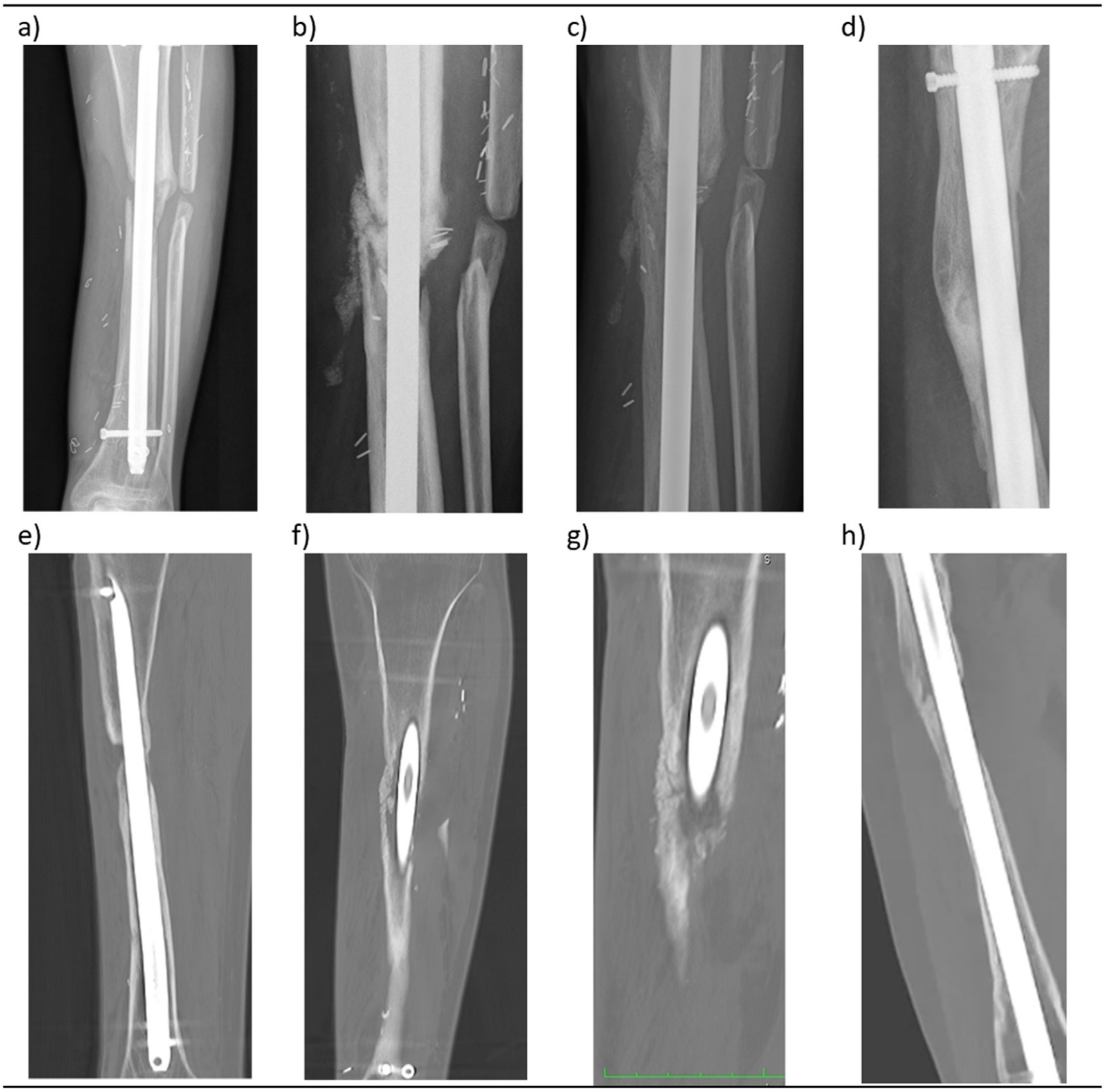

Fig. 1. Four stages of imaging were radiographically assessed in each of the cortical views including a) unchanged fracture from basal (stage 1); b) presence of initial, non-continuous callus (stage 2); c) presence of continuous callus with fracture still apparent (stage 3); d) presence of callus with same density as cortical (stage 4). Four stages were also assessed through CT in each of the cortical views including e) unchanged fracture from basal (stage 1); f) presence of initial, non-continuous callus (stage 2); g) presence of continuous callus with fracture still apparent (stage 3); h) presence of callus with same density as cortical (stage 4).

displays the 4 stages in radiographs. However, the osteosynthesis hardware (particularly if plates were used) may hide one cortical in radiographs. This cortical was then considered non-interpretable (0 points were then allocated). The evaluation of CT sections was used to validate cortical consolidation (Fig. 1e–h), avoiding any visual interference.

