The Journal of Foot & Ankle Surgery ■■ (2017) ■■–■■

Contents lists available at ScienceDirect

The Journal of Foot & Ankle Surgery

journal homepage: www.jfas.org

Original Research

Development and Validation of the Foot Union Scoring Evaluation Tool for Arthrodesis of Foot Structures

Timothy A. Karthas, DPM [1,2], Jeremy J. Cook, DPM, MPH, FACFAS [3,4], Michael R. Matthews, DPM [1,2], Michael L. Sganga, DPM [2], Daniel D. Hansen, DPM [2], Byron Collier, DPM [2], Philip Basile, DPM, FACFAS [3,5], Emily A. Cook, DPM, MPH, FACFAS [3,6]

[1]*Chief Resident, Division of Podiatric Surgery, Department of Surgery, Mount Auburn Hospital, Cambridge, MA*
[2]*Clinical Fellow, Department of Surgery, Harvard Medical School, Boston, MA*
[3]*Instructor, Department of Surgery, Harvard Medical School, Boston, MA*
[4]*Director of Research and Quality Assurance, Division of Podiatric Surgery, Department of Surgery, Mount Auburn Hospital, Cambridge, MA*
[5]*Chief, Division of Podiatric Surgery, Department of Surgery, Mount Auburn Hospital, Cambridge, MA*
[6]*Director of Resident Training, Division of Podiatric Surgery, Department of Surgery, Mount Auburn Hospital, Cambridge, MA*

ARTICLE INFO

Level of Clinical Evidence: 3

Keywords:
FUSET
fusion
radiographic union
scoring scale
tarsometatarsal

ABSTRACT

Reliable evaluation of osseous consolidation after pedal arthrodesis can be difficult, and the presence or absence of radiographic healing often dictates care. Plain radiographs remain the mainstay imaging tool owing to their cost, efficiency, and low radiation exposure. Applying radiographic parameters that can reliably determine osseous healing is essential. However, currently, no reliable or validated measures are available to determine osseous union of any joint in the foot or ankle. The purpose of the present study was to develop a radiographic healing scoring system that would enhance the diagnostic healing assessment after joint arthrodesis of the foot or ankle. We adapted several existing scales previously validated for fracture healing in the leg, because no study has attempted to apply this to a joint fusion model. A total of 150 cases were evaluated by 6 blinded assessors to test the interrater reliability of the subjective healing assessment compared with the proposed scoring system. The radiographs were classified by the postoperative period: ≤4 weeks, 5 to 12 weeks, and >12 weeks. The initial proposed scale was found to have high interrater reliability but was burdensome. Using a priori item reduction protocols, a limited 5-item scale further improved the internal consistency and reduced the burden. The result was excellent interrater reliability ($\alpha = 0.978$, standard deviation 0.02, 95% confidence interval 0.96 to 0.99) among all assessors compared with the reduced reliability ($\alpha = 0.752$) for subjective arthrodesis healing. Intrarater reliability was also found to be superior using a test–retest method. The reliability of this system appeared superior to the subjective assessment of arthrodesis healing, even in the absence of clinical correlates, after foot arthrodesis.

One question that all foot and ankle surgeons must face is when to advance patient mobility and weightbearing after both elective osseous procedures and trauma (1–7). To help answer this question, practitioners must rely on both clinical and radiographic findings during the postoperative period, because advanced imaging is costly and not routinely ordered. Clinically, surgeons consider patient signs and symptoms such as pain levels, edema, and the appearance and stability of the joints to gain an appreciation of healing. These have not always been reliable, however, because subjective differences and bias can be present on the part of the patients and practitioners (8–10). Therefore, foot and ankle surgeons must also rely on, and correlate their clinical findings with, the imaging findings, primarily from plain radiographs, to gain more information about osseous healing and fusion. This can also be difficult, especially in the foot, because the structure of the foot and the number of joints in a relatively small area can result in radiographic artifacts or even osseous and implant overlap of the bones and joints that must be evaluated. This potential misinterpretation could, in turn, contribute to clinical decisions about postoperative advancements that are inappropriate and possibly harmful (4,5,7,11,12).

Financial Disclosure: None reported.
Conflict of Interest: None reported.

Address correspondence to: Timothy A. Karthas, DPM, Division of Podiatric Surgery, Department of Surgery, Mount Auburn Hospital, 725 Concord Avenue, Suite 3600, Cambridge, MA 02138.
E-mail address: timothy.karthas@gmail.com (T.A. Karthas).

