Eur J Trauma Emerg Surg (2019) 45:13–19
https://doi.org/10.1007/s00068-017-0796-4

ORIGINAL ARTICLE

The Non-Union Scoring System: an interobserver reliability study

M. van Basten Batenburg[1] · I. B. Houben[1] · T. J. Blokhuis[2]

Received: 20 September 2016 / Accepted: 29 May 2017 / Published online: 2 June 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract
Purpose The Non-Union Scoring System (NUSS) aims to classify non-unions according to their severity and relate them to four treatment categories. The main purpose of this study was to evaluate the reliability of the NUSS. In addition we assessed its clinical validity.
Methods Forty-four Patients with a tibia non-union between 2005 and 2015 were included in this study. Data from all included patients were scored independently by three observers according to the NUSS criteria. The interobserver agreement was evaluated using the intraclass correlation coefficient (ICC). The interobserver agreement of the Weber–Cech system was assessed using Fleiss' kappa. Finally, the clinical validity of the NUSS was analysed by comparing outcomes of the actual treatment groups to the proposed treatment groups following from the NUSS scores.
Results Forty-four patients were included. The comparison of NUSS scores between observers showed substantial agreement [ICC; 0.78 (0.67–0.86)]. The comparison of the Weber–Cech classification between observers showed only fair agreement [Fleiss κ; 0.30 (0.17–0.42)]. The χ^2 test for the treatment groups according to the NUSS and the treatments at index procedure showed an independent relation ($\chi^2 = 5.794$, 6 degrees of freedom, p: 0.447). In contrast, the proposed treatment strategy corresponds well to the definitive treatment ($\chi^2 = 29.963$, 9 degrees of freedom, $p < 0.001$).
Conclusion We conclude that the NUSS is both a reliable and valid system to classify non-unions.

Keywords Tibia · Fracture · Non-union · NUSS · Interobserver reliability · Non-Union Scoring System

Introduction

Despite a growing palette of treatment options, non-unions of the tibia are an increasing economic burden of healthcare. Average costs for the treatment of non-unions of the tibia have been estimated ranging from $53,506 in the U.S.A. [1] to £29.204 in the United Kingdom [2]. In recent years, many factors contributing to the emergence of a fracture non-union have been identified [3]. Condition of the bone, soft tissues, systemic condition of the patient and environmental factors all play their part in the development of a non-union [4, 5]. In addition, many new treatment techniques have become available over the past decades, all aiming to improve patient outcome.

In contrast to the developments in treatment options, classification of non-unions has not shown much progress. The most commonly used classification system for non-unions still dates from the 1970s. In 1976, Weber and Cech introduced a system based on radiographic appearance of non-unions, which they related to the vascularity of the fracture [6]. In this classification, three types of non-union can be distinguished: hypertrophic, oligotrophic or atrophic non-union. A hypertrophic non-union implies a hypervascular fracture site with abundant callus formation. An oligotrophic non-union is characterised by adequate vascularisation with minimal callus formation.

✉ M. van Basten Batenburg
michielvanbb@gmail.com

[1] Department of Surgery, University Medical Center Utrecht, Voorstraat 98b, 3512 AV Utrecht, The Netherlands

[2] Department of Surgery, Maastricht University Medical Center, Maastricht, The Netherlands

Finally, an atrophic non-union is described as a fracture site with impaired vascularity, no callus formation and bone atrophy. Although this is the most frequently used system to date, the validity of the system has been questioned. Especially the theory of vascularity seems to be outdated since several studies have shown that atrophic non-unions are not avascular [7–9]. Over the years, other non-union classifications have been introduced based on different non-union properties. One system has been described by Ilizarov, who divided non-unions into two categories: either lax or stiff based on their mobility [10]. Paley et al. introduced a system based on lax and stiff fractures either with or without bone loss of more than 1 centimetre [11]. Although these systems aim to provide tools in the treatment of non-unions, their focus is mainly on the radiographic appearance in combination with the clinical stability of the fracture, thereby ignoring many other factors involved in the evolution of a non-union [3, 12].