Based on widely accepted parameters (callus formation and presence of fracture line, in each of the 4 cortical views), the REBORNE bone healing scale was generated with a score range of 0/16 to 16/16 points. The minimum 0/16 (if no cortical bone was evaluable) is not evaluable by radiographs (only by CTs), and then the minimum valid score was set at 4/16 (fracture unchanged in all 4 evaluated cortices). The maximum was set at 16/16 points (callus with same density as cortical in all 4 cortices). This range was condensed in a value from 0, or more precisely from 0.25 (0/16 to 4/16 points, no detected bone consolidation), to 1 (16/16 points, maximum detected consolidation).

A threshold for radiological consolidation was established from a REBORNE score of 0.6875 (11/16), that means at least bridging in 3 out of 4 cortices, and initial presence of callus in the last one (3 + 3 + 3 + 2). If two or more cortices are non-interpretable or non-visible, the REBORNE scale cannot be used with radiographs.

Cases

Data were obtained from the ORTHO-1 clinical trial (EudraCT 2011-005441-13), approved by the Ethics Committees and Regulatory Agencies of all participating centres and countries [9]. The REBORNE bone healing score (Figs. 1–3) was calculated based on each rating, at follow-up (FU) of 3, 6, and 12 months after surgery in radiographs, as long as at least 3 cortices were seen, and at 3 and 6 months in CTs. Twenty-six of 28 cases were included in this study (22 with studies at 3 and 6 months, 2 more with studies at 3 months, and 2 other cases with appropriate studies at 6 months). Two cases did not offer at least 3 cortical views to be fully evaluated in radiographs. All included cases had preoperative imaging and appropriate radiographs at 12 months postop.

