



ORIGINAL ARTICLE / *Musculoskeletal imaging*

Can we assess healing of surgically treated long bone fractures on radiograph?

V. Perlepe^{a,*}, P. Omoumi^b, A. Larbi^c, D. Putineanu^a,
J.-E. Dubuc^a, T. Schubert^a, B. Vande Berg^a

^a Cliniques universitaires Saint-Luc, avenue Hippocrate 10, 1200 Brussels, Belgium

^b University Hospital of Lausanne, Bugnon 46 CH-1011 Lausanne, Switzerland

^c Centre hospitalier régional universitaire de Nîmes, Nîmes, France

KEYWORDS

Bone;
Bone fracture;
Healing;
Radiographs;
Variability study

Abstract

Purpose: To determine the frequency and causes for limitations in the radiographic evaluation of surgically treated long bone fractures.

Materials and methods: Six readers separately scored 140 sets of antero-posterior (AP) and lateral radiographs of surgically treated long bone fractures, using a radiographic union score (RUS). We determined the rate of assessability of the fracture edges at each of the four cortical segments ($n = 560$) seen tangentially on the two radiographs and the causes for non-assessability. The rate of feasibility of the RUS (more than two fracture edges assessable per fracture) was determined and compared according to different parameters.

Results: Fracture edges were visible in 71% to 81% of the 560 cortical segments. Metal hardware superimposition was the most frequent cause for non-assessability (79–95%). RUS values could be calculated in 58% to 75% of fractures. Scoring was statistically significantly less frequently calculable in plated (31–56%) than in nailed fractures (90–97%), in distal (47–61%) than in proximal (78–89%) bones and in upper (27–49%) than in lower (76–91%) limb bones ($P \leq 0.01$).

Conclusions: The type of stabilization hardware is the main limiting factor in the radiographic assessment of surgically treated long bone fractures. Scoring was feasible in only 31% to 56% of plated fractures.

© 2018 Société française de radiologie. Published by Elsevier Masson SAS. All rights reserved.

Abbreviation: RUS, Radiographic union score.

* Corresponding author. Department of radiology, Cliniques universitaires Saint-Luc–UC Louvain, avenue Hippocrate 10/2942, 1200 Brussels, Belgium.

E-mail address: vasilikiperlepe@gmail.com (V. Perlepe).

<https://doi.org/10.1016/j.diii.2018.02.004>

2211-5684/© 2018 Société française de radiologie. Published by Elsevier Masson SAS. All rights reserved.

Please cite this article in press as: Perlepe V, et al. Can we assess healing of surgically treated long bone fractures on radiograph? *Diagnostic and Interventional Imaging* (2018), <https://doi.org/10.1016/j.diii.2018.02.004>

Introduction

Musculoskeletal trauma represents an important global health burden with approximately six million fractures each year in the United States [1]. As many as 10% of these may be complicated by non-union with a subsequent increase in medical costs and productivity loss [2]. The evaluation of fracture healing is mainly based on clinical and radiographic findings, both presenting some limitations [3,4]. Clinical evaluation remains a subjective process [5–7] and limitations in the radiographic evaluation have also been demonstrated [5,6,8,9], while plain X-ray remains the examination of choice for bone fracture assessment [10]. Despite the added value of scoring systems for the overall evaluation of radiographs [11–17], radiologists and orthopaedic surgeons frequently hesitate when assessing fracture edges on radiographs [4,5]. Although there is general agreement on the fact that radiographs are limited by overlapping metal hardware [5,6,8], we are not aware of any study attempting to determine the parameters that limit the contribution of the radiographs and the consequences in feasibility of radiographic scoring systems [11–17]. The current study aimed at defining the frequency and causes for limitations in the quantitative radiographic evaluation of surgically treated long bone fractures.

Methods and materials

Study population

We selected a series of patients aged from 18 to 65 years 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”, “forearm”, “thigh” and “leg” among examination 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. The study population consisted of 118 patients including 76 men and 42 women (mean age: 41.3 ± 13.3 [SD] 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 9 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 three weeks after fracture, between three weeks and three months after fracture, between three and six months after fracture and more than six months after fracture. 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).

