BMJ Open

The American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Score; Translation and Validation of the Dutch Language Version for ankle fractures

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017040
Article Type:	Research
Date Submitted by the Author:	28-Mar-2017
Complete List of Authors:	De Boer, A.; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery Tjioe, Roderik J.C.; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery van der Sijde, Fleur; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery Meuffels, Duncan; Erasmus MC, University Medical Center Rotterdam, Orthopaedic Surgery Den Hoed, P.; Ikazia Ziekenhuis, Surgery Van der Vlies, Cornelis; Maasstad Ziekenhuis Tuinebreijer, Wim; Erasmus MC, University Medical Center Rotterdam Verhofstad, Michael; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery Van Lieshout, Esther M.M.; Erasmus MC, University Medical Center Rotterdam AOFAS Study Group, .; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Emergency medicine, Rehabilitation medicine, Patient-centred medicine, Surgery, Sports and exercise medicine
Keywords:	Fracture, Reliability, Validity, Responsiveness, Ankle

The American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Score; Translation and Validation of the Dutch Language Version for ankle fractures

A. Siebe De Boer MD[1], R.J.C. Tjioe[1], F. Van der Sijde[1], Duncan E. Meuffels MD PhD[2], P. Ted Den Hoed MD PhD[3], Cornelis H. Van der Vlies MD PhD[4], Wim E. Tuinebreijer MD PhD[1], Michael H.J. Verhofstad MD PhD[1], Esther M.M. Van Lieshout MSc PhD[1*], AOFAS study Group*

*D.A. Newhall, E.N. Van der Velden – Macauley, L. Vellekoop, J. Romeo

[1]Trauma Research Unit Department of Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

[2]Department of Orthopaedic Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

[3]Department of Surgery, Ikazia Hospital, P.O. Box 5009, 3008 AA Rotterdam, The Netherlands

[4]Department of Surgery, Maasstad Hospital, P.O. Box 9100, 3007 AC Rotterdam, The Netherlands

[*]Corresponding author:

Esther M.M. Van Lieshout, MSc PhD

Trauma Research Unit Department of Surgery

Erasmus MC, University Medical Center Rotterdam

P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Email: e.vanlieshout@erasmusmc.nl

Telephone: + 31 10 7031050

Email addresses:

ASDB: a.deboer@erasmusmc.nl

RJCT: r.tjioe@erasmusmc.nl

FVDS: f.vandersijde@erasmusmc.nl

DEM: d.meuffels@erasmusmc.nl

PTDH: pt.hoed@ikazia.nl

CHVDV: VliesC@maasstadziekenhuis.nl

WET: w.tuinebreijer@erasmusmc.nl

MHJV: m.verhofstad@erasmusmc.nl

EMMVL: e.vanlieshout@erasmusmc.nl

Keywords: Ankle; Fracture; Reliability; Responsiveness; Validity.

ABSTRACT

Objectives: The American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Score is among the most commonly used instruments for measuring outcome of treatment in patients who sustained a complex ankle or hindfoot injury. It consists of a patient-reported and a physician-reported part. A validated, Dutch version of this instrument is currently not available. The aim of this study was to translate the instrument into Dutch and to determine the measurement properties of the AOFAS Ankle-Hindfoot Score Dutch Language Version (DLV) in patients with a unilateral ankle fracture.

Setting: Multicenter (two Dutch hospitals), prospective observational study.

Participants: In total 142 patients with a unilateral ankle fracture were included. Ten patients were lost to follow up.

Primary and secondary outcome measures: Patients completed the subjective (patient-reported) part of the AOFAS Ankle-Hindfoot Score-DLV. A physician or trained physician-assistant completed the physician-reported part. For comparison and evaluation of the measuring characteristics, the Foot Function Index (FFI) and the Short Form-36 (SF-36) were completed by the patient. Descriptive statistics (including floor and ceiling effects), reliability (*i.e.*, internal consistency), construct validity, reproducibility (*i.e.*, test-retest reliability, agreement, and smallest detectable change), and responsiveness were determined.

Results: The AOFAS-DLV and its subscales showed good internal consistency (Cronbach's alpha > 0.90). Construct validity and longitudinal validity were proven to be adequate (76.5% of predefined hypotheses were confirmed). Floor effects were not present. Ceiling effects were present from six months onwards, as expected. Responsiveness was adequate, with a smallest detectable change of 12.0 points.

Conclusions: The AOFAS-DLV is a reliable, valid, and responsive measurement instrument for evaluating functional outcome in patients with a unilateral ankle fracture. This implies that the questionnaire is suitable to compare different treatment modalities within this population or to compare outcome across hospitals.