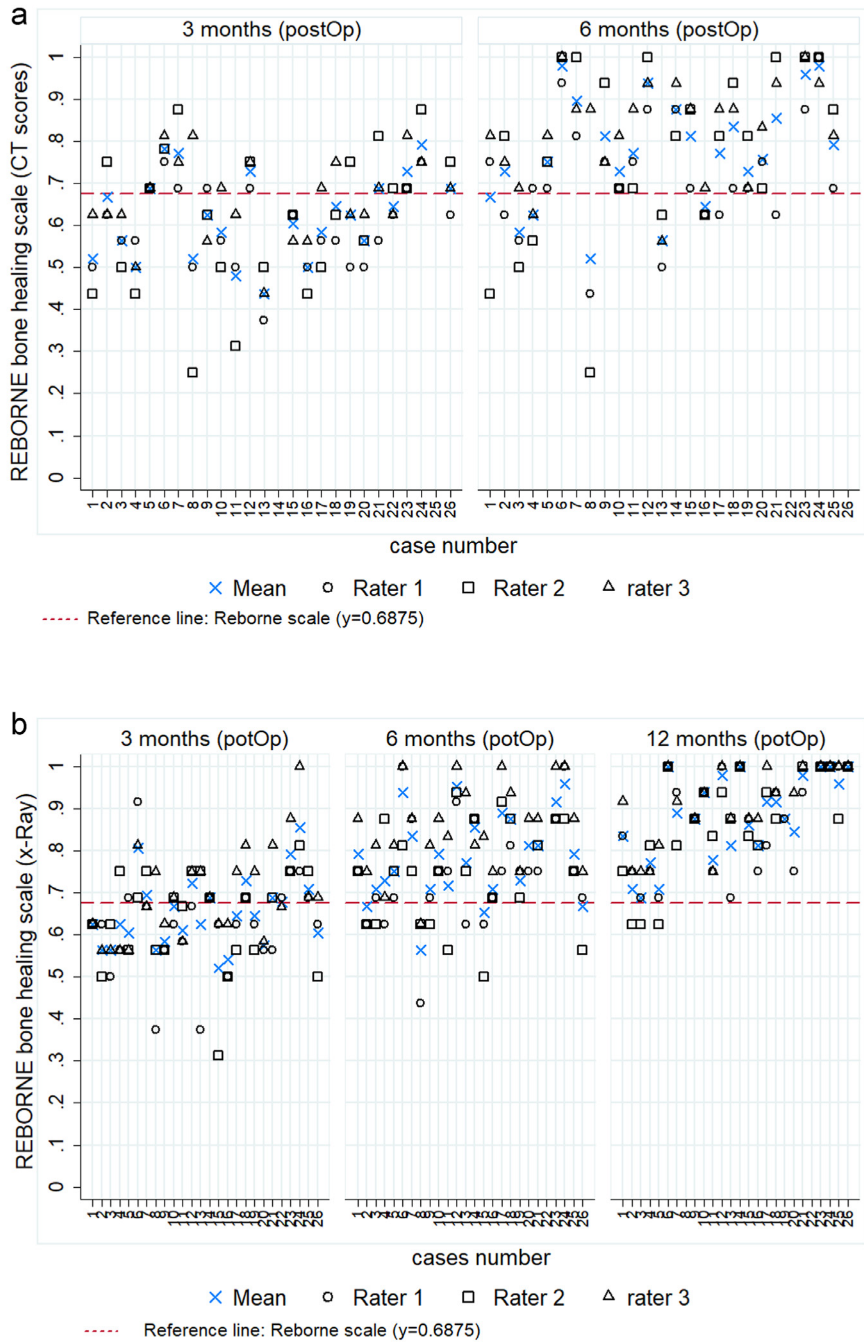

Fig. 2. Dispersion of mean of the REBORNE bone healing scale measurement, by raters and time since surgery. a) Distibution of individual mesurements and mean in CT images. At three and six months postOp, the ICC (inter-rater Reliability) was of 0.73(0.84) and 0.81(0.88), respectively. b) Distibution of individual mesurements and mean in x-Ray images. At three, six and twelve months postOp, the ICC (inter-rater Reliability) was of 0.62(0.78), 0.86(0.92) and, 0.91(0.94).

Statistical analysis

Data were described and analyzed using STATA software version 12 (Stata Corp. 2011). Data followed a normal distribution, then comparison of means was performed with the Student's *t*-test or with the analysis of variance (one-way ANOVA) test, as deemed appropriate.

The interclass correlation coefficients (ICC) for continuous variables were estimated based on a mean-rating ($k = 3$), absolute-agreement/multiple raters/measurements, two-way random effect model [10]. The analysis of variance (ANOVA) was conducted to identify the ICC estimators, score mean square and error mean square. The confidence intervals were assessed at 95%, based on Rosner's approach using the F-test. Root square of Cronbach's alpha coefficient was estimated to test inter-rater reliability. The level of agreement was considered "poor" if ICC value or reliability value were <0.5, "moderate" between 0.5 and 0.75, "good" between 0.75 and 0.9, and "excellent" if >0.9 [10].

Lin's concordance correlation coefficients (CCC) between mean CT scores and mean radiographic scores were estimated with a confidence of 95% to identify the degree to which scores were considered identical. Two more statistics were specifically revised, such as the Pearson correlation coefficient (r), and the Bias adjustment correlation (C_b) [11]. Bland and Altman's graphs were gen-