To further study fracture non-unions, it is essential that some comparison of non-unions is made possible. Moreover, it is critical for clinical decision making to be able to assess severity of a non-union. For clinical use, this assessment of severity should ideally be related to a treatment strategy. In 2008 Calori et al. proposed the Non-union Scoring System (NUSS), a new classification system which takes into account most factors involved in the development of a non-union, including the Weber–Cech classification [5]. The Non-Union Scoring System is shown in Table 1. In short, points are given for several risk factors and patient related factors, these scores are added up and multiplied by two, resulting in a score ranging from 0 to 100. These scores are then used to categorise non-unions into four separate groups reflecting the severity of the non-union. These severity categories are related to the treatment options [13].

For a classification system to be clinically relevant it has to be both reliable and valid [14, 15]. Since its publication the NUSS has been validated by Abumunaser et al. [16] and Calori et al. [13]. Both studies compared the actual treatment of non-unions to the proposed treatment according to the NUSS score. Both these studies showed excellent results using the predictive value of the NUSS and the patient outcome per non-union category. However, to our knowledge no reliability analysis was done so far. A reliable instrument is needed to get consistent results which can be reproduced [15]. Therefore, the main purpose of this study was to evaluate the reliability of the NUSS through the determination of interobserver agreement. In addition, we assessed the clinical validity of the NUSS by comparing the outcome of the actual treatment to the proposed NUSS treatment group.

Methods

All fractures of the tibia treated in a single level-I trauma centre between 2005 and 2015 were reviewed using the database of electronic medical records. Inclusion criteria were, patients with a radiographic non-union of the tibia according to the United States Food and Drug Administration (FDA) definition for non-union: “a fracture that is at least 9 months old and has not shown any signs of progression to healing for three consecutive months” [4]. The first intervention performed to procure healing at a minimum of 9 months after initial trauma was regarded as the index procedure upon which the NUSS was scored. Patients in whom no surgical intervention took place at least 9 months after trauma were excluded, as well as non-unions resulting from elective orthopaedic surgery. Data from all included patients were assessed independently by three observers: one senior orthopaedic trauma surgeon and two trauma surgery interns.

The Non-Union Scoring System (NUSS) criteria were retrieved from the medical records. Additional data were retrieved to determine patient demographics, fracture type, energy and circumstance of trauma, surgical history and clinical outcome. The criteria as described by Calori et al. [13] allow some subjective interpretation. Therefore, we specified all criteria in advance to be able to score the patients consistently. These specifications were added to the NUSS in Table 1. First of all, we scored all factors at or around the index procedure, except Gustilo classification and adequacy of primary surgery; these were scored based on records around the trauma date. Furthermore, within the number of previous interventions <2 was regarded as ≤2, and <4 was regarded as ≤4. Blood tests (C-reactive protein (CRP), white cell count (WCC) and erythrocyte sedimentation rate (ESR)) were scored positive if elevated within 3 months before the index procedure. Smoking was scored as a ‘Yes’ up to 6 months after cessation of smoking. Clinical infection status was scored septic if the patient had a leucocyte score >12.000/mm^3 and an active source of infection. Regarding drug use, both non-steroidal anti-inflammatory drugs (NSAIDs) and steroids were scored positively if used for more than 5 days within 3 months before or after the index procedure, regardless of dosage. In steroid use, any route of administration was scored positively, including inhaled steroids. All observers used a randomly selected set of ten non-unions from the database for training.

