UCLouvain
Secteur des sciences de la santé

IREC-IMAG

Validation of a scoring system for the semi-quantitative assessment of fracture healing

Vasiliki Perlepe, MD

A dissertation for the degree of Philosphiae Doctor (PhD) in Medical Sciences

July 2020

Promotor: Prof. Bruno Vande Berg

Jury Members:

Prof. Alain Poncelet, Prof. Frederic Lecouvet, Prof. Benoit Lengele, Prof. Emmanuel Coche, Prof. Christine Galant, Prof. Thomas Schubert

External Members:

Prof. Valérie Bousson (Hôpital Lariboisiere, Paris, France)

Prof. Alain Blum (CHU Nancy, France)

Remerciements

Au moment de terminer ce manuscrit, une évidence apparaît : beaucoup de gens ont contribué, de près ou de loin, à l'aboutissement de ce travail. Je tiens à les en remercier.

Je voudrais tout d'abord remercier mon promoteur, le Pr. Bruno Vande Berg, pour sa confiance, son encadrement rigoureux qui ont permis d'orienter mes travaux scientifiques. Merci encore pour sa générosité, sa disponibilité et son soutien. Bruno m'a fait partager sa passion pour l'imagerie musculo-squelettique et la recherche. Il a été le meilleur des guides et ce voyage n'aurait pas été possible sans lui.

Je voudrais adresser mes remerciements à tous les membres de mon jury de thèse pour leur intérêt et leurs conseils précieux lors des réunions du comité et pendant la défense privée. Merci également au Pr. Valerie Bousson et au Pr. Alain Blum qui ont accepté d'être membres du jury externe.

Merci à toute l'équipe du service de Radiologie des CUSL, le chef du département Pr. E. Coche, et merci tout particulier à mes collègues proches : Souad, Vassiliki, Frederic, Thomas, Isabelle et Latifa, pour leur aide à des degrés divers, leur bonne humeur et cette ambiance qui facilite le travail tous les jours. Merci à Perrine pour son aide logistique et à Nicolas pour son aide et ses explications dans l'analyse statistique des données.

Merci à mon frère et mes amis pour leur appui et leurs encouragements tout au long de cette aventure.

Je tiens à remercier tout spécialement mon cher conjoint Alexandros pour son écoute, son soutien constant à l'égard de mes travaux comme dans la vie en général. Sa présence, son écoute et son soutien constant m'assurent au quotidien des bases solides me permettant de persévérer et de me surpasser. Notre fille Ioli est encore trop petite pour comprendre mais j'espère qu'elle sera fière de sa maman un jour. Le bonheur que ma famille m'apporte a été le moteur dans ce long parcours qu'a été cette thèse.

Enfin, je dédie cette monographie à mes parents. J'ai pu compter sur leur amour et leur soutien indéfectibles pour réaliser mes projets en venant en Belgique pour faire ma spécialisation en Radiologie. Ils ont toujours poussé leurs enfants à donner le meilleur d'eux-mêmes. Ils sont loin mais toujours si présents et leur soutien m'a permis de surmonter les moments de doute et les difficultés. Vous avez toute ma gratitude !

TABLE OF CONTENTS

1. **LIST OF PUBLICATIONS**

Can we assess healing of surgically treated long bone fractures on radiograph?
Perlepe V, Omoumi P, Larbi A, Putineanu D, Dubuc JE, Schubert T, Vande Berg B.
Diagn Interv Imaging. 2018 Jun;99(6):381-386. doi: 10.1016/j.diii.2018.02.004. Epub 2018 Mar 17. *(Part I)*

Semi-quantitative CT assessment of fracture healing: How many and which CT reformats should be analyzed?
Perlepe V, Michoux N, Heynen G, Vande Berg B.
Eur J Radiol. 2019 Sep;118:181-186. doi: 10.1016/j.ejrad.2019.07.026. Epub 2019 Jul 19. *(Part II)*

Semi-quantitative assessment of nailed shaft fractures using computed tomography: comparison with radiographic and clinical scores during normal healing. Perlepe V, Michoux N,Kirchgesner T, Lecouvet F, Vande Berg B. (*Part III- paper submitted to European Journal Radiology in April 2020)*

Semi-quantitative CT assessment of tibial and femoral non-united fractures: Comparison with semi-quantitative radiographic assessment . Perlepe V, Michoux N,Kirchgesner T, Lecouvet F, Vande Berg B. *(Part IV- paper to be submitted)*

Value of a radiographic score for the assessment of healing of nailed femoral and tibial shaft fractures: A retrospective preliminary study.
Perlepe V, Cerato A, Putineanu D, Bugli C, Heynen G, Omoumi P, Berg BV.
Eur J Radiol. 2018 Jan;98:36-40. doi: 10.1016/j.ejrad.2017.10.020. Epub 2017 Nov 7. *(Part V)*

2. **ABBREVIATIONS**
RUS: Radiographic union score
TUS : Tomographic Union Score
CR: conventional radiography
CT: computed tomography
DTS : digital tomosynthesis

3. ABSTRACT

The objective of this thesis was to develop and evaluate scoring systems for the semi-quantitative assessment of fracture healing by using radiographs and CT scans of surgically treated long bones fractures.

In a preliminary study, we demonstrated the feasibility and substantial interobserver agreement among 6 observers of the scoring system in a series of radiographs of surgically treated upper and lower limb diaphyseal long bone fractures.

Second, we proposed a scoring system of CT data and showed that the analysis of a limited number of long-axis CT reformats was a valuable alternative to the analysis of the large number of CT reformats obtained in the three planes.

Third, after assessing the reproducibility of the scores, we compared the results of the radiographic and CT scoring systems in nailed tibial and femoral fractures in a group of patients during normal healing and in a group of patients with non-united fractures. CT provided different score values compared to x-ray in early and late phases of normal healing and in hypertrophic non-unions.

Finally, we demonstrated that the radiographic scoring system of nailed long bone fractures could contribute to detect fractures with delayed union at about 3 months after injury.

Funding sources:

This thesis was funded by IREC (Institut de recherche expérimentale et clinique) of the Université catholique de Louvain and by funds collected at continuing education (MAP en imagerie osteo-articulaire Lille-Bruxelles).

4. INTRODUCTION

4.1. BONE [1; 2]

Bone is a complex organ that serves many purposes. The first and foremost is the role of structure, but bone serves other functions such as mineral storage (for calcium and other ions such as sodium, phosphate and magnesium) and hematopoiesis as it stores marrow. Its structural role includes protection for the brain, spinal cord and chest organs, as well as rigid internal support for the limbs. The internal architecture of bone accounts for its lightness and high tensile strength which result from its hollow and tubular shape composed of a collagen matrix with mineral deposits.

Microscopically, two main types of bone can be observed. Woven bone, consisting of an irregular arrangement of collagen fibers that vary in diameter and are not stress-oriented in their direction, is only prominent in fetuses and in situations requiring rapid bone turnover such as during the healing phase of fractures and in disorders such as hyperparathyroidism and Paget disease. The majority of the skeleton in adults is composed of lamellar bone, which is more organized, stronger, and less flexible than woven bone, with collagen orientation that is stress-related. Lamellar architecture is found both in cortical and in cancellous bone.

Macroscopically, cortical bone is differentiated from cancellous bone. Cortical bone is a dense mineralized matrix of the lamellar type organized in structures called "osteons." Osteons are units constructed around Haversian canals and oriented parallel to the axis of the bone. Haversian structures are perpendicularly interconnected by Volkmann canaliculae. Osteocytes are disposed in a regular fashion within the structure. In contrast, cancellous or trabecular bone is a honeycombed framework of plates and struts of bone with intervening spaces. Cancellous bone represents a considerable exchange surface with interstitial liquids and therefore plays a major role in the phosphocalcic homeostasis. In adults, cortical bone typically accounts for 80% of the total bone mass and is located predominantly in the shafts of long bones, though cancellous bone is mainly found at the metaphysis of long bones, in short bones and in vertebrae.

Long bones are segmented into the epiphysis, metaphysis and diaphysis: the epiphyses are the extremities, the middle tubular part is the diaphysis and the metaphyses are the conical sections joining the epiphysis to the diaphysis.

A loss of continuity in the bone architecture defines a **fracture**. Cortical bone fractures can be caused by either the application of a single high- amplitude force (traumatic fracture), a subthreshold force applied to bone weakened from a pathologic process such as a neoplasm (pathologic fracture), or accumulative microdamage related to repetitive injury (stress fracture). A bone fracture is a pathological situation characterized by pain, swelling/hematoma, abnormal mobility/deformity of a bone segment and functional impairment.

4.2. FRACTURE HEALING

Repair and regeneration of bone fractures are complex processes requiring interaction of mesenchymal cells at different stages of differentiation and a coordinated cascade of growth factors and cytokines. Two different types of healing can occur [3; 4]:In **primary or direct healing**, such as occurs with rigid internal fixation such as compression plates with perfect reduction and stabilization of the bone fragments, there is an absence of a periosteal reaction, and healing occurs directly across the fracture site through the action of bone remodelling units known as ‘‘cutting cones’’. Direct fracture healing is characterized radiographically by a gradual disappearance of the fracture line without the formation of an external callus**. Secondary or indirect healing** results from less stable forms of fixation such as intramedullary nailing and external fixation that allow micromotion at the fracture site. Secondary fracture healing consists of both endochondral and intramembranous bone healing and is characterized radiographically by early bone resorption and subsequent callus formation. The repair process during secondary healing can be divided into five steps [5]:

1)Hematoma

- Activation of blood coagulation cascade
- Formation of provisional fibrin matrix
- Release of danger signal molecules
- Activation of local macrophages

2) Acute inflammation

- Recruitment of neutrophils and macrophages
- Clearance of necrotic tissue and provisional matrix
- Production of pro-inflammatory cytokines, chemokines, growth factors
- Recruitment and activation of MSCs, osteoprogenitors, fibroblasts

3) Granulation tissue

- Active proliferation of progenitor cells
- Deposition of immature fibrotic extracellular matrix
- Angiogenesis

4) Callus formation

- Differentiation of progenitor cells into chondrocytes and production of fibrocartilage (central areas)
- Differentiation of progenitor cells into osteoblasts and production of woven bone (periosteum)
- Fracture stabilization by fibrocartilage callus
- Apoptosis of chondrocytes, fibrocartilage calcification
- Vascular ingrowth, recruitment of osteoprogenitors cell
- Woven bone deposition on cartilage scaffold

5) Remodeling

- Formation of chondroclasts and osteoclasts
- Resorption of cartilage and woven bone
- Restoration of Haversian system

Adequate bone healing thus requires a careful combination of physiological and biomechanical factors. The diamond concept described by Giannoudis et al [6] determines the four key factors to bone regeneration after injury:

i) **Mechanical stability:** The bridging fracture callus allows load transmission and restoration of the architecture of the bone according to Wolff's law.

ii) **Osteogenic cells:** Local multipotent mesenchymal stem cells (MSC's) are recruited and migrate to the fracture callus. These cells will differentiate into the osteoblastic and chondroblastic phenotype during the callus maturation.

iii) **Growth factors:** Cytokines such as interleukin-1 and 6 (IL1, IL6), tumor necrosis factor-α (TNF-α), platelet-derived growth factor (PDGF), vascular endothelial growth factor (VEGF), insulin-like growth factor (IGF) or the transforming growth factor-β (TGF-β) superfamily members initiate a cascade of cellular events starting the bone healing process.

iv) **Adequate scaffold:** The extracellular matrix provides a natural support for cellular interaction.

Any disturbance in the delicate balance between these key factors will lead to complications such as non-unions.

4.3. ABNORMAL FRACTURE HEALING

According to the United States Food and Drug Administration (FDA), a **non-union** is established after a minimum of nine months after trauma with no visible progressive signs of healing for three months [7]. Tibia shaft is the most common site for non-union among long bones (up to 12% depending on type of treatment) [8]. Tibial shaft non-unions are associated with substantial healthcare resource use, common use of strong opioids, and high per-patient costs [9]. Femoral shaft non-unions have been reported in about 2-7% [10; 11]..

The Weber and Cech classification is the most popular classification system for non-unions [12]. At first, biological viable and reactive non-unions are characterized by well vascularized fragment ends and are divided due to radiological criteria: hypertrophic non-unions, so called elephant foot non- unions, are characterized by exuberant callus formation due to inadequate biomechanical stability; Horse's hoof non- unions are characterized by less callus formation and can be regarded as a milder form of elephant foot non-unions, which neither produce immovable stability nor allow interfragmentary mobility. Oligotrophic non-unions present viable fragment ends but no sign of callus is observed. Secondly, non-viable non-unions are regarded as biologically non-reactive as they show no biological activity on bone scans. This is mainly attributable to inadequate vascular supply to the fracture or non-union site. Atrophic non-unions are the final stage of non-viable non-unions with scar tissue in the former fracture gaps, osteoporosis, and atrophy of the main fragments close to the fracture site.

Fracture personalities and patient variables are important **risk factors** for non-union development. High-energy trauma with open fractures, vascular damage and large soft-tissue injury are associated with higher likelihood of non-union than closed ones. Fracture gaps larger than 10 mm showed significantly decreased healing than fracture gaps smaller than 3 mm [13]. There is evidence that reamed IM nailing of lower extremity long-bone fractures (tibia or femur) results in significantly fewer non-unions [11]. Other fracture-related risk factors for non-union include infection and complex comminuted diaphyseal fractures. Non-unions are found more frequent in elder patients (especially women), patients taking anticoagulants ,benzodiazepines, insulin ,antibiotics ,diuretics ,NSAID, and opioids [14] . Morbidities, such as high body mass index, smoking, increased alcohol consumption, diabetes mellitus type I and II, osteoarthritis and rheumatism, osteoporosis , vitamin D deficiency , and renal in- sufficiency seem to promote non-union development [15-17].

Treatment options for delayed unions and non-unions include nail dynamization, exchange nailing, bone grafts and bone morphogenetic proteins, cell therapy (MSCs isolated from bone marrow), external fixation techniques, augmentation plating [10]. For femoral and tibial diaphyseal shaft fractures, dynamization of the nail is an atraumatic, effective, and cheap surgical possibility to achieve bony consolidation, particularly in delayed non-unions before 24 weeks after initial surgery. In hypertrophic non-unions, biomechanical stability should be addressed by augmentation plating or exchange nailing. Hypotrophic or atrophic non-unions additionally require biological stimulation (by bone grafting or biologics) of bone healing for augmentation plating.

4.4. FRACTURE HEALING IN CLINICAL PRACTICE

Despite advances in treatment, the determination of when a fracture has healed remains subjective, constituting a dilemma for the orthopaedic surgeon. A survey of 444 orthopaedic surgeons a decade ago identified that there is a lack of consensus in defining delayed union and nonunion in tibial fractures among orthopaedic surgeons. There was considerable disagreement in both clinical and radiographic criteria to define fracture union as well as the average time required for diagnosis of delayed or failed union [18]. A systematic review of 92 studies published between year 2000 and 2006 showed similar trends of a lack of objective tools to radiographically or clinically assess fracture healing [19].

Corrales et al [20] in a meta-analysis of 123 articles found a wide variety in, and a lack of consensus about, the definition of fracture healing. Union was defined on the basis of a combination of clinical and radiographic criteria in 62% of the studies, on the basis of radiographic criteria only in 37%, and on the basis of clinical criteria only in 1%. The most commonly used radiographic and clinical criteria are presented in the following table:

Most commonly used criteria to define fracture union in long-bone fracture healing trials	
Radiographic criteria	**Clinical criteria**
Bridging of fracture by callus/bone/ trabeculae (53%)	Absence of pain or tenderness during weight-bearing (49%)
Bridging of the fracture site at three cortices (27%)	Absence of pain or tenderness on palpation or physical examination (39%),
Obliteration of the fracture line or cortical continuity (18%)	The ability to bear weight (18%)

Bhandari et al [21], in a survey of 335 orthopedic surgeons, demonstrated a lack of standardization in definitions for delayed unions and non-unions ; Nevertheless almost all agreed that defining a delayed union and nonunion should be done on the basis of both radiographic and clinical criteria (88%). Most respondents believed that the degree of soft-tissue injury, smoking history and vascular disease increased the risk of healing complications.

The clinical assessment of fracture union has proven to be challenging and highly subjective [22]. A study by Webb et al. [23] demonstrates that manual assessment of stiffness at fracture sites by orthopedic surgeons is no better than manual assessment by medical students. Pain on palpation is an even more unreliable and subjective measure, as individual and cultural differences in perception of pain and in pain tolerance vary so significantly. A patient's own assessment of function and pain level has increasingly become an important contributor to management decisions [24]. There are a number of generic outcome measures available for use in injured patients (such as SF-36 and SF-12 questionnaire // [25]) . Practical issues to consider in selecting a questionnaire include time for completion.

Routine blood tests such as a full blood count, urea, creatinine, electrolytes, liver function tests, bone profile and blood sugar are useful as a baseline assessment. Tests assessing systemic inflammatory activation including the white cell count, C-reactive protein and plasma viscosity should be performed but it is important to note that normal tests do not absolutely exclude infection and these markers are very non-specific. The possibility of a specific biological marker of delayed and non-union has provoked much research but no candidate has yet proved useful in routine clinical practice [26-29].

4.5. IMAGING OF FRACTURE HEALING

Radiology plays a key role in helping the orthopedic surgeon to form an impression on the state of fracture healing. Radiographs and CT represent the most commonly used modalities, though other modalities (including ultrasound and nuclear scintigraphy) are undergoing evaluation for their potential utility. Due to its low cost, wide availability, and relatively low radiation profile, radiography has classically been the most utilized technique in assessment of fracture healing [22]. Radiographic findings suggestive of a healing problem include persistent fracture lines, absence of bony bridging, lack of progressive healing on serial radiographs, progressive deformity and the presence of loose or broken implants [30]. Hypertrophic, oligotrophic and atrophic radiographic appearances allow the clinician to make inferences about the degree of fracture stability and the biologic viability of the fracture fragments while developing a treatment plan. **Conventional radiography** is generally used as the primary imaging modality, but its value for objective evaluation of bone healing has been reported to be limited [31-33].

When a patient is seen with symptoms of pain, dysfunction and a lack of evidence of healing on plain radiographs, the surgeon may choose to obtain a **computed tomography** scan to better delineate the bone anatomy. With advancements in the ability to create thinner slices and high- quality multiplanar reconstructions, CT is used as a second-line imaging modality either for better depiction of fracture anatomy or for complicated cases when abnormal healing or infection are suspected and non conclusive clinical or radiographic information are available. CT was reported to show 100% sensitivity for detecting non-union with high inter-observer agreement among readers [34]. However, computed tomography scanning had limited specificity and tended to lead to a diagnosis of non-union when, indeed, the fracture was healed in some patients. The role of CT continues to grow with the push towards less subjective means of assessment. CT images are more amenable to quantitative and volumetric measurements than radiographs, and CT can thus be used to evaluate macrostructural and micro- structural properties of healing bone [35]. One drawback of CT is that metal implants introduce artifacts and degrade image and reconstruction quality. Technical parameters to adapt in order to minimize metal hardware artifacts include appropriate focal spot size , increased kVp, increased mAs, decreased pitch value, decreased slice thickness , adequate patient positioning , increased reformat thickness , smooth reconstruction filters, large viewing windows [36].

Ultrasound is unable to penetrate cortical bone, but there is evidence that it is able to detect soft callus formation before radiographic changes are visible. Despite numerous publications over the last decade describing the early detection of callus on ultrasound prior to radiographs, all are limited by incomplete reporting of the reliability and reproducibility of their findings [37]. Moed et al. [38] conducted a large prospective study which showed that ultrasound findings at 6 and 9 weeks have a 97% positive predictive value and 100% sensitivity in determining fracture healing in patients with acute tibial fractures treated with locked intramedullary nailing. Quantitative ultrasound has also attracted interest in the monitoring of bone fracture healing, demonstrating that the ultrasound velocity is significantly different between healed bones and delayed unions or non-unions [39]. Sun et al [40]evaluated the validity of high-power vascular Doppler using *ex vivo* microangiography CT as a benchmark in an animal model. They found that the two modalities were strongly correlated and that power Doppler ultrasound was sufficiently reproducible. The vascularity of the fracture site decreased predictably with time in keeping with the developing callus and bone union.

In a prospective study by Ha et al [41], **digital tomosynthesis** was compared to radiography for the evaluation of fracture healing in a series of 51 wrist fractures treated with internal fixation. Eligible patients underwent radiographs, tomosynthesis and unenhanced CT on the same day. The authors found a significantly higher rate of detection of cortical fracture lines and less obscuration of fractures by hardware with tomosynthesis than with radiography, with consensus CT as the reference standard. The bony cortex was statistically significantly less obscured by metallic hardware on tomosynthesis images (8%) than on radiographs (15%) in contrast to the reference standard CT images (2%). They also found a weak but statistically significant correlation between tomosynthesis fracture assessment and patients' subjective pain levels. The authors concluded that given the lower cost and smaller radiation exposure than with CT, tomosynthesis should be considered an alternative imaging method for the assessment of healing.

Tc99m-MDP bone scan is a functional imaging technique in which radiotracer uptake corresponds to blood flow and new bone formation. In cases where callus is absent, the pattern of uptake on bone scan may be helpful in differentiating viable from nonviable non-union, as a photon deficient area is suggestive of nonviability, and the lack of a photon deficient area suggests viability y[42]. **[^{18}F]FDG PET/CT** is reported as a promising tool for diagnoses of infected non-unions [43]. However, PET/CT has not been used in routine fracture follow-up primarily due to its high radiation profile compared with other imaging modalities.

5. AIMS OF THE THESIS

The objectives of this thesis were to develop and evaluate scoring systems for the semi-quantitative assessment of fracture healing by using radiographs (Radiographic Union Score- RUS) and CT scans (Tomographic Union Score- TUS) of surgically treated long bones diaphyseal fractures.

The following were our specific aims:

1. To determine the feasibility and interobserver agreement of a radiographic scoring system in a series of radiographs of surgically treated long bone fractures (part I: RUS).
2. To determine which combination of CT reformats number and planes would represent the best surrogate for the semi-quantitative analysis of the entire CT data (part II: PRE-TUS).
3. To analyse the feasibility and reproducibility of radiographic and CT scoring systems in a series of nailed tibial and femoral fractures during normal healing and to compare scoring results (part III: RUSTUS normal).
4. To analyse the feasibility and reproducibility of radiographic and CT scoring systems in a series of nailed tibial and femoral non-unions and to compare scoring results (part IV: RUSTUS Non-Unions).
5. To investigate the accuracy of the radiographic scoring system to predict fracture non-union (part V: predictive RUS).

6. Part I: **Semi-quantitative radiographic assessment of fracture healing in surgically treated long bone fractures: A feasibility study.**

6.1. Introduction

Conventional radiography remains the examination of choice for bone fracture assessment [44] but limitations in the radiographic evaluation of fracture healing have been demonstrated [31-33]. Over the years, several scoring systems have been developed for a more standardised and objective evaluation of fracture healing [45-48]. More recently Whelan et al developed and validated a radiographic union score for the evaluation of nailed tibial fracture (RUST - Radiographic Union Score for Tibial fracture) [49], which was subsequently applied to other type of bones (see Apendix A).

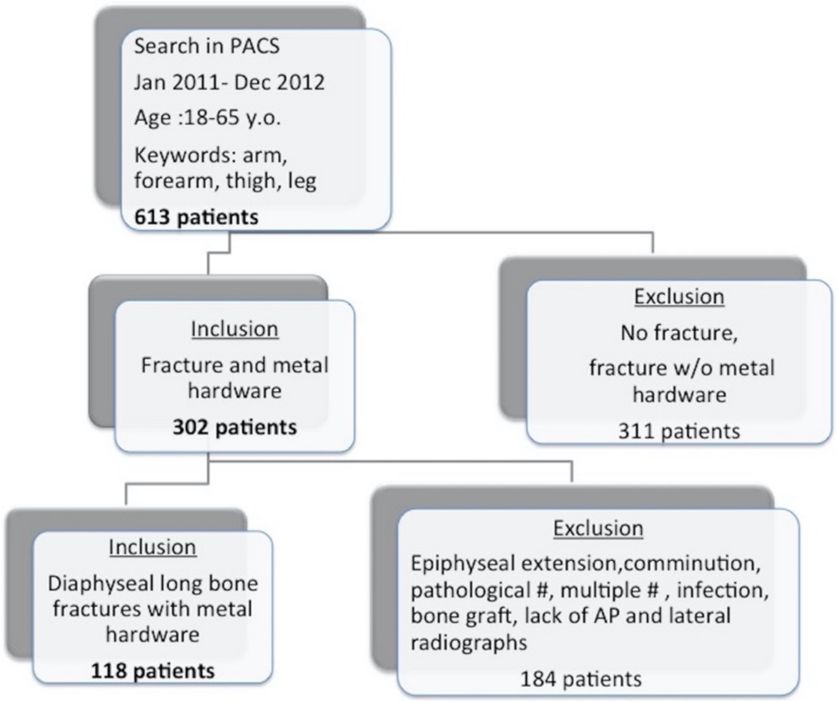
The current retrospective study aimed at defining limitations in the semi-quantitative radiographic evaluation of surgically treated long bone fractures by using radiographs and a Radiographic Union Score (RUS). We hypothesized that the feasibility of this scoring system could be influenced by the type of surgical fixation but not by the age of fracture. The current study also aimed at assessing the interobserver agreement among orthopaedic surgeons and radiologists.

6.2. Methods and materials

6.2.1. Study population

We selected a series of patients aged between 18 and 65 years of age with a surgically treated long bone fracture by searching in our picture archiving and communication systems (PACS) (Carestream Client version 11.3; Carestream Health, Rochester, NY, USA) using the keywords "arm", fore-arm", "thigh" and "leg” among radiographic examinations performed during a 2-year period. The search yielded a series of 613 sets of radiographs. A last-year radiology resident (VP) and a musculoskeletal radiologist with 20 years of experience (BVB) reviewed all radiographs to select patients with a surgically treated diaphyseal fracture and to exclude patients with one of the following exclusion criteria, including lack of surgically treated fracture, pathological fracture, pre-existing metallic hardware or allograft in fractured bone segment, signs of infection or of delayed-union, or missing radiographs (figure 1). The study population consisted of 118 patients including 76 men and 42 women (mean age +/- SD: 41.3 +/- 13.3 years; range: 18-65 years). There were 140 fractures (19 humeri, 19 radius, 17 ulnas, 26 femurs, 42 tibias and 17 fibulas). Thirteen patients had radial and ulnar fractures and nine patients had tibial and fibular fractures. There were 63 nailed fractures (8 humeri, 1 ulna, 21 femurs, 31 tibias, 2 fibulas) and 77 plated fractures (11 humeri, 19 radius, 16 ulnas, 5 femurs, 11 tibias, 15 fibulas). For each of the 140 fractures, a data manager of our department selected one radiographic set among all available sets for each fracture to obtain an equivalent number of radiographic sets obtained at different time delays after the fractures for each type of bone. Time categories were defined as follows: between 0 and 3 weeks after fracture (early stage), between 3 weeks and 3 months after fracture (intermediate stage), between 3 and 6 months after fracture (late stage) and more than 6 months after fracture (healed stage). The 140 sets of radiographs had been obtained at a mean delay of 114 days after fracture (standard deviation 93 days; range 10 - 340 days). There were 35 early, 35 intermediate, 34 late and 36 healed stage radiographs. The authorization of the ethical committee of our institution had been obtained for this retrospective study.

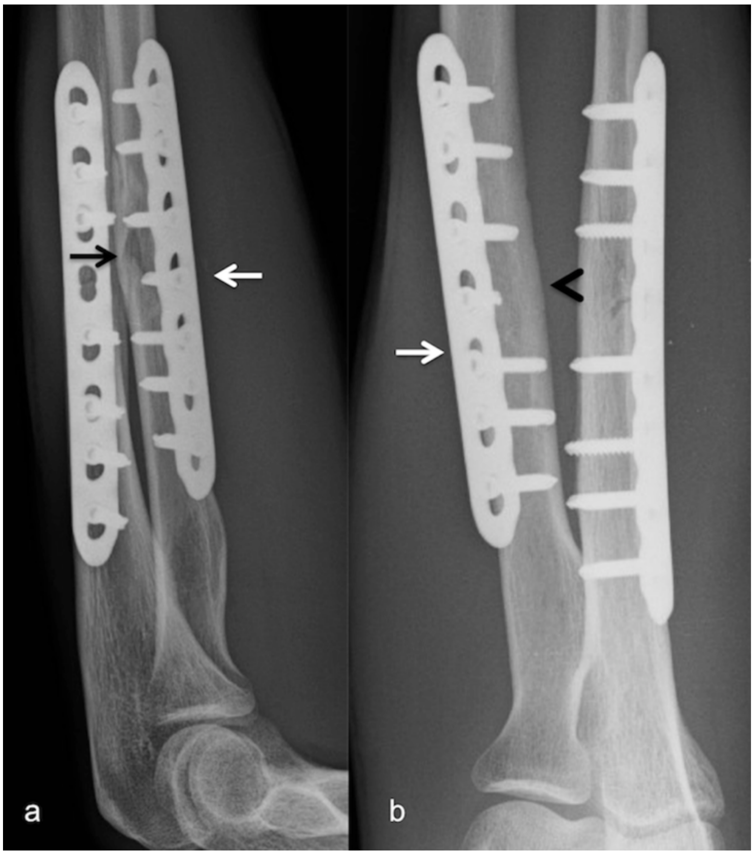
Figure 1: Flowchart for patient selection



6.2.2. Analysis of the radiographs

Analysis of the radiographs was performed by six readers including three radiologists (a last year resident (VP) and two MSK radiologists with three (AL) and twenty (BVB) years of experience) and three surgeons (a last-year resident (TS), and two orthopaedists with five (DP) and twenty-four (JED) years of experience). All readers blinded to clinical findings separately analysed each set of radiographs on PACS workstations. Readers were asked to grade each of the four cortical edges of the fracture at which the X-ray beam was tangent as visible or not on the radiographs. If the fracture edges were visible, cortical scoring was performed as described below. If the fracture edges were not visible, the reader selected at least one of the following causes: (a) superimposed metal hardware, (b) superimposed adjacent bone (leg and fore-arm), (c) another reason such as blurring of the radiograph or inadequate exposure (figure 2). Non-assessability of the cortex due to superimposed metal hardware, or bone were defined respectively by the lack of visibility of the cortical bone to which the X-ray beam was tangent due to superimposed metal, or another bone. The readers were allowed to select metal hardware and cortical bone when these two causes were simultaneously present.