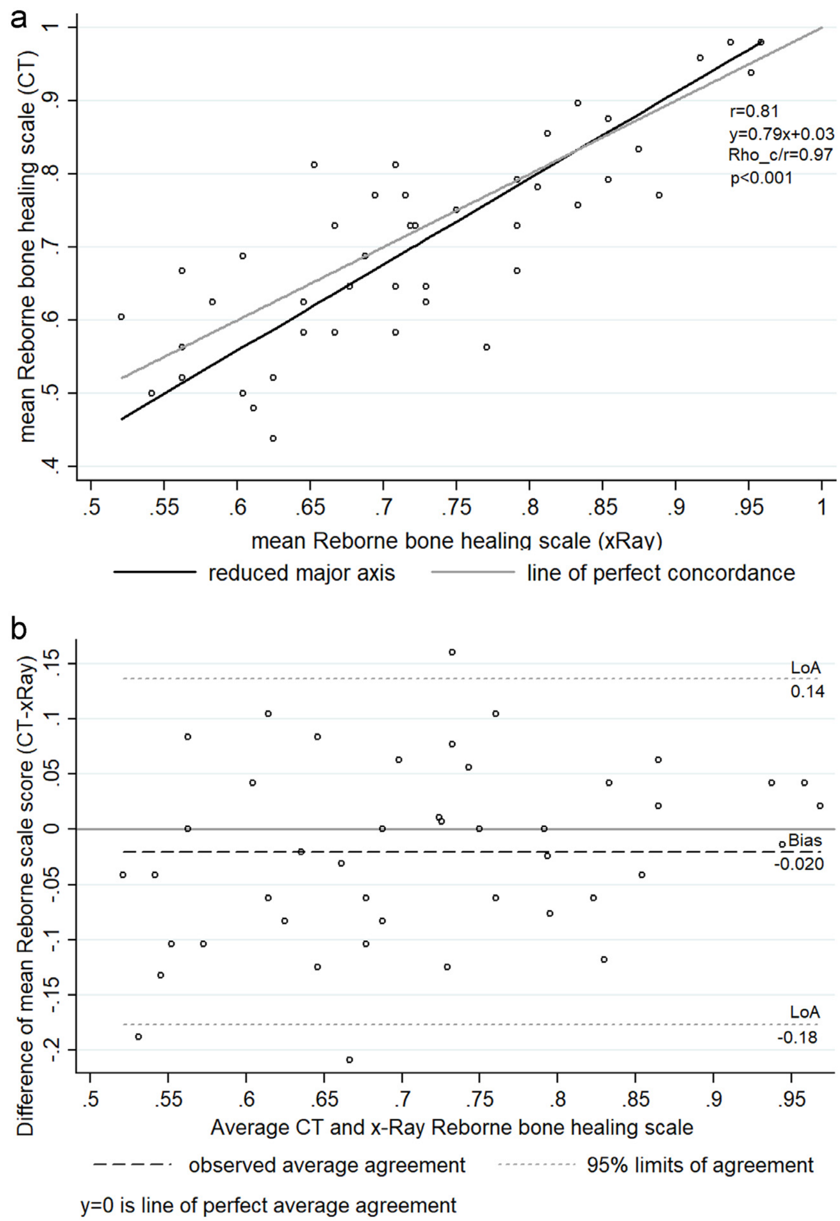

Fig. 3. Agreement between CT and radiographs in the REBORNE Bone healing scale. a) Scatter plot showing perfect correlation between X-Ray and CT Reborne scale. r= Pearson correlation; y=Lin's concordance correlation equation; Cb=bias adjustment correlation. b) Bland-Altman plot showing dispersion of measurements between CT and x-Ray tests and the Limits of Agreement (LoA). Bias from perfect agreement is of −0.02. Correlation between difference and mean=0.26; Bradley-Blackwood $F<0.001$.

erated to provide the limits of agreement between both tests. The agreement (degree to which both REBORNE scores are identical) was considered "slight" if CCC was between 0 and 0.20, "fair" between 0.21 and 0.40, "moderate" between 0.41 and 0.60, "substantial" between 0.61 and 0.80, and "perfect" if >0.80, following Landis and Koch interpretation [12].

CT and radiographic scores were dichotomized as bone consolidation, if the REBORNE score was equal or above 0.6875 points, or not consolidated bone if below this score. A logistic regression was modelled using bone consolidation in the CT (yes/no) as dependent variable. The independent variables were bone consolidation in radiographs (yes/no), time since surgery (3 and 6 months), and rater (#1, #2, and #3). The anatomical bone with the non-union (femur, humerus, tibia) was also included as an independent variable to control the systematic error. Sensitivity and specificity of the REBORNE scale were assessed in radiographs (against CT), and the Area Under the Receiver Operating Characteristics (ROC) Curve was estimated (ROC –AUC) with 95% CI.

Results

CT evaluation

The mean (CI, 95%) REBORNE score in CT imaging (Fig. 2a) was 0.65 (0.62–0.69), 0.69 (0.63–0.75), and 0.74 (0.70–0.78) for rater 1, 2 and 3 respectively. Data suggest that mean scores increase with follow-up. Rater 1 mean (CI, 95%) score changed from 0.59 (0.55–0.63) to 0.72 (0.66–0.77), from 3 to 6 months of FU; rater 2 mean score changed from 0.61 (0.54–0.68) to 0.78 (0.69–0.86); and rater 3 mean score changed from 0.66 (0.62–0.70) to 0.83 (0.77–0.87).

A *t*-test statistically confirmed the significance of the mean score difference between 3 and 6 months ($p < 0.001$), for each rater.