Trial Registration: Netherlands Trial Register (NTR5613; 05-jan-2016).

ARTICLE SUMMARY

Strengths and limitations of this study:

- It is a prospective, multicenter, observational study with a strong methodologic design
- It shows substantial, previously unknown information about the performance of the American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Score.
- The topic of the clinical study is relevant for orthopedic trauma surgeons, since there is growing need for translated and validated patient reported outcome measures that can be used for determining functional outcome over time.
- Statistical analyses complied with the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines.
- Although the study is mostly relevant for the Dutch-speaking regions, it is also informative for other regions.

BACKGROUND

Ankle fractures are common injuries with a reported incidence rate of 187 fractures per 100,000 people each year (1). Due to an increasing number of people involved in sports and the growing elderly population, this rate is rising significantly in many industrialized countries (1). Ankle fractures can cause a temporary loss of function and quality of life. In order to monitor recovery after treatment, Patient-Reported Outcome Measures (PROMs) are increasingly used in clinical practice and clinical research. They enable detailed evaluation of functional outcome and quality of life after (non-)operative treatment of musculoskeletal injuries from a patient's perspective.

The clinical rating system published by the American Orthopaedic Foot and Ankle Society, the AOFAS Ankle-Hindfoot Score, is one of the mostly used assessment tool in foot surgery (2). This clinical rating system, developed by Kitaoka *et al.*, combines subjective scores of pain and function provided by the patient and objective scores based on the physician's physical examination (*i.e.*, gait, sagittal motion, hindfoot motion, ankle-hindfoot stability, and alignment of the ankle-hindfoot) (3). The questionnaire includes nine items that can be divided into three subscales (pain, function, and alignment). Each of the nine items is scored, accumulating to a total score ranging from 0 points (indicating severe pain and impairment) to 100 points (no symptoms or impairment).

The AOFAS Ankle-Hindfoot Score as a complete scale has been shown to be responsive and valid in its original language version (3-6). The patient-reported part of the scale has been shown to be valid and reliable (7). Reliability of the objective (physician-reported) portion of the scale has not been published. Previous studies involved a wide spectrum of diagnoses, such as general ankle-hindfoot complaints (5), pending ankle or foot surgery (7), surgically treated calcaneal fractures (6), and end-stage ankle osteoarthritis (4).

A validated Dutch version of the AOFAS Ankle-Hindfoot score is not available. The aim of this study was to translate the questionnaire into Dutch and to culturally adapt it to the Dutch population. The next aim was to determine the measurement properties of the AOFAS Ankle-Hindfoot Score Dutch Language Version (AOFAS-DLV) in patients who sustained an ankle fracture.

METHODS

Study design and ethics statement

This study followed a multicenter, prospective, observational study design (*i.e.*, case series) and was performed at two Dutch hospitals. The study is registered at the Netherlands Trial Register (NTR5613). A detailed study protocol is published elsewhere (8). The study was approved by the Medical Research Ethics Committees or Local Ethics Boards of all participating centers. All patients provided informed consent.

Translation

First, the American (original) version of the AOFAS Hindfoot-Ankle Score was translated and cultural adapted into Dutch according to the guideline for Cross Cultural Adaptation of Self-Report Measures by Beaton *et al*. (9), as described in detail in the published study protocol (8). In the last stage of this guideline the pre-final Dutch version was tested in a group of 20 patients, presenting themselves with various foot/ankle problems in one of the participating hospitals. Since there were no ambiguities or misunderstandings of the questions in this group, the translated questionnaire was considered the final AOFAS Ankle-Hindfoot Score-DLV.

Validation

Patient recruitment

Patients were recruited from May 1, 2014 to March 29, 2016. Patients were identified from hospital records, based upon their ICD-10 (International Coding of Diseases, 10th revision) code or Diagnosis Related Group (DRG; in Dutch, DBC) code. Inclusion criteria were; 1)

unilateral ankle fracture; 2) age of 18 years or older; and 3) provision of informed consent by the patient. Treatment should have been started between six weeks and three months and/or between seven and nine months prior to the start of the study. Exclusion criteria were; 1) multiple trauma (only if functional recovery of additional injuries was not achieved at time of enrolment, as that likely affects the outcome scores); 2) pathological fracture; 3) severe physical comorbidity (*i.e.*, American Society of Anaesthesiologists (ASA) ≥3); 4) patient was non-ambulatory prior to the injury; 5) insufficient comprehension of the Dutch language to understand and complete the questionnaires; and 6) expected problems of maintaining follow-up.