Figure 2: (a) Lateral and (b) AP radiograph of a plated radial fracture. Anterior and lateral cortices (white arrows) cannot be analyzed because of superimposition by metal hardware. Posterior cortex (black arrow) cannot be analyzed because of superimposition by the ulna.Cortical score for medial cortex (arrowhead) was 4 because of remodeled bridging callus (and absence of cortical discontinuity).



A cortical score from 1 to 4 was given to each cortical segment of the fracture site to which the X-ray beam was tangent according to the scoring system by Litrenta et al [50]: score 1: lack of callus; score 2: non-bridging callus; score 3: bridging callus; score 4: remodelled bridging callus.

SCORE	**Callus PRESENT**	**Callus BRIDGING**	**Callus REMODELED**
1	**NO**	**-**	**-**
2	**YES**	**NO**	**ANY**
3	**YES**	**YES**	**NO**
4	**YES**	**YES**	**YES**

For each fracture, the RUS value was calculated by adding the four cortical scores (figures 3-5). Missing data were calculated as the mean of the available data. The RUS was considered to be feasible when at least three of the four cortical scores could be assessed. The score was considered to be non-reliable and therefore not calculated when less than three cortical score could be assessed (figure 6).

Twelve months after the initial analysis, two radiologists (B.V. and V.P.) repeated the same analysis to determine intra-observer agreement.

Figure 3: (a) AP and (b) lateral radiographs of a nailed femur. The score of each cortical area (arrows) was 1 because of the visibility of cortical discontinuity and the lack of callus. RUS value was 4.

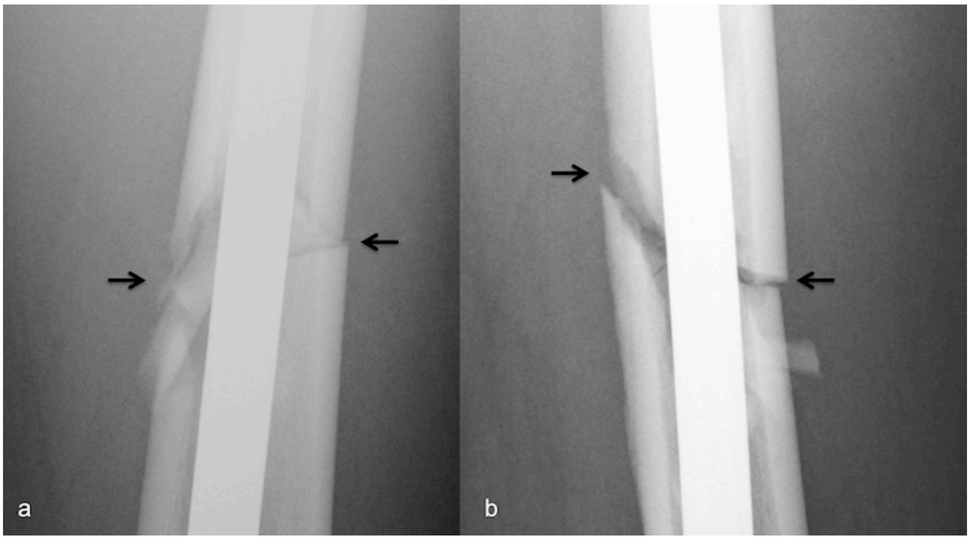


Figure 4: (a) AP and (b) lateral radiographs of a nailed femur. The score of each cortical area (arrows) was 2 because of the visibility of cortical discontinuity and the presence of non-bridging callus. RUS value was 8.

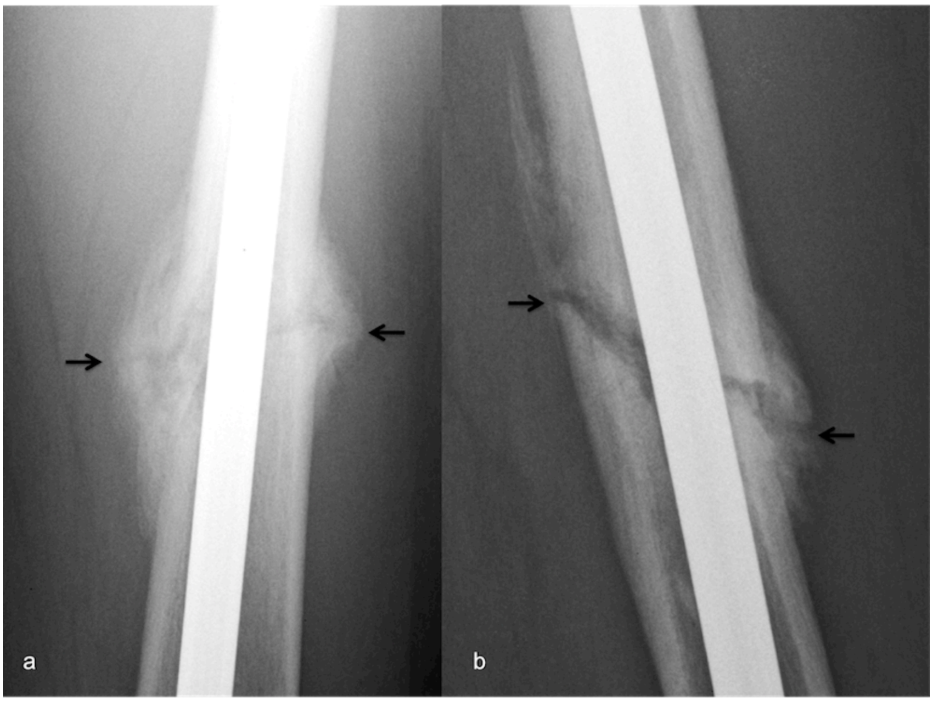


Figure 5: (a) AP and (b) lateral radiograph of a nailed femur. Cortical scores for medial/ lateral cortices (black arrows) and anterior/ posterior cortices (white arrows) were 4 because of presence of remodeled bridging callus. RUS value was 16.

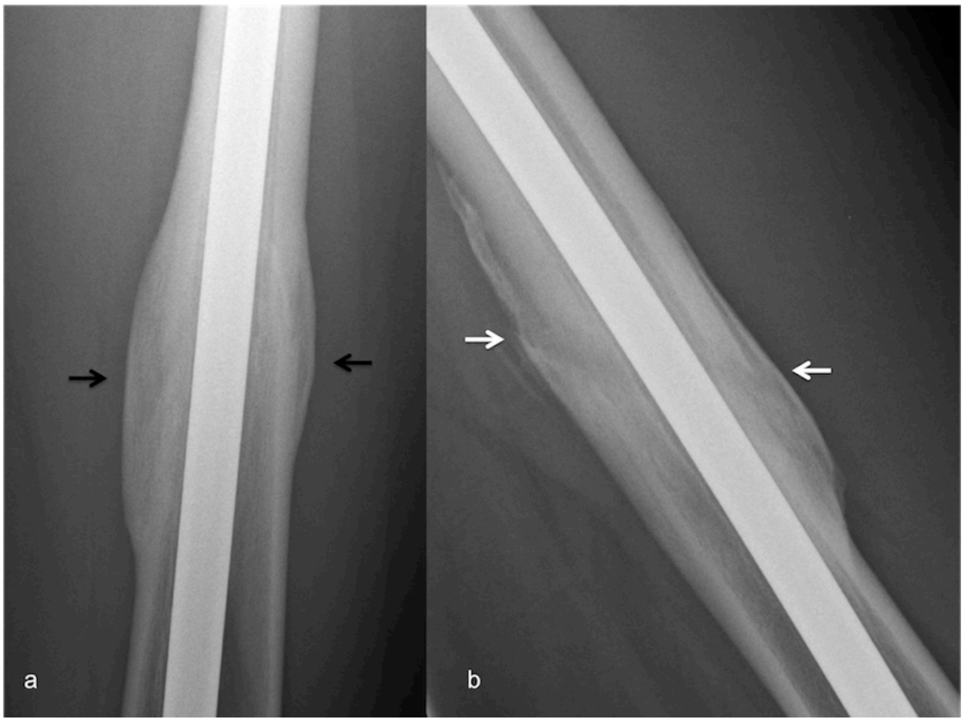
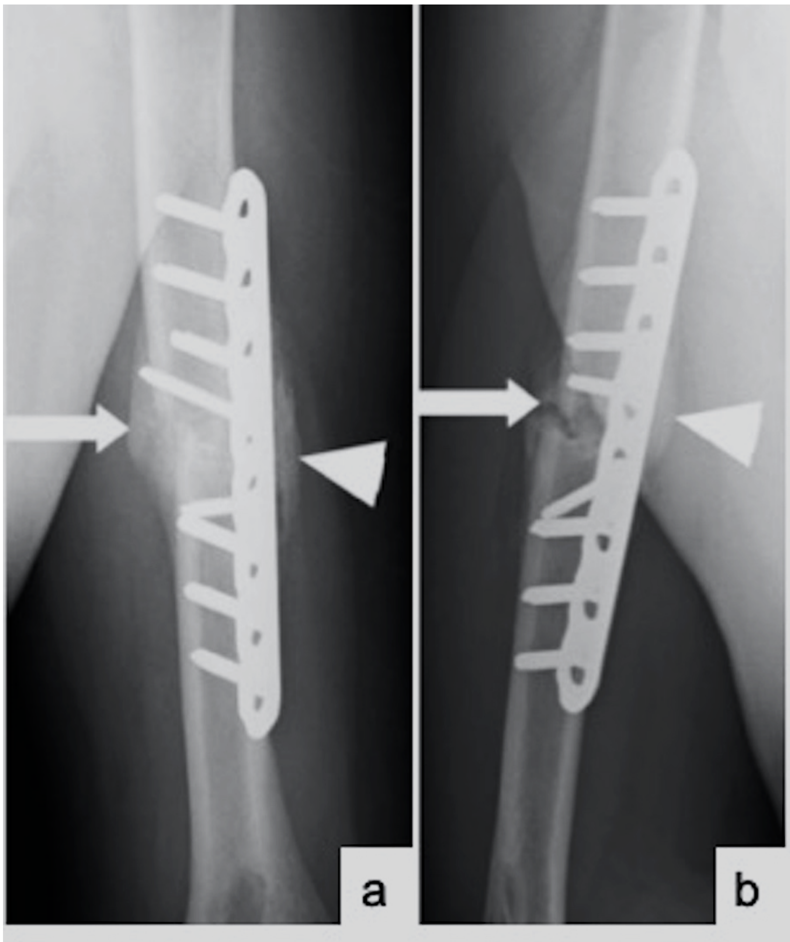


Figure 6: (a) Antero-posterior and (b) lateral radiographs of an arm with a plated humeral shaft fracture. The medial and posterior cortical segments (arrows in a and b) but not the lateral and anterior cortical segments (arrowheads in a and b) of the fracture were assessable. Cortical scores for the medial and posterior cortices were 3 (bridging callus) and 2 (non-bridging callus) respectively. The RUS value was not calculated because of the availability of two out of four cortical scores.



6.2.3. Statistical analysis

The rate of assessability of the fracture edges to which the X-ray beam was tangent in the 140 fractures was calculated for each reader. The rate of feasibility of the scoring was calculated for each reader and compared according to patient's gender and age, fracture age, nailed vs. plated fractures, upper vs. lower limbs and proximal vs. distal bones by using a Chi-Square test. We performed a regression analysis to evaluate the influence of the aforementioned parameters considered as independent variables on the feasibility of the RUS by using stepwise regression analysis. The interobserver agreement was evaluated using intraclass correlation coefficients (ICC) for the assessability of the fracture edges, for the feasibility of the RUS feasibility as well as for RUS values. The intraobserver agreement for the RUS feasibility was evaluated for two radiologists using ICCs. The interobserver and intraobserver agreement were interpreted according to Landis and Koch [51] with ICC values <0.00 defined as poor, 0.00-0.20 defined as slight, 0.21-0.40 defined as fair, 0.41-0.60 defined as moderate, 0.61-0.80 defined as substantial, and values of 0.81-1.00 defined as almost perfect [52]. The RUS scores were compared between readers as a surrogate of the quality of the analysis performed by the six readers. Data were processed using a statistical software package (MedCalc version 11.6; MedCalc Software, Mariakerke, Belgium). For all tests, P-values < 0.05 were considered to be statistically significant.

6.3. Results

6.3.1. Assessability of the fracture edges on radiographs

The rate of assessability of the 560 fracture edges to which the X-ray beam was tangent on the AP and lateral radiographs of the 140 fractures ranged between 73% to 76% among radiologists and between 71% to 81% among surgeons (the fracture edges were non-assessable in 19 to 29% of cases)(Table 1). Non-assessability of the cortical segments was due to overlapping metal hardware in 79% to 95% and to overlapping cortical bone in 5% to 13% according to readers (Table 1). Interobserver agreement for the assessability of the cortical segments was substantial for the radiologists (ICC=0.69 (95%CI : 0.66-0.73)), the surgeons (ICC=0.69 (95%CI :0.65-0.73)) and for all readers (ICC=0.69 (95%CI :0.66-0.72)).

Table 1: Frequency of assessability and of non-assessability of the 560 fracture edges on the 140 radiographic sets for three radiologists (R1, R2, R3) and three orthopedic surgeons (O1, O2, O3). Frequency of causes for non-assessability of cortical fracture segments are given for each reader.

	Frequency of		**Causes for non-assessability**		
	Assessability	**Non-assessability**	**Metal hardware superimposed**	**Cortical bone superimposed**	**Metal and bone superimposed**
R1	407 (73%)	153 (27%)	120 (79%)	8 (5%)	25 (16%)
R2	409 (73%)	151 (27%)	143 (95%)	8 (5%)	0 (0%)
R3	426 (76%)	134 (24%)	110 (82%)	7 (5%)	17 (13%)
O1	420 (75%)	140 (25%)	112 (80%)	10 (7%)	18 (13%)
O2	398 (71%)	162 (29%)	141 (87%)	19 (12%)	2 (1%)
O3	452 (81%)	108 (19%)	94 (87%)	14 (13%)	0 (0%)

Results are raw data followed by proportions (%) in parentheses.

R1: last year resident in Radiology; R2: MSK radiologist with 3 years of experience; R3: MSK radiologist with 20 years of experience; O1: last year resident in orthopedic surgery; O2 orthopedic surgeon with 5 years of experience; O3: orthopedic surgeon with 24 years of experience.

6.3.2. Feasibility of the radiographic score

The radiographic scoring RUS was feasible (more than two fracture edges assessable per fracture) in 59% to 69% of the 140 fractures according to the radiologists and in 58% to 75% according to the surgeons (Table 2). Interobserver agreement for the feasibility of the RUS was substantial for the radiologists (ICC: 0.74 (95% CI: 0.67- 0.80)), the surgeons (ICC :0.66 (95% CI: 0.57- 0.74)) and for all readers (ICC :0.72 (95% CI: 0.66- 0.78). Intraobserver agreement for the feasibility of the RUS was substantial for the two readers (ICC=0.74 (95%CI :0.63-0.85) and ICC=0.74 (95%CI: 0.62-0.86), for reader 1 and 2 respectively).

For all readers, the feasibility of the RUS was statistically significantly higher in nailed (90-97%) than in plated (31-56%) fractures ($p<0.001$), in proximal (78-89%) than in distal (47-61%) bones (all $p\leq0.01$), and in lower (76-91%) than in upper (27-49%) limb fractures ($p<0.001$)(Table 2). The feasibility of the RUS did not vary with fracture age (all $p\geq52$), patient's age (all $p\geq0.09$) and sex (all $p>0.05$). According to the logistic regression analysis, the unique parameter that had a statistically significant influence on the feasibility rate of the RUS was the type of stabilizing material ($p<0.001$ for all readers).

Table 2: Frequency (%) of feasibility of the RUS for three radiologists and three orthopedic surgeons in the entire series (140 fractures) and in fracture groups according to the type of stabilizing material, topography of bones and type of limb. The RUS was considered to be feasible when more than 2 cortical segments were assessable on the radiographs.

	All	**Type of material**			**Topography of bone**			**Limb segment**		
	(n= 140)	**Nails (n=63)**	**Plates (n=77)**	***p***	**Proximal (n=45)**	**Distal (n=95)**	***p***	**Upper (n=55)**	**Lower (n=85)**	***p***
R1	61 %	92%	35%	$<10^{-3}$	78%	52%	$<10^{-2}$	35%	78%	$<10^{-3}$
R2	59 %	90%	32%	$<10^{-3}$	80%	48%	$<10^{-2}$	31%	76%	$<10^{-3}$
R3	69 %	92%	49%	$<10^{-3}$	84%	61%	$= 10^{-2}$	42%	86%	$<10^{-3}$
O1	61 %	95%	32%	$<10^{-3}$	80%	52%	$<10^{-2}$	31%	80%	$<10^{-3}$
O2	58 %	90%	31%	$<10^{-3}$	80%	47%	$<10^{-2}$	27%	78%	$<10^{-3}$
O3	75 %	97%	56%	$<10^{-3}$	89%	67%	$<10^{-2}$	49%	91%	$<10^{-3}$

6.3.3. Interobserver agreement for cortical scores and RUS values

Interobserver agreement for cortical scores was substantial for the radiologists (ICC=0.63 (95%CI:0.59-0.67)), the surgeons (ICC=0.65 (95%CI:0.61-0.68)) and for all readers (ICC=0.63 (95%CI:0.60-0.66)). Interobserver agreement for RUS values was almost perfect for the surgeons (ICC=0.80 (95%CI:0.73-0.86)) and substantial for the radiologists (ICC=0.76 (95%CI:0.68-0.83)) and for all readers (ICC=0.79 (95%CI:0.72-0.83)).The interobserver agreement for the RUS values was substantial for surgeons and radiologists in almost all fracture groups according to different parameters (Table 3). ICC values for the intraobserver agreement of RUS values were 0.68 (95% CI: 0.60-0.76) for reader 1 and 0.64 (95% CI: 0.56-0.72) for reader 2.

Table 3: Interobserver agreement of RUS values according to parameters among orthopedic surgeons, radiologists and among all readers.

ICC	Age		Gender		Fracture Age			
	<50	>50	male	female	early	Interm	Late	healed
Surgeons	0.82	0.76	0.80	0.82	0.77	0.65	0.74	0.76
Radiologists	0.75	0.78	0.76	0.78	0.59	0.65	0.77	0.58
All readers	0.79	0.78	0.79	0.80	0.71	0.67	0.78	0.66

ICC	Surgical Material		Type of bone		Type of Limb	
	nails	plates	prox	distal	upper	Lower
Surgeons	0.80	0.82	0.78	0.81	0.78	0.81
Radiologists	0.80	0.66	0.75	0.80	0.84	0.75
All readers	0.80	0.76	0.77	0.80	0.82	0.79

6.3.4. Correlation of RUS values with fracture features

The RUS was statistically significantly different between subgroups of fractures of different fracture age with an increase in score value according to fracture age ($p<0.001$ for all readers, correlation coefficients varying from 0.39 to 0.55). There was no statistically significant difference in the RUS values according to patient's gender, age, and type of surgical material, bone and limb.

6.4. Discussion

> Our study of 140 nailed and plated long bone fractures and six readers identified three important findings:
>
> - Feasibility rates of the Radiographic Union Score (RUS) varied between 58% to 75% among readers.
> - Feasibility was influenced by the type of stabilizing hardware with a higher rate in nailed than in plated fractures.
> - Interobserver agreement when using RUS was substantial.

Assessment of the degree of fracture healing is an every-day challenge in clinical practice that requires clinical and radiological criteria [3; 21; 53; 54]. Several scales and classification systems have been proposed [45-49] but the feasibility of radiographic scoring has barely been addressed in the literature[45]. Given the development of new treatment options, a quantitative system of fracture healing evaluation is warranted in order to monitor treatment's efficiency.

The current retrospective study of 140 surgically treated long bone fractures identified several important findings. First, we demonstrated that rigorous analysis of the 4 cortical segments at which the X-ray beam was tangent on the AP and lateral radiographs was not feasible in about 25% of cases of all cortical segments with variations from 19% to 29% among readers. To the best of our knowledge, our study is the first to objectively quantify this limitation. Most likely, this relatively high rate of non-feasibility partially accounts for difficulties in the analysis of radiographs in routine clinical practice[55; 56].

Second, our study highlighted the fact that impaired visibility of cortical segments was mainly related to overlapping metal hardware, with limited role of overlapping cortical bone. The relative frequency of these 2 causes was likely related to the type of bones and of stabilizing material, as discussed below. Patel et al [57] found a moderate interobserver agreement among surgeons in the evaluation of plated distal radial fractures and noted that the difference in inter-observer agreement among surgical and nonsurgical cases can likely be explained by the presence of hardware that can obscure bony detail and interfere with assessment of callus and fracture lines.

Third, our study demonstrated that the RUS was feasible in 58% to 75% of cases among readers. In the current study, the RUS was considered to be feasible if at least 3 cortical segments were visible among the 4 expected cortical segments to which the X-ray beam is tangent. It is generally accepted that bridging of fracture at three cortical points implies fracture healing [20; 53; 58] indicating that at least three segments need to be analyzed for the assessment of fracture healing. Comparison with previous studies is limited, as feasibility of the RUS has not been formally addressed.

Fourth, the feasibility rate of the RUS was statistically significantly higher in nailed than in plated fractures, in proximal than in distal bones and in lower than in upper limb bones. Fracture age did not appear to influence the feasibility rate. According to the logistic regression analysis, the type of surgical material i.e. nails or plates was the only parameter that had a statistically significant impact on the feasibility rate of the RUS. The most likely explanation is that medullary nails are centrally located and therefore do not interfere with the projection of the most eccentrically located cortical segments to which the incident X-ray beam is tangent. On the other hand, plates are excentrally positioned on the cortical surface and may interfere with its visibility on radiographs depending on the width of the plate and of the plated bone. This finding highlights the essential role of the type of fixation in the application of a global scoring system in the assessment of healing. Litrenta et al [50] also discussed that the difference in agreement and score values at union between plated and nailed fractures in his study likely reflects differences in fixation material (the presence of a plate obscures the evaluation of one or more cortices).

Finally, interobserver agreement in determining the cortical scoring, the RUS values and the RUS feasibility among orthopedic surgeons, radiologists and among all readers was substantial (ICC=0.63-0.64 for cortical scoring, 0.76-0.80 for the RUS, 0.66-0.74 for the RUS feasibility). Intraobserver agreement for RUS values and RUS feasibility was also substantial for the two radiologists who analyzed twice the radiographs. Our global ICC value of 0.79 for the RUS compares favorably with other studies in which inter-observer agreement was analyzed. Whelan et al calculated an ICC value of 0.86 in a study where 7 surgeons analyzed 45 nailed tibial fractures [49]. Kooistra et al mentioned an ICC of 0.84 in a series of 549 tibial fractures analyzed by 3 orthopedic surgeons [59]. Chiavaras et al calculated an ICC of 0.66 in scoring 150 intertrochanteric femoral fractures with 3 radiologists and 3 orthopedic surgeons using a scoring system based on similar principles [60; 61]. Oliver et al [62] calculated an ICC of 0.79 in scoring 60 humeral shaft fractures analysed by 2 surgeons. We acknowledge that the comparison of our results with those of cited ones is of limited value because of different study material. Actually, the current study included different bone type, fracture age and surgical treatment whereas the other studies focused on well-defined subtypes of fractures treated with a single surgical or conservative procedure. Nevertheless, interobserver agreement was substantial in all studies despite the differences.

The current study on surgically treated fractures deliberately included various types of bone, of fracture type and age and of stabilizing metal hardware, reflecting the diversity of clinical practice. We acknowledge some inconsistency in applying a callus-based scoring system in plated fractures. Actually, periosteal callus formation occurs in nailed fractures and is rather an unexpected feature in plated fractures because of different biomechanical environment [4; 63-65]. The rigid plate fixation may not result in callus formation due to the lack of micro- movement between fracture ends. In plated fractures, bridging occurs through direct cortical bone formation with limited, if any, contribution from the periosteum. Furthermore, if there is any callus, it may be inhibited in the cortical segment under the plate. Therefore, the limited feasibility of the scoring method in plated fractures might be of limited clinical significance.

In the current study, RUS values could not be compared with healing status as determined by clinical evaluation because of the retrospective design of our study. The fact that the RUS values positively correlated with fracture age might indicate that this scoring system could be used for fracture healing evaluation. Further prospective study is warranted to assess the RUS in clinical practice.

Several limitations to the current study should be noted. First, it is a retrospective study with a limited number of patients. Second, our results are based on the analysis of AP and lateral radiographs of surgically treated diaphyseal long bone fractures in patients aged between 18 and 65 years. Results could be possibly different in conservatively treated fractures and in other patient population. It must be emphasized that several fracture subtypes such as diaphyseal fractures with epiphyseal extension and comminuted diaphyseal fractures were excluded from this series and must be further studied. The added value of oblique radiographs in selected situations was not addressed.

In conclusion, the current preliminary study confirmed our hypothesis that the feasibility of scoring was limited on radiographs in surgically treated long bone fractures and that the type of stabilizing hardware was the predominant limiting factor with a higher feasibility rate in nailed than in plated fractures. As a consequence, the value of two orthogonal radiographs for the assessment of plated fractures must be questioned and the added value of oblique radiographs or of tomographic techniques should be addressed. The current study also confirmed a substantial agreement between three orthopedic surgeons and three musculoskeletal radiologists, even with different level of experience. Further studies are warranted to evaluate the feasibility and interobserver agreement of scoring by using other imaging modalities. A tomographic Union Score (TUS) based on similar morphological criteria as the RUS is developed in order to optimize the semi-quantitative assessment of fracture healing. Further studies are needed to correlate the quantitative imaging data with clinical information.

7. Part II: **Semi-quantitative CT assessment of fracture healing: how many and which CT reformats should be analyzed?**

7.1. Introduction

Determining when a fracture is healed remains a challenge in orthopedic practice. In a survey of 444 orthopedic surgeons, Bhandari et al. demonstrate a lack of consensus regarding both clinical and radiologic criteria used to assess healing as well as significant differences in definitions of delayed union and malunion [18; 21]. Although variability in definitions exists, surgeons agree that the assessment of healing requires both radiographic and clinical criteria [18; 20; 66]. Analysis of radiographs provides an overall estimation of fracture healing albeit with limited inter-observer and intra-observer reproducibility [33; 53; 58] most likely due to limitations associated with the projectional nature of radiographic image formation and overlapping metal hardware [67]. Computed tomography (CT) enables to assess the presence and extent of bone callus at fracture edges thanks to its multiplanar capacity and high spatial resolution at the cost of an additional examination and increased ionizing radiation burden [22; 34; 45; 68; 69]. Based on only two imaging dimensions, overlapping structures (eg, bones, metal hardware, cast) might lead to a missing of persistent fractures and a reduced ability to assess the extent of fracture healing and stability by CR. Compared to radiography, CT overcomes the problem of potentially overlapping structures, visualizes bone bridging directly, and enables a more sensitive detection of callus formation. Direct imaging of bone bridging and callus formation has become possible.