The overall interclass correlation (ICC, 95%CI) was 0.88 (0.80–0.94), which reflects a "good" agreement of the measurements with an "excellent" reliability index of 0.91. The split analysis showed an ICC (95%CI) of 0.73 (0.56–0.86) at three months from surgery, and of 0.81 (0.67–0.90) at six months. The reliability coefficient was 0.84 and 0.88 at 3 and 6 months postoperative, respectively.

Radiographic evaluation

The mean (CI, 95%) of the REBORNE scale in radiographs (Fig. 2b) was 0.65 (0.62–0.68) at 3 months, 0.78 (0.75–0.81) at 6 months, and 0.88 (0.85–0.91) at 12 months after surgery. Individuals' scores are reflected in Fig. 2b. Rater 1 mean (CI, 95%) score changed from 0.62 (0.57–0.66) to 0.75 (0.69–0.80), and to 0.87 (0.82–0.92), from 3 to 6 and to 12 months, respectively; rater 2 mean score changed from 0.63 (0.59–0.68) to 0.74 (0.69–0.79) and to 0.86 (0.81 −0.92); and rater 3 mean score changed from 0.69 (0.65–0.74) to 0.86 (0.83–0.90), and to 0.91 (0.87–0.95). A one-way ANOVA test confirmed the difference in the mean scores ($p < 0.001$) along the 3 postoperative timings, for each rater.

The overall interclass correlation (ICC, 95%CI) was of 0.92 (0.86–0.96) that reflects an "excellent" agreement of the measurements with an "excellent" reliability index of 0.95. Results by timing of measurement showed an ICC (CI, 95%) of 0.62 (0.41–0.79) at three months postoperatively, of 0.86 (0.75–0.93) at six months, and of 0.91 (0.85–0.96) at twelve months. The reliability coefficients were of 0.78, 0.92, and 0.96 at 3, 6, and 12 months postoperative, respectively.

Validation of the radiographic REBORNE score

Validation of the REBORNE bone healing scale in radiographic measurements compared with CT is shown in Fig. 3, observing "substantial" agreement with CCC of 0.79 (CI 95%: 0.73–0.85) and low bias between both types of imaging ($r = 0.81$; $p<0.001$; bias -0.02 ± 0.08). The bias adjustment correlation was 0.97 ("perfect" conformity of radiographic score with CT). The concordance correlations relation Lin's (CI 95%)/Pearson/Bias-adjustment were 0.60(0.35–0.85)/0.62/0.97 at three months postop, and 0.74(0.56–0.92)/0.76/0.96 at six months postoperatively.

The categorical variable of bone consolidation in radiographic measurements was associated with bone consolidation in the CT with an odds ratio of 5.4, adjusted by months since acute fracture, rater, and anatomical bone of the non-union (Table 2). Seventy-seven% of the evaluations classifying the non-union as consolidated with radiographs were correctly classified. The sensitivity of the REBORNE bone healing scale to correctly detect consolidation in non-unions was 82%, and to detect non consolidation, 69%. The overall accuracy of the Reborne scale with radiographs was of 83%, according the AUC-ROC curve (Fig. 4).

Table 2
Association between CT bone consolidation and x-Ray bone consolidation.

Variables	Odds ratio	95% CI
Bone consolidation (x-Ray)		
No	1	
Yes	5.43***	2.25–13.12
Months since surgery		
3 months	1	
6 months	3.73**	1.60–9.13
Rater		
R1	1	
R2	1.44	0.52–3.97
R3	2.43	0.85–6.91
Anatomical bone		
Femur	1	
Humerus	12.21**	2.02– 73.58
Tibia	1.19	0.49–2.88
Cons	0.14**	0.05–0.43

Logistic regression using as dependent variable the Bone Consolidation from CT scores (Yes/No).
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
*Model validation: psdR^2= 0.28****.

Discussion

The main results in this validation study of a scale that could interpret bone healing in long bone non-unions after treatment were the confirmed interobserver agreement in the radiographic and the CT-based scores, increased with time (better at 6 months, and even better in the radiographs at 12 months), and the concordance analysis of CT and radiographic scores. High sensitivity and notable specificity were also found for the bone healing threshold, set at 0.6875 (11/16 consolidated cortices).