The AOFAS Ankle-Hindfoot Score-DLV, the Foot Function Index (FFI-DLV), and the Short Form Health Survey (SF-36-DLV) questionnaires could be completed in total on three occasions: at 2 months (t=1), 7 months (t=2), and 7.5 months (t=3) after trauma. The time between the recordings was 5-6 months (responsiveness, t=1 and t=2) and/or 2-3 weeks (test-retest, t=2 and t=3) in between. Patients were allowed to participate in both the responsiveness and test-retest part, and if so, the questionnaires at t=2 were also used as first questionnaire for test-retest reliability.

Questionnaires and data collection

The FFI is a scoring system developed to measure the impact of foot pathology. It consists of 23 items, which are grouped into the subscales pain, difficulty, and activity limitation. Scores for all (sub)scales range from zero (no disability) to 100 (highest level of disability) (10).

The SF-36 Health Survey is a generic measure of health status (11-18). It consists of 36 items, representing eight domains that are grouped into a Physical Component Summary (PCS) and a Mental Component Summary (MCS).

A research physician or research assistant performed the physical examination that is part of the physician-reported part of the AOFAS Ankle-Hindfoot Score-DLV using a standardized protocol. Patients completed the patient-reported part, as well as the FFI and SF-36. Demographic, injury and treatment data were collected from the patient's medical files.

Statistical analysis

Statistical analyses were performed using the Statistical Package for Social Sciences (SPSS, version 21). Data are reported following the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) (19). Since raw data for individual items were analyzed, missing data were not imputed. Descriptive statistics was used in order to describe the main characteristics of the study participants and the questionnaire scores at the different time points. Measurement properties of the AOFAS-DLV (sub)scales were determined by comparing these (sub)scales with the FFI and SF-36 (sub)scales. They were determined in compliance with the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines (20). A detailed description of the measurement properties and statistical analysis is shown in the published study protocol (8). A summary is given below.

Floor and ceiling effects are present if more than 15% of the study population rates the lowest or highest possible score (8, 21, 22). Data for each time point were evaluated separately.

Internal consistency (measure of reliability) was considered adequate if the Cronbach's alpha value is between 0.70 and 0.95, provided that the scale is unidimensional (21). For reasons of heterogeneity in scores, data for t=1 were used.

Construct validity was assessed by determining the correlation of the AOFAS-DLV (sub)scales with (sub)scales of the FFI and SF-36. Spearman's Rho (rank correlation)

coefficients (*r*) were calculated since data were non-parametric. Data of t=1 were used. Strength of correlation was categorized as high ($r > 0.6$), moderate ($0.3 < r < 0.6$), or low ($r < 0.3$) (23). Construct validity was considered adequate if at least 75% of the results were in line with the predefined hypotheses in a (sub)sample of at least 50 patients (21). Expected correlations are given in Supplemental Table 1.

Evaluation of the test-retest reliability was performed by calculating the intraclass correlation coefficient ($ICC_{agreement}$) of (sub)scales administered at t=2 and t=3. ICC is reported with 95% confidence interval (CI). Reliability was given a positive rating when the ICC is at least 0.70 in a sample size with a minimum of 50 patients (21).

The degree of absolute agreement was expressed as the standard error of measurement ($SEM_{agreement}$). For individual patients, the smallest detectable change (SDC) was calculated as $1.96 \times \sqrt{2} \times SEM$ (21). The SDC measurable in a group of people (SDC_{group}) was calculated by dividing the SDC in individuals (SDC_{ind}) by $\sqrt{n}$ (24, 25). Finally, the reliable change index (RCI) was calculated, representing the SDC as a percentage of the maximum obtainable score.

The degree of absolute agreement was also determined with a Bland and Altman analysis (26). The limits of agreement equal the mean change in scores of repeated measurements ($mean_{change}$) ± 1.96 x standard deviation of these changes (SD_{change}) (21). Zero falling outside this interval indicates bias in the measurements.

Analogous to construct validity, longitudinal validity (a measure of responsiveness) was assessed by testing predefined hypotheses (Supplement Table 1B) about expected correlations between changes in AOFAS Ankle-Hindfoot scale-DLV (sub)scales versus changes in FFI and SF-36 (sub)scales (21). Change scores were calculated from t=1 to t=2. Since data were non-parametric, Spearman's rank correlation coefficients were calculated.

Longitudinal validity was considered adequate if at least 75% of the results were in line with the predefined hypotheses in a (sub)sample of at least 50 patients (21).

The effect size (ES) and standardized response mean (SRM) were determined as measures of the magnitude of change over time, using the data of t=1 and t=2. ES was calculated as change in score (t=2 - t=1)/SD_{T1} (21). SRM was calculated as change in score (t=2 – t=1)/SD_{change} (21). Values of 0.2-0.4 were considered a small effect, 0.5-0.7 a moderate, and 0.8 or higher a large effect (27). Large effect sizes were expected a priori, since at t=1 patients were expected to have functional limitations, whereas at t=2 full recovery was expected for most patients.