Image analysis

Analysis of the radiographs was performed by six readers including three radiologists (a last-year resident [VP] and two musculoskeletal radiologists with three [AL] and 20 [BVB] years of experience) and three surgeons (a last-year resident [TS], and two orthopedists with five [DP] and 24 [JED] years of experience). All readers blinded to clinical findings separately analyzed each set of radiographs on a PACS workstation. 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. When the fracture edges were visible, cortical scoring was

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 percentages in parentheses. R1: last-year resident in radiology; R2: MSK radiologist with three years of experience; R3: MSK radiologist with 20 years of experience; O1: last-year resident in orthopedic surgery; O2 orthopedic surgeon with five years of experience; O3: orthopedic surgeon with 24 years of experience.

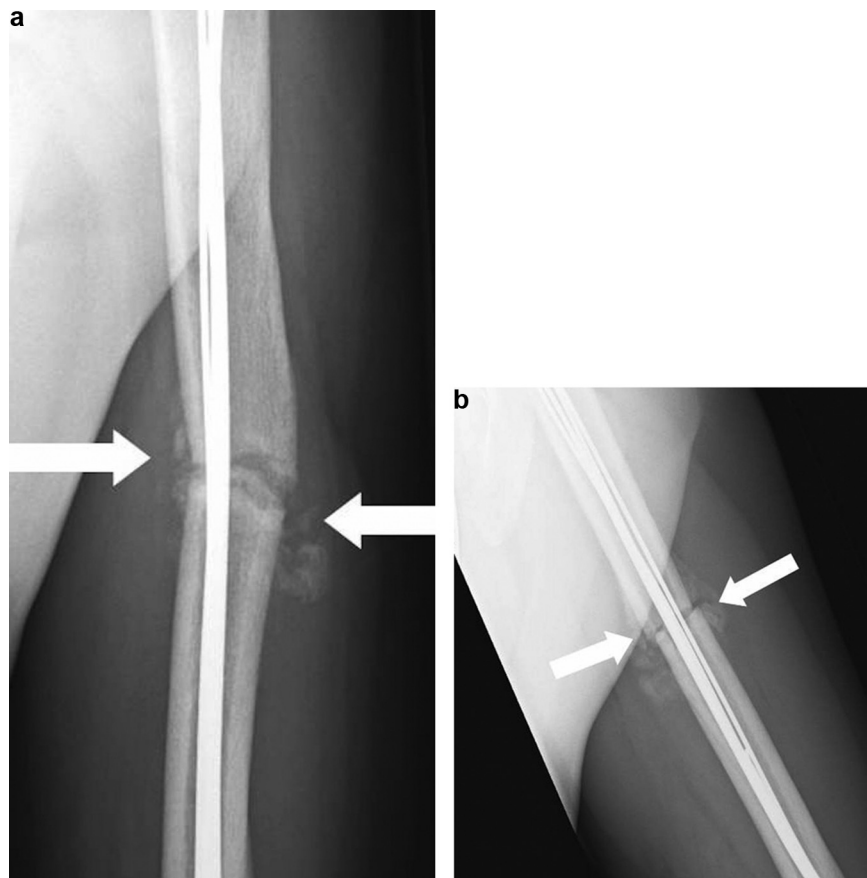


Figure 1. (a) Antero-posterior and (b) lateral radiographs of an arm with a nailed humeral shaft fracture. The four cortical segments of the fracture to which the X-ray beam was tangent (arrows) were assessable and could be scored by all readers. The RUS score could be calculated.

performed as described below. When the fracture edges were not visible, the reader selected at least one of the following potential 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. 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.

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 of Litrenta et al.: score 1: lack of callus; score 2: non-bridging callus; score 3: bridging callus; score 4: remodelled bridging callus [16]. For each fracture, the radiographic union score (RUS) value was calculated by adding the four cortical scores. 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.