After the observers scored all patients independently, the patients were assigned to treatment groups based on the mean NUSS score from the three observers. These four treatment groups as described by Calori et al. [5]

Table 1 Non-Union Scoring System with specifications

The bone	Score	Max. score
Quality of the bone		
Good	0	
Moderate (e.g. mildly osteoporotic)	1	3
Poor (e.g. severe porosis or bone loss)	2	
Very poor (Necrotic, appears avascular or septic)	3	
Primary injury open or closed fracture		
Closed	0	
Open grade 1	1	5
Open grade 2–3A	3	
Open grade 3 B–C	5	
Number of previous interventions to procure healing[a]		
None	1	
≤2	2	4
≤4	3	
>4	4	
Invasiveness of previous interventions		
Minimallyinvasive: closed surgery (screws, *k*-wires, …)	0	
Internal intra-medullary nailing	1	3
Internal extra-medullary nailing	2	
Any osteosynthesis which includes bone grafting	3	
Adequacy of primary surgery		
Inadequate stability	0	
Adequate stability	1	1
Weber & Cech group		
Hypertrophic	1	
Oligotrophic	3	5
Atrophic	5	
Bone alignment		
Non-anatomic alignment	0	
Anatomic alignment	1	1
Bone defect—Gap		
0.5–1 cm	2	
1–3 cm	3	5
>3 cm	5	
Soft tissues		
Status		
Intact	0	
Previous uneventful surgery, minor scarring	2	6
Previous treatment of soft tissue defect (e.g. skin loss, local flap cover, multiple incisions, compartment syndrome, old sinuses)	3	
Previous complex treatment of soft tissue defect (e.g. free flap)	4	
Poor vascularity: absence of distal pulses, poor capillary refill, venous insufficiency	5	
Presence of actual skin lesion/defect (e.g. ulcer, sinus, exposed bone or plate)	6	
The patient		
ASA grade		
1 or 2	0	
3 or 4	1	1
Diabetes		
No	0	
Yes—well controlled (HbA1c < 10)	1	2
Yes—poorly controlled (HbA1c > 10)	2	
Blood tests: FBC, ESR, CRP[b]		
FBC: WCC > 12	1	
ESR > 20	1	3
CRP > 20	1	
Clinical infection status		
Clean	0	
Previously infected or suspicion of infection	1	4
Septic[c]	4	
Drugs[d]		
Steroids	1	
NSAIDs	1	2
Smoking status[e]		
No	0	
Yes	5	5

[a] Including first surgical intervention

[b] Scored positive if elevated within 3 months before the index procedure

[c] Leucocyte score > 12.000/mm^3 and an active source of infection

[d] Scored positively if used for more than 5 days within 3 months before or after the index procedure, regardless of dosage

[e] Scored positive up to 6 months after cessation

reflect the severity of the non-union and indicate what type of treatment, ranging from simple procedures, such as nail exchange to amputation, will most likely be indicated in the according group as definitive treatment. Group 1 includes NUSS scores ranging from 0 to 25, group 2 from 26 to 50, group 3 from 51 to 75 and group 4 from 76 to 100.

To assess the clinical validity of the NUSS, the treatment as deemed indicated based upon the NUSS was compared with treatment at the index procedure. We divided the patients based on the actual received treatment at index procedure into four treatment groups according to the guidelines as proposed by Calori et al. in 2014 [13]. Our patients with a simple procedure such as change of fixation system, were placed in group 1 (0–25). Patients who underwent a more invasive procedure which included resection of bone without the need for bone grafting were included in group 2 (26–50). Patients who

had a resection of the fracture site treated with bone grafts were placed in group 3 (51–75). Finally, patients who underwent an amputation were scored in group 4 (76–100).

If additional surgical intervention was required after the index procedure to procure bone healing, this was regarded as the 'definitive procedure'. The definitive procedure was categorised and compared with the proposed NUSS treatment in the same manner as the index procedure.

Finally, the treatment outcomes were determined. Treatment failure was defined as a reoperation performed to procure healing after the index procedure, or a non-union at final follow-up without additional procedures. Final outcome at last follow-up was assessed using radiographs which were scored according to the RUST criteria [17]. Patients with a follow-up of <6 months after the index procedure were regarded as lost to follow-up and were excluded from the outcome analysis.