In clinical practice, CT is used as a second-line imaging modality either for better depiction of fracture anatomy or for complicated cases when abnormal healing or infection are suspected with non-conclusive clinical or radiographic information. CT was reported to show 100% sensitivity for detecting non-union with high inter-observer agreement among readers [34]. However, computed tomography scanning had limited specificity and tended to lead to a diagnosis of nonunion when, indeed, the fracture was healed. Grigoryan et al performed a qualitative and quantitative assessment of fracture healing in a longitudinal prospective study using x-ray and CT imaging [45]. Qualitative assessment included fracture line, fracture margins, fracture gap, first appearance of callus, bridging of involved cortices and fracture union based on bridging of at least ¾ cortices. Quantitative assessment of CT density changes (HU-Hounsfield units) in the fracture gap was performed in fractures without metal hardware. The study demonstrated that the radiographic evaluation of fracture line and edges was limited because of fixation hardware superimposition. They demonstrated that CT images allowed for more complete and detailed visualization of healing compared to CR. Krestan et al analyzed coronal and sagittal CT reconstructions and radiographs to determine bone healing using a semiquantitative approach (complete: bone bridges with no gap; partial: some bone bridges with gaps between; and no bridging, no osseous bridges) [69]. In 16/43 patients, MDCT altered the diagnosis initially made on digital radiography: in eight patients (19%) digital radiographs underestimated the extent of bone healing, whereas in another eight patients (19%) they overestimated the degree of fusion. However, the methodology to assess the large amount of data afforded by CT has not been clearly defined yet [70].

We hypothesized that the analysis of a subset of CT reformats could be sufficient to provide a representative overview of the entire fracture. The current study aimed at determining the best combinations of CT reformat numbers and planes that would provide a reliable overview of non-united fractures in a series of patients with non-united surgically treated femoral or tibial fractures.

7.2. Material and methods

7.2.1. Study population

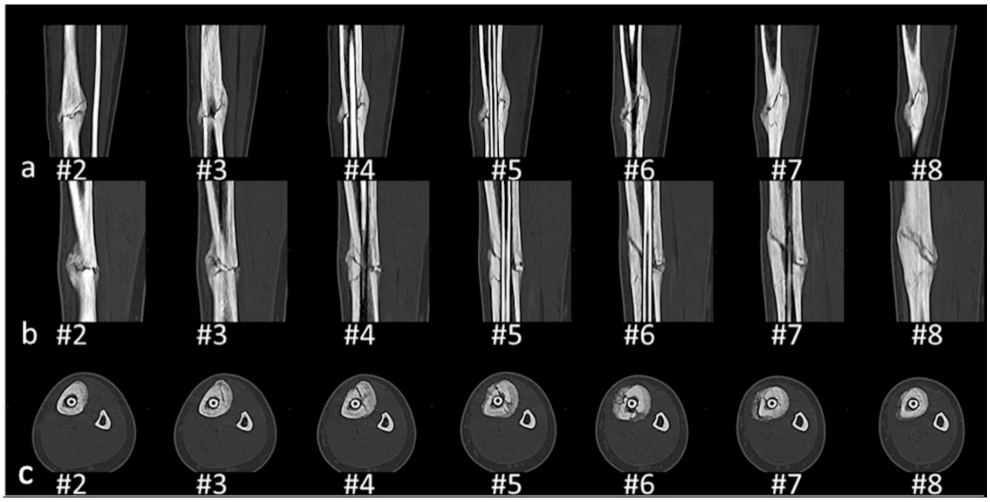
Authorization of our institutional ethic committee for this retrospective study was obtained and authorization from patients was waived. A search over a period of four years was performed in our picture archiving and communication systems (PACS)(Carestream Client version 11.3; Carestream Health, Rochester, NY, USA) by a fellow in musculoskeletal radiology (VP) with 2 years of experience to select patients aged between 18 and 65 years with a surgically treated tibial or femoral shaft fracture in whom CT had been performed and in whom non-union was later demonstrated at surgery. Exclusion criteria were (a) healed fractures, (b) fractures of less than 6 months, (c) non-united fractures that had been treated with allo- or auto-grafts and (d) fractures with imaging or biological signs of infections. Our study population consisted of 17 male and 12 female patients (mean +/- standard deviation (SD): 44.44 years +/- 14.4; range 19-64 years). There were 29 cases of non-unions in 12 nailed and four plated tibial fractures and in 12 nailed and one plated femoral fractures. The mean time delay (+/-SD) between fracture occurrence and CT examination was 595 days (+/- 482) (range: 180 - 1983 days). The mean (+/- SD) time delay between the CT examination and surgery was 168 days (+/- 241) (range: 1 – 1063 days).

7.2.2. Imaging protocols

All unenhanced CT examinations were performed on a 40-detector row CT scanner (Somatom Definition AS; Siemens Healthcare, Forchheim, Germany), patients lying supine with the feet first. Data acquisition was obtained over the entire fractured bone using the following tube parameters: tube voltage, 120 kVp; reference tube current-time product, 275 mAs. Effective current time-product, 178-305 mAs by applying automated tube current modulation (Care Dose 4D; Siemens Healthcare); detector configuration, 40X0.6 mm; pitch, 0.8; gantry rotation time, 1 s. CT images were reconstructed with a section width and increment of 0.75 mm using standard bone (B70h) and soft tissue (B41s) deconvolution kernels and feed-back retroprojection calculation. Field-of-view varied from 12x12cm to 20x20 cm. No iterative reconstruction algorithms were used. All reconstructions were stored in our PACS.

Between January and March 2015, the same fellow obtained sagittal, coronal and transverse reformats in a standardized manner as follows. For each case, the axial images obtained with the bone deconvolution kernel were uploaded in the multiplanar reconstruction software available in our PACS. A stack of nine 1.5 mm-thick reformats in the sagittal, coronal and transverse planes were obtained according to anatomic landmarks. For the coronal and sagittal planes, the first (#1) and last (#9) reformats were placed in contiguity with the target bone which was only visible on reformats #2-8. For the transverse plane, the first (#1) and last (#9) reformats were positioned above and below the fracture level and fracture sites were visible only on reformats #2-8 (figure 1). The slice intergap (mean+/- standard deviation) subsequent to the positioning of the stack of reformats in order to include the entire fracture was 2.2 mm +/- 0.4 for the long axis reformats and 8.7mm+/-4.9 for the transverse reformats. A data manager anonymized all the stacks, labelled them in a random order by using numericals and stored them in a dedicated research PACS.

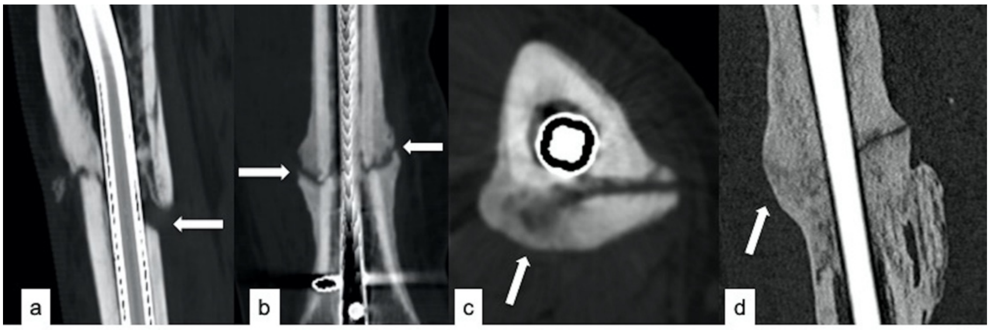
Figure 1: (a) Coronal, (b) sagittal and (c) transverse CT reformats of a nailed non-united tibial fracture in a 55-year-old man. Lateral (#2 and #8), paralateral (#3 and #7), paramedian (#4 and #6) and median (#5) CT reformats were obtained in the three orthogonal planes.



7.2.3. Image analysis

Between May and October 2016, the fellow in musculoskeletal radiology and a musculoskeletal radiologist with 24 years of experience independently scored the CT reformats on a PACS workstation with adapted window width and level values. Coronal, sagittal and transverse CT reformats were analysed separately during three different reading periods separated by a one-month interval. On each of the seven CT reformats #2 to #8, readers scored two fracture sites by using a Tomographic Union Score (TUS) based on the callus appearance on the CT reformat. The TUS was derived from the previously described Radiological Union Score (RUS) by Litrenta et al who scored fracture healing on radiographs [50]. The score at the fracture edges was defined as follows: score 1: no callus; score 2: discontinuous callus, score 3: continuous bridging immature callus; score 4: continuous remodelled callus (figure 2).

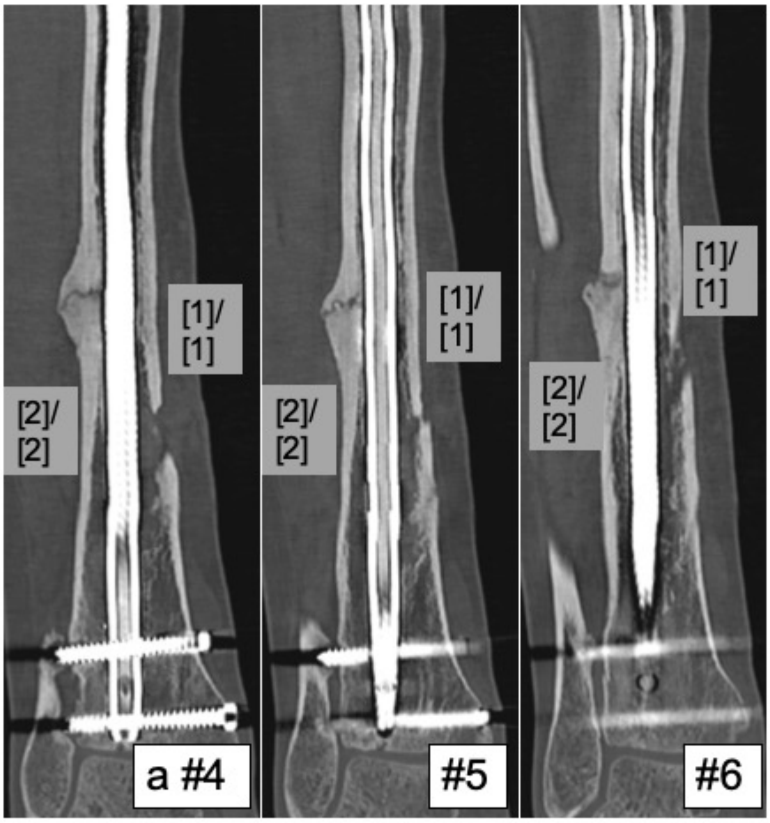
Figure 2: The Tomographic Union Score (TUS)- score 1: no callus (a); score 2: discontinuous callus (b), score 3: continuous bridging immature callus (c); score 4: continuous remodelled callus (d) as indicated in arrows.

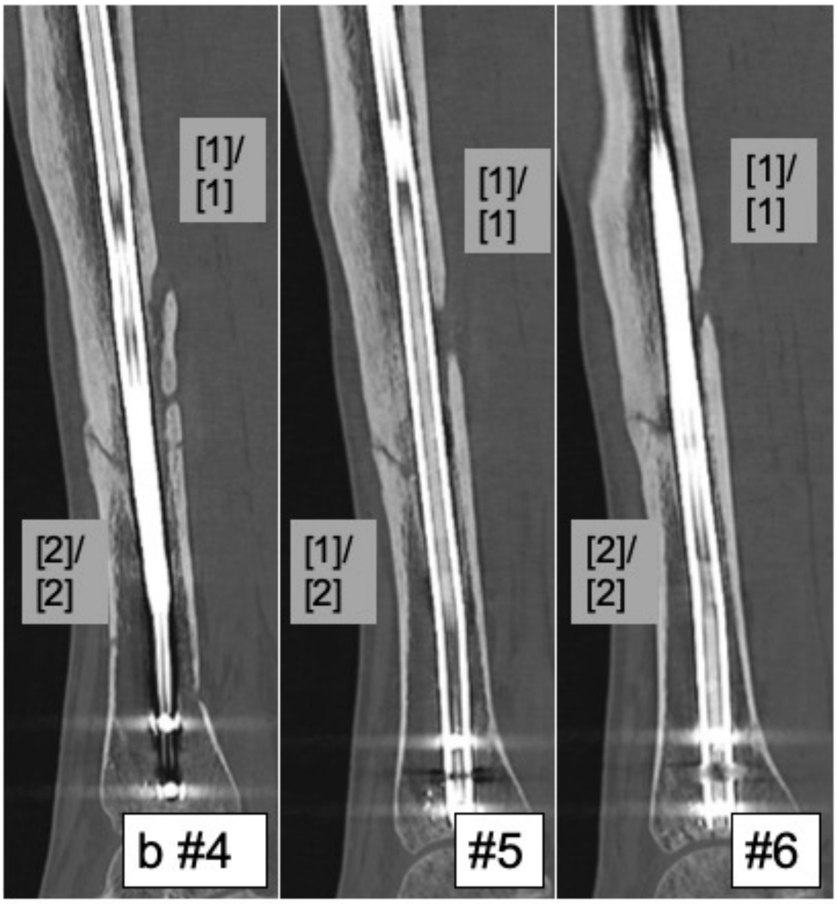


On each CT reformat, two fracture sites had to be scored (one score for the anterior and one for the posterior cortex on the sagittal reformats, one score for the medial and one for the lateral cortex on the coronal reformat, two scores for the transverse reformat). When several fracture sites were detected, the two least healed sites were scored. Scoring was considered to be non-feasible when metal artifacts or metal hardware obscured the fracture, when one of the fracture edges was not visible due to bone deformity or when the fracture plane was not visible mainly when the orientation of the reformats was parallel to that of the fracture (figures 3-4-5). For each case, a TUS^{Sa} was derived from the scoring of the seven sagittal reformats, a TUS^{Co} was derived from the seven coronal reformats and a TUS^{Tr} was derived from the seven transverse reformats. TUS could range from 14 to 56 because 14 scores ranging from 1 to 4 were derived from the seven reformats and added. In our study, the standard of reference for each fracture was the mean of TUS^{Sa} , TUS^{Co} and TUS^{Tr}.

In January 2017, the same two readers reviewed separately only the median sagittal, coronal and transverse CT reformats without seeing the other reformats of the same stack or of the other planes to assess intra-observer and inter-observer agreement in scoring.

Figure 3: (a) Median (#5) and paramedian (#4 and #6) coronal, (b) sagittal and (c) transverse CT reformats of a nailed non-united tibial fracture. Numbers in brackets indicate scores given by reader 1 / reader 2, respectively. The ratio between the sum of the scores and the maximum score obtained on the three median reformats (9/24 for both readers) is equal to that obtained on the two paramedian reformats (18/48 for both readers).




[1]/
[1]
[2]/
[2]
b #4
[1]/
[1]
[1]/
[2]
#5
[1]/
[1]
[2]/
[2]
#6

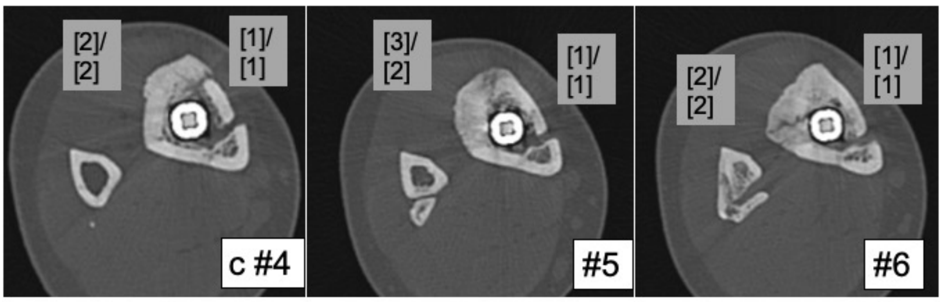
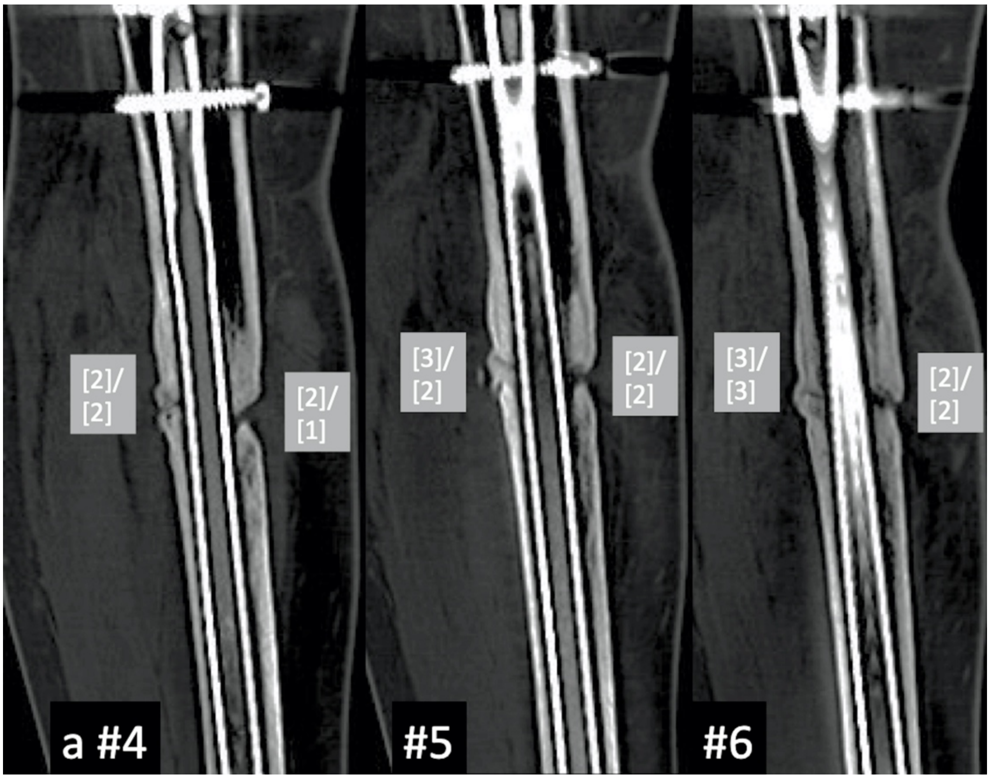
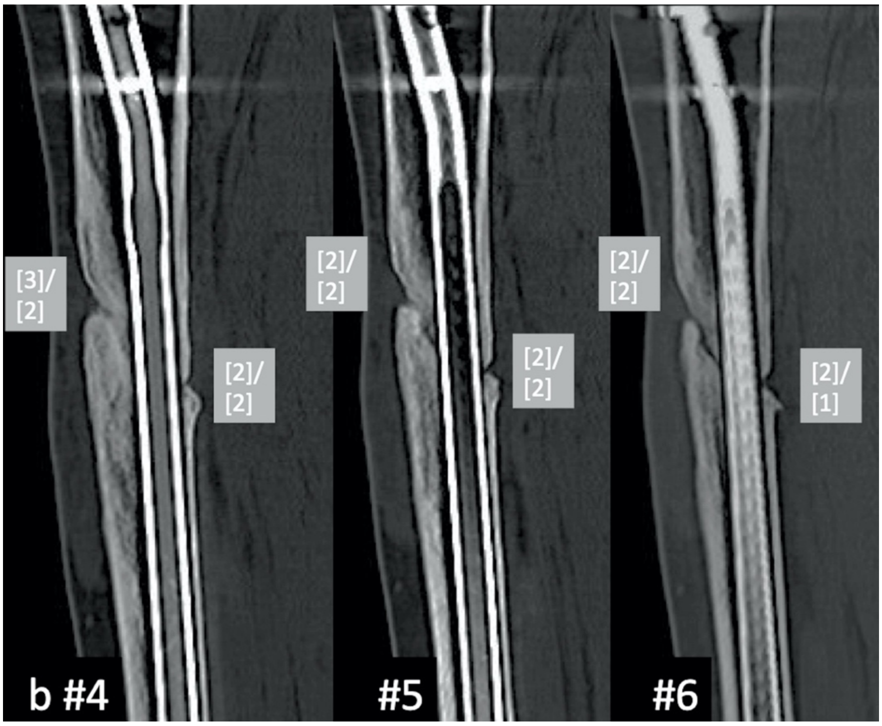

[2]/
[2]
[1]/
[1]
c #4
[3]/
[2]
[1]/
[1]
#5
[2]/
[2]
[1]/
[1]
#6

Figure 4: (a) Median (#5) and paramedian (#4 and #6) coronal, (b) sagittal and (c) transverse CT reformats of a nailed non-united tibial fracture. Numbers in brackets indicate scores given by reader 1 / reader 2, respectively. The ratio between the sum of the scores and the maximum score obtained on the three median reformats (9/16 for reader 1 and 8/16 for reader 2) is very close to that obtained on the two paramedian reformats (22/48 for reader 1 and 20/48 for reader 2). Non feasible (NF) scores are noted on transverse reformats because of metal artifacts.




[3]/
[2]
[2]/
[2]
[2]/
[2]
[2]/
[2]
[2]/
[2]
[2]/
[1]
b #4
#5
#6

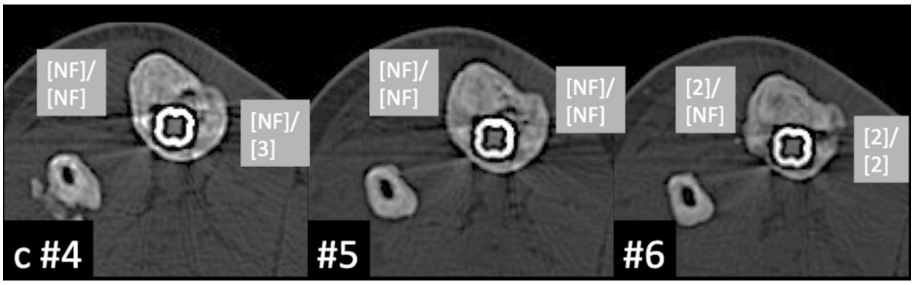
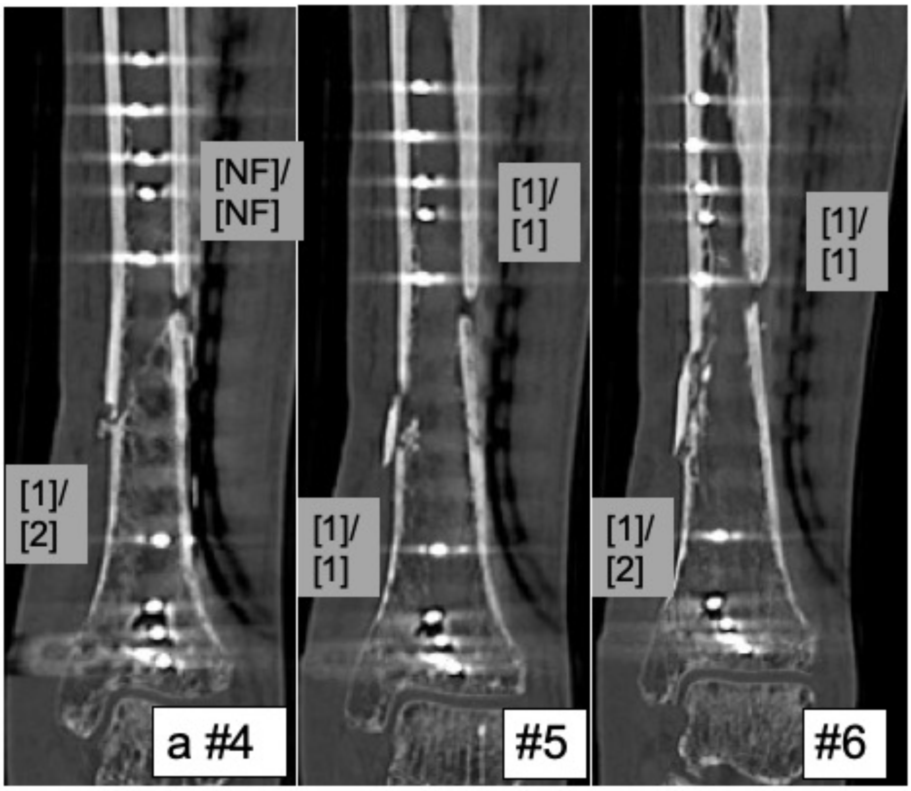
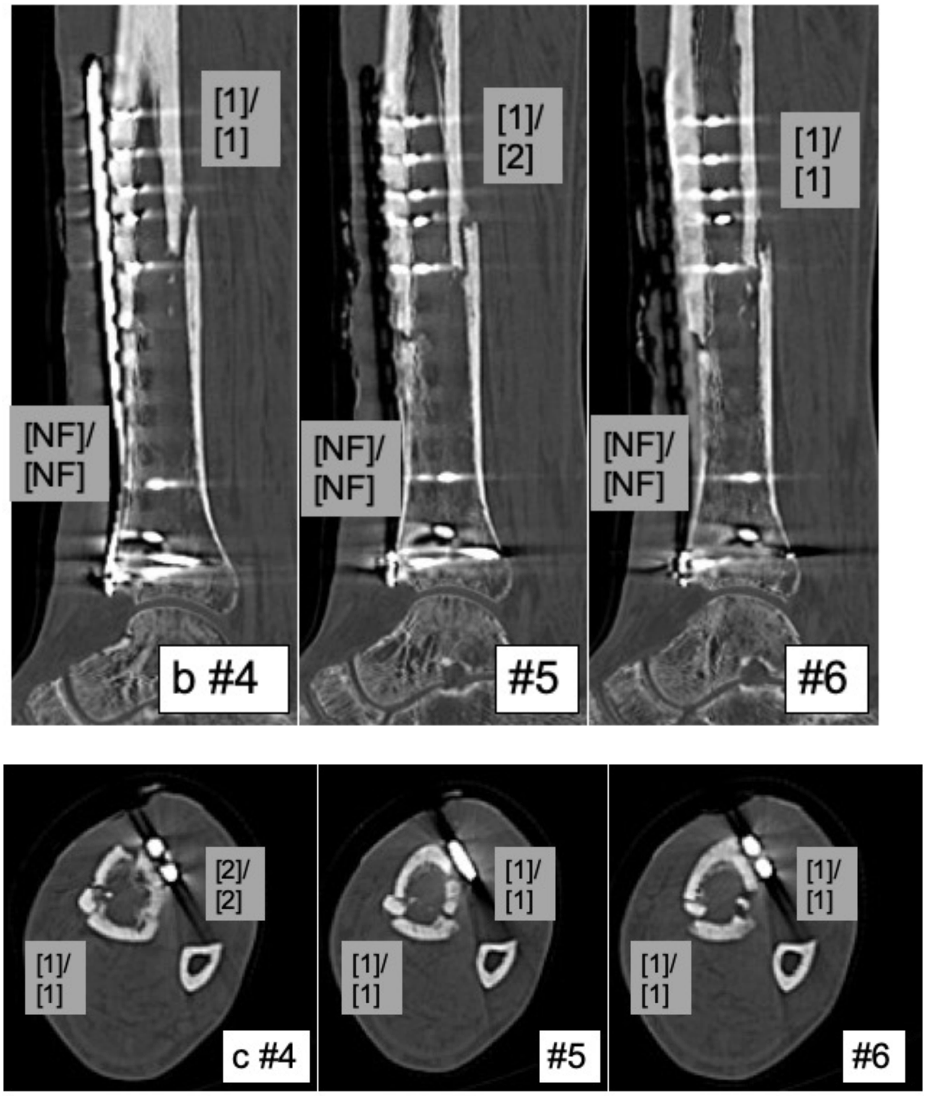

[NF]/
[NF]
[NF]/
[3]
[NF]/
[NF]
[NF]/
[NF]
[2]/
[NF]
[2]/
[2]
c #4
#5
#6

Figure 5: (a) Median (#5) and paramedian (#4 and #6) coronal, (b) sagittal and (c) transverse CT reformats of a plated non-united tibial fracture. Numbers in brackets indicate scores given by reader 1 / reader 2, respectively. The ratio between the sum of the scores and the maximum score obtained on the three median reformats (5/20 for reader 1 and 6/20 for reader 2) is very close to that obtained on the two paramedian reformats (10/36 for reader 1 and 12/36 for reader 2). Non feasible (NF) scores are noted on coronal and sagittal reformats because of metal artifacts et metal hardware superimposition.




[1]/
[1]
[NF]/
[NF]
b #4
[1]/
[2]
[NF]/
[NF]
#5
[1]/
[1]
[NF]/
[NF]
#6
[2]/
[2]
[1]/
[1]
c #4
[1]/
[1]
[1]/
[1]
#5
[1]/
[1]
[1]/
[1]
#6

7.2.4. Statistical analysis

The rate of feasibility of the cortical scoring was determined for the two readers on the three planes and differences in frequency of non-feasibility were compared between the three planes. Intra- and inter-observer reproducibility of the cortical scoring of the median CT reformats of the three planes was assessed by determining the intra-class correlation coefficient (ICC) based on a two-way model after exclusion of the non-assessable areas. Inter-observer agreement for TUS^{Sa}, TUS^{Co} and TUS^{Tr} was assessed by determining the intra-class correlation coefficient (ICC) based on a two-way model after exclusion of the non-assessable areas for all slices obtained in each plane.

Thirty-three models including different combinations of one to seven reformats and one to three planes were constructed (Table 1,2). The TUS calculated for each of these models were compared with the TUS derived from the scoring of all reformats in the three planes by calculating ICC and median difference in total score. Missing cortical scores were calculated as the mean of the available scores from the same plane.

Finally, statistically significant differences in ICC between the models combining one (#5), two (#4 and #6) or three (#4, #5 and #6) CT reformats obtained in the coronal and sagittal planes with those combining either the transverse and sagittal or the transverse and coronal planes were investigated following a resampling procedure based on 200 samples of 14 patients randomly drawn from the whole series. A Wilcoxon signed rank test was performed to compare differences between the models. A Bonferroni-type correction for performing nine comparisons between models was applied and a p-value < 0.0056 was regarded as statistically significant.