RESULTS

In total 142 individual participants were included, 70 completed t=1 and t=2, 132 completed t=2 and t=3 (Figure 1). During the course of the study ten patients were lost to follow up. One patient, who participated in the test-retest part, had to be removed from the analysis; due to removal of osteosynthesis material, the patient reported a change in function between both recordings.

The median age was 46 years (P_{25} -P_{75} 35-60), see Table 1. The majority of patients (N=75; 52.8%) were male. Most ankle fractures were unimalleolar (N=100; 70.4%), and the majority (N=84; 59.2%) were treated operatively.

The changes over time in AOFAS-total, FFI-total, SF-36 PCS and SF-36 MCS are shown in Figure 2. The AOFAS and SF-36 (PCS and MCS) show an increase in scores in the period from t=1 to t=2. The FFI, focusing on disabilities rather than function, shows a decrease in score. Scores at t=2 and t=3 were similar for all instruments.

Floor and ceiling effects

A floor effect was only present in two SF-36 subscales; namely SF-36 RP subscale at t=1; 58.6% of the patients reported the minimum score, at t=2 (19.7%) and t=3 (17.6%), and the SF-36 RE subscale at t=1 (28.6%); Figure 3a).

A ceiling effect was present in several (sub)scales, and became more evident at longer follow-up (Figure 3b). The AOFAS pain subscale had a ceiling effect from the t=1 onwards, where 22.9% of patients reported the maximum score. From t=2 onwards, ceiling effects were also noted for AOFAS function (27.0%) and alignment (65.9%) subscales, FFI pain (16.7%) and limitation (21.0%) subscales, and SF-36 BP (21.9%) and PF (19.5%) subscales.

The AOFAS as a total scale only showed a ceiling effect at t=3; 17.7% of patients reported the maximum score.

Reliability

Internal consistency

The Cronbach's alpha for the AOFAS total scale and function subscale were 0.947 and 0.927, respectively, representing adequate internal consistency (Table 2). The value for the total scale should be interpreted carefully as it contains three subscales. Cronbach's alpha could not be calculated for AOFAS pain and alignment subscales, since these have one item only.

The FFI total scale (α = 0.649) and pain subscale (α = 0.687) did not show adequate internal consistency. For the total scale, this may be explained by the fact that it is not unidimensional. All SF-36 (sub)scales showed adequate internal consistency, with the exception of the subscales general health (α = 0.621) and vitality (α = 0.648).

Construct validity

Spearman's rank correlations regarding construct validity are shown in Table 3. Construct validity was adequate for all AOFAS (sub)scales; out of 17 correlations, 14 (82.4%) were in line with predefined hypotheses for the total scale, 13 (76.5%) for the pain subscale, 15 (88.2%) for the function subscale, and 16 (94.1%) for the alignment subscale.

Reproducibility

Test-Retest reliability

The intraclass correlation coefficient indicates the reliability of each (sub)scale (Table 4). The calculated ICC for the total AOFAS (sub)scales ranged from 0.85 to 0.93, indicating adequate

test-retest reliability. The ICC was also proven to be adequate (> 0.70) for all FFI and SF-36 (sub)scales, with the exception of SF-36 subscale General Health perceptions (ICC = 0.64).

Agreement and Smallest Detectable Change

The level of agreement is indicated by the SDC and the corresponding RCI, as listed in Table 4. The SDC was 12.0 (RCI: 12.0%) for the AOFAS total scale, 16.4 (RCI: 16.4%) for the FFI total scale, 10.7 (RCI: 15.3%) for the SF-36 PCS subscale, and 11.36 (RCI: 14.6%) for the SF-36 MCS subscale.

The Bland and Altman analysis (Figure 4 and Table 4) there is no bias in measurements, as the 95% Limits of Agreement for the mean change in scores contains zero for every single (sub)scale.

Responsiveness

Spearman's rank correlation coefficients for longitudinal validity are shown in Table 5. Longitudinal validity was adequate for all AOFAS (sub)scales; out of 17 correlations, 15 (88.2%) were in line with predefined hypotheses for the total scale, 14 (82.5%) for the AOFAS pain subscale, 13 (76.5%) for function subscale, and 17 (100%) for alignment subscale.

The Standardized Response Mean (SRM) and the Effect Size (ES) of the instruments are presented in Table 6. The AOFAS total scale (SRM 1.07, ES 0.89) and function subscale (SRM 1.29, ES 1.06) had a large magnitude of change. The one-item subscales showed a moderate effect size for pain (SRM 0.27) and a small effect size for alignment (SRM < 0.2).