Observer and sample size calculation

The sample size was calculated based upon the guidelines for a reliability study as described by Walter et al. [18]. This technique determines the number of observers needed to achieve adequate reliability as well as the number of subjects (observations) needed. According to these guidelines we determined an intraclass correlation $\rho 0$ (H0 lowest acceptable rate of reliability) of 0.6 and a $\rho 1$ (H1 expected outcome of ICC) of 0.8. This resulted in an optimal number of observers (n) of 3. To calculate the sample size the same guideline by Walter et al. was used. Using a $\rho 0$ of 0.6, $\rho 1$ of 0.8 and $n = 3$, the sample size was calculated to be at least 26.1 subjects using a significance (α) = 0.05 and a power ($1 - \beta$) of 0.80.

Statistical analysis

The intraclass correlation coefficient (ICC) with a 95% confidence interval was used to assess interobserver reliability of agreement for the overall NUSS scores. The ICC is used for continuous data and is the equivalent of the quadratically weighted κ for categorical data [19]. Therefore the reliability analysis of the Weber–Cech class was performed using the weighted κ model for multiple observers as described by Fleiss et al. [20]. For the ICC, a two-way mixed effect model was used with absolute agreement, assuming the sample effects are randomly assigned with a fixed number of three raters, rating all patients. Values between 0.2 and 0.4 were considered "fair agreement," between 0.4 and 0.6 "modest agreement," between 0.6 and 0.8 "substantial agreement" and over 0.8 as "excellent agreement" as described by Landis and Koch [21]. The χ^2 test was used to compare the actual treatment groups at index and definitive treatment to the treatment groups proposed by the average NUSS scores. All tests were performed using SPSS version 22.0 (IBM SPSS Statistics for Windows, Version 22.0, Armonk, NY).

Results

Data from 44 patients were included in this analysis. 30 patients were male and 14 were female. The mean age ± standard deviation at the index procedure was 41 years ± 15 (range 19–82 years). The mean time from trauma until index procedure was 12.6 months ± 4.8 (range 9–31 months). In 35 cases the fracture was due to a high-energy trauma, in 7 cases due to a low-energy trauma and in 2 cases the trauma mechanism was unknown. 28 fractures were due to a road traffic accident, 5 were work related, 6 had other causes, and in 5 cases the circumstances of trauma were unknown. Following the index procedure 17 patients (38.6%) underwent further surgery to procure healing. Within this group a mean of 3.4 (range 1–10) surgeries was performed after the index procedure. The follow-up period for all patients after the index procedure averaged 24.37 months ± 19.5. An overview of the demographics is shown in Table 2.

The comparison of NUSS scores between observers showed substantial agreement [ICC; 0.78 (0.67–0.86)]. The comparison of treatment groups following from the NUSS scores also showed substantial agreement [Fleiss κ; 0.62 (0.50–0.75)]. The comparison of Weber–Cech classification between raters showed only fair agreement [Fleiss κ; 0.30 (0.17–0.42)].

After the index procedure the outcome of this treatment was determined. Seven patients were lost to follow-up (<6 months after index procedure) and were, therefore, excluded for this outcome analysis. This resulted in 37 patients available for outcome analysis. According to

Table 2 Patient demographics

Gender	30 Male/14 female
Mean age	41 (range 19–82)
Mean duration of non-union from trauma until index procedure (months)	12.6 (range 9–31)
Trauma mechanism[a]	35 HET/7 LET/2 unknown
Cause of trauma[b]	28 RTA/5 work/6 other
Mean NUSS score ± SD	43.09 ± 16.1 (range 20–84)
Mean follow-up (months)	24.37 (range 0–138)