7.3. Results

7.3.1. Feasibility and agreement of the scoring

The rate of feasibility of the scoring for reader 1/reader 2 was 85% / 85% for the transverse reformats, 97%/100% for the coronal reformats and 98.5%/99% for the sagittal reformats. Missing scores among the expected 1218 scores were due to presence of metal artifacts (n=13 and 5 for reader 1 and 2), absence of one of the fracture edges on the reformat (n=8 and 5 for reader 1 and 2) and non-visibility of the fracture (n= 49 and 52 for reader 1 and 2). Missing scores were statistically significantly more frequent in the transverse than in the coronal and sagittal planes ($p<10^{-4}$).

The intra-observer reproducibility for the scoring performed on the median reformat was moderate on the transverse plane (ICC:+0.54 [+0.24;+0.75]for R1 and +0.60 [+0.30;+0.79] for R2) and good on the coronal (ICC:+0.72 [+0.48;+0.86]for R1 and +0.62 [+0.28;+0.81]for R2) and sagittal plane (ICC:+0.76 [+0.60;+0.89] for R1 and +0.61 [+0.26;+0.81] for R2). The inter-observer reproducibility for the scoring performed on the medial reformat was fair on the transverse plane (ICC:+0.33 [-0.02;+0.61]for R1 and +0.26 [-0.07;+0.55] for R2), moderate on the coronal plane (ICC:+0.46 [+0.12;+0.71]for R1 and +0.27 [-0.06;+0.56] for R2)and good on the sagittal plane (ICC:+0.65 [+0.33;+0.82]for R1 and +0.66 [+0.41;+0.83] for R2).

7.3.2. Reference TUS values and agreement

The TUS values of reference for each plane ranged from 23 to 26 with very good inter-observer reproducibility in the transverse plane (ICC : 0.83 [+0.66;+0.91]) and good inter-observer reproducibility in the sagittal (ICC:+0.72 [+0.38;+0.87] and coronal (+0.72 [+0.48;+0.86]) planes.

7.3.3. Comparison between tested models and reference TUS values

a. Models combining three planes

All tested models based on three planes had an ICC value equivalent or superior to 0.90 and difference in median values between -1.5 and +0.5 in comparison with the standard of reference for both readers without any statistically significant difference in value (Table 1,2).

b. Models combining two planes

Five of the nine tested models based on two planes had a ICC value equivalent or superior to 0.90 and difference in median values between -1.5 and + 2.5 in comparison with the standard of reference for both readers without any statistically significant difference in value (Table 1,2).

c. Models with one plane

Only four (reader 1) and one (reader 2) of the 21 models based on one plane had a ICC value equivalent or superior to 0.90 and difference in median values between -1.0 and +1.0 in comparison with the standard of reference without any statistically significant difference in value (Table 1,2).

Table 1: Correlation coefficients (ICC) and median differences in score value between the reference standard and 33 models combining different planes and image numbers for reader 1. Numbers in [] are 95%CI.

	Three planes	Two planes		
Slice number	**Tr, Co, Sa**	**Co Sa**	**Tr Co**	**Tr Sa**
#5	0.90 [0.80; 0.95] -1.5 [-4.0; +1.5]	0.91 [0.81; 0.95] +0.5 [-1.5; +2.5]	0.81[0.64; 0.91] -1.0[-4.5; +2.5]	0.80 [0.62; 0.90] -2.0 [-6.0; +2.5]
#4 # 6	0.97 [0.93; 0.98] -1.0 [-2.5; +0.5]	0.97[0.93; 0.98] -0.5 [-2.5; +1.0]	0.94 [0.88; 0.97] -1.5[-3.5; +1.0]	0.90 [0.81; 0.95] -1.5[-4.5; +0.5]
#4 #5 #6	0.96 [0.91; 0.98] -1.5 [-3.0; +0.5]	0.96 [0.92; 0.98] -0.0 [-2.0; +1.5]	0.93 [0.85; 0.96] -1.0 [-3.5; +2.0]	0.89 [0.79; 0.95] -2.0 [-4.5; +1.0]
#3 #7	NA	NA	NA	NA
#3 #5 #7	NA	NA	NA	NA
#2 #4 #6 #8	NA	NA	NA	NA
#2-#8	NA	NA	NA	NA

Tr: transverse; Co: Coronal; Sa:Sagittal; #5: median CT reformat; #4 and# 6: paramedian CT reformats, #3 and #7: paralateral CT reformats; #2 and #8 : lateral CT reformats; NA: not assessed.

	One plane		
Slice number	**Tr**	**Co**	**Sa**
#5	0.57 [0.26; 0.77] -5.0 [-11.0;+0.5]	0.73 [0.50; 0.86] +0.5 [-3.5; +4.5]	0.83 [0.67; 0.92] +0.0 [-4.0; +4.0]
#4 # 6	0.83 [0.66; 0.91] -3.0[-6.0; +0.0]	0.85 [0.71; 0.93] +0.0 [-2.5; +3.0]	0.83 [0.67; 0.92] -1.0 [-4.5; +3.5]
#4 #5 #6	0.79[0.61; 0.90] +3.0 [-6.0; +0.0]	0.83 [0.66; 0.91] +0.5 [-3.0; +3.0]	0.85 [0.70; 0.93] -1.0 [-4.5; +3.0]
#3 #7	0.87 [0.74;0.94] +1.5 [-2.5;+4.5]	0.87 [0.75;0.94] +1.0 [+2.5;+4.5]	0.85 [0.71;0.93] +0.5 [-3.0;+4.0]
#3 #5 #7	0.93 [0.87;0.97] -0.5 [-3.0;+1.5]	0.86 [0.72;0.93] +0.5 [-3.0;+4.0]	0.87 [0.75;0.94] +0.5 [-3.0;+3.5]
#2 #4 #6 #8	0.81 [0.63;0.90] +2.5 [-2.0;+7.0]	0.87 [0.73;0.93] -0.5 [-2.5;+2.5]	0.90 [0.80;0.95] -1.0 [-4.5;+2.0]
#2-#8	0.90 [0.80;0.95] +1.0 [-2.0;+4.5]	0.87 [0.74;0.94] -0.0 [-2.5;+2.5]	0.90 [0.79;0.95] -0.5 [-3.5;+2.0]

Tr: transverse; Co: Coronal; Sa:Sagittal; #5: median CT reformat; #4 and# 6: paramedian CT reformats, #3 and #7: paralateral CT reformats; #2 and #8 : lateral CT reformats; NA: not assessed.

Table 2: Correlation coefficients (ICC) and median differences in score value between the reference standard and 33 models combining different planes and image numbers for reader 2. Numbers in [] are 95%CI. Numbers in [] are 95%CI.

	Three planes	Two planes		
Slice number	**Tr, Co, Sa**	**Co Sa**	**Tr Co**	**Tr Sa**
#5	0.93 [0.86; 0.97] -1.5 [-3.5; +1.0]	0.93 [0.85; 0.96] +2.0 [-0.5; +4.5]	0.82[0.61; 0.92] -5.0 [-9.0; -2.0]	0.86[0.73; 0.93] -2.0 [-4.5; +2.0]
#4 # 6	0.97 [0.94; 0.99] +0.5 [-1.0; +1.5]	0.95 [0.89; 0.98] +2.5 [-0.0; +4.0]	0.93[0.85; 0.97] -2.0 [-4.0; +0.5]	0.91 [0.82; 0.96] +0.5[-2.0; +3.5]
#4 #5 #6	0.98 [0.95; 0.99] -0.5 [-1.5; +1.0]	0.96 [0.91; 0.98] +2.0 [+0.0; +3.5]	0.93 [0.81; 0.97] -3.0 [-5.0; -0.5]	0.91 [0.82; 0.96] +0.0 [-3.0; +3.0]
#3 #7	NA	NA	NA	NA
#3 #5 #7	NA	NA	NA	NA
#2 #4 #6 #8	NA	NA	NA	NA
#2-#8	NA	NA	NA	NA

Tr: transverse; Co: Coronal; Sa:Sagittal; #5: median CT reformat; #4 and# 6: paramedian CT reformats, #3 and #7: paralateral CT reformats; #2 and #8 : lateral CT reformats; NA: not assessed.

	One plane		
Slice number	**Tr**	**Co**	**Sa**
#5	0.64[0.19; 0.84] -9.5 [-165; -4.5]	0.84 [0.68; 0.92] -1.0 [-.05; +2.5]	0.73 [0.48; 0.86] +5.0 [-0.0; +11]
#4 # 6	0.75 [0.51; 0.88] -4.0[-8.5; +0.5]	0.88 [0.75; 0.94] +0.5 [-4.0; +2.5]	0.79[0.59; 0.90] +3.5[+0.0; +7.5]
#4 #5 #6	0.76 [0.44; 0.89] -6.0 [-10.0; -2.0]	0.89 [0.78; 0.95] +0.0 [-3.5; +2.0]	0.80[0.59; 0.90] +4.0 [+0.5;+8.5]
#3 #7	0.80 [0.62;0.90] -0.5 [-5.5;+2.0]	0.80 [0.61;0.90] +1.0 [-4.0;+4.0]	0.76 [0.52;0.88] +4.0 [-0.0;+8.0]
#3 #5 #7	0.85 [0.65;0.93] -4.0 [-7.0;-0.5]	0.86 [0.72;0.93] -0.0 [-3.5;+3.0]	0.80 [0.58;0.91] +4.0 [+0.5;+8.5]
#2 #4 #6 #8	0.80 [0.61;0.90] -1.0 [-4.5;+2.5]	0.89 [0.77;0.95] -2.5 [-5.5;+0.5]	0.89 [0.78;0.95] +3.0[+0.0;+6.0]
#2-#8	0.84 [0.69;0.92] -2.5 [-6.0;+0.5]	0.90 [0.81;0.95] -1.0 [-4.5;+1.0]	0.87 [0.72;0.94] +3.5[+0.5;+6.0]

Tr: transverse; Co: Coronal; Sa:Sagittal; #5: median CT reformat; #4 and# 6: paramedian CT reformats, #3 and #7: paralateral CT reformats; #2 and #8 : lateral CT reformats; NA: not assessed.

7.3.4. Comparison between best models

For both readers, the median differences in ICC were statistically significantly closest to the reference standard with the models combining coronal and sagittal CT reformats than those combining coronal and transverse, or sagittal and transverse CT reformats (Table 3). The median differences in ICC were the lowest for the models that combined the two paramedian reformats for the two readers (Table 3).

Table 3: Comparison for the two readers of the performance of the models combining one (#5), two (#4 and #6) or three (#4,#5 and #6) CT reformats obtained in the coronal and sagittal planes with those combining either the transverse and sagittal or the transverse and coronal planes.

Plane combinations	**Reader 1**		**Reader 2**	
CoSA* vs TrSa*	**P value**	**Median difference in ICC [95%CI]**	**P value**	**Median difference in ICC [95%CI]**
#5**	< 10^{-4}	+0.104 [+0.101; +0.107]	< 10^{-4}	+0.039 [+0.037; +0.042]
#4 and #6**	< 10^{-4}	+0.031 [+0.030; +0.033]	< 10^{-4}	+0.039 [+0.038; +0.040]
#4, #5 and #6	< 10^{-4}	+0.072 [+0.070;+0.073]	< 10^{-4}	+0.051 [+0.050;+0.052]
CoSA* vs TrCo*	**P value**	**Median difference in ICC [95%CI]**	**P value**	**Median difference in ICC [95%CI]**
#5**	< 10^{-4}	+0.095 [+0.092; +0.98]	< 10^{-4}	+0.078 [+0.076; +0.080]
#4 and #6**	< 10^{-4}	+0.007 [+0.006; +0.008]	< 10^{-4}	+0.020 [+0.019; +0.020]
#4, #5 and #6	< 10^{-4}	+0.038 [+0.038;+0.039]	< 10^{-4}	+0.031 [+0.030; +0.032]

The median difference in ICC was statistically significantly closest to the reference standard with the models combining the coronal and sagittal imaging planes and difference was the lowest for the model combining the two paramedian reformats. * CoSa combines coronal and sagittal planes; TrSa combines transverse and sagittal planes; TrCo combines transverse and coronal planes.**#5 is median CT reformat; #4 and #6 are the two paramedian CT reformats.

7.4. Discussion

> Our study of 29 tibial and femoral non-unions by two readers identified three important findings:
>
> - Analysis of a large number of CT reformats is not mandatory to assess non-united fractures.
> - Scoring of transverse reformats is not mandatory if sagittal and coronal reformats are scored.
> - Two valuable alternatives to the scoring of all reformats were to score either the median reformats of each plane (three reformats) or two paramedian coronal and sagittal reformats (four reformats) .

First, we demonstrated that the scores derived from the analysis of the median sagittal, coronal and transverse CT reformats or from the two paramedian sagittal and coronal CT reformats had a very high correlation coefficient (> 0.90) with the scores obtained at the analysis of all sagittal, coronal and transverse CT reformats. Most likely, the scoring of all reformats from the three planes leads to data oversampling as the same cortical areas are analyzed on the three stacks i.e. the fracture area located in the most lateral segment of the diaphysis can be assessed on the most lateral sagittal CT reformat and on the medial coronal CT reformat.

Second, we demonstrated that the scores derived only from a limited number of sagittal and coronal CT reformats appeared to have a higher correlation coefficient with the reference values than when the analyzed reformats were derived from the transverse CT reformats and one of the two long-axis CT reformats. Actually, on the transverse reformats, analysis of the transverse component of the fractures was more difficult due to partial volume artifacts because cortical discontinuity is transverse. In addition, analysis of the fracture sites was more often non feasible on short- than on long-axis reformats due to more metal-related artifacts on transverse images.

Finally, the analysis of all reformats from only one plane provided results that poorly correlated with the standard of reference. This observation could derive from the complex and variable spatial orientation of cortical interruption in fractures that, depending on their relation with the orientation of the plane may or may not be conspicuous on one of the plane.

Several difficulties interfered with the feasibility of the analysis including metal-related artifacts, non-visibility of one of the fracture edges due to bone deformity on long-axis images and limited conspicuity of the fracture line due to partial volume artifacts mainly on transverse images. Although metal-related artifacts could have been reduced by using specific post-processing software, their use could have induced some loss in spatial resolution and they were not used in this study [71].

The results of the current study could pave the way to the use of digital tomosynthesis (DTS) for semi-quantitative fracture healing evaluation. Tomographic images obtained at DTS enable to detect non-displaced fractures and to evaluate the healing process [41; 72-74]. The fact that DTS does not provide transverse tomographic images due to the translational nature of X-ray source displacement and to its limited multiplanar capabilities could therefore have limited implication in the evaluation of fracture healing with DTS since short -axis images that are lacking at DTS seem to be of limited interest.

Several limitations of the current study should be noted. First, a limited number of patients were studied and we focused on surgically proven non-united tibial and femoral shaft fractures. It is possible that focusing on delayed unions, we favored the inclusion of fractures in which cortical discontinuity is transverse or transverse oblique. Actually, the presence of a same cortical discontinuity orientated in a longitudinal plane instead of a transverse plane would not have been included in the study as a longitudinal bone defect does not have identical biomechanical implications as a transverse bone defect. Therefore, the limited value of the transverse imaging plane could be inherent to our selected population. Second, we used a semi-quantitative scoring system derived from a radiographic scoring system that has been validated for the analysis of fracture healing on antero-posterior and lateral radiographs [50]. The current study did not assess the value of this tomographic scoring system and did not compare it with results obtained on radiographs of the same lesion. Third, the reproducibility of the score calculated on a single reformat was fair to moderate whereas it was good to very good when numerous reformats were assessed indicating that the analysis of one reformat may be influenced by that of the adjacent reformats. Finally, readers were not blinded to the planes of the reformats during their analysis.

In conclusion, the current study performed in a series of tibial and femoral shaft non-unions demonstrated that the analysis of the median coronal, sagittal and transverse CT reformats or of the two paramedian sagittal and coronal CT reformats was a valuable alternative to the analysis of a large number of CT reformats obtained in the three orthogonal planes. The observation that short-axis reformats were not mandatory could pave the way to the use of DTS - that only provides long-axis reformats- to semi-quantitatively evaluate the healing process of long bone fractures.

8. Part III: **Semi-quantitative assessment of nailed shaft fractures using computed tomography: comparison with radiographic and clinical scores during normal healing.**

8.1. Introduction

Assessment of fracture healing remains challenging in orthopedic practice. Surgeons agree on the fact that the assessment of healing requires both radiographic and clinical information. In a survey of 444 orthopedic surgeons, Bhandari et al. demonstrated a lack of consensus regarding the value of clinical and radiologic criteria to assess healing [18; 21] . Since 2008, several authors reported the value of a Radiographic Union Score (RUS) that enables more reproducible evaluation of fracture healing in tibial (RUST), femoral (RUSH) and humeral fractures (RUSHU)[49; 50; 57; 60-62; 75; 76].

Conventional radiography is the primary imaging modality to assess fracture healing, but its value for objective quantification of bone healing is limited [31; 32] partially due to overlapping structures (eg, bones, metal hardware, cast). Multi-detector Computed Tomography (MDCT) overcomes the radiographic limitations due to overlapping structures, and enables a more sensitive detection of callus formation and of cortical continuity [22; 68] at the cost of an additional imaging examination with higher cost and radiation burden [34; 45]. A tomographic union score (TUS) has been proposed to semi-quantitatively assess fracture healing according to the callus-based RUS score [50]. We aimed at determining the feasibility, reproducibility and repeatability of the TUS in nailed tibial and femoral fractures and at comparing TUS with RUS and a clinical score. The current prospective study included patients with tibial or femoral nailed fractures that normally healed at nine months follow-up and in whom CT was obtained for the study purpose at the same moment as a routine radiographic examination.

8.2. Material and Method

Authorization of the institutional ethic committee for this prospective study and written authorizations from patients were obtained. Between January 2015 and December 2016, recruitment was performed by a musculo-skeletal radiologist with 4 years of clinical experience (VP) when she was reading follow-up radiographs at the orthopaedic consultation. Eligibility criteria were (a) age between 18 and 75 years, (b) nailed tibial or femoral fracture and (c) less than seven months of fracture age. Patients with fractures that had been treated with allo- or auto-grafts, with imaging or biological signs of infections or pathological fractures were not eligible. Pregnant women or women in a reproductive age not using a contraceptive method were not considered for inclusion. The radiologist met eligible patients and asked them to participate to our study by answering a clinical questionnaire, undergoing a CT examination of the fracture and authorizing the radiologist to assess healing at nine months follow-up by consulting their medical files. Before performing the CT examination, eligible patients filled in a clinical questionnaire (Fix-it score) which assesses the ability to bear weight (maximum 6 points) and importance of pain at the fracture site (maximum 6 points). The maximum score of 12 indicates normal function and absence of pain of the limb [77]. The radiographic and CT examinations and the clinical score were obtained on the same day. Criterion for definite inclusion was the demonstration of clinical and radiological healing at nine months follow-up according to medical files. Study participants who showed non-union at nine months follow-up were excluded. Lack of Fix-it score completion was not an exclusion criterion.

8.2.1. Imaging protocols

All unenhanced CT examinations were performed on a 40-detector row CT scanner (Somatom Definition AS; Siemens Healthcare, Forchheim, Germany), patients lying supine with the feet first. Data acquisition was obtained over the entire fractured bone using the following tube parameters: tube voltage, 120 kVp; reference tube current-time product, 275 mAs. Effective current time-product , 178-305 mAs by applying automated tube current modulation (Care Dose 4D; Siemens Healthcare); detector configuration, 40X0.6 mm; pitch, 0.8; gantry rotation time , 1 s. The CT raw data were reconstructed by using a conventional feed-back retro-projection algorithm. The following image reconstruction parameters were used: field-of-view, 12X12 cm to 20X20 cm; section/thickness increment, 0.75/.075 mm; soft tissue (B41s) and bone (B70h) convolution kernels. All reconstructions were stored in our PACS. Between January and March 2018, the recruiting radiologist reformatted nine sagittal and nine coronal CT images from the CT data obtained by using the bone convolution kernel. Each stack of reformats contained nine 1.5 mm-thick reformats with variable slice intergap depending on the bone diameter. The first (image #1) and last (image #9) reformats were placed in contiguity with the target bone which was only depicted on reformats (images #2-#8). A data manager anonymized all the stacks in a random order per plane and stored them in a dedicated research PACS (Carestream Client version 11.3; Carestream Health, Rochester, NY, USA). For each fracture, one A-P radiograph and one lateral radiograph were available for reading.

8.2.2. Image analysis

In June 2018, the recruiting radiologist (R2:VP) and two independent musculoskeletal radiologists (R1: BV, R3: FL) with 28 and 23 years of experience analysed the sagittal and coronal CT reformats on a PACS workstation with adapted window width and level values. R1 and R2 had already performed semi-quantitative assessment of fractures in previous studies; R3 had no previous experience and a 1-hour training session was performed before reading by using radiographs and CT of fracture cases that were not included in the study. Readers, blinded to radiographic and clinical data analysed separately the coronal and sagittal CT reformats at one month time interval. The readers scored the fracture edges seen on the two para-median (images #4 and #6) sagittal and the two paramedian coronal CT reformats by using a Tomographic Union Score (TUS) based on the callus appearance [78] . The cortical score was defined as follows: score 1: no callus; score 2: discontinuous callus; score 3: continuous bridging immature callus; score 4: continuous remodelled callus (Figures 1,2). On the four analysed CT reformats, two fracture sites had to be scored (one score for the anterior and one for the posterior cortex on the sagittal reformats, one score for the medial and one for the lateral cortex on the coronal reformat). When several fracture sites were detected, the least healed site was scored. Cortical scoring was considered to be non-feasible when metal artefacts obscured the fracture or when the reformat plane was identical to that of the fracture that was no more visible. The TUS was the half of the sum of the four cortical scores obtained on the sagittal reformats and of the four cortical scores obtained on the coronal reformats. The TUS ranged from four to 16. TUS was considered to be feasible when at least one cortical score per area (anterior and posterior areas on the sagittal reformats and medial and lateral areas on the coronal reformats) in at least three of the four areas was available.

Two months later, all readers separately scored each set of radiographs on PACS workstations by using the RUS (Radiographic Union Score) (Figures 1,2). Four cortical segments had to be scored (one anterior and one posterior on the lateral radiograph and one medial and one lateral on the antero-posterior radiograph). If the cortical segment was visible, cortical scoring was performed by using the same score as described above. If the cortical segment was not visible, scoring was considered to be not feasible and the reader had to select at least one of the three following explanations: (a) superimposed metal hardware, (b) superimposed adjacent bone, (c) another reason such as blurring of the radiograph or inadequate exposure. For each set of radiographs, the RUS was calculated by adding the four cortical scores. The RUS ranged from four to 16. In fractures in which three cortical scores only were available the cortical score of the not assessable segment was calculated as the mean of the three other scores and the corresponding RUS was calculated. In fractures in which less than three cortical scores were available, RUS was not calculated and considered to be non-feasible. A second reading of all radiographs and CT examinations were performed by R1 and R2 six months later without knowledge of the results of their initial readings.

Figure 1: Left tibial fracture in a 50-year-old man 57 days after fracture. (a,b) Antero-posterior and lateral radiographs demonstrated discontinuous callus (black arrow; score 2) at medial and posterior fracture edges and no callus (white arrow; score 1) at lateral and anterior fracture edges. RUS was 6. (c,d) Paramedian coronal and (e,f) sagittal CT reformats demonstrated discontinuous callus (black arrow; score 2) and no callus (white arrow; score 1). TUS was 6,5. RUS and TUS are not statistically significantly different in early healing.

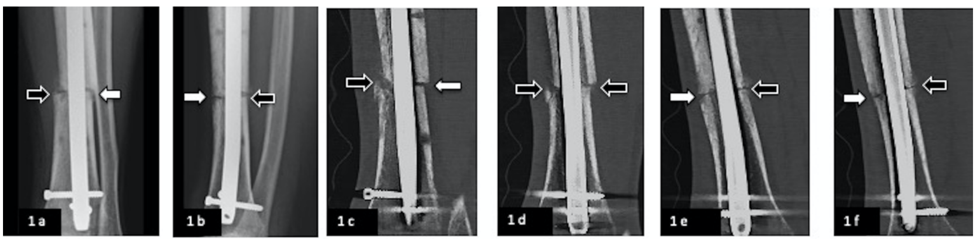
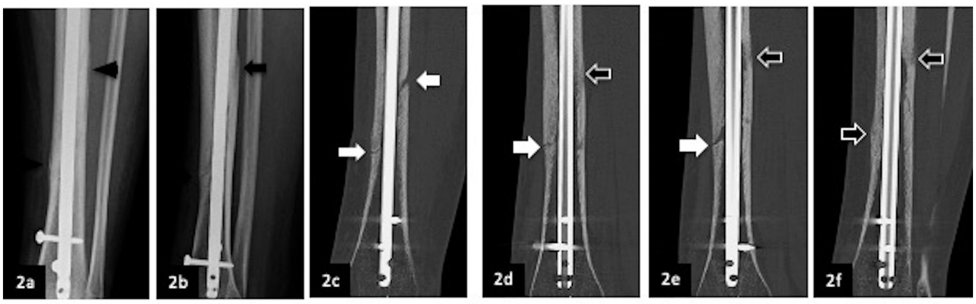


Figure 2 : Left tibial fracture in a 46-year-old woman 121 days after fracture. (a,b) Antero-posterior and lateral radiographs demonstrated continuous remodelled callus (black arrowhead; score 4) at medial and lateral fracture edges and immature continuous callus (black arrow; score 3) at anterior and lateral fracture edges. RUS was 14. (c,d) Paramedian coronal and (e,f) sagittal CT reformats demonstrated no callus (white arrow; score 1), discontinuous callus (black arrow; score 2), continuous immature callus (white arrowhead; score 3) and continuous remodelled callus (black arrowhead; score 4). TUS was 8. In advanced healing, TUS is statistically significantly lower than RUS.



8.2.3. Statistical analysis

Intra-observer and inter-observer agreement on RUS and TUS scores were assessed using the Intraclass Correlation Coefficient (ICC) and its associated 95% confidence interval [79]. Strength of agreement was interpreted as follows: ≤ 0.20 = poor, 0.21-0.40 = fair, 0.41-0.60 = moderate, 0.61-0.80 = good and ≥ 0.81 = very good [51]. Kendall rank correlation coefficient was used to assess statistical associations between TUS score and RUS score, FIX-it score, type of bone, patient age, sex and fracture age. Strength of association was interpreted as follows: strong ($t_b \geq 0.9$), substantial ($0.7 \leq t_b < 0.9$), moderate ($0.5 \leq t_b < 0.7$) or weak ($t_b < 0.5$)[80]. Finally, potential differences between RUS and TUS scores were assessed according to different threshold values of RUS score, fracture age (≤ 3 months and ≥ 3 months), FIX-it score (< 7 and > 6) and type of bone (femur and tibia) using boxplots and non-parametric Wilcoxon signed ranks test. A *p*-value < 0.05 was regarded as statistically significant [81] . Statsdirect software (https://www.statsdirect.com) was used for the statistical analysis.

8.3. Results

8.3.1. Study population

Radiographs and CT examinations were obtained in 28 patients with tibial or femoral nailed fractures at a delay between one and seven months after fracture. Four patients (two tibial and two femoral fractures) showed non-union at nine months follow-up and were excluded from the study. The study population consisted of 24 study participants (17 men and seven women; median age (95%CI)): 47,5 y.o. (95% CI : 33-54)) with 19 nailed tibial and five femoral fractures. The mean time delay (+/-SD) between fracture occurrence and CT examination was 109 days (+/- 57, range: 42-204 days). There were twelve 42-A, four 42-B, three 42-C tibial fractures and two 32-A, two 32-B and one 32-C femoral fractures according to the AO classification [82]. Eighteen study participants answered the Fix-it questionnaire.

8.3.2. Feasibility and agreement of semi-quantitative scores

The rate of feasibility of cortical scoring was 99% (95/96) for all readers on radiographs and 100% (192/192), 99% (190/192) and 99.5% (191/191) for R1, R2 and R3 respectively on CT reformats. RUS and TUS were feasible in all cases for all readers. Intra- and inter-observer agreement was very good for RUS and TUS (table 1).

Table 1: Intra and inter-observer agreement of TUS and RUS

Agreement	TUS	RUS
Intra-observer ICC for R1	0,90 [0,77 -0,95]	0,94 [0,87-0,98]
Intra-observer ICC for R2	0,91 [0,70 -0,97]	0,97 [0,94 -0,99]
Inter-observer ICC (R1, R2, R3)	0.89 [0.76- 0.95]	0.83 [0.57- 0.93]

8.3.3. Correlations between scores and clinical parameters

Imaging (RUS, TUS) and clinical (Fix-it) scores did not statistically significantly correlated with sex, patient age or type of bone. Correlation between RUS and fracture age was moderate for all readers (p=0.0002 for R1 and R2 and p=0.0005 for R3). Correlation between TUS and fracture age was weak for R1 (p=0.001) and moderate for R2 and R3 (p=0.0007 and p=0.0004 respectively) (table 2). FIX-it score moderately correlated with fracture age (p = 0.0003).