Bone healing interpretation in treated non-unions is a difficult task without significant related studies. An accepted radiographic assessment of bone healing is crucial not only to make clinical decisions but also, and more specifically, when bone consolidation is an end-point in a clinical trial [13]. In view of this, fracture bone healing studies and scores may be the reference. In a series of 208 tibial fractures with 166 treated non-operatively [6], the authors evaluated the callus formation and the fracture line to define achievement of bone union in fractures (callus leading to homogenous bone structure or at least bone trabeculae crossing the fracture line, and also obliterated or barely discernible fracture line). However, they included a category (number 3, with apparent bone formation and bridging, but discernable fracture line) in which the bone union was uncertain. This uncertain bone healing is the main issue, given the implications of false positives and false negatives, and the radiographs were found to correctly predict union in only 50% of the fractures. The interobserver agreement of the Hammer scale was evaluated as only "moderate" (kappa=0.60) [14]. No mention was made to non-union bone healing.

In an intent to improve the evaluation of bone healing, particularly after intramedullary fixation, the RUST score was proposed [14], and improved substantially the available tools to define bone consolidation. The RUST score has been considered a simplified 4-cortical score (3 stages: callus absent and fracture line visible, callus present and fracture line visible, and callus present and fracture line invisible) in the AP and lateral radiographic views. However, all these studies were performed in tibial fracture orthogonal radiographs, after intramedullary nailing, and compared to baseline postoperative radiographs. These are part of the RUST limitations, as no comparison has been performed with CT, no threshold has been established to define bone healing at a certain timing (required to ascertain end-points in clinical trials), and no application of the RUST score to non-unions has been communicated. Particularly, CT imaging is considered the current gold standard to assess bone healing. The sensitivity of CT to detect non-union has been found up to 100% [15], with ICC of 0.89 and accuracy of 89.9%. Of note, the ICC obtained in our study is similar for CT. Although with lower specificity after intraoperative visualization in the mentioned study, it is agreed that CT is superior to radiographs to assess bone healing. These are the reasons that forced us to define and validate a score comparable in CT and radiographs, with an established threshold to define bone healing, and to apply the score to non-union treatment with different fixation methods, particularly with incorporated biomaterials than may jeopardize the image.

The proposed REBORNE score differentiates from the RUST in the intermediate categories. While the RUST simplified in 3 stages