DISCUSSION

The results of this study showed that the AOFAS Ankle-Hindfoot scale Dutch Language Version (AOFAS-DLV) is a valid, reliable, and responsive instrument for measuring symptoms and disability in patients who suffered an ankle fracture.

Floor effects were not present for the AOFAS-DLV in this study. Ceiling effects, on the other hand, did occur. The AOFAS total scale showed a ceiling effect at t=3. Ceiling effects were expected to occur at follow-up moments t=2 and t=3, as most patients were expected to have achieved full recovery (and thus the maximum score) at those follow-up moments. Ceiling effects have been reported in another study for the same reason (13). The AOFAS subscales pain and alignment also showed a ceiling effect at t=1. As the AOFAS scale has never been evaluated by its independent subscales, this is a new finding. It might be explained by the fact that also minor ankle fractures were enrolled in this study.

Evaluating the predictions about Spearman's rank correlations between all (sub)scales, the AOFAS scale as a total showed adequate construct validity. This is in correspondence with previous research, conducted by Ibrahim *et al*. (7). Construct validity also showed to be adequate for all AOFAS subscales separately. The correlations between the AOFAS total score and the SF-36 did show to be higher than the correlations found by SooHoo *et al.* (28). Instead of a high correlation, they found the SF-36 subscales bodily pain, and physical functioning to have a moderate correlation with the AOFAS total scale. The difference in correlation was even bigger for the SF-36 PCS, which SooHoo *et al.* found to have a low, instead of a high correlation with the AOFAS total scale in this study (28). A possible explanation for these differences is the difference in study population, as this study only focused on ankle fractures and SooHoo *et al.* included all injuries of the ankle and hindfoot (28).

As far as conclusions can be drawn, the AOFAS Ankle-Hindfoot scale DLV appears to have adequate internal consistency. Cronbach's α for the AOFAS-scale as a total is 0.947. This value however, should be interpreted carefully as this scale is not unidimensional. Pinsker *et al.* also did find Cronbach's α to be adequate (α = 0.84) for the five patient-reported items of the AOFAS Ankle-Hindfoot scale in the original language (29).

The reliability of the AOFAS DLV is proven to be sufficient, as the ICC for the total AOFAS scale was 0.93. The separate subscales also showed to be reliable on an independent level, with ICC of > 0.70 for all AOFAS subscales. Validation studies for the Portuguese and Turkish version of the AOFAS Ankle-Hindfoot scale found similar ICC values of 0.92 ($p < 0.001$) and 0.89 ($p= 0.001$), respectively (30, 31).

Responsiveness of the AOFAS-DLV, considered being a product of longitudinal validity and magnitude of change, was adequate in this study. Concerning longitudinal validity, > 75% of all hypothesized correlations for Spearman's Rho were confirmed, indicating adequate longitudinal validity. Magnitude of change for the outcome measures was high for the AOFAS Ankle-Hindfoot scale DLV as a whole, with an SRM of 1.07 and ES of 0.89. This is comparable to the magnitude of change for the total FFI (SRM -0.93, ES -0.74) and the SF-36 subscales with the highest magnitude of change (PCS, PF, RP and SF) in our study. Values for SRM and ES of the AOFAS-DLV found in this study are in correspondence with the values found in previous research by SooHoo *et al.* (5), regarding the original AOFAS Ankle-Hindfoot scale (SRM 1.10, ES 1.12).

The level of agreement of the AOFAS total scale compared well to the FFI and SF-36 in this study. The SEM for the AOFAS-DLV was 4.3 points. The SDC was 12.0 points. Similar values for SEM and SDC were found in the validation study of the AOFAS Ankle-Hindfoot score in Turkish (SEM, 4.8 points and SDC 13.3 points) (30).

The Bland and Altman analysis showed there is no bias in measurements, as the 95% Limits of Agreement for the mean change in scores contained zero for every single (sub)scale. As the AOFAS-DLV shows sufficient reliability and the level of agreement is equivalent to the level of agreement of the SF-36 and FFI (which are both validated PROMs), the reproducibility of the questionnaire is proven to be acceptable.

CONCLUSION

This study evaluated the measurement properties of the AOFAS Ankle-Hindfoot scale Dutch Language Version and confirmed it is a reliable, valid, and responsive measurement instrument for evaluating functional outcome in Dutch patients with a unilateral ankle fracture. This makes the questionnaire suitable for comparing outcome in future studies and after different treatment modalities within this study population or for comparing outcome across hospitals or between patient groups.