1067-2516/$ - see front matter © 2017 by the American College of Foot and Ankle Surgeons. All rights reserved.
https://doi.org/10.1053/j.jfas.2017.11.033

A number of validated orthopedic radiographic scoring scales have been developed, such as the radiographic union scale, radiographic union scale for the hip, and the radiographic union scale for tibial fractures (6,13–15). These, just as with most of the reported data on osseous union, pertain to traumatic bone healing and not elective osteotomy or arthrodesis. In addition, 2 radiographic scoring systems for elective osseous healing in the foot have recently been developed. One for healing of long bones of the foot and one for healing of the irregular bones of the foot (16,17). Combined, these are valuable tools for foot and ankle surgeons to use when evaluating postoperative or posttraumatic radiographs, because they can help improve clinical decision-making and patient care. Additionally, these scales can provide a standardized method of evaluating radiographs that could be implemented in future studies or in developing set postoperative protocols. Despite the inherent value of these scales to both patients and physicians, only these 2 scoring systems have been designed for elective osteotomies of the foot and none have been designed specifically for arthrodesis of the bones of the foot or ankle (2,8,16–18). Furthermore, it is unclear whether the previous scoring systems can be extrapolated directly to arthrodesis procedures, because an inherent difference could exist in the radiographic characteristics after arthrodesis compared with fracture or osteotomy healing within a single bone. This seems likely considering the joint preparation work and presence of a dense subchondral plate during elective arthrodesis that is not encountered in osteotomies or trauma. Also, the healing times and risk of nonunion could be increased after arthrodesis in foot and ankle surgery, for which nonunion rates can range from 0% to 47%, depending on the joint and several patient factors. Combined, these factors make it all the more important to develop a validated scoring system for arthrodesis of the foot and ankle, because patient advancement in weightbearing and the interval to initiating physical therapy are often directly linked to signs of osseous healing (17,19,20).

The primary aim of the present study was to develop a reliable radiographic healing scoring system of foot and ankle arthrodesis without the use of computed tomography (CT). We hypothesized that the final instrument would have greater inter- and intrarater reliability than the standard subjective physician radiographic assessment alone in determining osseous union of a foot joint.

Patients and Methods

A prospective reviewer-blinded study was undertaken to evaluate the reliability of a newly proposed radiographic osseous union scoring instrument for arthrodesis of foot joints. Previous research on this subject has been focused on establishing radiographic healing features for a single bone (e.g., irregular bones, such as the calcaneus, and long bones, such as the metatarsal) (16,17). These contributed to the process of identifying radiographically relevant findings that led to the more difficult challenge of assessing fusion of an irregular and long bone. Instrument development proceeded with a modified Delphi approach to determine the items that should be included in the scoring system and represented radiographic signs of healing. The development team first identified existing radiographic bone healing scales or instruments previously validated in the orthopedic data for other bones of the leg and hip (radiographic union score, radiographic union scale for hip, radiographic union scale for tibial fractures, and long bones of the foot) (6,13,15,17). Members of the development team then adapted portions of these scales to generate an initial scoring scale that would be subjected to a validation process (Fig. 1). The inter- and intrarater reliability of the initial scoring scale components were previously confirmed in a recent study assessing radiographic osteotomy healing in the long bones of the foot and irregular bones of the hindfoot (16,17).

A validation process for assessing radiographic union after foot arthrodesis was subsequently designed using methods previously described for foot and ankle bone healing assessment (16,17). To limit minimize the signal-to-noise in the final analysis, elective first tarsometatarsal arthrodesis was the procedure used as the proxy for foot arthrodesis. After institutional review board approval, study cases were obtained. The study cases included standardized anteroposterior (AP) and lateral radiographs for each case were obtained from a picture archive and communication system of all consecutive elective first tarsometatarsal arthrodeses from January 2000 to April 2015, identified using an institution billing database and Current Procedural Terminology (American Medical Association, Chicago, IL) codes 28297, 28730, 28735, and 28740. Patients were excluded if any of the following features were present: elective fusions not involving the first tarsometatarsal joint, poor image quality, incomplete radiographs (missing AP or lateral films), <1 year minimum of radiographic follow-up data, and signs on the radiographs that made it obvious the patient was in the immediate postoperative period, including overlying dressings, casts, or staples, that could not be stripped from the images.