TUS substantially correlated with RUS (p<0.0001 for all readers). For R1 and R3, RUS and TUS moderately correlated with FIX-it scores (p=0.0002 for TUS/R1, p=0.0003 for RUS/R1, p=0.0004 for TUS/R3, p=0.002 for RUS/R3). For R2, RUS and TUS correlated moderately (p=0.0003) and substantially (p<0.0001) with FIX-it scores (table 2).

Table 2: Correlations between TUS, RUS and clinical parameters

[NS for non- significant correlation. When significant correlation, the Kendall's tau b value (95% CI) shows the strength of the association between variables. (p<0,05, **p<0,001, ***p<0,0001)]*

	TUS R1	RUS R1	TUS R2	RUS R2	TUS R3	RUS R3
Sex	NS	NS	NS	NS	NS	NS
Patient age	NS	NS	NS	NS	NS	NS
Fracture age	b:0,46 (0,21-0,71)*	b:0,55 (0,38-0,72)*	b : 0,51 (0,31-0,71)**	b :0,55 (0,37-0,73)**	b : 0,52 (0,28-0,72)**	b : 0,51 (0,32-0,71)**
Type of bone	NS	NS	NS	NS	NS	NS
Fix-it score	b:0,64 (0,41-0,87)**	b:0,63 (0,42-0,83)**	b :0,69 (0,51-0,88)**	b :0,63 (0,43-0,84)**	b : 0,62 (0,35-0,80)**	b : 0,53 (0,04-0,74)*
RUS	b:0,73 (0,54-0,91) ***	NA	b :0,76 (0,59-0,93) ***	NA	b : 0,67 (0,22-0,84)***	NA

8.3.4. Comparison between RUS and TUS according to different parameters

In fractures with RUS < 8, TUS was not statistically significantly different from RUS (figure 3). In fractures with RUS≥8 or ≥9, TUS was statistically significantly lower than RUS for R2 and R3 but not for R1 (Figure 3). In fractures with RUS ≥10, TUS was statistically significantly lower than RUS for all readers (figure 3). TUS was statistically significantly lower than RUS for R2 and R3 but not for R1 in patients with Fix-it scores > 6 (figure 4) and in patients with fracture age > 3 months (figure 5).

Figure 3: Comparison between RUS and TUS according to different RUS threshold values.

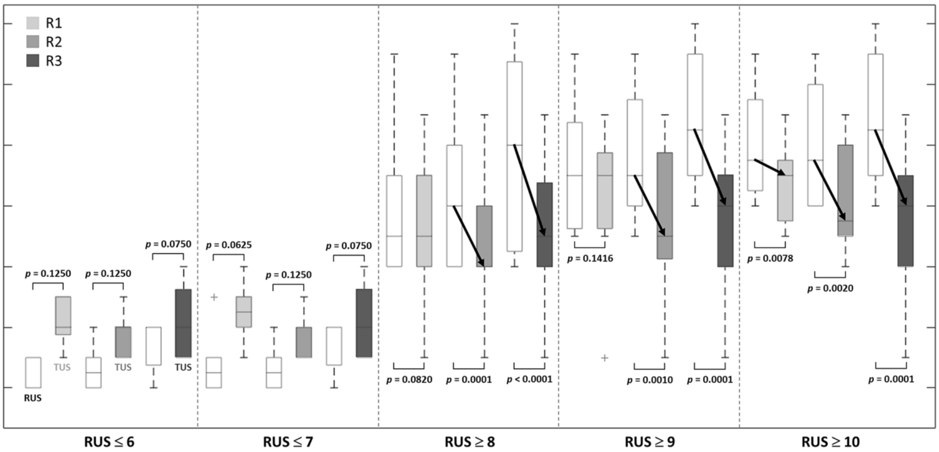


Figure 4: Comparison between RUS and TUS according to different Fix-it score threshold values.

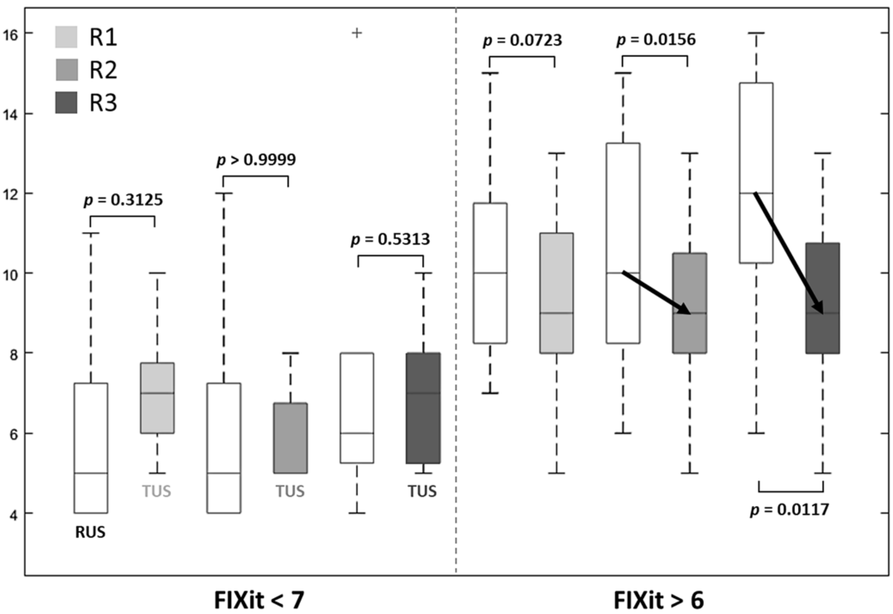
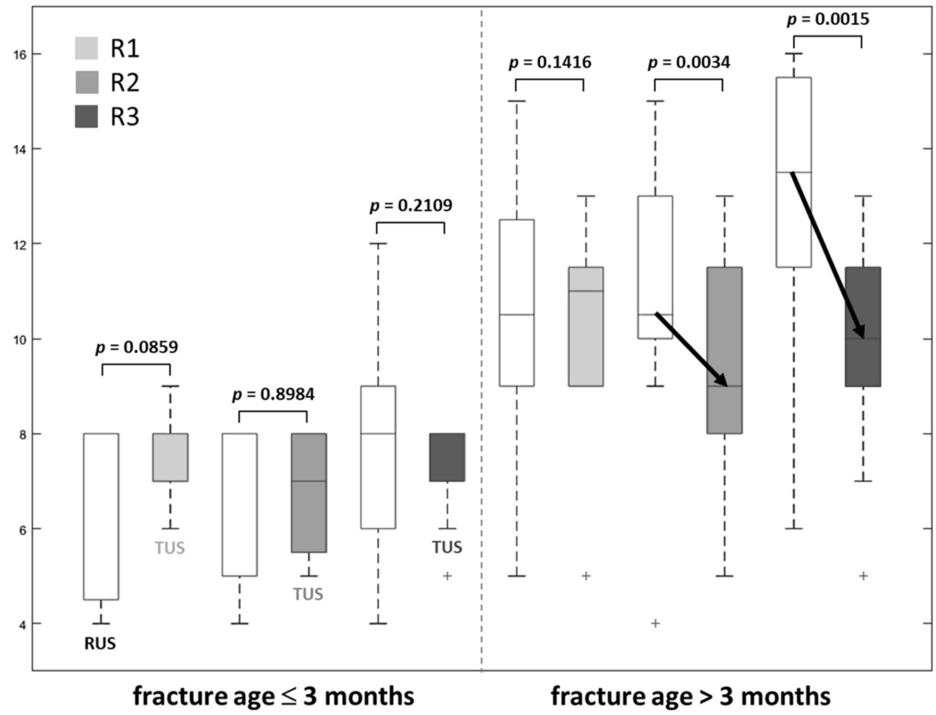


Figure 5: Comparison between RUS and TUS according to different fracture age threshold values.



8.4. Discussion

Our study of 24 nailed tibial and femoral shaft fractures with normal healing and three readers identified three important findings:

- Semi-quantitative assessment of fractures with normal healing is feasible and reproducible.
- TUS and RUS correlated with clinical score at least moderately.
- In later phases of healing, TUS was statistically significantly lower than RUS.

First, we demonstrated that semi-quantitative CT assessment of fracture healing by using a callus-based scoring system was feasible and reproducible among three separate readers, including one reader without previous experience in fracture scoring.

Second, the clinical score and the two imaging scores obtained in the current study of fractures with normal healing statistically significantly positively correlated with fracture age indicating that they enable to assess the normal healing process. The fact that TUS and RUS only moderately correlated with the Fix-it score is in line with the general agreement that both clinical and imaging data are necessary to assess fracture healing. A perfect correlation between imaging and clinical data would have questioned their complementary value.

Finally, TUS substantially correlated with RUS in the entire population, but correlations differed when stratification was applied according to radiological score, clinical score and fracture age. Actually, TUS was statistically significantly lower than RUS in fractures with important callus formation (high RUS), with high clinical score (almost healed fracture), and with fracture age >3 months. This observation most likely derived from the fact that CT reformats better delineate the osseous architecture to a very fine degree, and occasionally depict cortical discontinuity not seen on the x-ray thus downgrading a radiological score of three (continuous cortex) to two (discontinuous cortex). On the contrary, in patients with low RUS value, indicating limited callus formation without cortical continuity, TUS was higher than RUS, but this difference was not statistically significant, most likely partially due to the limited number of cases. Grigoryan et al. found that early manifestations of healing, including blurring of fracture margins and formation of external callus, were observed earlier with CT scan than with radiographs [45]. Compared to radiography, CT overcomes the problem of potentially overlapping structures and enables a more sensitive detection of callus formation, therefore upgrading a radiological score of 1 (no callus) to 2 (presence of callus).

Most previous quantitative CT evaluations of fracture healing were based on callus density or volume. Grigoryan et al [45] performed a quantitative assessment of CT density changes (HU) in the fracture gap (medullary callus) in fractures without metal hardware with good reproducibility and a statistically significant correlation to qualitative changes. Callus density cannot be reliably determined in metalized fractures due to metal-related artifacts. In addition, callus volume has been shown to be non-relevant with respect to fracture healing as it can be important despite persistent cortical discontinuity as in hypertrophic pseudarthrosis. TUS was based on callus presence and callus continuity which most closely correlates with the biomechanical value of the callus [83; 84]. Experimental studies comparing TUS with CT volumetry and histopathological findings would be useful in order to validate the current semi-quantitative evaluation.

In the current study, nailed tibial and femoral shaft fractures that healed normally were investigated by using radiographs and CT at different time delays after fracture to evaluate bone changes that develop during normal healing. Further studies should be performed to assess clinical situations in which CT may contribute to the better assessment of fracture healing when healing is not reached within accepted time delays. By using a qualitative approach to analyze CT data (complete healing: bone bridges with no gap; partial healing: some bone bridges with gaps between; and no bridging, no osseous bridges), Krestan et al [69] showed that MDCT altered the radiographic diagnosis by demonstrating different degree of fracture healing than radiographs in 37% of cases in a series of 43 fractures. Our results suggest that CT is unlikely to add important information in atrophic delayed or non-united fractures and is more likely to contribute to a better assessment of fracture when hypertrophic non-union is suspected on radiographs.

Several limitations to the current study should be noted. First, only a limited number of patients agreed upon participating to our study and our patient series was not consecutive. Therefore, recruitment bias could have occurred as more symptomatic patients could have been more inclined to participate in our study. Nonetheless, clinical and radiological follow-up was obtained to include only patients with definite normal healing process and four patients in whom healing was not obtained at nine months follow-up were excluded. Second, readers analyzed only two sagittal and two coronal CT reformats per fracture as a CT study performed in non-united fractures had demonstrated that analysis of transverse reformats or of more than four longitudinal reformats was not mandatory [78]. We acknowledge that it has not been demonstrated that this alternative method could also be applied to fractures during normal healing. Finally, there is no histological evidence that cortical discontinuity or continuity as seen on CT imaging actually correspond to continuous callus at histology or at micro-CT imaging.

In conclusion, semi-quantitative CT assessment of fracture healing using a callus-based score was feasible and reproducible in nailed shaft fractures and correlated positively with fracture age. TUS yielded lower values than RUS in advanced healing suggesting that more callus discontinuity is visible on CT than on radiographs.

9. Part IV: Semi-quantitative CT assessment of tibial and femoral non-united fractures: Comparison with semi-quantitative radiographic assessment

9.1. Introduction

Fractured bones have a tremendous potential for repair and regeneration to restore osseous continuity and function. Non-union occurs in about 10% of fractures, mostly in long bone shafts [8; 10; 11]. Radiographs contribute to the diagnosis of abnormal healing [18] by depicting persistent fracture lines, absence of bony bridging, lack of progressive healing on serial radiographs, progressive deformity and the presence of loose or broken implants [30]. To overcome the poor contributive value of overall evaluation of radiographs, a semi-quantitative assessment of radiographs by using a radiographic union score (RUS) has recently been developed and was proven to increase the reproducibility and accuracy in the assessment of abnormal fracture healing on radiographs [49; 57; 60-62; 75; 76].

Computed tomography (CT) can be used as a second-line imaging modality when clinical or radiographic information are not conclusive. By using a qualitative approach to analyze CT data (complete healing: bone bridges with no gap; partial healing: some bone bridges with gaps between; and no bridging, no osseous bridges), Krestan et al [69] showed that CT altered the radiographic diagnosis by demonstrating different degree of fracture healing than radiographs in 37% of cases in a series of 43 fractures. However, the value of CT remains debated [70]. A retrospective study of 32 non-unions [70] demonstrated that the interobserver reliability of MDCT scan is not greater than conventional radiographs for determining non-union indicating the need for a more rigorous analysis of CT images. In the current retrospective study, we aimed at comparing semi-quantitative CT and radiographic callus-based scores in a series of surgically proven tibial and femoral non-unions in which radiographs and CT examinations had been obtained for clinical purpose.

9.2. Material and Method

9.2.1. Study population

Authorization of the institutional ethic committee for this retrospective study was obtained and authorization from patients was waived. A search over a period of four years was performed in our picture archiving and communication systems (PACS)(Carestream Client version 11.3; Carestream Health, Rochester, NY, USA) by a musculoskeletal radiologist (VP) with 4 years of experience to select patients. Inclusion criteria were (a) patients aged between 18 and 65 years, (b) nailed tibial or femoral fracture without healing more than five months after fracture, (c) availability of CT and radiographic examinations of the fracture bone obtained within 30 days of delay and more than five months after fracture, (d) subsequent surgical demonstration of abnormal healing. Patients with fractures that had been treated with allo- or auto-grafts, with imaging or biological signs of infections or pathological fractures were not eligible. Pregnant women or women in a reproductive age not using a contraceptive method were not considered for inclusion. The study population consisted of 23 patients (14 men and 9 women; median age (95%CI)): 49 y.o. (95% CI: 28-56)) with 12 nailed tibial and 11 femoral fractures. The mean time delay (+/- SD) between fracture occurrence and CT examination was 565 days (+/- 519, range: 180-1983). CT examinations were obtained at a mean time delay after radiographs of seven +/- 10 days (range:0-30 days; median delay: 0 days). Surgery was performed at a mean +/-SD delay of 178 +/- 171 days after CT performance (range: 2-451 days; median: 96 days).

9.2.2. Imaging protocols

All unenhanced CT examinations were performed on a 40-detector row CT scanner (Somatom Definition AS; Siemens Healthcare, Forchheim, Germany), patients lying supine with the feet first. Data acquisition was obtained over the entire fractured bone using the following tube parameters: tube voltage, 120 kVp; reference tube current-time product, 275 mAs. Effective current time-product, 178-305 mAs by applying automated tube current modulation (Care Dose 4D; Siemens Healthcare); detector configuration, 40X0.6 mm; pitch, 0.8; gantry rotation time, 1 s. The CT raw data were reconstructed by using a conventional Feed-back retro-projection algorithm. The following image reconstruction parameters were used: field-of-view, 12X12 cm to 20X20 cm; section/thickness increment, 0.75/.075 mm; soft tissue (B41s) and bone (B70h) convolution kernels. All reconstructions were stored in our PACS. Between January and March 2018, we obtained sagittal and coronal reformats in a standardized manner. In our previous study (part 2 [78]) we demonstrated that short-axis reformats were not mandatory if the long-axis reformats were scored. We therefore only used data from sagittal and coronal reformats. Each stack of reformats contained nine 1.5 mm-thick reformats with variable slice intergap depending on the extent of the fracture. The first (#1) and last (#9) reformats were placed in contiguity with the target bone which was only depicted on reformats #2-8. A data manager anonymized all the stacks in a random order per imaging plane and stored them in a dedicated research PACS (Carestream Client version 11.3; Carestream Health, Rochester, NY, USA). For each fracture, one A-P radiograph and one lateral radiograph was available for reading.

9.2.3. Image analysis

In June 2018, two musculoskeletal radiologists with 24 (BV-reader1) and four (VP-reader 2) years of experience independently scored the CT reformats on a PACS workstation with adapted window width and level values. Coronal and sagittal CT reformats of each fracture were read separately in a random order. The readers scored the two para-median sagittal and coronal CT reformats according to a previously described method (part 2 [78]). The other reformats of the same plane but not of the other planes were available while scoring. Readers scored each cortical fracture site by using a Tomographic Union Score (TUS). In July 2018, the same readers separately scored each set of radiographs on PACS workstations using RUS (Radiographic Union Score).

On radiographs, readers scored the four cortical segments of the fracture at which the X-ray beam was tangent on the medial, lateral, anterior and posterior cortices. If the cortical segment with fracture was visible, cortical scoring was performed as follows score 1: no callus; score 2: discontinuous callus, score 3: continuous bridging immature callus; score 4: continuous remodelled callus (Figures 1,2). If the cortical segment was not visible, the reader had to select at least one of the three following explanations for non-feasibility: (a) superimposed metal hardware, (b) superimposed adjacent bone, (c) another reason such as blurring of the radiograph or inadequate exposure. For each set of radiographs, the RUS was calculated by adding all four cortical scores. The RUS ranged from four to 16. In fractures in which three cortical scores only were available due to lack of visibility of one cortical segment, the cortical score of the not assessable segment was calculated as the mean of the three other scores and the corresponding RUS was calculated. In fractures in which less than three cortical scores were available, RUS was considered to be non-feasible.

On CT reformats, readers scored each cortical fracture site by using a Tomographic Union Score (TUS) based on the callus appearance as previously described [78]. The cortical score was defined as follows: score 1: no callus; score 2: discontinuous callus; score 3: continuous bridging by immature callus; score 4: continuous remodelled callus (Figure 1,2). On each CT reformat, two fracture sites had to be scored (one score for the anterior and one for the posterior cortex on the two sagittal reformats, one score for the medial and one for the lateral cortex on the two coronal reformat). When several fracture sites were detected, the least healed site was scored. Cortical scoring was considered to be non-feasible when metal artefacts obscured the fracture or when the fracture plane was parallel to that of the reformat. The TUS was the half of the sum of the 4 cortical scores obtained on the two sagittal reformats and of the four cortical scores obtained on the two coronal reformats. TUS was considered to be feasible when at least one cortical score per area (anterior, posterior for sagittal and medial, lateral for coronal reformats) in at least three of the four areas was available.

Figure 1: Paramedian (#4 and #6) coronal (a,b), sagittal (c,d) CT reformats and AP and lateral radiographs (e,f) of a nailed tibial non-union with low RUS. The score-ratio between the sum of the cortical scores and the maximum score obtained on the CT reformats (11/32) is equal to that obtained on the x-rays (5/16).

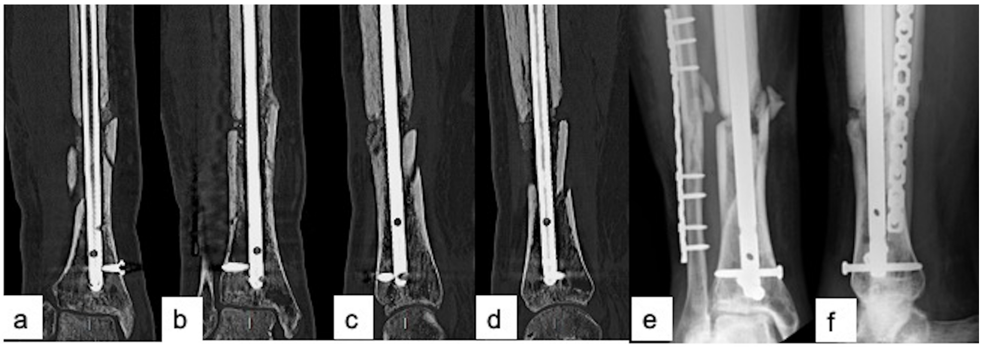
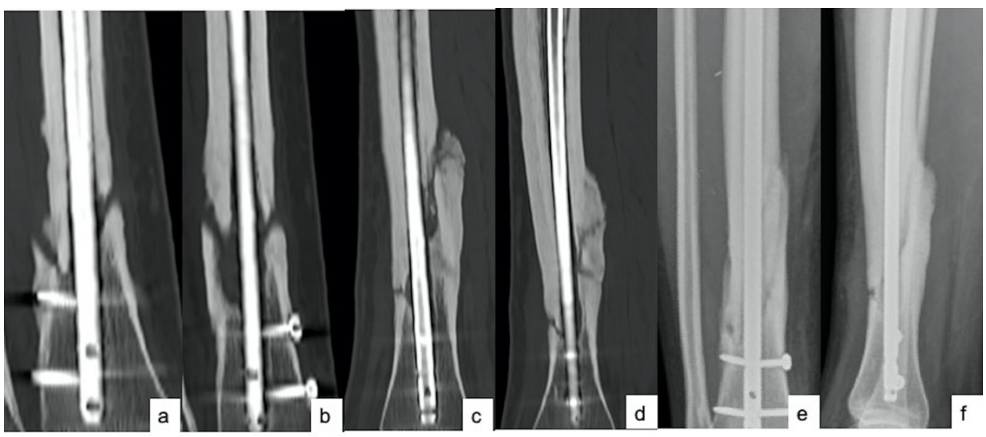


Figure 2: Paramedian (#4 and #6) coronal (a,b), sagittal (c,d) CT reformats and AP and lateral radiographs (e,f) of a nailed tibial non-union with high RUS. The score-ratio between the sum of the cortical scores and the maximum score obtained on the CT reformats (11/36) is lower than that obtained on the x-rays (10/16).



9.2.4. Statistical analysis

Intra-observer and inter-observer agreement on RUS & TUS scores were assessed using the Intraclass Correlation Coefficient (ICC) and its associated 95% confidence interval [79]. Strength of agreement was interpreted as follows: ≤ 0.20 = poor, 0.21-0.40 = fair, 0.41-0.60 = moderate, 0.61-0.80 = good and ≥ 0.81 = very good [51]. Kendall rank correlation coefficient was used to assess statistical associations between TUS score and RUS score, type of bone (femur vs tibia), patient age, sex and fracture age. Strength of association was interpreted as follows: strong ($t_b \geq 0.9$), substantial ($0.7 \leq t_b < 0.9$), moderate ($0.5 \leq t_b < 0.7$) or weak ($t_b < 0.5$) [80]. Finally, potential differences between RUS and TUS scores were assessed according to different threshold values of RUS score and the type of bone using boxplots and non-parametric Wilcoxon signed ranks test. A *p*-value < 0.05 was regarded as statistically significant [81]. Statsdirect software (https://www.statsdirect.com) was used for the statistical analysis.

9.3. Results

9.3.1. Feasibility, agreement and mean score values

The rate of feasibility of cortical scoring was 99% (91/92) for both readers on radiographs and 100% (184/184) for both readers on CT reformats. RUS and TUS were feasible in all cases. Intra- and inter-observer agreement was good for RUS and TUS (table 1).

Table 1: Intra and inter-observer agreement (ICC) of TUS and RUS

Agreement	TUS	RUS
Intra-observer ICC R1	0,75 [0,25 -0,91]	0,76 [0,51-0,89]
Intra-observer ICC R2	0,73 [0,47 -0,87]	0,78 [0,54 -0,90]
Inter-observer ICC (R1,R2)	0,70 [0,43 -0,86]	0,72 [0,44 -0,87]

The mean value(+/-SD) for TUS was 7.1(+/-1.7) for R1 and 6.8(+/-1.5) for R2.

The mean value(+/-SD) for RUS was 8(+/-1.9) for R1 and 8.1(+/-2.2) for R2.

9.3.2. Correlations between imaging scores and clinical parameters.

TUS fairly correlated with RUS (p=0.003 for R1 and p=0.004 for R2). For R1, TUS only weakly correlated with bone type (p= 0.007). There was no statistically significant correlation between imaging scores and patient age, sex and fracture age for R1. For R2, TUS weakly correlated with bone type (p=0.02) and fracture age (p=0.04) and RUS weakly correlated with patient age (p=0.01) (table 2).

Table 2: Correlations between TUS, RUS and clinical parameters

[NS for non- significant correlation. When significant correlation, the Kendall's tau b value (95% CI) shows the strength of the association between variables.(p<0,05, **p<0,001, ***p<0,0001)]*

	TUS R1	**RUS R1**	**TUS R2**	**RUS R2**
Sex	NS	NS	NS	NS
Bone type	b: 0,41 (-0,02-0,66)*	NS	b: 0,35 (-0,14-0,60)*	NS
Patient age	NS	NS	NS	b : -0,37 (-0,61- -0,06)*
Fracture age	NS	NS	b: 0,30 (0,12-0,55)*	NS
RUS	b: 0,45 (0,15-0,67)*		b : 0,43 (0,03-0,70)*	

9.3.3. Comparison between RUS and TUS according to different parameters

In fractures with RUS ≤ 8, TUS was not statistically significantly different from RUS (figure 3). In fractures with RUS ≥8, TUS was statistically significantly lower than RUS for R1 and R2 (figure 4). In tibial but not in femoral fractures, TUS was statistically significantly lower than RUS for R1 and R2 ($p<10^{-2}$).

Figure 3: Comparison between RUS and TUS according to different low RUS values

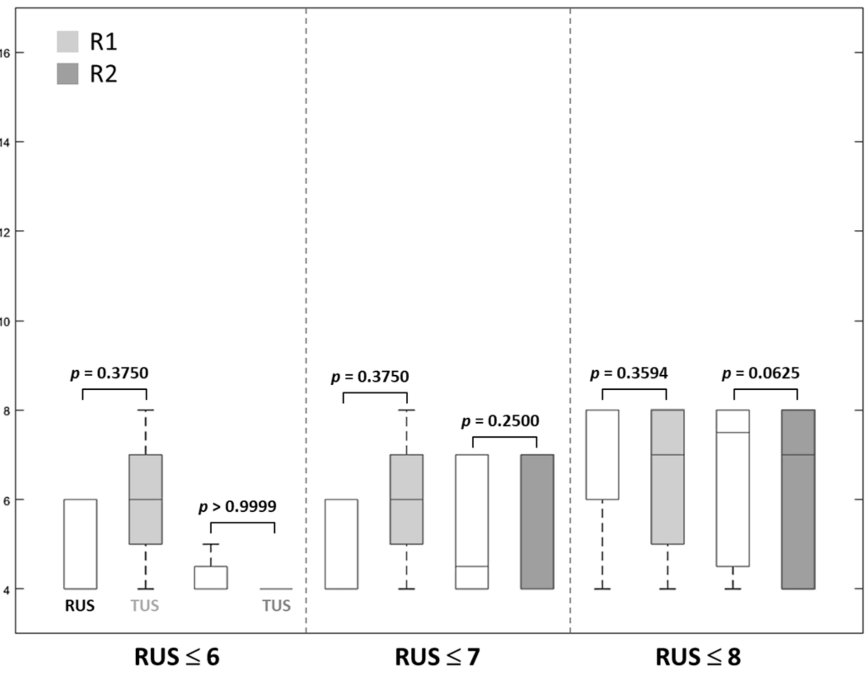
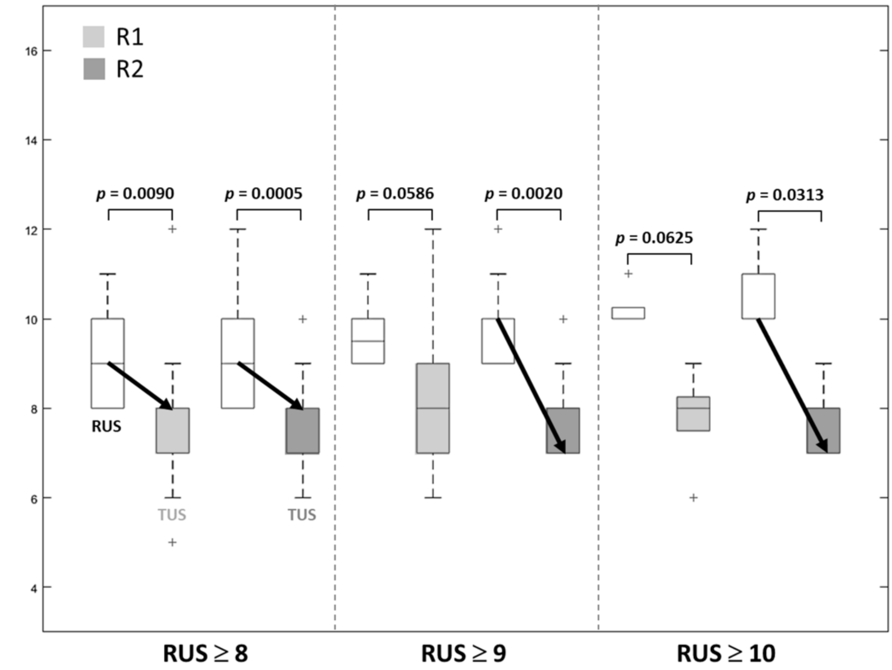


Figure 4: Comparison between RUS and TUS according to different high RUS values



9.4. Discussion

Our study of 23 nailed tibial and femoral non-unions by two readers identified three important findings:

- TUS and RUS are feasible and reproducible imaging scores to assess nailed fractures with abnormal healing.
- TUS only fairly correlated with RUS.
- TUS was statistically significantly lower than RUS in hypertrophic but not in atrophic non-unions.