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 χ^2 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 [18] 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 [19]. The RUS scores were compared between readers as a surrogate of the quality of the analysis performed by the six

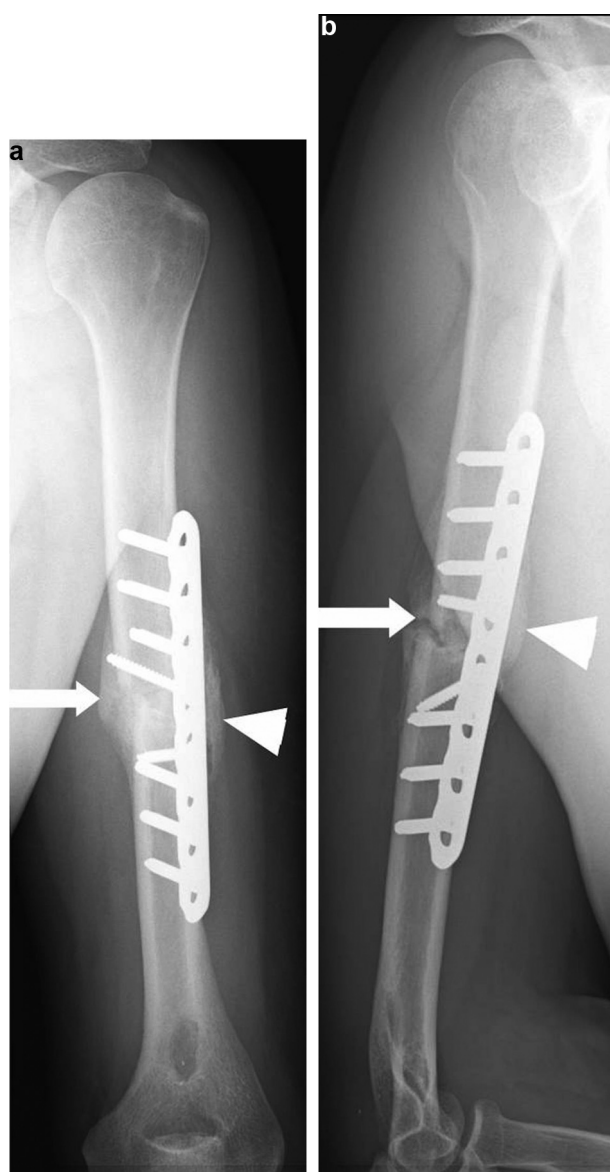


Figure 2. (a) Antero-posterior and (b) lateral radiographs of an arm with a plated humeral shaft fracture. The medial and posterior cortical segments (arrows in a and b) but not the lateral and anterior cortical segments (arrowheads in a and b) of the fracture were assessable. The RUS score was not calculated because of the availability of two out of four cortical scores.

readers. Data were processed using a statistical software package (MedCalc version 11.6; MedCalc Software, Mariakkerke, Belgium). For all tests, a P -value < 0.05 was considered as statistically significant.

Results

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 anteroposterior and lateral radiographs of the 140 fractures ranged between 73%

to 76% among radiologists and between 71% to 81% among surgeons (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) (Figs. 1–3). 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]).

Feasibility of a radiographic score

The radiographic scoring 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, the surgeons and for all readers (ICC: 0.66–0.74). 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 $P \leq 0.01$), 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 $P \geq 52$), patient's age (all $P \geq 0.09$) and gender (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).

Interobserver agreement for cortical scorings 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]).

Discussion

The current study addressing the limitations of radiography for the evaluation of healing of surgically treated long bone fractures demonstrated several findings. First, 19 to 29% of the 560 cortical segments on the AP and lateral radiographs of 140 long bone fractures were not assessable. In 79 to 95% of radiographic sets, overlapping metal hardware impeded the analysis of the fracture edges, while superimposed cortical bone from an adjacent bone prevented the assessability in only 5 to 13% of radiographic sets. This relatively high rate of non-assessability of the fracture edges could partially account for the previously mentioned limitations in the analysis of radiographs in clinical practice [17,20,21].