ETHICS

This study has been exempted by the medical research ethics committee (MREC) Erasmus MC (Rotterdam, The Netherlands). Each participant provided written consent to participate and remained anonymized during the study. The study is registered at the Netherlands Trial Register (NTR5613; 05-jan-2016).

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. There are no statements to declare relating contributorship, data sharing, or ethics approval.

CONTRIBUTORSHIP STATEMENT

EMMVL, ASDB, DEM, CHVDV, PTDH, WET, and MJHV developed the study. ASDB and EMMVL drafted the manuscript. EMMVL acted as trial principal investigator. ASDB, RJCT, FVDS, CHVDV, PTDH, DEM, and MHJV participated in patient inclusion and outcome assessment. ASDB, WET, and EMMVL performed statistical analysis of the study data. All authors have read and approved the final manuscript.

DATA SHARING STATEMENT

All data is processed in this manuscript. There are no further unpublished data from this study available.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing interests.

LIST OF ABBREVIATIONS

AOFAS, American Orthopaedic Foot and Ankle Society; BP, bodily pain; ES, effect size; FFI, Foot Function Index; GH, general health perceptions; ICC, intraclass correlation; MCS, mental component summary; MH, general mental health; PCS, physical component summary; PF, physical functioning; RCI, reliable change index; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SDC, smallest detectable change; SEM, standard error of measurement; SF, social functioning; SF-36, Short Form-36; SRM, standardized response mean; VT, vitality, energy, or fatigue.