First, we demonstrated that semi-quantitative assessment of callus was feasible and reproducible on CT reformats. Previous studies showed that the callus-based scoring of radiographs by using RUS was reliable with substantial agreement among surgeons and radiologists for semi-quantitative analysis of fractures [49; 57; 60-62; 75; 76] and to determine union [49; 61]. To the best of our knowledge, this is the first study to evaluate semi-quantitative CT scores in a cohort of patients with non-united shaft fractures. A more quantitative CT evaluation of fracture healing based on callus absorptiometry (mineralized callus volume and total callus volume) and callus volume measurements could not have been performed as in animal models [84-86]. In a unique study comparing RUS and micro-CT quantitative parameters in guinea pig femoral fractures stabilized by a radiolucent polyetheretherketone non-locking plate, Fiset et al demonstrated that RUS statistically significantly correlated with micro-CT parameters and with biomechanical parameters indicative of fracture healing [87]. In humans, Grigoryan et al [45] demonstrated that quantitative assessment of CT density changes in medullary cavity of fractures without metal hardware statistically significantly correlated with qualitative changes. In our study, metal artifacts associated with nails precluded the use of CT density measurements in callus. Rather than callus volume, it could be its continuity and density that most closely correlates with its biomechanical value [83; 84]. The potential of the TUS, a callus-based score integrating callus presence and continuity should be assessed in experimental studies comparing TUS with CT and biomechanical quantitative parameters.

Second, we demonstrated that TUS did not replicate RUS and only weakly correlated with RUS. Limitations of radiographs in the assessment of fracture healing have consistently been reported [21; 32; 46; 88]. CT was reported to show 100% sensitivity but limited specificity for detecting nonunion with high inter-observer agreement among readers [34]. CT delineates the osseous architecture to a very fine degree, and it can occasionally depict clefts in bone that are of uncertain clinical importance. This limitation to CT may partially explain the weak correlation between RUS and TUS with a trend to lower values for TUS than for TUS that did not reach statistical significance in our limited series.

Finally, comparison between RUS and TUS differed according to RUS values. Actually, TUS was statistically significantly lower than RUS in patients with RUS≥8 only and not in patients with RUS < 8. Actually, fracture non-union is broadly divided into hypertrophic non-union (with exuberant callus formation) or atrophic non-union (no callus and bone resorption) [12]. In cases with important callus formation (high RUS), suggesting hypertrophic non-union, CT tended to underestimate the degree of healing in comparison to radiographs. Compared to radiography, CT overcomes the problem of potentially overlapping structures, visualizes bone bridging directly, and enables a more sensitive detection of clefts in callus, not seen on the radiographs. In other terms, CT could downgrade a radiographic score of three (continuous callus) to a CT score of two (discontinuous callus). On the contrary, in patients with low RUS value suggesting rather atrophic non-union, TUS was not statistically different from RUS. Therefore, the role of a second-line CT could be less contributive when there is little or no callus formation on the radiographs. Krestan et al [69] analyzed coronal and sagittal CT reconstructions and radiographs to determine bone healing using a semi-quantitative approach and showed that MDCT altered the initial radiographic diagnosis by demonstrating different degree of fracture healing than radiographs in 37% of cases in a series of 43 fractures [69].

The current study did not aim at assessing the contribution of CT in the detection of abnormal healing as this would require a prospective study with another study population and inclusion of additional clinical parameters. The results of the current study pave the way to a more reliable evaluation of CT data whose added value in assessing fracture healing remains to be assessed in comparison with radiographs.

Several limitations to the current study should be noted. First, a limited number of patients with tibial and femoral nailed shaft non-unions were included. A study including not only a larger number of patients, but also other fixation devices should be performed. Actually, plates could interfere with the feasibility of the scoring and appearance of periosteal callus in plated fractures could indicate instability rather than normal healing. Second, readers knew they were analyzing non-united fractures as no control group was added in this preliminary study. Third, readers analyzed only a limited number of coronal and sagittal CT reformats according to our previous study (part II). Finally, feed-back retroprojection software without any additional postprocessing aiming at reducing metal-related artifacts was used in this study. The consequence on CT scoring of the use of iterative reconstruction and metal artifact reduction techniques should be assessed.

In conclusion, semi-quantitative CT assessment of nailed non-united shaft fractures by using a callus-based score is feasible and reproducible. TUS differ from RUS in cases with important callus formation.

10. Part V: **Assessment of healing of surgically treated long bone fractures using a radiographic score: Prediction of delayed union**

10.1. Introduction

Non-union of a fracture represents a challenging clinical scenario for the orthopaedic trauma surgeon and the radiologist. Diagnosis in itself can be difficult and treatment consumes considerable time and resources. Currently, there is no reliable way to predict which patients will suffer from nonunion. The development of a method to do so would allow for early intervention to reduce the complications of nonunion and the associated morbidity. Furthermore, such early intervention could shorten the recovery time significantly, resulting in a significant impact on health care system costs and society.

Several published studies showed that non-unions can be diagnosed early in the course of healing [89-91]. Bhandari et al [92] identified a set of three prognostic variables that can assist surgeons in predicting reoperation following operative treatment of tibial shaft fractures: the presence of an open fracture wound, lack of cortical continuity between the fracture ends following fixation and the presence of a transverse fracture. Fong et al. [93] retrospectively studied 200 patients treated with IM nails or plating and found that less than 25% cortical continuity was the only factor predictive of nonunion in their series. Lack et al [94] showed that any cortical bridging within four months was an excellent predictor of final healing. Salih et al [95] described the callus fracture sign (extension of the fracture line beyond the original cortical boundary but not to the periphery of the fracture callus) as a useful radiological predictor of progression to non-union. A prognostic radiologic scoring system capable of identifying fractures as likely or unlikely to progress to union during the first few months after injury would be potentially more useful in clinical practice. The RUST score proposed by Whelan et al. [49] in 2010 has drawn the attention of the musculoskeletal community as indicated by the number of studies published on the subject over the last few years [50; 57; 60-62; 75; 76; 96; 97]. In our previous studies we demonstrated that a modified version of the score, the Radiographic Union Score (RUS) , was a feasible assessment tool of fracture healing with substantial interobserver agreement among radiologists and surgeons.

The aim of this retrospective longitudinal study was assess changes in RUS value over time during fracture healing in a series of surgically treated tibial and femoral shaft fractures and to assess the accuracy of the RUS value in the early detection of fracture non-union.

10.2. Material and methods

10.2.1. Selection of patients

The authorization of the ethical committee of our institution had been obtained for this retrospective study. In July-August 2015, a musculoskeletal radiologist with 3 years of experience (V.P.) and a final year resident in radiology (A.C.) retrieved the radiographs of thighs and legs obtained in our emergency department between January 2007 and January 2015 in patients aged between 18 and 80 years by using the keywords "thigh" and "leg". The keywords “pelvis” and “hip” were not used to avoid detection of patients with possible hip fractures in our search. They detected 482 patients in whom radiographs with femoral or tibial fractures had been obtained at the emergency department. Subsequently, the entire corresponding radiological files were examined by the same radiologists to select patients according to the following inclusion criteria : (a) diaphyseal femoral or tibial fracture, (b) surgical fixation by using an internal fixation hardware (nail or plate) and (c) availability of at least 3 sets of antero-posterior and lateral radiographs obtained at baseline after surgery until 6-months after fracture with at least one radiographic set obtained between 15 and 24 weeks after fracture. Four hundred and four tibial or femoral fractures were excluded because of conservative treatment (n= 35), associated epiphyseal fracture (n= 277), multiple fractures (n=7), pathological fracture/ septic fracture/ pre-existing metallic hardware (n= 36) and lack of 3 sets of radiographs (n= 49). Seventy-eight fractures in 78 patients were eligible. In our PACS, there were 347 radiographic sets obtained at follow-up in these 78 fractured bones according to our guidelines that recommend serial radiographs at hospital discharge and at approximately 1-, 3- and 6-months after fracture.

10.2.2. Selection of radiographs

In September 2015, a data manager encoded the 347 radiographic examinations in an excel spreadsheet and added a categorical variable according to the delay between the fracture and the moment at which the radiographs had been obtained. The 6-month follow-up time was divided in 7 periods defined as follows: 1-2 weeks after fracture (T0); 3-4 weeks (T1); 5-7 weeks (T2); 8-10 weeks (T3); 11-14 weeks (T4); 15-18 weeks (T5) and 19-24 weeks (T6). The 347 sets of radiographs were anonymized, re-identified after randomization by an independent statistician (C.B.) and stored in our PACS. No quality control was performed on the included radiographs. Radiographs that demonstrated staples or casts were included despite the fact that these features could provide hints for the fracture age.

10.2.3. Clinical data

In February 2016, the orthopedic surgeon with 15 years of experience responsible of the trauma section in the orthopedic department reviewed the files of the 78 patients to determine for each patient the smoking habit, the fracture type according to the AO classification, the type of fixation device and the clinical outcome. According to the AO classification, there were 42 type A fractures, 15 type B fractures and 21 type C fractures. The type of metallic hardware used for fixation was classified in two categories. Nailed fractures were stabilized only by medullary hardware including intramedullary nails and rods. Plated fractures were stabilized by material placed around or within the cortex including plates and free screws. No fractures were stabilized simultaneously by plates and nails.

The clinical outcome of the fracture was determined by the same orthopedist by reviewing the medical files completed by the surgeon in charge of the patient at a mean delay of 193 days +/- 35 days (range: 117-283 days) after fracture. Healing evaluation had been performed according to standards of good practice based on clinical and radiological findings at that time. Normal healing was defined by the ability of the patient to walk without assistive device with any pain at the fracture site and by the overall estimation by the surgeon of the presence of bridging callus on radiographs. Delayed union was defined by the presence of pain when bearing weight or upon palpation on physical examination and by the presence of unsatisfactory appearance of the radiographs. Delayed union was also considered when surgical revision with screw ablation was performed during the first 6 months. Skin debridement was not considered as revision surgery.

10.2.4. Study population

The study population included 44 men and 31 women (mean age +/- SD: 49.73 +/- 17.51) including 21 smokers. There were 61 (81,3%) tibial (47 nailed and 14 plated fractures) and 14 (18,7%) femoral fractures (11 nailed and 3 plated fractures). At clinical follow-up, normal healing was noted in 61 (81.3%) fractures and delayed union was noted in 14 (18.7%) fractures including 8 (17,9 %) nailed and 2 (14.2%) plated tibial fractures and 2 (18.2 %) nailed and 2 (66 %) plated femoral fractures. Mean age, sex ratio and frequency of smokers in the group of patients with normal healing were not statistically significantly different from those in the group of patients with delayed union.

10.2.5. Analysis of the radiographs

The 347 radiographic examinations were read separately within a period of 4 weeks by three radiologists including the two radiologists (V.P., A.C.) who had included the patients and a musculo-skeletal radiologist with 25 years of experience (B.V.). The readers were blinded to all clinical information and they were asked not to discuss the cases within the reading period. A unique reading was performed per reader. They were asked to score each set of AP and lateral radiographs on a PACS workstation (Ecrans Eizo Radiforce RX240 21.3'' (54 cm) 2MP (1200 x 16)). The readers had to analyse the automatic cropping of the radiographic image without access to the zoom modality. They could adjust the window level and width at their convenience. Similar luminance conditions were reproduced in the reading room (20 lux). Each reader was asked to decide whether the scoring of each of the 4 cortical areas was feasible or not. If feasible, scoring was performed using the Radiographic Union Score (RUS) as described in part 1. In fractures in which one cortical score was not feasible, the RUS was calculated by adding the 3 determined scores and their calculated mean as missing data. In fractures in which less than 3 cortical scores were available, RUS was not calculated and was considered to be non-feasible. After this initial reading session, we excluded patients in whom almost all radiographic sets were found not assessable by at least two readers. We also excluded radiographic sets of included patients in whom one or two radiographic sets were found not assessable by at least two readers. Finally, in 12 cases, a radiographic set was found not assessable by only one reader; the readers were asked by our data manager in charge of the study to perform a second reading and to score the fracture, knowing that he/she was the unique reader who had considered the corresponding radiographic set to be not assessable.

10.2.6. Statistical analysis

All summary tables, data listings and graphs are produced using the statistical software JMP v12. All statistical tests were conducted at the significance level of 5%, unless otherwise stated. Descriptive statistics were done for each variable: continuous variables are presented using the mean, standard deviation (SD), standard error of the mean (SEM) and 95%-confidence interval. For categorical variables, the population size (N), the number of events, the number of subjects with events (n) and the proportion of subjects with events (p=n/N) are reported.

Differences in patient characteristics between the two clinical endpoints were tested using 2- sided Mann-Whitney test for continuous variables and chi-squared test for categorical variables. Mean RUS score between the two clinical endpoints were tested using 2-sided Mann-Whitney test.

The intra-class correlation coefficient (ICC) adapted to the random factorial design was computed to assess the agreement between the three readers. The interobserver agreement was interpreted according to Landis and Koch, with ICC values < 0.00 defined as poor, 0.00-0.20 defined as slight, 0.21-0.40 defined as fair, 0.41-0.60 defined as moderate, 0.61-0.80 defined as substantial, and values of 0.81 -1.00 defined as almost perfect [51].

Binary logistic regression was used to estimate the probability of healing in order to find a threshold for the total RUS value to predict healing. Threshold was evaluated using a ROC (receiver operating characteristic) curve and the area under the curve (AUC) was reported to evaluate performance of the prediction.

A two-way ANOVA with random effects was performed to assess (i) differences in scores and, (ii) determine which factor (time after injury or type of fixation device) or combination of factors (time * type of fixation device) was associated with the difference. Both time (in days) and type of fixation device were considered to represent a fixed factor, while readers were considered to represent a random sample drawn from a larger population of observers (random factor). A third term was added to the ANOVA mixed model to account for potential interaction between both factors. In case where the p-value associated with that term were not statistically significant, the interaction term was to be removed and the ANOVA mixed model to be repeated with only 2 main effects. In case where the p-value associated with the interaction term were statistically significant, individual effects would not be interpreted and their associated p-values given).

10.3. Results

10.3.1. Feasibility of the scoring system

The three readers were able to score 323 of the 347 radiographic sets in 75 of the 78 patients. Three patients with plated tibial fractures were excluded because of non-feasibility of the scoring on all their 14 available radiographic sets. Ten out of the 333 radiographic sets available in the 75 other patients were also excluded for the same reason.

10.3.2. Comparison of the mean RUS values between fractures with normal healing or delayed union

For each reader, the mean RUS values calculated at each time point showed a progressive increase over time for nailed and plated fractures with normal healing (figures 1,2,3) but not for those with delayed union at follow-up (figures 4,5) (table 1). For all readers confounded, the mean RUS increased significantly over time for fractures with normal healing ($p<0.001$) and was significantly higher for nailed than for plated fractures ($p=0.0012$) (figure 6). Interobserver agreement for the RUS values varied from substantial to almost perfect (ICC: 0.79 for all fractures, ICC: 0.81 for nailed and ICC: 0.72 for plated fractures).

In nailed fractures, the mean RUS values were statistically significantly higher in fractures with normal healing than in fractures with delayed union at timepoints T4 (from 11 to 14 weeks after fracture) and timepoint T6 for all readers ($p<0.032$) and also at timepoint T5 for reader 2 ($p=0.01$) (table 1). In plated fractures, there was no statistically significant difference in the mean RUS values between the 2 groups of fractures, except for reader 1 at timepoint T3 ($p=0.04$) (table 1).

Table 1: Mean RUS values (+/- standard deviation) at each time point for the 3 readers according to healing status at 6-months follow-up in nailed and plated fractures. [p-value : * significant at p< 0.05, ** significant at p< 0.01, *** significant at p< 0.001, **** significant at p<0,0001]

	NAILS		**PLATES**	
T0	Normal healing (n= 36)	Abnormal healing (n= 7)	Normal healing (n= 6)	Abnormal healing (n= 2)
Mean RUS (+/-SD) Reader 1	4 (+/-0)	4 (+/-0)	4 (+/-0)	4 (+/-0)
Mean RUS (+/-SD) Reader 2	4,31 (+/-0,72)	4,57 (+/-1,13)	4 (+/-0)	4 (+/-0)
Mean RUS (+/-SD) Reader 3	4,39 (+/-1,02)	4 (+/-0)	4 (+/-0)	4 (+/-0)
T1	Normal healing (n= 22)	Abnormal healing (n= 4)	Normal healing (n= 5)	Abnormal healing (n= 2)
Mean RUS (+/-SD) Reader 1	4,36 (+/-1,18)	4 (+/-0)	4 (+/-0)	4 (+/-0)
Mean RUS (+/-SD) Reader 2	4,56 (+/-1,23)	4,25 (+/-0,5)	4 (+/-0)	4,65 (+/-0,92)
Mean RUS (+/-SD) Reader 3	4,95 (+/-1,68)	4 (+/-0)	4 (+/-0)	4 (+/-0)
T2	Normal healing (n= 35)	Abnormal healing (n= 9)	Normal healing (n= 6)	Abnormal healing (n= 4)
Mean RUS (+/-SD) Reader 1	4,8 (+/-1,45)	4,11 (+/-0,33)	4,22 (+/-0,53)	5 (+/-1,29)
Mean RUS (+/-SD) Reader 2	5,39 (+/-1,49)	4,4 (+/-0,89)	5,32 (+/-2,21)	5,2 (+/-1,90)
Mean RUS (+/-SD) Reader 3	5,43 (+/-2,05)	4,33 (+/-0,71)	5,14 (+/-1,95)	5,25 (+/-1,89)
T3	Normal healing (n= 23)	Abnormal healing (n= 3)	Normal healing (n= 6)	Abnormal healing (n= 1)
Mean RUS (+/-SD) Reader 1	5,5 (+/-1,73)	4,33 (+/-0,58)	4 (+/-0)	6,7 ***** NA
Mean RUS (+/-SD) Reader 2	6,33 (+/-1,92)	5,17 (+/-1,61)	5,22 (+/-1,60)	6,6 NA

Mean RUS (+/-SD) Reader 3	6,33 (+/-1,82)	4,33 (+/-0,58)	5,14 (+/-1,77)	8 NA
T4	Normal healing (n= 35)	Abnormal healing (n= 8)	Normal healing (n= 8)	Abnormal healing (n= 4)
Mean RUS (+/-SD) Reader 1	6,96 (+/-2,51)	4,75 * (+/-0,89)	6,66 (+/-3,72)	5,85 (+/-1,28)
Mean RUS (+/-SD) Reader 2	7,43 (+/-2,05)	5,08 * (+/-1,08)	7,29 (+/-2,52)	6,3 (+/-1,67)
Mean RUS (+/-SD) Reader 3	7,07 (+/-1,94)	4,25 *** (+/-0,46)	5,63 (+/-2,26)	5,75 (+/-2,06)
T5	Normal healing (n= 21)	Abnormal healing (n= 4)	Normal healing (n= 7)	Abnormal healing (n= 2)
Mean RUS (+/-SD) Reader 1	7,71 (+/-2,44)	5,25 (+/-1,89)	5,86 (+/-3,08)	6,65 (+/-3,75)
Mean RUS (+/-SD) Reader 2	9,26 (+/-2,69)	5,65 * (+/-1,68)	8,1 (+/-3,03)	6,15 (+/-1,20)
Mean RUS (+/-SD) Reader 3	8,12 (+/-2,55)	5,5 (+/-2,38)	8 (+/-3,90)	6,5 (+/-3,54)
T6	Normal healing (n= 43)	Abnormal healing (n= 9)	Normal healing (n= 8)	Abnormal healing (n=3)
Mean RUS (+/-SD) Reader 1	10,35 (+/-3,34)	5,67 *** (+/-1,35)	7,63 (+/-3,85)	7,77 (+/-1,66)
Mean RUS (+/-SD) Reader 2	10,66 (+/-2,78)	5,83 **** (+/-1,17)	9,51 (+/-2,76)	7,7 (+/-0,98)
Mean RUS (+/-SD) Reader 3	9,91 (+/-2,70)	5,61 **** (+/-1,45)	8,3 (+/-3,86)	7,67 (+/-0,58)

Figure 1: Graph showing the mean RUS values determined in nailed and plated fractures that are healed at 6-month follow-up at different timepoints (T0: 1-2 weeks after injury; T1: 3-4 weeks; T2: 5-6 weeks;T3: 7-9 weeks;T4: 10-13 weeks;T5: 14-17 weeks and T6: 18-24 weeks) by three readers.

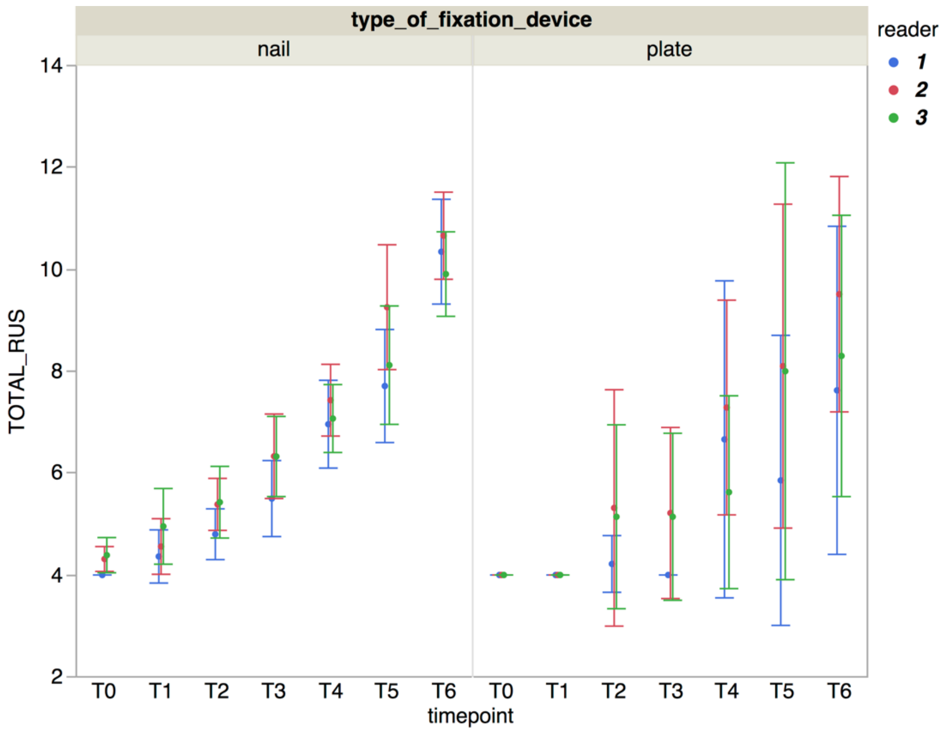


Figure 2: Normal healing: Serial radiographs of a nailed tibial fracture obtained a,b)3 weeks, c,d) 12 weeks and e,f) 23 weeks after injury demonstrate progressive appearance of callus. The corresponding RUS values as determined by Reader 1/ Reader 2/Reader 3 were 4/4/4 in T1, 8/9/8 in T4 and 14/15/14 in T6.

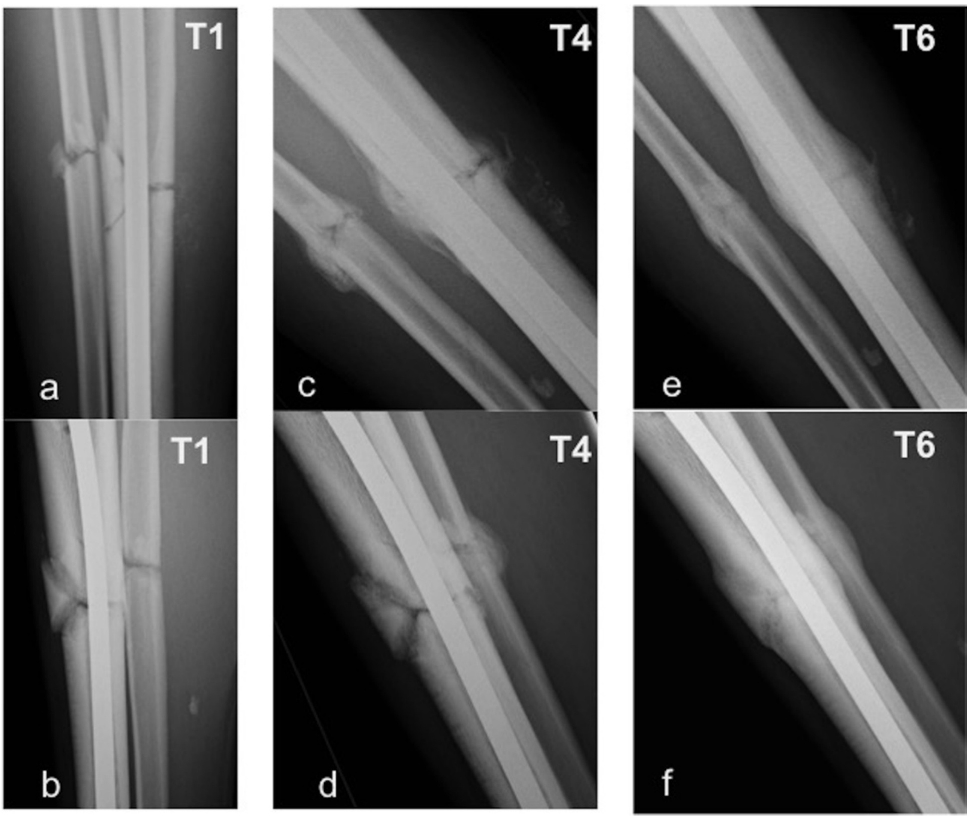


Figure 3: Normal healing: Serial radiographs of a plated femoral fracture obtained a) 3 weeks, b) 12 weeks and c) 23 weeks after injury demonstrate progressive appearance of callus. The corresponding RUS values as determined by Reader 1/ Reader 2/Reader 3 were 4/4/4 in T1, 8/8/8 in T4 and 12/14/14 in T6.

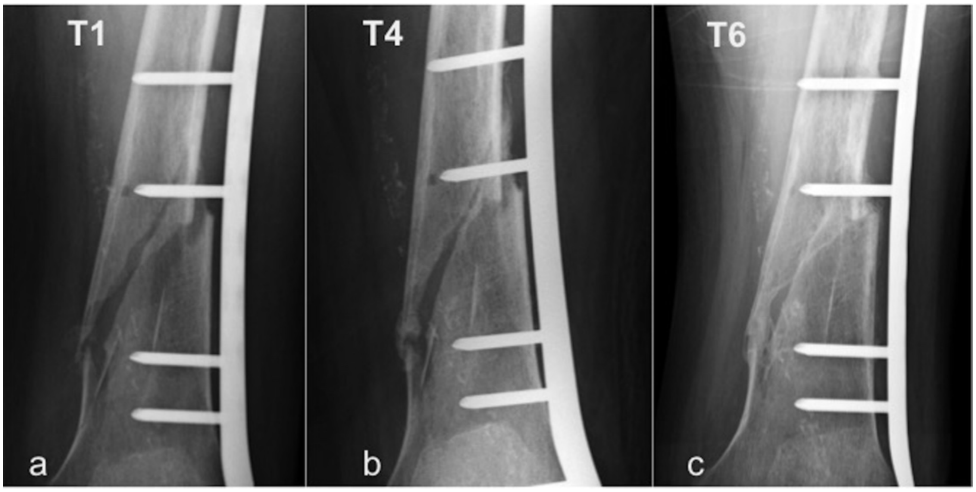


Figure 4: Graph showing the mean RUS value determined in nailed and plated fractures with delayed union at 6-month follow-up at different timepoints (T0: 1-2 weeks after injury; T1: 3-4 weeks; T2: 5-6 weeks;T3: 7-9 weeks;T4: 10-13 weeks;T5: 14-17 weeks and T6: 18-24 weeks) by three readers.