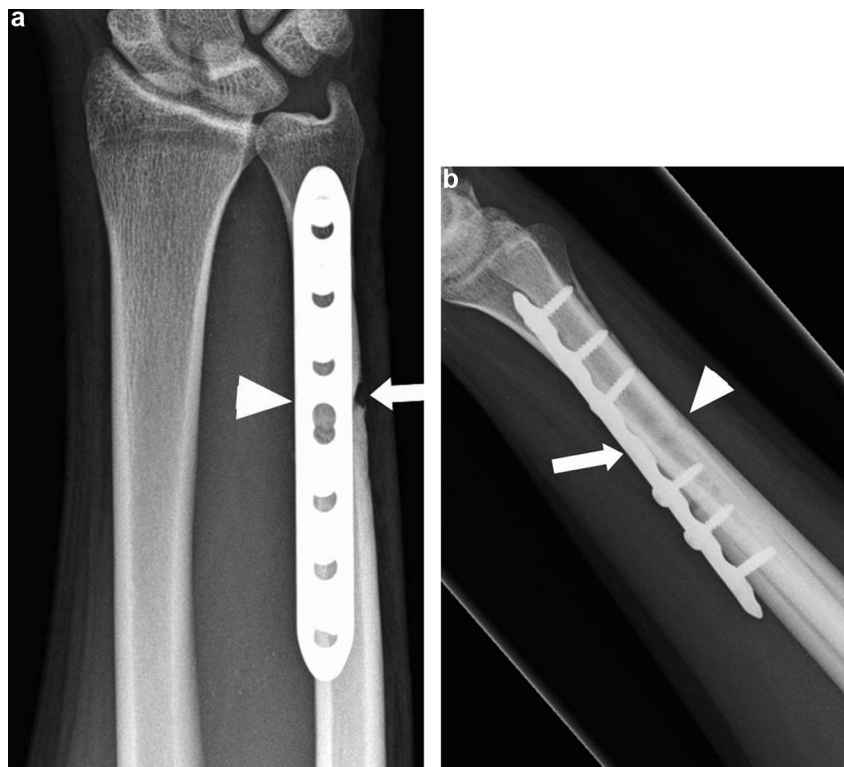


Figure 3. Plated ulnar shaft fracture. (a) On the AP radiograph, the medial aspect of the fracture (arrow) could be assessed. There is overlapping metal hardware on the lateral aspect (arrowhead) of the fracture. (b) On the lateral radiograph, the anterior aspect of the fracture was not assessable due to overlapping bone and metal. There was overlapping cortical bone on the posterior aspect of the fracture.

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 two cortical segments were assessable on the radiographs.

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

In our study, the radiographic score of fracture healing could be calculated in only 58% to 75% of the 140 fractures (at least three out of four cortical scores assessable). This number of three cortical scores is usually considered to be mandatory for data reliability because it is generally accepted that fracture healing is reached when cortical bridging of the fracture can be demonstrated in at least three of the four fracture edges on a radiographic set [4,6]. Finally, the feasibility rate of the RUS was statistically significantly lower in plated than in nailed fractures, in distal than in proximal bones and in upper than in lower limb bones. According to logistic regression analysis, the type of surgical material (nails vs. plate) was the only parameter that had a statistically significant impact on the feasibility rate of the fracture scoring. The demonstrated limitation in the radiographic evaluation of fracture callus in plated fractures is

inherent to the projection nature of radiographs and to the interaction between the X-ray beam and atoms with high atomic numbers found in fixation devices. 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 eccentrically 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.

Our study has some limitations. First, it is a retrospective study with a limited number of patients. Second, AP and lateral radiographs were analyzed and we did not assess the values of additional oblique radiographic projections. Third, we used one scoring system that relies on callus evaluation [16]. The rate of assessability is unlikely to be influenced by the type of scoring systems because all radiographic

scoring systems are based on the scoring of cortical segments to which the X-ray beam is tangent. Therefore, the assessability of the cortical segment is key to all scoring systems, independent from the analyzed parameters. Finally, the value of the scoring system to assess fracture healing was not addressed in the current study. The scoring system was only used as a surrogate to check the quality of the readings. Actually, the agreement between our readers was similar to that observed by other authors [11–17].

In conclusion, the assessability of the fracture edges on the radiographs of long bone shaft fractures is frequently limited in plated but not in nailed 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.

Funding source

This work was funded by IREC (Institut de recherche expérimentale et clinique) of the Université catholique de Louvain.

Disclosure of interest

The authors declare that they have no competing interest.