[a] HET: high-energy trauma, LET: low-energy trauma

[b] RTA: road traffic accident, Work: work-related injury

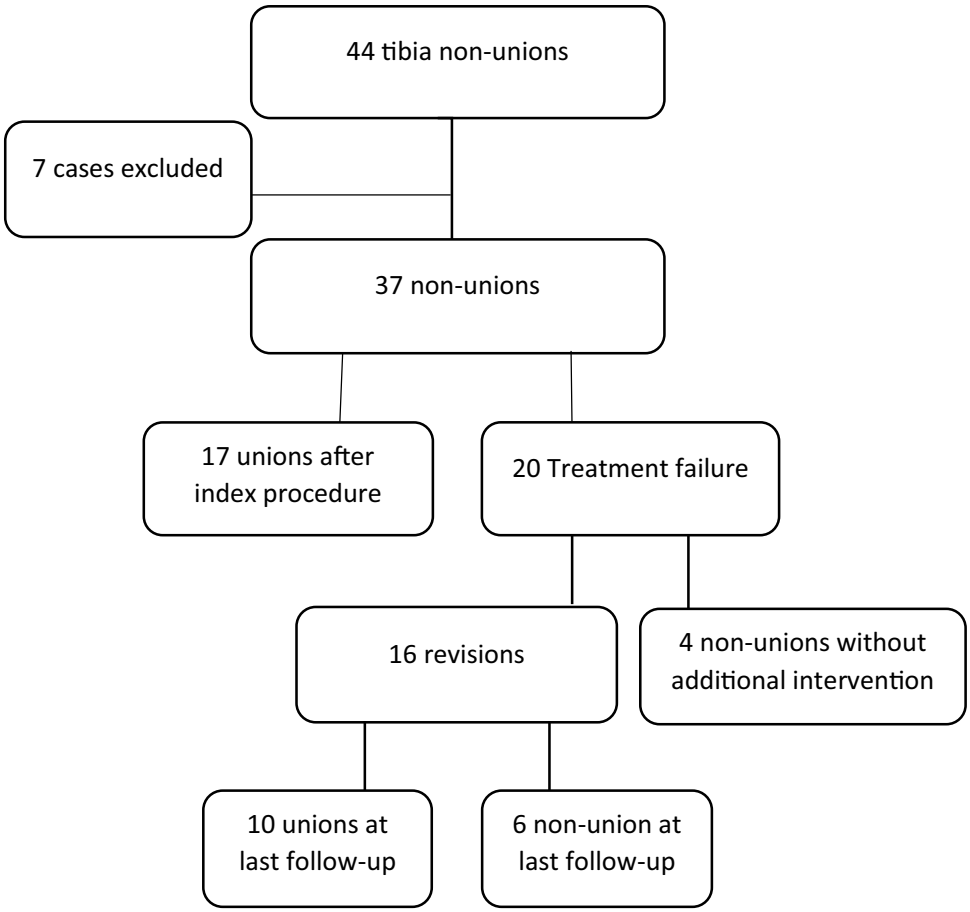

Fig. 1 Flow chart of outcome analysis of index procedure. Flow chart describing the number of patients included in the outcome analysis of the index procedure. Exclusion of seven cases occurred because of follow-up <6 months

the abovementioned definition, union was achieved in 17 patients and treatment failure occurred in 20. Of the treatment failure group, 16 were reoperated upon and 4 had no reoperations but a persistent non-union and at last follow-up. A flow chart of the patients is shown in Fig. 1.

Of the 20 patients showing treatment failure, 14 (70%) underwent index procedures classified lower than the proposed treatment as suggested by the NUSS score; 3 (15%) were treated accordingly; and 3 (15%) were treated according to a higher treatment group. Of the patients who underwent additional surgical procedures, 6 out of the 16 patients underwent procedures in a higher category than at the index procedure. At final follow-up this resulted in 6 patients remaining undertreated according to the NUSS, 3 treated in the same category as proposed according to the NUSS, and 7 were treated in a higher category than the NUSS. Of the 17 patients who achieved union after the index procedure, 7 underwent treatment classified lower than the proposed treatment, 3 were treated accordingly, and 7 were treated in a higher treatment group. At final follow-up radiographic union was achieved in 27 out of the 37 patients (73%). The relation between NUSS score, index procedure, definitive treatment and outcome is illustrated in Figs. 2 and 3.