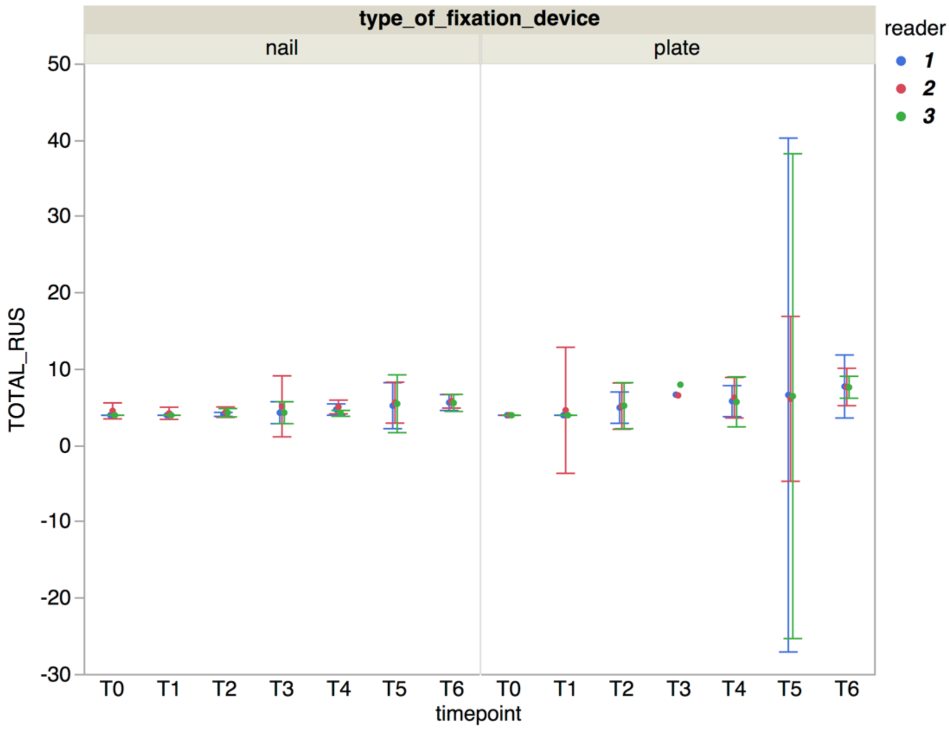


Figure 5: Abnormal healing: Serial radiographs of a nailed tibial fracture obtained a,b) 3 weeks, c,d) 12 weeks and e,f) 22 weeks after injury demonstrate lack of appearance of callus. The corresponding RUS values as determined by Reader 1 /Reader 2/Reader 3 were 4/4/4 in T1, 4/4/4 in T4 and 4/6/5 in T6.

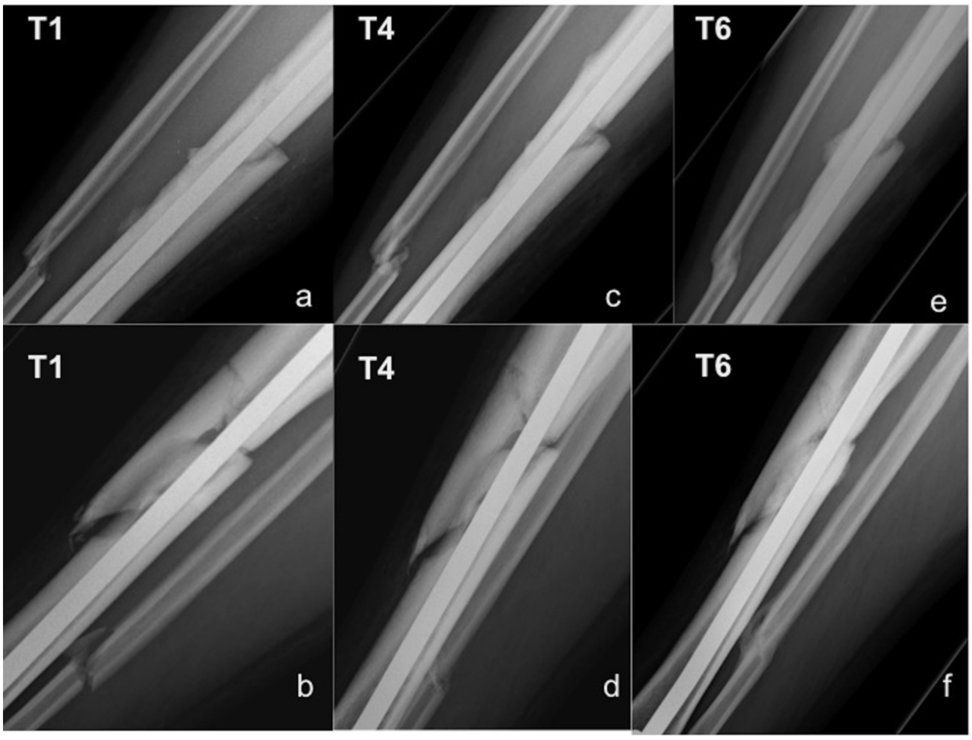
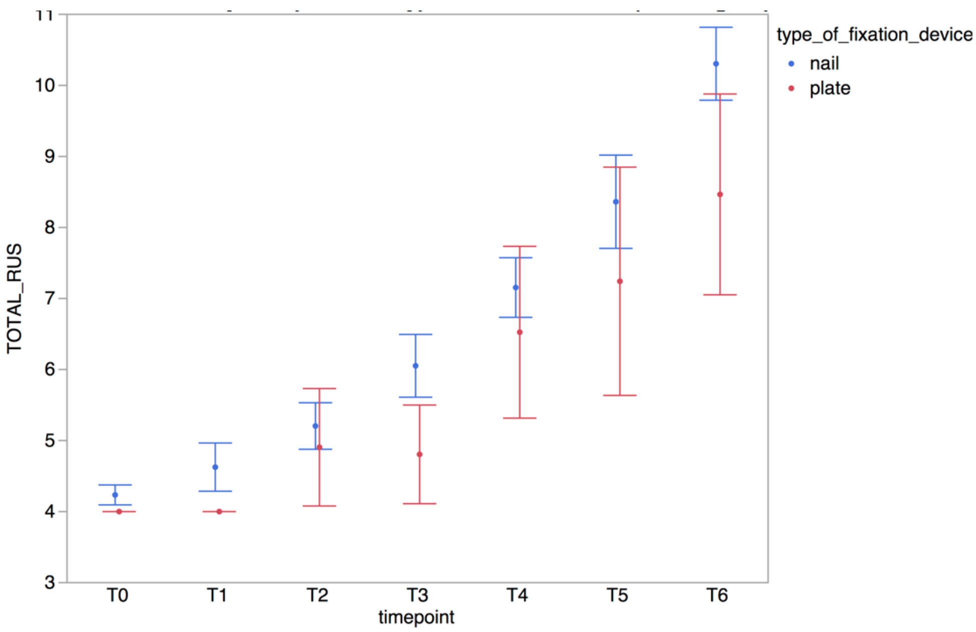


Figure 6: Graph showing the evolution of the mean mRUS value for the three readers confounded at different timepoints in nailed and plated fractures with normal healing at 6-month follow-up.



10.3.4. Definition of threshold RUS value for prediction of delayed union

The RUS threshold, sensitivity and specificity values for detecting delayed union in nailed fractures at each timepoint were derived from ROC curves (Table 2). At T4 (from 11 to 14 weeks after injury), a RUS score lower than 7 (reader 1), 7.5 (reader 2) or 5.5 (reader 3) enabled detection of patients with delayed union with a sensitivity that ranged from 0.63 to 0.77 and a specificity of 1 (AUC: 0.77- 0.88).

Table 2: Sensitivity, specificity and AUC derived from ROC curves to predict delayed union for nailed fractures at each time point.

Delay	Reader 1				Reader 2			
	RUS	Se	Spe	AUC	RUS	Se	Spe	AUC
T0	5.5	0.82	0.21	0.5	6	0.97	0.14	0.55
T1	8	0.09	1	0.55	5.3	0.18	1	0.51
T2	7	0.2	1	0.60	5.3	0.51	0.89	0.70
T3	5.5	0.43	1	0.67	8	0.39	1	0.67
T4	7	0.63	1	0.77	7.5	0.66	1	0.83
T5	5.3	0.81	0.75	0.79	5.5	1	0.75	0.90
T6	8	0.84	0.89	0.91	7	0.93	0.89	0.95

Delay	Reader 3			
	RUS	Se	Spe	AUC
T0	5	0.17	1	0.58
T1	6	0.27	1	0.64
T2	6	0.37	0.89	0.65
T3	7	0.57	1	0.79
T4	5.5	0.77	1	0.89
T5	6.5	0.81	0.75	0.77
T6	8	0.84	0.89	0.93

The RUS threshold, sensitivity and specificity values for detecting delayed union in plated fractures at each time point were derived from ROC curves (Table 3) but no threshold value could be proposed because of the insufficient number of plated fractures.

Table 3: Sensitivity, specificity and AUC derived from ROC curves to predict abnormal healing for plated fractures at each time point

Delay	Reader 1				Reader 2			
	RUS	Se	Spe	AUC	RUS	Se	Spe	AUC
T0	NA	NA	NA	0,5	NA	NA	NA	0,5
T1	NA	NA	NA	0,5	4	1	0,5	0,75
T2	4	0,83	0,5	0,69	9,3	0,17	1	0,46
T3	NA	NA	NA	NA	6	0,83	1	0,83
T4	8	0,375	1	0,47	7,5	0,5	0,75	0,59
T5	8	0,8571	0,5	0,57	6	1	0,5	0,75

Delay	Reader 3			
	RUS	Se	Spe	AUC
T0	NA	NA	NA	0,5
T1	NA	NA	NA	0,5
T2	4	0,71	0,5	0,57
T3	5	0,86	1	0,86
T4	4	0,63	0,5	0,5
T5	5	0,58	0,5	0,58

10.4. Discussion

Our study of 75 nailed and plated tibial and femoral shaft fractures by three readers identified three important findings:

- The mean RUS increased significantly over time in fractures with normal healing but not in fractures with delayed union.
- In nailed fractures only, the RUS values, determined on radiographs obtained at about 3 months after injury ,were statistically significantly higher in fractures with normal healing than in fractures with delayed union in F/U.
- For each of the 3 readers, the RUS value determined at approximately 3 months after injury enables to detect fracture with delayed union with a sensitivity that ranged from 0.63 to 0.77 and a specificity of 1.0.

First, the RUS calculated in fractures with normal healing increased statistically significantly over time. Little information on changes in score value over time was available in fractures with normal healing up to now. In a longitudinal study of hip fractures, Bhandari et al showed that there was an increase in the score values with time. The fact that readers were allowed to score by comparing serial radiographs limits the robustness of that study [61]. In a transverse study of 35 radiographs obtained at different time delays from injury in 35 distal radius fractures with normal healing, Patel et al showed that mean RUS values also increased with time [57]. To the best of our knowledge, the current study demonstrated for the first time the evolution of the RUS values over time for tibial and femoral shaft fractures with a robust methodology because there were 3 readers who were blinded to all data including the other radiographs and the outcome at follow-up. In addition, we showed that mean RUS values were higher in nailed than in plated fractures. This difference is likely to be related with the use of a callus-based scoring system. Actually, periosteal callus formation is expected to occur in nailed but not in plated fractures because of different biomechanical environment [4; 63; 64; 98; 99]. The rigid plate fixation may not result in callus formation due to the lack of micro-mobility between fracture edges. In plated fractures, bridging occurs through direct cortical bone formation with limited, if any, contribution from the periosteum.

Second, in nailed fractures only, the mean RUS values determined on radiographs obtained within a time frame of 11 to 14 weeks after injury were statistically significantly higher in fractures with normal healing than in fractures that developped delayed union at follow-up. The possibility to predict fracture healing at about three months after surgery has already been suggested but not demonstrated in several studies. According to Bhandari et al, surgeons tend to endorse early signs of non-union by the 14th week for tibial and femoral shaft fractures [21]. In 2013, another study demonstrated that tibial nonunion can be predicted at three months postoperatively using clinical and radiographic data including mechanism of injury and absence of callus , however with a poor interobserver agreement [90]. In our series, no difference in RUS values could be detected between the two groups of fracture at a delay shorter than 11 weeks after injury. In the time frame of 15 to 18 weeks after injury, mRUS valus were higher in fractures with normal than with abnormal healing for the three readers but statistical significance was reached for only one reader probably due to the small number of radiographs available in that time period. Lack et al demonstrated in a retrospective study that any cortical bridging at 18 weeks after fracture was an excellent predictor of normal healing (accuracy=99%, AUC = 0.995, p< 0.0001) [94]. The added value of a scoring system over an overall estimation of fracture healing on radiographs should be further assessed but several studies have already demonstrated a higher reproducibility by using a scoring system instead of an overall estimation [58; 61]. As expected, in plated fractures, no statistically significant difference in RUS values was noted between fractures with normal or abnormal healing except for one reader at a time delay between 8 and 10 weeks after injury. The most likely explanation might be that, as indicated earlier, this scoring system might not be appropriate to evaluate fracture healing in plated fractures because of different biomechanical environment [4; 63; 64; 98; 99].

Finally, in the group of nailed fractures, we demonstrated the existence of reader-dependent threshold RUS values that enabled with a high specificity the detection of fractures that would later demonstrate delayed union. Other studies also aimed at defining a threshold values in order to predict non-union. Frank et al [97] using the radiographic union score for hip fractures (RUSH) suggested that a threshold score below 18 was associated with 100% specificity and a positive predictive value of 100% for radiographic non-union at 6 months post-injury. Litrenta et al [50] suggested that a standard RUST of 8.5 and a modified RUST of 11.4 correlated with union but they did not specify their time delay after injury. To the best of our knowledge, the current study is the first to suggest a threshold RUS value to predict abnormal healing in nailed tibial and femoral shaft fractures at a delay of about 3 months after fracture. Patel et al [57] found RUS values greater than or equal to 6 being predictive of fracture union in distal radial fractures but they did not neither specify their time delay after injury. A recent study in 2019 by Oliver et al [62] showed that RUS was a reliable and effective tool in identifying patients at risk of developing nonunion at six weeks after injury, with a RUS of < 8 significantly predictive of nonunion in a series of humeral shaft fractures with conservative treatment. To the best of our knowledge, the current study is the first to suggest a threshold RUS value to predict abnormal healing in nailed tibial and femoral shaft fractures at a delay of about 3 months after fracture.

Our study has several important limitations. First, it is retrospective and patients may have been subjected to selection biases for the obtention of follow-up radiographs. Patients in whom less than 3 sets of radiographs were available were not included. Second, the healing status that is normal healing or delayed union was defined by an orthopaedic surgeon based on files content including notes of the treating surgeon and the reports of the radiographs. Subsequently, there was a large variation in the time delay at which the outcome status was defined. Third, we could not avoid diagnostic review bias because the imaging modality that we evaluated, a radiographic set, was also used in the determination of the clinical outcome, standard of reference in our study. However, reporting radiologists and treating surgeons used an overall estimation of the radiographs in clinical practice and did not perform any scoring. Finally, two of the three readers were also involved in the selection process based on X-rays and all readers belonged to the same radiological department.

In conclusion, the RUS values determined on serial radiographs of surgically treated tibial and femoral fractures increased during the first 6 months after injury in fractures with normal healing. Threshold values of RUS enabled detection of fractures with abnormal healing at approximately three months after injury. Our results await further confirmation in larger prospective and multicentric studies. In addition, the added value of this scoring system with respect to clinical evaluation remains to be determined.

11. SUMMARY OF RESULTS AND PERSPECTIVES

Musculoskeletal trauma represents an important global health burden with approximately six million fractures each year in the United States. As many as 10% of these may be complicated by non-union with a subsequent increase in medical costs and productivity loss [100]. There is a need for an accurate and reliable assessment of radiographic healing given the wide range of parameters used for fracture union and the large body of research being conducted on novel fracture-healing therapies.

In part I, we addressed the limitations of conventional radiography for the evaluation of healing in a series of surgically treated long bone fractures. We showed variable feasibility rates for the Radiographic Union Score (RUS) depending on the type of stabilizing hardware, with statistically significantly higher rates for nailed fractures. Interobserver agreement when using RUS was substantial among radiologists and orthopedic surgeons.

In part II, we proposed a CT scoring system (TUS) and showed that the analysis of a limited number of long-axis CT reformats was a valuable alternative to the analysis of the large number of CT reformats obtained in the three planes.

Third, we analyzed the utility and reliability of the scoring system in nailed tibial and femoral fractures with normal healing (part III) and non-union (part IV) using radiographs and CT scans. TUS and RUS were feasible and reproducible imaging tools of semi-quantitative assessment of fracture healing in both cohorts of fractures. CT provided different score values compared to x-ray in later phases of normal healing and in hypertrophic non-unions.

Finally, in part V, we studied the predictive value of the radiographic scoring system and demonstrated that RUS could contribute to detect fractures with delayed union at about 3 months after injury in nailed fractures.

The following questions remain to be answered and open new perspectives for future research projects:

1) Is there a place of other imaging modalities in the evaluation of fracture healing?

First, recent development in CT imaging using metal artifact CT techniques and iterative reconstructions should be included in future studies.

Second, our results could pave the way to the use of digital tomosynthesis (DTS) for quantitative fracture healing evaluation. DTS does not provide transverse tomographic images, but according to our analysis, short -axis CT images were shown to be less accurate or of uncertain interpretation in the evaluation of fracture healing.

Third, the role of texture analysis and arteficial intelligence in the quantitative assessment of fracture healing could be adressed in future studies.

Six years later after the start of the current thesis, Artificial Intelligence (AI) is about to disrupt the diagnostic field of imaging by introducing new ways to exploit imaging data. By the end of March 2020, a pubmed research using the terms “fracture” “detection” “deep” “learning” yielded 18 articles. What about a pubmed research using the terms “fracture” “healing” “deep” “learning” ? Not a single article is retrieved, so far! “Fracture detection” is a preliminary step of the AI adventure in trauma (diagnosis). “Fracture healing” is definitely a limit-free domain where the potential of AI should be developed with three main questions addressing:

(1) Prediction of slow or delayed healing at the onset of tibial fracture,

(2) Monitoring of healing,

(3) Determination of healing and of sequelae.

There is a general agreement on the predictive value for delayed union of several parameters related to the fracture including (a) fracture anatomy, (b) adjacent soft tissue damage, and (c) infection. Parameters related to the patient have also been identified including (a) age, (b) smoking habit, (c) systemic disorders such as diabetus mellitus. Definitely, AI is likely to add several parameters derived from a more extensive medical history of the patient, from a complete recording of the trauma circumstances and from photographs of the involved leg at the entrance of the emergency department. What if the circumference of the contralateral uninvolved leg combined with the hypodermis thickness was also a predictive parameter?

To illustrate the challenges that prospective studies will face, it is worth mentioning a list of events that limit patient inclusion in a large monocentric study as performed by Ross et al who attempted to develop a model to predict delayed union for tibial shaft fractures [101]. A total of 987 patients with acute tibial shaft fractures were identified at the onset of fracture. 664 (67%) patients were excluded from the study because of inadequate follow-up (n=336), additional operative interventions (n=118), missing radiographs (n=58), age below 18 (n=50), and miscellaneous events (premature death, amputations, definitive fixation at an outside institution, missing chart data, pathological fractures, periprosthetic fractures, pregnancy, indeterminate fracture healing, nonoperative treatment, stress fracture). The remaining 323 patients were included in the study. Most likely, AI will contribute to analyze prospective cohorts of patients without the need to exclude the vast majority of them.

Monitoring of healing will most likely benefit from permanent recording of gait and weight-bearing patterns through dedicated connected shoes, that is likely to reflect influxes from deep bone pain originating from fracture site and adjacent soft tissue. What will be the residual value of clinical and radiographic examination once remote patient monitoring will become efficient?

Finally, definition of healing remains debated with limited data to determine when a fracture is healed. Restauration of bone continuity on radiographs and lack of symptoms are the most widely accepted criteria despite challenges in their evaluation. An AI definition of fracture healing could be a restauration of normal lower limb function as assessed by permanent patient monitoring. Once again, how will “normal” be defined?

AI and remote patient monitoring will definitely upgrade the collection and the analysis of data derived from patients with tibial shaft fractures, most likely downgrading the value of clinical and radiological examinations. Fracture scoring has only been one of the desperate attempts of humans to understand fracture healing in the pre-AI era.

2) Are RUS and TUS validated imaging tools to quantify fracture healing and predict delayed union?

Our results indicated that RUS and TUS were feasible and reproducible imaging tools in the semi-quantitative evaluation of nailed tibial and femoral fractures in particular. Moreover we determined the predictive value of RUS in identifying lower limb fractures with delayed union.

Further prospective longitudinal studies should be conducted with larger sample size to correlate the imaging scores with clinical information, including lower and upper limb fractures with surgical and conservative treatment.

Future experimental studies comparing TUS with CT volumetry and histopathological findings would be useful in order to validate TUS in the monitoring of fracture healing.The predictive value of TUS in identifying fractures with delayed union also needs to be evaluated.

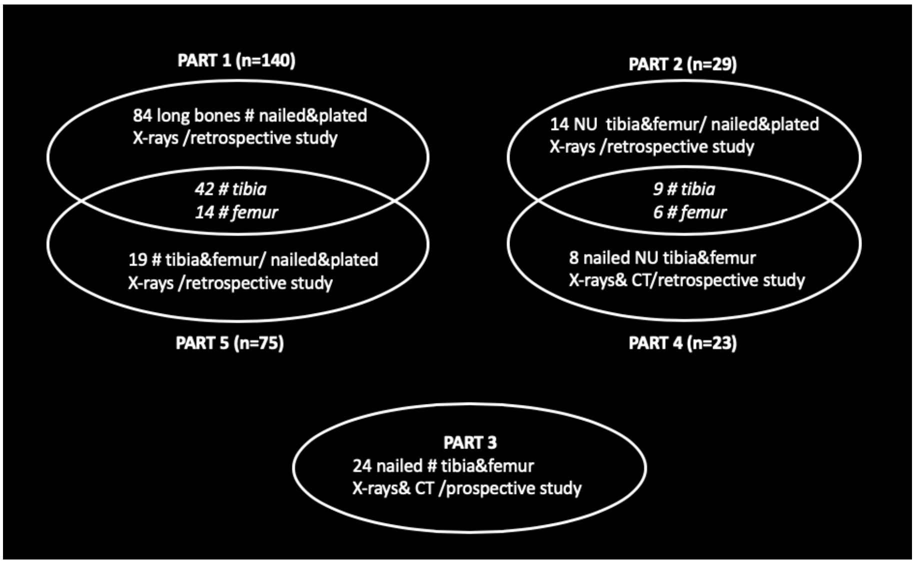
3) Can we establish a standardized imaging strategy for the assessment of fracture healing?

Even if one chooses a specific fracture, there is no consensus within the orthopedic community regarding the appropriate schedule for follow-up imaging. One multi-institutional survey of 40 hand surgeons demonstrated a high degree of variability in the number of radiographic series (ranging from 1 to 6 series per patient) obtained during the course of postoperative follow-up of distal radial fractures. The study demonstrated a mean of 2.6 radiographic series per patient (± 1.0) with imaging obtained at 74% of visits [102].

The convention at many institutions for surveillance imaging of most internally fixed fractures includes radiographs at the following postoperative intervals: baseline post-op, 2 weeks, 6 weeks, 3 months, 6 months, and 12 months. Additional radiographs or CT scan are obtained if there is clinical concern for complication or delayed union.

Our study demonstrated that RUS values increase over time after injury in fractures with normal healing but not with delayed union and that RUS could predict delayed union at about 3 months after injury in nailed fractures. If the predictive value of RUS is validated in larger prospective studies with correlation to clinical data, the 3-month radiograph could become a critical timepoint for the orthopedic surgeon in the evaluation of fracture healing. In case of a low RUS value (threshold to determine in larger studies), a CT scan at 3 months can be useful.

12. APPENDIX A: Summary of study population
Venn diagrams showing common number of patients between Part 1&5 and between Part 2&4.



13. APPENDIX B

Radiological semi-quantitative evaluation of fracture healing: review of the literature

The correlation between radiological appearances and mechanical strength of a fracture was addressed early in experimental osteotomies by Panjabi et al in 1985 [55], who reported that the best single predictor of mechanical strength was cortical continuity. Hammer et al in 1985 [46] proposed a **5-point scale of fracture healing** and aimed at comparing the radiographic assessment with mechanical strength. Seven radiologists evaluated 127 tibial fractures treated nonoperatively. The results of this study showed that there was poor correlation between mechanical strength of the fracture and radiological appearances across it. Conventional RX examinations as a means of assessing the stage of union are generally inconclusive.

Hammer et al scoring system of fracture healing	
Grade 1	homogeneous bone structure, fracture line obliterated, stage of union achieved
Grade 2	massive callus formation with bone trabeculae crossing fracture line (barely discernible). Stage of union achieved
Grade 3	apparent bridging callus, fracture line discernible, uncertain stage of union
Grade 4	trace of callus, not bridging, fracture line distinct, union not achieved
Grade 5	no callus, distinct fracture line, union not achieved

Other authors [47; 103] have also produced similar scoring systems to evaluate fracture healing, based on a combination of the presence of callus, bridging of callus, the presence of lucent fracture lines and remodeling ,without assessing intra and inter-observer reliability. Eastaugh-Waring in 2009 [48] assessed the validity of using serial measurements of callus index as a measure of fracture healing in 45 patients with tibial fractures treated by external fixation, intramedullary nailing, and casting. **Callus index** is defined as the ratio of the maximum callus diameter to bone diameter at the same level as the callus. The results of this study suggested that it is possible to identify an endpoint of healing by measuring the callus index and that the time of occurrence of this end point will be different for different fracture treatments.

Grigoryan et al [45] performed a qualitative and quantitative assessment of fracture healing in a longitudinal prospective study using x-ray and CT imaging. Qualitative assessment included fracture line, fracture margins, fracture gap, first appearance of callus , bridging of involved cortices and fracture union based on bridging of at least ¾ cortices.

Quantitative assessment of CT density changes (HU) in the fracture gap was performed in fractures without metal hardware. The study demonstrates that the evaluation of fracture line and margins on x-rays was limited because of fixation hardware superimposition. Overall, CT images allowed for more complete and detailed visualization of healing compared to CR. Quantitative CT evaluation showed good reproducibility and a statistically significant correlation to qualitative changes.

Whelan et al in 2002 [58] examined the inter- and intraobserver agreement of the healing of tibial fractures among four orthopaedic trauma surgeons who, on two separate occasions eight weeks apart, independently assessed the radiographs of 30 patients with fractures of the tibial shaft which had been treated by intramedullary nail fixation. The surgeons scored the radiographs using the scale by Hammer and then provided an overall rating of fracture healing (healed, probably healed, indeterminate, probably not healed, not healed) . Agreement was moderate in assessing fracture healing whether readers used the criteria of Hammer or their general impression. Substantial agreement occurred when surgeons recorded the number of cortices bridged by callus. The authors suggest that measurement of the number of cortices (0 to 4) bridged by callus is the most reliable method for assessing the progression of the healing of fractures after intramedullary nailing.

The Radiographic Union Scale for Tibia fractures (RUST) was developed by Whelan et al in 2010 [49] to aid in semi-quantitatively evaluating fracture healing using radiographs. The study aimed at evaluating its inter and intraobserver reliability among seven orthopedic surgeons in a series of 45 nailed tibial shaft fractures at various stages of healing.

The **RUST** system assigns a score to a given set of anteroposterior and lateral radiographs based on the assessment of healing at each of the four cortices visible on these projections (i.e., medial and lateral cortices on the anteroposterior X-ray, anterior and posterior cortices on the lateral X-ray) . The individual cortical scores are added to give a total for the set of films with 4 being the minimum score indicating that the fracture is definitely not healed and 12 being the maximum score indicating that the fracture is definitely healed .

RUST according to Whelan et al	
Score 1	fracture line visible, no callus
Score 2	fracture line visible , callus visible
Score 3	no fracture line, bridging callus

Overall interobserver agreement among all reviewers was substantial, with ICC=0.86 (95% CI, 0.79 – 0.91). The overall intraobserver ICC for the group of reviewers was again substantial (ICC= 0.88; 95% CI, 0.80 – 0.96).

Several other studies have demonstrated substantial inter and intra-observer reliability when using RUST in the evaluation of fracture healing [59; 96; 104; 105]. Leow et al [96] showed that agreement (ICC) slightly increased from 0.75 to 0.79 when the radiographs which are scored can be compared with the baseline post-operative radiographs. RUST has also been evaluated in a series of 41 nailed tibial fractures and showed significant correlation with widely used physical and pain scoring systems (SF-36, VAS and Karlström-Olerud) [76].

Chiavaras et al in 2013 [60] developed the radiographic union score for hip (RUSH), similar to RUST. Three radiologists and three orthopedic surgeons independently assessed a total of 100 radiographs (blinded to the time interval after surgery) of intertrochanteric fractures at various stages of healing treated with either sliding hip screw or intramedullary nail (one set of x-rays per fracture). A random sample of 50 cases from the first assessment was re-administered to the same reviewers at 4 weeks to evaluate intra-rater agreement. Reviewers first assessed whether the fracture was healed (scored as "Healed" or "Not Healed") based upon their general assessment of the radiograph and then applied **RUSH.** Inter-rater agreement for the overall subjective impression of fracture healing between reviewer groups was only fair (intraclass coefficient [ICC] = 0.34, 95 % CI: 0.11–0.52. Use of the RUSH score improved overall agreement between groups to substantial (ICC=0.66, 95 % CI: 0.53–0.75). Intra-observer agreement after 4 weeks was almost perfect across both radiologists and surgeons for the RUSH score (ICC = 0.86 vs. 0.84, respectively).