References

- [1] Einhorn T. Enhancement of fracture healing. *J Bone Joint Surg Am* 1995;77:940–56.
- [2] Grigoryan M, Lynch JA, Fierlinger AL, Guermazi A, Fan B, MacLean DB, et al. Quantitative and qualitative assessment of closed fracture healing using computed tomography and conventional radiography. *Acad Radiol* 2003;10:1267–73.
- [3] Morshed S, Corrales LA, Genant H, Miclau 3rd T. Outcome assessment in clinical trials of fracture-healing. *J Bone Joint Surg Am* 2008;90:62–7.
- [4] Corrales LA, Morshed S, Bhandari M, Miclau 3rd T. Variability in the assessment of fracture-healing in orthopaedic trauma studies. *J Bone Joint Surg Am* 2008;90:1862–8.
- [5] Bhandari M, Guyatt GH, Swiontkowski MF, Tornetta 3rd P, Sprague S, Schemitsch EH. A lack of consensus in the assessment of fracture healing among orthopaedic surgeons. *J Orthop Trauma* 2002;16:562–6.
- [6] Dijkman BG, Sprague S, Schemitsch EH, Bhandari M. When is a fracture healed? Radiographic and clinical criteria revisited. *J Orthop Trauma* 2010;24:76–80.
- [7] Dijkman BG, Busse JW, Walter SD, Bhandari M. The impact of clinical data on the evaluation of tibial fracture healing. *Trials* 2011;12:237.
- [8] Whelan DB, Bhandari M, McKee MD, Guyatt GH, Kreder HJ, Stephen D, et al. Interobserver and intraobserver variation in the assessment of the healing of tibial fractures after intramedullary fixation. *J Bone Joint Surg Br* 2002;84:15–8.
- [9] Davis BJ, Roberts PJ, Moorcroft CI, Brown MF, Thomas PB, Wade RH. Reliability of radiographs in defining union of internally fixed fractures. *Injury* 2004;35:557–61.
- [10] Molière S, Dosch JC, Bierry G. Pelvic, acetabular and hip fractures: what the surgeon should expect from the radiologist. *Diagn Interv Imaging* 2016;97:709–23.
- [11] Eastaugh-Waring SJ, Joslin CC, Hardy JR, Cunningham JL. Quantification of fracture healing from radiographs using the maximum callus index. *Clin Orthop Relat Res* 2009;467:1986–91.
- [12] Whelan DB, Bhandari M, Stephen D, Kreder H, McKee MD, Zdero R, et al. Development of the radiographic union score for tibial fractures for the assessment of tibial fracture healing after intramedullary fixation. *J Trauma* 2010;68:629–32.
- [13] Bhandari M, Chiavaras MM, Parasu N, Choudur H, Ayeni O, Chakraverty R, et al. Radiographic union score for hip substantially improves agreement between surgeons and radiologists. *BMC Musculoskelet Disord* 2013;25:70.
- [14] Chiavaras MM, Bains S, Choudur H, Parasu N, Jacobson J, Ayeni O, et al. The Radiographic Union Score for Hip (RUSH): the use of a checklist to evaluate hip fracture healing improves agreement between radiologists and orthopedic surgeons. *Skeletal Radiol* 2013;42:1079–88.
- [15] Leow JM, Clement ND, Tawonsawatruk T, Simpson CJ, Simpson AH. The radiographic union scale in tibial (RUST) fractures: reliability of the outcome measure at an independent centre. *Bone Joint Res* 2016;5:116–21.
- [16] Litrenta J, Tornetta P, Mehta S, Jones C, O'Toole RV, Bhandari M, et al. Determination of radiographic healing: an assessment of consistency using RUST and modified RUST in metadiaphyseal fractures. *J Orthop Trauma* 2015;29:516–20.
- [17] Frank T, Osterhoff G, Sprague S, Garibaldi A, Bhandari M, Slobogean GP, et al. The Radiographic Union Score for Hip (RUSH) identifies radiographic nonunion of femoral neck fractures. *Clin Orthop Relat Res* 2016;474:1396–404.
- [18] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [19] Farzin B, Gentric JC, Pham M, Tremblay-Paquet S, Brosseau L, Roy C, et al. Agreement studies in radiology research. *Diagn Interv Imaging* 2017;98:227–33.
- [20] Fayad LM, Patra A, Fishman EK. Value of 3D CT in defining skeletal complications of orthopedic hardware in the postoperative patient. *Am J Roentgenol* 2009;193:1155–63.
- [21] McClelland D, Thomas PB, Bancroft G, Moorcroft CI. Fracture healing assessment comparing stiffness measurements using radiographs. *Clin Orthop Relat Res* 2007;457:214–9.