The comparison between the proposed treatment according to the NUSS score and the treatment at index procedure showed that there was no significant relation between these two. The Chi-square test for the treatment groups according to the NUSS and the treatments at index procedure showed an independent relation with a χ^2 of 5.794 with 6 degrees of freedom and p: 0.447. In contrast, the proposed treatment strategy corresponds well to the definitive treatment ($\chi^2 = 29.963$, 9 degrees of freedom, $p < 0.001$).

Discussion

The aetiology of non-unions is still not fully understood [9]. Despite advances in treatment options, non-unions remain challenging and cause debate amongst orthopaedic trauma surgeons. Accordingly, consensus is lacking in definition and classification of non-unions. [22]. Radiographic appearance, although frequently used, has no relation to the biological status of a non-union and therefore has limited use in determining treatment strategy [9]. The Non-Union Scoring System was described by Calori

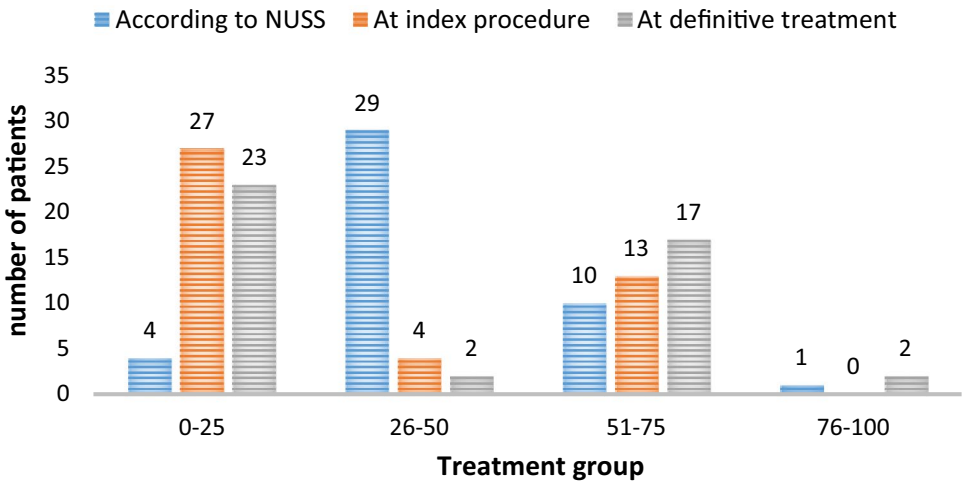

Fig. 2 Distribution of patients per treatment group according to NUSS, index procedure and definitive treatment. The treatment groups are shown on the *X*-axis, where 0–25 corresponds to group 1 of the NUSS score as described by Calori et al. [13], the category with relatively simple non-unions; 26–50 corresponds to group 2, 51–75 corresponds to group 3, 76–100 to group 4

et al. to compare and evaluate non-unions more objectively. It takes the most relevant factors in their aetiology into account [5]. The NUSS was then validated by Abumunaser et al. in 2011 and by Calori et al. in 2014 [13, 16]. In this study the reliability was investigated, showing substantial agreement between observers, and a clear relation between outcome of definitive treatment and the severity category as determined with the NUSS score.

With the ICC showing substantial agreement between observers, the NUSS can be regarded as a reliable tool to assess fracture non-unions. Since our study was performed by one orthopaedic trauma surgeon and two trauma surgery interns, we feel that even with limited clinical experience the NUSS score is still a reliable instrument to evaluate fracture non-unions. In contrast, the Weber–Cech classification showed only fair agreement with a Fleiss κ of 0.30, indicating a very limited reliability [18]. Therefore, the NUSS seems to be a more consistent system to assess non-unions.