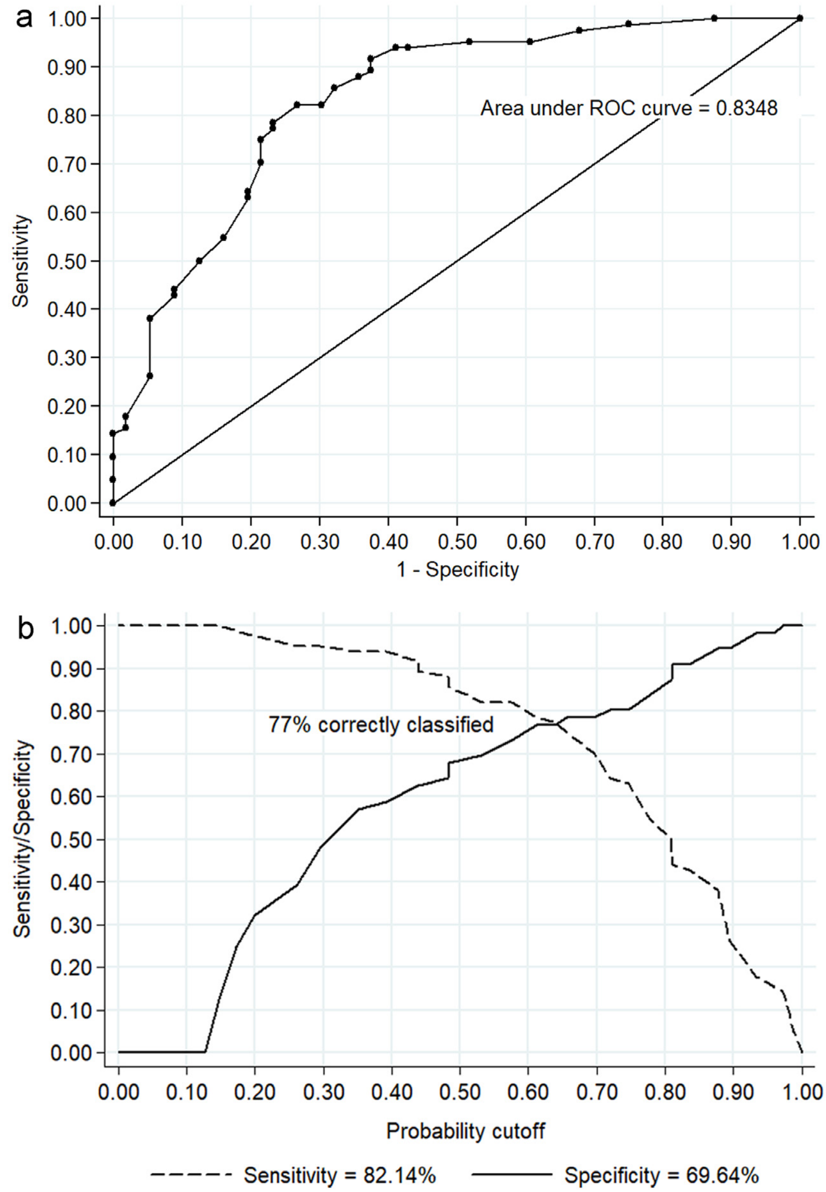

Fig. 4. Sensitivity and specificity against threshold used to distinguish bone consolidation, from not, at 3 and 6 months postop. Positive predictive value (80.23%); Negative predictive value (72.22%). False positive rate for true bone consolidation (30.36%); False negative rate for true bone consolidation (17.86%); False positive rate for classified positive (19.77%); False negative rate for classified negative (27.78%).

the bone healing, with a central category of "callus present, fracture visible" that does not considered the fracture as consolidated, it was further strengthened by adding clinical outcomes to better define fracture consolidation [13,16]. However, this intermediate category may be seen as non-consolidated in fresh fractures, but it could still be seen as "uncertain" in the non-unions. In this context, we feel that this intermediate bone healing in non-unions needs to differentiate if callus is ongoing and if bridging is significant, which may even be considered as consolidated. In those cases showing active callus formation and significant bridging, CT scans may be determinant, and also longer follow-up (12 months). Therefore, we decided to split the intermediate category in two, that could be confirmed with the CT: callus non-continuous with fracture still present (not healed), and callus continuous but fracture line still seen (healed).

The strengths of this REBORNE score, as studied in this paper, include the comparison to a gold standard evaluation (CT), a sequential evaluation of bone healing (from preop into 6 and 12 months) with increasing sensitivity with time, and the established and validated threshold for bone healing in non-unions.

Compared with other scales to assess tibial fracture healing, an advantage of the RUST score relays on its reliability, repeatability and validity that have been widely verified [13,17,18]. Improved reliability was observed with the RUST score, from ICC of 0.67 to ICC of 0.79 when radiographs were serially evaluated, from the post-operative films to the index radiograph [18]. Significant correlation with clinical outcomes of the fracture also sustained its reliability [17].

The ICC that was obtained with the REBORNE score in a difficult bone healing scenario (treated non-unions) with CTs was 0.88, adjusted by follow-up time, with a reliability of 0.84 at 3 months, and of 0.88 at 12 months. Therefore, longer follow-up in radiographs helps clarifying the bone healing adjudication, while CTs may support earlier definition of bone consolidation. The agreement between radiographic and CT scores was substantial (0.79), and data confirmed the concordance, correlation, and bias adjustment of these scores. Scores through the radiographic evaluation were found equivalent to scores through the CTs at 6 months FU. Furthermore, the radiographic threshold that was set showed a high capacity to detect consolidated fractures (82%), particularly at longer times (over 6 months).

Limitations in our study include a slight linear trend observed in Bland-Altman graphs (Fig. 3b), with a correlation between difference and mean of 0.26, suggesting a systematic error not controlled with the studied variables (at 3 months, the trend is 0.18, and at 6 months, 0.36). When adding the bone site of the non-union to the logistic model, a strong association was found between Humerus ($n = 4$) and the CT Bone consolidation variable. Femur ($n = 10$) and Tibia ($n = 12$) showed similar association. We hypothesized that the trend in the concordance correlation is related to the bone site variable, but a higher sample will be needed to test it. This would be particularly true for the group with less cases (humeral non-unions). Also, we did not plan CT at 12 months in the trial, as we originally considered that at 6 months, the cases would have been considered healed or not healed. Sometimes, this may not be the case and a longer follow-up may be required in non-union cases where slow progression occurs. Clear evaluation of non-union treatment is better obtained with radiographs and CT at least 6 months after treatment, and even 12 months.

Conclusion

The REBORNE scale, measured with CT or radiographic images, is reliable among raters and follow-up time for long bone non-union fractures, with a biomaterial included in the surgical treatment. More definite results are observed at 6 and 12 months after treatment.