All patient identifiers and indicators of date and time were removed from all images. All images were standardized to size, scale, and background and then placed in a neutral digital format that permitted viewing but not adjustment of magnification, contrast, or any other image manipulation (Fig. 2). AP and lateral radiograph sets were collected and organized into 3 distinct postoperative periods: ≤4 weeks, 5 to 12 weeks, and >12 weeks. To minimize bias, each radiograph set from a given period was assigned a number, and a random number generator was used to assign the order in which the blinded assessors viewed the cases.

An a priori power calculation was then performed to determine the necessary minimum required radiographs and assessors needed to complete the present study. Using the formula ($2 \times$ [number of assessors]^2), it was determined that a minimum of 72 patient cases with 6 blinded assessors would provide a high degree of precision for diagnostic validation of the scoring scale (6). To enhance the robustness of the results, the first 150 cases (300 radiograph assessments) meeting the selection criteria were included.

Each of the 6 blinded assessors were instructed to evaluate each AP and lateral radiographic image set for healing in 2 stages (stage 1, subjective score; and stage 2, arthrodesis assessment tool). In the first stage (subjective score), the assessors were asked to evaluate the images using their own experience and expertise and to indicate whether they believed the fusion site had healed by simply indicating “yes” for healed and “no” for not healed. For the proposed scale, each of the components were scored with a 3-point Likert scale: 0, no healing evident; 1, some healing evident; and 2, complete healing evident. This is a generalized example but the specifics applied for each component are shown in Fig. 1. Callus formation, hardware lucency, and image quality assessments were all graded using a binary scale (Fig. 1).

In the next stage, the assessors were given the same images in a newly randomized order and instructed to reassess each AP and lateral radiograph set again using the proposed arthrodesis assessment tool developed using the modified Delphi process. Assessors were not permitted to go back and change answers after moving onto the next radiograph in the sequence. This prevented assessors from changing their response if they recognized the same case in a different period later in the evaluation sequence. Each assessor evaluated 150 cases of paired AP and lateral radiographs according to the instructions.

Validation of the scale proceeded with assessments of the interrater reliability and sensitivity analyses of the radiographic imaging evaluations. Interrater agreement of the binary subjective score was first analyzed (binary subjective assessment of whether an osteotomy had healed using the assessor’s own experience and expertise). Agreement of the binary subjective score was then compared with the agreement of the arthrodesis assessment tool. Interrater reliability was determined through the use of Cronbach’s α correlation coefficient, which is useful in examining sets of items within an instrument (21) (Table 1).

The individual components of arthrodesis union from the radiographic assessment tool were then analyzed for the possibility of testing domain redundancy, because this could lead to item reduction to decrease instrument burden. Sensitivity analyses were conducted by removing various components of the scores to determine whether item reduction resulted in improved instrument stability. A second sensitivity analysis determined how many arthrodeses from ≤4 weeks postoperatively were designated as “healed” by the assessors’ binary subjective assessment.

To assess intrarater reliability, a test–retest method was selected. This specifically quantifies the internal consistency of each reviewer. This phase of the study was performed 2 weeks after the initial assessment to minimize the risk of recall bias. A total of 50 cases were selected from the original 150 cases using a random number generator and placed into a new evaluation sequence. The blinded assessors were asked to evaluate osseous healing using the same 2-stage assessment (stage 1, subjective score; stage 2, arthrodesis assessment tool) but this time using the optimized arthrodesis assessment tool. Intrarater agreement was assessed using the Spearman rank correlation coefficient (22,23). Again, a sensitivity analysis was used to assess the accuracy of binary healing. All analyses were calculated using SAS, version 9.1.3 (SAS Institute, Cary, NC). A p value of $< .05$ was considered statistically significant, and 95% confidence intervals were used as appropriate.

Table 1
Interpretation of internal consistency for interrater reliability (Cronbach’s α) (21–23)

Cronbach’s α	Correlation Coefficient
Excellent, $\alpha \geq 0.9$	Very high, $r \geq 0.9$
Good, $\alpha \geq 0.7$ but <0.9	High, $r \geq 0.70$ but <0.90
Acceptable, $\alpha \geq 0.6$ but <0.7	Moderate, $r \geq 0.50$ but <0.70
Poor, $\alpha \geq 0.5$ but <0.6	Low, $r \geq 0.30$ but <0.50
Unacceptable, $\alpha <0.5$	Negligible, $r <0.30$