Apart from the classification of non-unions, the NUSS aims to support treatment strategy by categorising the non-union severity and by proposing a treatment strategy for each severity category. Our results indicate that the NUSS score is indeed useful for this purpose. The index procedure was initiated without the use of the NUSS score. When comparing the index procedure to the proposed treatment strategy as indicated by the NUSS score, it becomes clear that a majority of patients who developed a treatment failure were undertreated at the index procedure. This means that implementation of the NUSS score, including the proposed treatment strategy, would have resulted in more aggressive or invasive surgical procedures. Although outcome of a more extensive treatment remains hypothetical for these patients, our data show that use of the NUSS score has direct consequences for choosing a treatment strategy. This is in accordance with the earlier studies on validity by Calori and Abumunaser [13, 16].

In the current study some limitations of the NUSS were encountered. Specifically the definition of some of the parameters in the NUSS is not very strict, allowing subjective interpretation. To be as consistent as possible, we defined some of these factors in advance. It is arguable whether this is allowed in testing interobserver reliability of a certain classification system, since doing so increases consistency and thus improves the test results. However, if the NUSS is to be further studied or applied in clinical decision making, it is very important to specify these conditions.

Furthermore, our retrospective design introduced some weaknesses to our study. Not all factors described by Calori et al. were available in our medical records for all patients. Especially ESR, and to a lesser extent the other blood tests (CRP, WCC), were not performed in a significant number of cases since these values are only determined when clinical signs of infection are present. Another possible limitation of the present study is that no radiologist was consulted to assess the radiographs. Because clinical decision making in our institution is done without participation of the radiologists, and since the radiographic evaluation is only a minor aspect of the total NUSS score, not consulting a radiologist was decided to mimic clinical judgement as closely as possible. The used sample size is relatively small although it is comparable to the validation study by Abumunaser et al. [16]. Also, one has to keep in mind that the inclusion of cases was done using a strict definition of non-union (9 months after trauma) and that only tibia non-unions were included. This limits the translation of our results to, for example, a delayed union of the femur. Still, the number of patients is higher than the minimum number of cases as determined in the sample size calculation for this reliability study.

In conclusion, the validity of the NUSS score was already tested in other studies, and these results are confirmed in the present study. Moreover, evaluation of reliability of the NUSS score shows substantial agreement

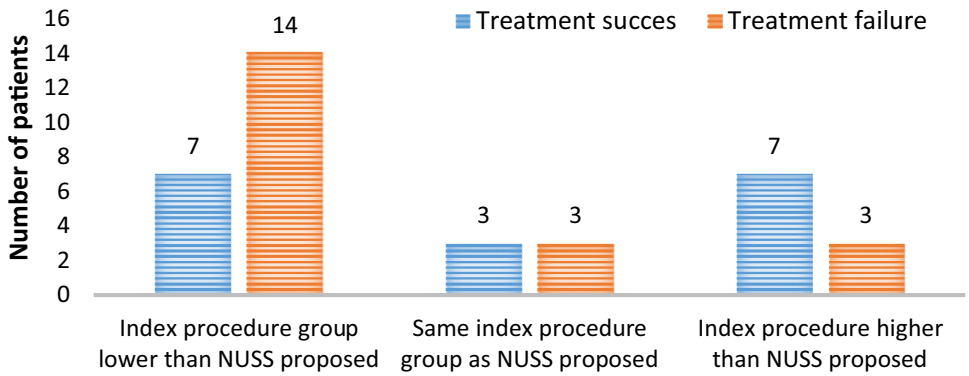

Fig. 3 Comparison between treatment group at index procedure and proposed treatment group according to NUSS is related to the outcome of the index procedure. The *X*-axis shows the treatment groups at the index procedure compared with the NUSS proposed treatment groups. This is related to the outcome of the index procedure with the *blue bar* depicting the patients with treatment success, and the *orange bar* depicting the patients with treatment failure

between observers. For these reasons we conclude that the NUSS gives us both a reliable and valid system to classify non-unions.