In a following similar study in 2013 by Bhandari et al [61], the same panel of six reviewers (three orthopedic surgeons and three radiologists) independently assessed fracture healing using sequential radiographs from 100 patients with femoral neck fractures and 100 patients with intertrochanteric fractures. Each patient had a baseline post-operative radiograph and three to five radiographs at various time points within 18 months of their hip fracture (longitudinal study). Reviewers first assessed whether the fracture was healed (yes or no) based upon their overall assessment of the radiograph and then performed **RUSH** evaluation. The agreement for the overall RUSH score was near perfect between radiologists and orthopedic surgeons with little difference between the femoral neck (ICC = 0.85, 95% CI: 0.82-0.87) and the intertrochanteric fracture (ICC = 0.88, 95% CI: 0.86-0.90) . Compared to the results of the previous study, the introduction of serial radiographs in which the time from surgery was known significantly improved agreement between the reviewers for both the overall impression of fracture healing and the RUSH score. The authors also observed an asymptotic increase in the RUSH score toward the maximum score of 30 with increasing time from the post- operative baseline visit.

Frank et al in 2016 [97] sought to investigate the utility of the **RUSH s**core to define femoral neck fracture nonunion at 6 months post-injury. Two reviewers independently assigned a RUSH score to the 6-month postinjury radiographs of 250 femoral neck fractures surgically treated. Fracture healing was determined by two independent methods: (1) concurrently by the treating surgeon using both clinical and radiographic assessments as per routine clinical care; and (2) retrospectively by a Central Adjudication Committee using complete obliteration of the fracture line on radiographs alone. A threshold score of < 18 was associated with a 100% specificity (95% CI, 97%-100%) and a positive predictive value of 100% (95% CI, 73%-100%) for radiographic nonunion.

Patel et al in 2014 [57] evaluated the intra- and interobserver reliability of a scoring system for distal radius fracture union **(RUSS) .** Two hand surgeons and 3 musculoskeletal radiologists analysed two sets of 35 AP and lateral radiographs (one set for patients treated non surgically and one set for those treated with internal plate fixation). Reviewers were asked to provide an overall assessment of each case as united or not united and they also applied RUSS. RUSS derives from RUST: Each cortex received a score of 0 points if there was a fracture line with no callus formation, 1 point if there was callus present but a fracture line was still visible, and 2 points if there was bridging callus with no evidence of a fracture line. The results of this study were threefold:

-For nonsurgically treated fractures, substantial agreement in union scores was found with regard to both intra- and interobserver reliability. For fractures treated with ORIF, substantial agreement was found in union scores with regard to intraobserver reliability and moderate agreement with regard to interobserver reliability.

-Mean RUSS scores increased with progressive time after fracture (p<0,001) for both cohorts.

-When using the reviewers' overall assessment of union as a reference standard, RUSS was found to have a statistically significant predictive value in being able to differentiate between united and not united fractures, with RUSS scores greater than or equal to 6 being predictive of fracture union.

Litrenta et al [50] in 2015 aimed at determining the reliability of the Radiographic Union Scale for Tibia (RUST) and a new **modified RUST** score in quantifying healing and at defining a value for radiographic union in a large series of metadiaphyseal fractures treated with plates or intramedullary nails. This study proposed a modified RUST in which a fracture with callus is further subdivided as simply present or bridging:

Modified RUST according to Litrenta et al	
Score 1	no callus
Score 2	callus present
Score 3	bridging callus
Score 4	remodeled callus, fracture not visible

Initially, 12 orthopedic trauma surgeons evaluated 27 sets of AP and lateral radiographs of distal femurs treated with either nail or plate fixation in various stages of healing (2-210 weeks). The reviewers scored each cortex with RUST and modified RUST and then made a determination of whether the fracture was united or not at that time point. In the second part of the study, data from two prospective studies were analyzed including 46 proximal tibial fractures (22 plates and 24 nails) and 81 distal femoral fractures (37 plates and 44 nails). Radiographs were taken at 3, 6 and 12 months of F/U. Only the investigator at the individual site scored the radiographs and provided an assessment of union.

The study demontrated a substantial and slightly higher agreement for the modified RUST than the standard RUST in metadiaphyseal fractures. Score value for definite union would be 10 on the RUST and 13 on the modified RUST with over 90% of reviewers assigning union.

Two experimental studies evaluated the use of RUST and mRUST scores to assess bone healing compared with the use of micro-CT scans and biomechanical tests:

-Cooke et al [106] showed that the mRUST and RUST score correlated strongly with several biomechanical properties, such as callus bone mineral density (BMD) , bone volume fraction ,callus strength stiffness and rigidity. The reviewers considered all fractures with an average RUST of at least 10 or mRUST of at least 13, to be healed.

-Fiset et al [87] showed strong statistical correlations of RUST and mRUST with micro-CT and biomechanical parameters. Intraclass correlation coefficients (ICCs) were 0.89 (95% confidence interval [CI]: 0.78 to 0.94) for RUST and 0.86 (95% CI: 0.74 to 0.93) for mRUST, higher than those in almost all previous studies. A RUST score of 10 or an mRUST score of 15 may be considered the threshold above which a plated bone is “healed”.

At last, Oliver et al in 2019 [62] presented the Radiographic Union Score for Humeral shaft fractures **(RUSHU),** based on RUST (Score 1: no callus ; Score 2: Callus present, non-bridging ; Score 3: Bridging callus). A group of 60 patients (non operative management of #) with radiographs six weeks after injury (40 with fractures that united and 20 with fractures that developed nonunion at 6 months), were scored by two blinded observers using RUSHU. RUSHU was proven a reliable and effective tool in identifying patients at risk of developing nonunion at six weeks after injury, with a RUSHU of < 8 significantly predictive of nonunion.

The following table presents a summary of the studies on recent Radiographic Scoring Systems:

Authors	**Year of study**	**Number& type of #**	**Type of treatment**	**Readers**	**Scoring system**	**Aims of the study**	**Results**
Whelan et al	2010	45 tibial shaft #	IM nail	7 surgeons	RUST	Agreement	Almost perfect inter & intra-observer agreement
Chiavaras et al	2013	100 intertrochanteric #	IM nail or hip screw	3 radiologists & 3 surgeons	RUSH	Agreement	Substantial inter & almost perfect intra-observer agreement
Bhandari et al	2013	100 femoral neck & 100 intertrochanteric # (serial radiographs for each patient)	IM nail or hip screw	3 radiologists & 3 surgeons (same panel as in previous study)	RUSH	Agreement & score values	-Higher agreement than previous study - RUSH values increase with time after injury
Patel et al	2014	70 distal radial #	35 casts & 35 plates	2 surgeons	RUSS	Agreement & threshold score values that indicate union	-moderate to substantial agreement - RUSS values increase with time after injury -RUSS ≥6 predictive of # union
Litrenta et al	2015	a)27 femoral # b) 46 tibial & 81 femoral # metaphyso-diaphyseal #	IM nail or plate	a)12 surgeons b) 1 surgeon	Modified RUST (mRUST)	a) Agreement b) Threshold score values that indicate union	a)higher agreement of mRUST (substantial for nails &moderate for plates) b) RUST ≥10 & mRUST ≥13 predictive of # union

Frank et al	2016	250 femoral neck #	IM nail or hip screw	2 surgeons	RUSH	Threshold score values indicating non-union at 6 months post-injury	threshold score of < 18 indicates NU
Oliver et al	2019	60 humeral shaft #	cast	2 surgeons	RUSHU	Agreement & predictive value for non-union at 6 weeks	-Substantial inter & almost perfect intra-observer agreement - RUSHU < 8 predictive of nonunion at 6 weeks post injury

14. APPENDIX C

The **Function IndeX for Trauma (FIX-IT)** assessment by Bhandari et al [77] provides a simple standardized approach to the measurement of weight-bearing and pain assessment in patients with lower extremity fractures, specifically tibia and femoral fractures. The FIX-IT score (table below) is a clinical outcomes measure ranging from 0 to 12 points in 2 domains: the ability to bear weight (maximum 6 points) and pain at the fracture site (maximum 6 points) In their preliminary study in 2013 ,the FIX-IT score had high interrater agreement across multiple examiners and it correlated with the physical scores of the SF-36 questionnaire.

WEIGHTBEARING				
Score	Single leg stand score	Actual score	Ambulation score	Actual score
0	Unable to bear full body weight on affected leg (includes subjects requiring assistive device or support to stand)		Unable to bear any body weight on affected leg as evidenced by requirement for wheelchair or 2 crutches to ambulate	
1	Able to bear full body weight on affected leg as evidenced by ability to stand on affected leg momentarily (< 10 seconds)		Able to bear modest body weight on affected leg as evidenced by ability to ambulate with one crutch, cane, or walker	
2	Able to bear full body weight on affected leg as evidenced by ability to stand on affected leg momentarily (10 to 30 seconds)		Able to bear modest body weight on affected leg as evidenced by ability to ambulate without assistive device, but with clear difficulty (eg, abnormal gait, limp)	
3	Able to bear full body weight on affected leg as evidenced by ability to stand on affected leg > 30		Able to bear full body weight on affected leg as evidenced by ability to ambulate without	

	seconds		assistive device and restoration of prefracture gait	
Total single leg stand score			Total ambulation score	
Total score for ability to bear weight on fractured limb				/6
PAIN				
Score	Palpation score	Actual score	Stress score	Actual score
0	Pain without palpation		Pain without stress	
1	Pain with light palpation		Pain with mild stress	
2	Pain with deep palpation		Pain with strong stress	
3	No pain with deep palpation		No pain with strong stress	
Total palpation score			Total stress score	
Total score for pain at fracture site				/6
TOTAL SCORE				/12

15. REFERENCES

1 Schubert T (2013) Adipose Mesenchymal Stem Cells for bone tissue engineeringSSS/IREC. Université Catholique de Louvain

2 Pathria MN, Chung CB, Resnick DL (2016) Acute and Stress-related Injuries of Bone and Cartilage: Pertinent Anatomy, Basic Biomechanics, and Imaging Perspective. Radiology 280:21-38

3 Axelrad TW, Einhorn TA (2011) Use of clinical assessment tools in the evaluation of fracture healing. Injury 42:301-305

4 Marsell R, Einhorn TA (2011) The biology of fracture healing. Injury 42:551-555

5 Loi F, Cordova LA, Pajarinen J, Lin TH, Yao Z, Goodman SB (2016) Inflammation, fracture and bone repair. Bone 86:119-130

6 Giannoudis PV, Einhorn TA, Marsh D (2007) Fracture healing: the diamond concept. Injury 38 Suppl 4:S3-6

7 Frolke JP, Patka P (2007) Definition and classification of fracture non-unions. Injury 38 Suppl 2:S19-22

8 Nandra R, Grover L, Porter K (2015) Fracture non-union epidemiology and treatment. Trauma 18:3-11

9 Antonova E, Le TK, Burge R, Mershon J (2013) Tibia shaft fractures: costly burden of nonunions. BMC Musculoskelet Disord 14:42

10 Rupp M, Biehl C, Budak M, Thormann U, Heiss C, Alt V (2018) Diaphyseal long bone nonunions - types, aetiology, economics, and treatment recommendations. Int Orthop 42:247-258

11 Tzioupis C, Giannoudis PV (2007) Prevalence of long-bone non-unions. Injury 38 Suppl 2:S3-9

12 Weber BG, Cech O (1973) Pseudarthrosen : Pathophysiologie, Biomechanik, Therapie, Ergebnisse. Huber, Bern

13 Claes L, Augat P, Suger G, Wilke HJ (1997) Influence of size and stability of the osteotomy gap on the success of fracture healing. J Orthop Res 15:577-584

14 Giannoudis PV, MacDonald DA, Matthews SJ, Smith RM, Furlong AJ, De Boer P (2000) Nonunion of the femoral diaphysis. The influence of reaming and non-steroidal anti-inflammatory drugs. J Bone Joint Surg Br 82:655-658

15 Santolini E, West R, Giannoudis PV (2015) Risk factors for long bone fracture non-union: a stratification approach based on the level of the existing scientific evidence. Injury 46 Suppl 8:S8-S19

16 Schemitsch EH, Bhandari M, Guyatt G et al (2012) Prognostic factors for predicting outcomes after intramedullary nailing of the tibia. J Bone Joint Surg Am 94:1786-1793

17 Calori GM, Albisetti W, Agus A, Iori S, Tagliabue L (2007) Risk factors contributing to fracture non-unions. Injury 38 Suppl 2:S11-18

18 Bhandari M, Guyatt GH, Swiontkowski MF, Tornetta P, 3rd, Sprague S, Schemitsch EH (2002) A lack of consensus in the assessment of fracture healing among orthopaedic surgeons. J Orthop Trauma 16:562-566

19 Morshed S (2014) Current Options for Determining Fracture Union. Adv Med 2014:708574

20 Corrales LA, Morshed S, Bhandari M, Miclau T, 3rd (2008) Variability in the assessment of fracture-healing in orthopaedic trauma studies. J Bone Joint Surg Am 90:1862-1868
21 Bhandari M, Fong K, Sprague S, Williams D, Petrisor B (2012) Variability in the definition and perceived causes of delayed unions and nonunions: a cross-sectional, multinational survey of orthopaedic surgeons. J Bone Joint Surg Am 94:e1091-1096
22 Fisher JS, Kazam JJ, Fufa D, Bartolotta RJ (2019) Radiologic evaluation of fracture healing. Skeletal Radiol 48:349-361
23 Webb J, Herling G, Gardner T, Kenwright J, Simpson AH (1996) Manual assessment of fracture stiffness. Injury 27:319-320
24 Beaton DE, Schemitsch E (2003) Measures of health-related quality of life and physical function. Clin Orthop Relat Res. 10.1097/01.blo.0000079772.06654.c8:90-105
25 Shearer D, Morshed S (2011) Common generic measures of health related quality of life in injured patients. Injury 42:241-247
26 Cox G, Einhorn TA, Tzioupis C, Giannoudis PV (2010) Bone-turnover markers in fracture healing. J Bone Joint Surg Br 92:329-334
27 Coulibaly MO, Sietsema DL, Burgers TA, Mason J, Williams BO, Jones CB (2010) Recent advances in the use of serological bone formation markers to monitor callus development and fracture healing. Crit Rev Eukaryot Gene Expr 20:105-127
28 Moghaddam A, Muller U, Roth HJ, Wentzensen A, Grutzner PA, Zimmermann G (2011) TRACP 5b and CTX as osteological markers of delayed fracture healing. Injury 42:758-764
29 Hankenson KD, Zimmerman G, Marcucio R (2014) Biological perspectives of delayed fracture healing. Injury 45 Suppl 2:S8-S15
30 Hak DJ, Fitzpatrick D, Bishop JA et al (2014) Delayed union and nonunions: epidemiology, clinical issues, and financial aspects. Injury 45 Suppl 2:S3-7
31 Blokhuis TJ, de Bruine JH, Bramer JA et al (2001) The reliability of plain radiography in experimental fracture healing. Skeletal Radiol 30:151-156
32 Panjabi MM, Lindsey RW, Walter SD, White AA, 3rd (1989) The clinician's ability to evaluate the strength of healing fractures from plain radiographs. J Orthop Trauma 3:29-32
33 Davis BJ, Roberts PJ, Moorcroft CI, Brown MF, Thomas PB, Wade RH (2004) Reliability of radiographs in defining union of internally fixed fractures. Injury 35:557-561
34 Bhattacharyya T, Bouchard KA, Phadke A, Meigs JB, Kassarjian A, Salamipour H (2006) The accuracy of computed tomography for the diagnosis of tibial nonunion. J Bone Joint Surg Am 88:692-697
35 Firoozabadi R, Morshed S, Engelke K et al (2008) Qualitative and quantitative assessment of bone fragility and fracture healing using conventional radiography and advanced imaging technologies--focus on wrist fracture. J Orthop Trauma 22:S83-90
36 Vande Berg B, Malghem J, Maldague B, Lecouvet F (2006) Multi-detector CT imaging in the postoperative orthopedic patient with metal hardware. Eur J Radiol 60:470-479

37 Nicholson JA, Tsang STJ, MacGillivray TJ, Perks F, Simpson A (2019) What is the role of ultrasound in fracture management?: Diagnosis and therapeutic potential for fractures, delayed unions, and fracture-related infection. Bone Joint Res 8:304-312
38 Moed BR, Subramanian S, van Holsbeeck M et al (1998) Ultrasound for the early diagnosis of tibial fracture healing after static interlocked nailing without reaming: clinical results. J Orthop Trauma 12:206-213
39 Protopappas VC, Vavva MG, Fotiadis DI, Malizos KN (2008) Ultrasonic monitoring of bone fracture healing. IEEE Trans Ultrason Ferroelectr Freq Control 55:1243-1255
40 Sun MH, Leung KS, Zheng YP et al (2012) Three-dimensional high frequency power Doppler ultrasonography for the assessment of microvasculature during fracture healing in a rat model. J Orthop Res 30:137-143
41 Ha AS, Lee AY, Hippe DS, Chou SH, Chew FS (2015) Digital Tomosynthesis to Evaluate Fracture Healing: Prospective Comparison With Radiography and CT. AJR Am J Roentgenol 205:136-141
42 Niikura T, Lee SY, Sakai Y, Nishida K, Kuroda R, Kurosaka M (2014) Comparison of radiographic appearance and bone scintigraphy in fracture nonunions. Orthopedics 37:e44-50
43 Sollini M, Trenti N, Malagoli E et al (2019) [(18)F]FDG PET/CT in non-union: improving the diagnostic performances by using both PET and CT criteria. Eur J Nucl Med Mol Imaging 46:1605-1615
44 Moliere S, Dosch JC, Bierry G (2016) Pelvic, acetabular and hip fractures: What the surgeon should expect from the radiologist. Diagn Interv Imaging 97:709-723
45 Grigoryan M, Lynch JA, Fierlinger AL et al (2003) Quantitative and qualitative assessment of closed fracture healing using computed tomography and conventional radiography. Acad Radiol 10:1267-1273
46 Hammer RR, Hammerby S, Lindholm B (1985) Accuracy of radiologic assessment of tibial shaft fracture union in humans. Clin Orthop Relat Res:233-238
47 Tower SS, Beals RK, Duwelius PJ (1993) Resonant frequency analysis of the tibia as a measure of fracture healing. J Orthop Trauma 7:552-557
48 Eastaugh-Waring SJ, Joslin CC, Hardy JR, Cunningham JL (2009) Quantification of fracture healing from radiographs using the maximum callus index. Clin Orthop Relat Res 467:1986-1991
49 Whelan DB, Bhandari M, Stephen D et al (2010) Development of the radiographic union score for tibial fractures for the assessment of tibial fracture healing after intramedullary fixation. J Trauma 68:629-632
50 Litrenta J, Tornetta P, 3rd, Mehta S et al (2015) Determination of Radiographic Healing: An Assessment of Consistency Using RUST and Modified RUST in Metadiaphyseal Fractures. J Orthop Trauma 29:516-520
51 Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159-174
52 Farzin B, Gentric JC, Pham M et al (2017) Agreement studies in radiology research. Diagn Interv Imaging 98:227-233

53 Dijkman BG, Sprague S, Schemitsch EH, Bhandari M (2010) When is a fracture healed? Radiographic and clinical criteria revisited. J Orthop Trauma 24 Suppl 1:S76-80
54 Dijkman BG, Busse JW, Walter SD, Bhandari M, Investigators T (2011) The impact of clinical data on the evaluation of tibial fracture healing. Trials 12:237
55 Panjabi MM, Walter SD, Karuda M, White AA, Lawson JP (1985) Correlations of radiographic analysis of healing fractures with strength: a statistical analysis of experimental osteotomies. J Orthop Res 3:212-218
56 McClelland D, Thomas PB, Bancroft G, Moorcraft CI (2007) Fracture healing assessment comparing stiffness measurements using radiographs. Clin Orthop Relat Res 457:214-219
57 Patel SP, Anthony SG, Zurakowski D et al (2014) Radiographic scoring system to evaluate union of distal radius fractures. J Hand Surg Am 39:1471-1479
58 Whelan DB, Bhandari M, McKee MD et al (2002) Interobserver and intraobserver variation in the assessment of the healing of tibial fractures after intramedullary fixation. J Bone Joint Surg Br 84:15-18
59 Kooistra BW, Dijkman BG, Busse JW, Sprague S, Schemitsch EH, Bhandari M (2010) The radiographic union scale in tibial fractures: reliability and validity. J Orthop Trauma 24 Suppl 1:S81-86
60 Chiavaras MM, Bains S, Choudur H et al (2013) 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 42:1079-1088
61 Bhandari M, Chiavaras MM, Parasu N et al (2013) Radiographic union score for hip substantially improves agreement between surgeons and radiologists. BMC Musculoskelet Disord 14:70
62 Oliver WM, Smith TJ, Nicholson JA et al (2019) The Radiographic Union Score for HUmeral fractures (RUSHU) predicts humeral shaft nonunion. Bone Joint J 101-B:1300-1306
63 McKibbin B (1978) The biology of fracture healing in long bones. J Bone Joint Surg Br 60-B:150-162
64 Aro HT, Chao EY (1993) Bone-healing patterns affected by loading, fracture fragment stability, fracture type, and fracture site compression. Clin Orthop Relat Res:8-17
65 Greenbaum MA, Kanat IO (1993) Current concepts in bone healing. Review of the literature. J Am Podiatr Med Assoc 83:123-129
66 Harwood P, Newman J, Michael A (2010) An update on fracture healing and non-union. Orthopaedics and Trauma 24:9-23
67 Perlepe V, Omoumi P, Larbi A et al (2018) Can we assess healing of surgically treated long bone fractures on radiograph? Diagn Interv Imaging 99:381-386
68 Kuhlman JE, Fishman EK, Magid D, Scott WW, Jr., Brooker AF, Siegelman SS (1988) Fracture nonunion: CT assessment with multiplanar reconstruction. Radiology 167:483-488
69 Krestan CR, Noske H, Vasilevska V et al (2006) MDCT versus digital radiography in the evaluation of bone healing in orthopedic patients. AJR Am J Roentgenol 186:1754-1760

70 Kleinlugtenbelt YV, Scholtes VA, Toor J et al (2016) Does Computed Tomography Change our Observation and Management of Fracture Non-Unions? Arch Bone Jt Surg 4:337-342
71 Wellenberg RHH, Hakvoort ET, Slump CH, Boomsma MF, Maas M, Streekstra GJ (2018) Metal artifact reduction techniques in musculoskeletal CT-imaging. Eur J Radiol 107:60-69
72 Ottenin MA, Jacquot A, Grospretre O et al (2012) Evaluation of the diagnostic performance of tomosynthesis in fractures of the wrist. AJR Am J Roentgenol 198:180-186
73 De Silvestro A, Martini K, Becker AS et al (2018) Postoperative imaging of orthopaedic hardware in the hand and wrist: is there an added value for tomosynthesis? Clin Radiol 73:214.e211-214.e219
74 Blum A, Noel A, Regent D, Villani N, Gillet R, Gondim Teixeira P (2018) Tomosynthesis in musculoskeletal pathology. Diagn Interv Imaging 99:423-441
75 Bhandari M, Chiavaras M, Ayeni O et al (2013) Assessment of radiographic fracture healing in patients with operatively treated femoral neck fractures. J Orthop Trauma 27:e213-219
76 Cekic E, Alici E, Yesil M (2014) Reliability of the radiographic union score for tibial fractures. Acta Orthop Traumatol Turc 48:533-540
77 Bhandari M, Wasserman SM, Yurgin N, Petrisor B, Sprague S, Dent RE (2013) Development and preliminary validation of a Function IndeX for Trauma (FIX-IT). Can J Surg 56:E114-120
78 Perlepe V, Michoux N, Heynen G, Vande Berg B (2019) Semi-quantitative CT assessment of fracture healing: How many and which CT reformats should be analyzed? Eur J Radiol 118:181-186
79 McGraw K, Wong SP (1996) Forming Inferences About Some Intraclass Correlation Coefficients. Psychological Methods 1:30-46
80 Howell DC (2009) Stastical Methods for psychology, Thomson Wadsworth edn, Belmont
81 Conover WJ (1999) Practical nonparametric statistics, Wiley edn, New York
82 I.C.C.F.C. (2018) Fracture and dislocation-Compendium 2018. journal of Orthopaedic Trauma 32
83 Markel MD, Morin RL, Wikenheiser MA, Lewallen DG, Chao EY (1991) Quantitative CT for the evaluation of bone healing. Calcif Tissue Int 49:427-432
84 den Boer FC, Bramer JA, Patka P et al (1998) Quantification of fracture healing with three-dimensional computed tomography. Arch Orthop Trauma Surg 117:345-350
85 Riegger C, Kropil P, Jungbluth P et al (2012) Quantitative assessment of bone defect healing by multidetector CT in a pig model. Skeletal Radiol 41:531-537
86 Kropil P, Hakimi AR, Jungbluth P et al (2012) Cone beam CT in assessment of tibial bone defect healing: an animal study. Acad Radiol 19:320-325
87 Fiset S, Godbout C, Crookshank MC, Zdero R, Nauth A, Schemitsch EH (2018) Experimental Validation of the Radiographic Union Score for Tibial Fractures (RUST) Using Micro-Computed Tomography Scanning and Biomechanical Testing in an in-Vivo Rat Model. J Bone Joint Surg Am 100:1871-1878

88 Watanabe Y, Nishizawa Y, Takenaka N, Kobayashi M, Matsushita T (2009) Ability and limitation of radiographic assessment of fracture healing in rats. Clin Orthop Relat Res 467:1981-1985
89 Squyer ER, Dikos GD, Kaehr DM, Maar DC, Crichlow RJ (2016) Early prediction of tibial and femoral fracture healing: Are we reliable? Injury 47:2805-2808
90 Yang JS, Otero J, McAndrew CM, Ricci WM, Gardner MJ (2013) Can tibial nonunion be predicted at 3 months after intramedullary nailing? J Orthop Trauma 27:599-603
91 Busse JW, Bhandari M, Guyatt GH et al (2012) Development and validation of an instrument to predict functional recovery in tibial fracture patients: the Somatic Pre-Occupation and Coping (SPOC) questionnaire. J Orthop Trauma 26:370-378
92 Bhandari M, Tornetta P, 3rd, Sprague S et al (2003) Predictors of reoperation following operative management of fractures of the tibial shaft. J Orthop Trauma 17:353-361
93 Fong K, Truong V, Foote CJ et al (2013) Predictors of nonunion and reoperation in patients with fractures of the tibia: an observational study. BMC Musculoskelet Disord 14:103
94 Lack WD, Starman JS, Seymour R et al (2014) Any Cortical Bridging Predicts Healing of Tibial Shaft Fractures. J Bone Joint Surg Am 96:1066-1072
95 Salih S, Blakey C, Chan D et al (2015) The callus fracture sign: a radiological predictor of progression to hypertrophic non-union in diaphyseal tibial fractures. Strategies Trauma Limb Reconstr 10:149-153
96 Leow JM, Clement ND, Tawonsawatruk T, Simpson CJ, Simpson AH (2016) The radiographic union scale in tibial (RUST) fractures: Reliability of the outcome measure at an independent centre. Bone Joint Res 5:116-121
97 Frank T, Osterhoff G, Sprague S et al (2016) The Radiographic Union Score for Hip (RUSH) Identifies Radiographic Nonunion of Femoral Neck Fractures. Clin Orthop Relat Res 474:1396-1404
98 Rand JA, An KN, Chao EY, Kelly PJ (1981) A comparison of the effect of open intramedullary nailing and compression-plate fixation on fracture-site blood flow and fracture union. J Bone Joint Surg Am 63:427-442
99 Henderson CE, Lujan T, Bottlang M, Fitzpatrick DC, Madey SM, Marsh JL (2010) Stabilization of distal femur fractures with intramedullary nails and locking plates: differences in callus formation. Iowa Orthop J 30:61-68
100 Einhorn TA (1995) Enhancement of fracture-healing. J Bone Joint Surg Am 77:940-956
101 Ross KA, O'Halloran K, Castillo RC et al (2018) Prediction of tibial nonunion at the 6-week time point. Injury 49:2075-2082
102 Bohl DD, Lese AB, Patterson JT, Grauer JN, Dodds SD (2014) Routine imaging after operatively repaired distal radius and scaphoid fractures: a survey of hand surgeons. Journal of wrist surgery 3:239-244
103 Lane JM, Sandhu HS (1987) Current approaches to experimental bone grafting. Orthop Clin North Am 18:213-225

104 Richards BS, Wilkes D, Dempsey M, Nurenberg P (2015) A radiographic scoring system to assess healing in congenital pseudarthrosis of the tibia. J Pediatr Orthop B 24:118-122
105 Tawonsawatruk T, Hamilton DF, Simpson AH (2014) Validation of the use of radiographic fracture-healing scores in a small animal model. J Orthop Res 32:1117-1119
106 Cooke ME, Hussein AI, Lybrand KE et al (2018) Correlation between RUST assessments of fracture healing to structural and biomechanical properties. J Orthop Res 36:945-953