The REBORNE scale, measured in radiographs, is valid to assess consolidation against CT measurements, with a concordance correlation of 79%, a capacity to correctly classify the bone consolidation of 77%, and an accuracy based on ROC curves of 83%.

Declaration of Competing Interest

The authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

CRediT authorship contribution statement

Enrique Gómez-Barrena: Formal analysis, Data curation, Writing - original draft, Writing - review & editing. **Norma G. Padilla-Eguiluz:** Formal analysis, Writing - original draft, Writing - review & editing. **Eduardo García-Rey:** Data curation, Formal analysis, Writing - review & editing. **Pablo Hernández-Esteban:** Data curation, Formal analysis, Writing - review & editing. **José Cordero-Ampuero:** Data curation, Formal analysis, Writing - review & editing. **Juan C. Rubio-Suárez:** Data curation, Formal analysis, Writing - review & editing.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7-HEALTH-2009) with the REBORNE Project (under G.A. 241876); and the European Union's Horizon 2020 Programme (H2020-SC1 2016-2017), with the ORTHOUNION Project (under G.A. 7333288).

References

[1] Morshed S. Current options for determining fracture union. Adv Med 2014;2014:708574.
[2] Cunningham BP, Brazina S, Morshed S, Miclau T 3rd. Fracture healing: a review of clinical, imaging and laboratory diagnostic options. Injury 2017;48(Suppl 1):S69–75.
[3] Bhandari M, Fong K, Sprague S, Williams D, Petrisor B. Variability in the definition and perceived causes of delayed unions and nonunions: a cross–sectional, multinational survey of orthopaedic surgeons. J Bone Jt Surg Am 2012;94(15):e1091–6.
[4] Bhandari M, Guyatt GH, Swiontkowski MF, Tornetta P, Sprague S 3rd, Schemitsch EH. A lack of consensus in the assessment of fracture healing among orthopaedic surgeons. J Orthop Trauma 2002;16(8):562–6.
[5] Gomez-Barrena E, Rosset P, Lozano D, Stanovici J, Ermthaller C, Gerbhard F. Bone fracture healing: cell therapy in delayed unions and nonunions. Bone 2015;70:93–101.
[6] Hammer RR, Hammerby S, Lindholm B. Accuracy of radiologic assessment of tibial shaft fracture union in humans. Clin Orthop Relat Res 1985(199):233–8.
[7] 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.
[8] Regenerating bone defects using new biomedical engineering approaches (REBORNE), 2010. https://cordis.europa.eu/project/rcn/92715/factsheet/en. (Accessed 7 April 2019).
[9] Gomez-Barrena E, Padilla-Eguiluz NG, Avendano-Sola C, Payares-Herrera C, Velasco-Iglesias A, Torres F, et al. A multicentric, open-label, randomized, comparative clinical trial of two different doses of expanded hBM-MSCs plus biomaterial versus iliac crest autograft, for bone healing in nonunions after long bone fractures: study protocol. Stem Cells Int 2018;2018:6025918.
[10] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–63.
[11] Liu J, Tang W, Chen G, Lu Y, Feng C, Tu XM. Correlation and agreement: overview and clarification of competing concepts and measures. Shanghai Arch Psychiatry 2016;28(2):115–20.
[12] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.
[13] Kooistra BW, Sprague S, Bhandari M, Schemitsch EH. Outcomes assessment in fracture healing trials: a primer. J Orthop Trauma 2010;24(Suppl 1):S71–5.
[14] 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 Jt Surg Br 2002;84(1):15–18.
[15] Bhattacharyya T, Bouchard KA, Phadke A, Meigs JB, Kassarjian A, Salamipour H. The accuracy of computed tomography for the diagnosis of tibial nonunion. J Bone Jt Surg Am 2006;88(4):692–7.
[16] Morshed S, Corrales L, Genant H, Miclau T 3rd. Outcome assessment in clinical trials of fracture-healing. J Bone Jt Surg Am 2008;90(Suppl 1):62–7.
[17] Cekic E, Alici E, Yesil M. Reliability of the radiographic union score for tibial fractures. Acta Orthop Traumatol Turc 2014;48(5):533–40.
[18] Leow JM, Clement ND, Tawonsawatruk T, Simpson CJ, Simpson AH. The radiographic union scale in tibial (RUST) fractures: reliability of the outcome measure at an independent centre. Bone Jt Res 2016;5(4):116–21.