Compliance with ethical standards

Conflict of interest M. van Basten Batenburg, I. B. Houben and T.J. Blokhuis declare that they have no conflict of interest.

Research involving human and animal participants Since this study solely used retrospective data, neither humans nor animals participated actively in the study.

Informed consent The retrieval of data was covered by the general institutional review board (IRB) approval for data analysis in our institution.

References

1. Antonova E, Le TK, Burge R, Mershon J. Tibia shaft fractures: costly burden of nonunions. BMC Musculoskelet Disord. 2013;26(14):42.
2. Patil S, Montgomery R. Management of complex tibial and femoral nonunion using the Ilizarov technique, and its cost implications. J Bone Joint Surg Br. 2006;88(7):928–32.
3. Santolini E, West R, Giannoudis PV. Risk factors for long bone fracture non-union: a stratification approach based on the level of the existing scientific evidence. Injury. 2015;46(Suppl 8):S8–19.
4. Harwood PJ, Newman JB, Michael ALR. (ii) An update on fracture healing and non-union. Orthop Trauma. 2010;24(1):9–23.
5. Calori GM, Phillips M, Jeetle S, Tagliabue L, Giannoudis PV. Classification of non-union: need for a new scoring system? Injury. 2008;39(Suppl 2):S59–63.
6. Weber BG, Cech O. Pseudarthrosis. New York: Grune and Stratton; 1976.
7. Reed AA, Joyner CJ, Brownlow HC, Simpson AH. Human atrophic fracture non-unions are not avascular. J Orthop Res. 2002;20(3):593–9.
8. Brownlow HC, Reed A, Simpson AH. The vascularity of atrophic non-unions. Injury. 2002;33(2):145–50.
9. Panteli M, Pountos I, Jones E, Giannoudis PV. Biological and molecular profile of fracture non-union tissue: current insights. J Cell Mol Med. 2015;19(4):685–713.
10. Catagni MA, Bianci-Maiocchi A. Treatment of fractures, nonunions, and bone loss of the tibia with the Ilizarov method. Milan: Medicalplastic; 1998.
11. Paley D, Catagni MA, Argnani F, Villa A, Benedetti GB, Cattaneo R. Ilizarov treatment of tibial nonunions with bone loss. Clin Orthop Relat Res. 1989;241(241):146–65.
12. Giannoudis PV, Einhorn TA, Marsh D. Fracture healing: the diamond concept. Injury. 2007;38(Suppl 4):S3–6.
13. Calori GM, Colombo M, Mazza EL, Mazzola S, Malagoli E, Marelli N, et al. Validation of the Non-Union Scoring System in 300 long bone non-unions. Injury. 2014;45(Suppl 6):S93–7.
14. Kooistra BW, Sprague S, Bhandari M, Schemitsch EH. Outcomes assessment in fracture healing trials: a primer. J Orthop Trauma. 2010;24(Suppl 1):S71–5.
15. Jackowski D, Guyatt G. A guide to health measurement. Clin Orthop Relat Res. 2003;413(413):80–9.
16. Abumunaser LA, Al-Sayyad MJ. Evaluation of the calori et Al nonunion scoring system in a retrospective case series. Orthopedics. 2011;34(5):359.
17. Whelan DB, Bhandari M, Stephen D, Kreder H, McKee MD, Zdero R, et al. Development of the radiographic union score for tibial fractures for the assessment of tibial fracture healing after intramedullary fixation. J Trauma. 2010;68(3):629–32.
18. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17(1):101–10.
19. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas. 1973;33:613–9.
20. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971;76(5):378–82.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
22. Bhandari M, Fong K, Sprague S, Williams D, Petrisor B. Variability in the definition and perceived causes of delayed unions and nonunions: a cross-sectional, multinational survey of orthopaedic surgeons. J Bone Joint Surg Am. 2012;94(15):e1091–6.