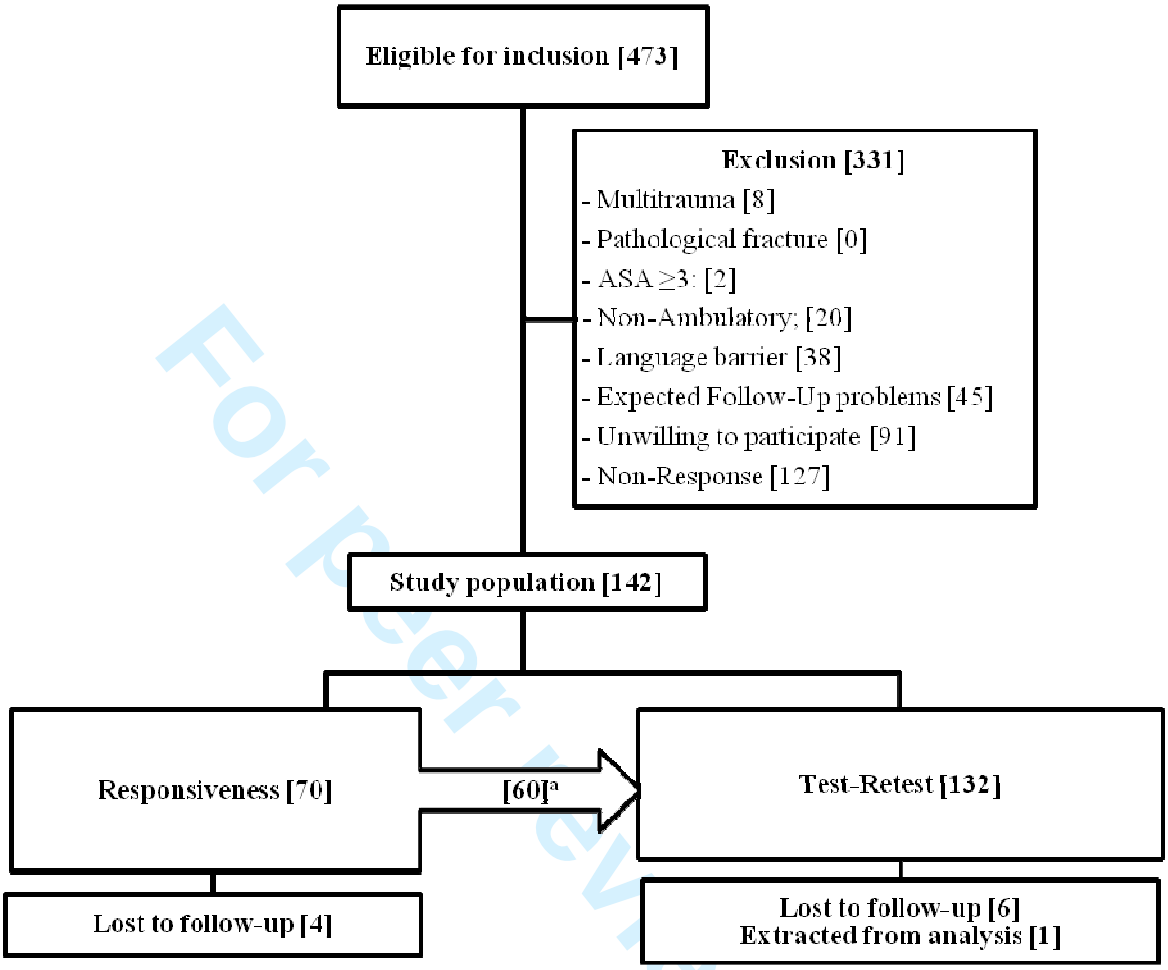
REFERENCES

1. Daly PJ, Fitzgerald RH, Jr., Melton LJ, Ilstrup DM. Epidemiology of ankle fractures in Rochester, Minnesota. Acta Orthop Scand. 1987;58(5):539-44.
2. Schneider W, Jurenitsch S. Normative data for the American Orthopedic Foot and Ankle Society ankle-hindfoot, midfoot, hallux and lesser toes clinical rating system. Int Orthop. 2016;40(2):301-6.
3. Kitaoka HB, Alexander IJ, Adelaar RS, Nunley JA, Myerson MS, Sanders M. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. Foot Ankle Int. 1994;15(7):349-53.
4. Madeley NJ, Wing KJ, Topliss C, Penner MJ, Glazebrook MA, Younger AS. Responsiveness and validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle Hindfoot Score, and Foot Function Index in end stage ankle arthritis. Foot Ankle Int. 2012;33(1):57-63.
5. SooHoo NF, Vyas R, Samimi D. Responsiveness of the foot function index, AOFAS clinical rating systems, and SF-36 after foot and ankle surgery. Foot Ankle Int. 2006;27(11):930-4.
6. Westphal T, Piatek S, Halm JP, Schubert S, Winckler S. Outcome of surgically treated intraarticular calcaneus fractures--SF-36 compared with AOFAS and MFS. Acta Orthop Scand. 2004;75(6):750-5.
7. Ibrahim T, Beiri A, Azzabi M, Best AJ, Taylor GJ, Menon DK. Reliability and validity of the subjective component of the American Orthopaedic Foot and Ankle Society clinical rating scales. J Foot Ankle Surg. 2007;46(2):65-74.
8. Van Lieshout EMM, De Boer AS, Meuffels DE, Den Hoed PT, Van Der Vlies CH, Tuinebreijer WE, et al. The American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Score; Study Protocol for the Translation and Validation of the Dutch Language Version. BMJ Open 2016:In press.
9. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186-91.
10. Budiman-Mak E, Conrad KJ, Roach KE. The Foot Function Index: a measure of foot pain and disability. J Clin Epidemiol. 1991;44(6):561-70.

11. de Haan J, den Hartog D, Tuinebreijer WE, Iordens GI, Breederveld RS, Bronkhorst MW, et al. Functional treatment versus plaster for simple elbow dislocations (FuncSiE): a randomized trial. BMC Musculoskelet Disord. 2010;11:263.
12. Den Hartog D, Van Lieshout EM, Tuinebreijer WE, Polinder S, Van Beeck EF, Breederveld RS, et al. Primary hemiarthroplasty versus conservative treatment for comminuted fractures of the proximal humerus in the elderly (ProCon): a multicenter randomized controlled trial. BMC Musculoskelet Disord. 2010;11:97.
13. Mahabier KC, Den Hartog D, Theyskens N, Verhofstad MHJ, Van Lieshout EMM, Hummer Trial Investigators. Reliability, validity, responsiveness, and minimal important change of the Disablities of the Arm, Shoulder and Hand and Constant-Murley scores in patients with a humeral shaft fracture. J Shoulder Elbow Surg. 2016.
14. Obremskey WT, Dirschl DR, Crowther JD, Craig WL, 3rd, Driver RE, LeCroy CM. Change over time of SF-36 functional outcomes for operatively treated unstable ankle fractures. J Orthop Trauma. 2002;16(1):30-3.
15. Pan SL, Liang HW, Hou WH, Yeh TS. Responsiveness of SF-36 and Lower Extremity Functional Scale for assessing outcomes in traumatic injuries of lower extremities. Injury. 2014;45(11):1759-63.
16. Schep NW, De Haan J, Iordens GI, Tuinebreijer WE, Bronkhorst MW, De Vries MR, et al. A hinged external fixator for complex elbow dislocations: a multicenter prospective cohort study. BMC Musculoskelet Disord. 2011;12:130.
17. van Tetering EA, Buckley RE. Functional outcome (SF-36) of patients with displaced calcaneal fractures compared to SF-36 normative data. Foot Ankle Int. 2004;25(10):733-8.
18. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol. 1998;51(11):903-12.
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9.
20. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49.

21. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.
22. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4(4):293-307.
23. Hinkle DE WW, Jurs SG. Applied statistics for the behavioral sciences. Boston USA: Houghton MiZin. 1998.
24. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care. 2001;17(4):479-87.
25. de Boer MR, de Vet HC, Terwee CB, Moll AC, Volker-Dieben HJ, van Rens GH. Changes to the subscales of two vision-related quality of life questionnaires are proposed. J Clin Epidemiol. 2005;58(12):1260-8.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
27. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale (NJ): Lawrence Erlbaum Associates. 1998.
28. SooHoo NF, Shuler M, Fleming LL, American Orthopaedic F, Ankle S. Evaluation of the validity of the AOFAS Clinical Rating Systems by correlation to the SF-36. Foot Ankle Int. 2003;24(1):50-5.
29. Pinsker E, Inrig T, Daniels TR, Warmington K, Beaton DE. Reliability and validity of 6 measures of pain, function, and disability for ankle arthroplasty and arthrodesis. Foot Ankle Int. 2015;36(6):617-25.
30. Analay Akbaba Y, Celik D, Ogut RT. Translation, Cross-Cultural Adaptation, Reliability, and Validity of Turkish Version of the American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Scale. J Foot Ankle Surg. 2016;55(6):1139-42.
31. Rodrigues RC MD, Mizusaki JM, Imoto AM, Peccin MS, Cohen M, Alloza JFM. Translation, cultural adaptation and validity of the American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Scale. Acta Ortop Bras 2008;16:107–11.

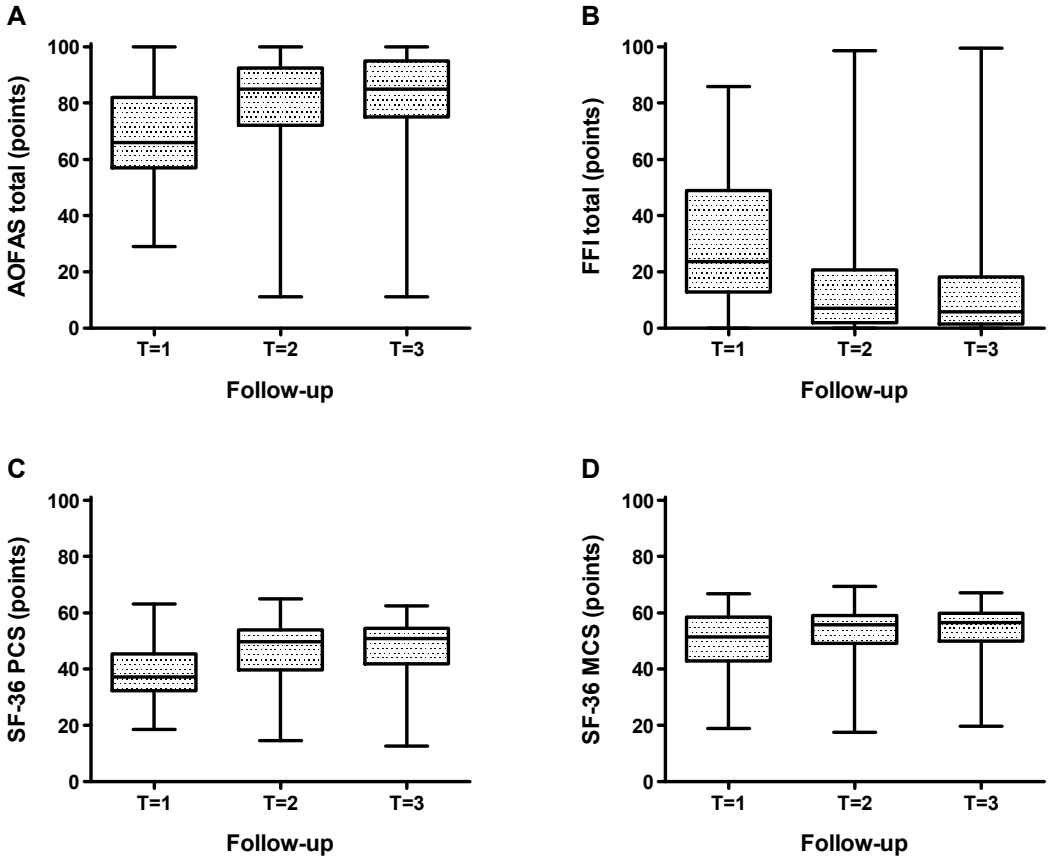
Figure 1: Flowchart



The number of patients in each particular group is shown between square brackets.

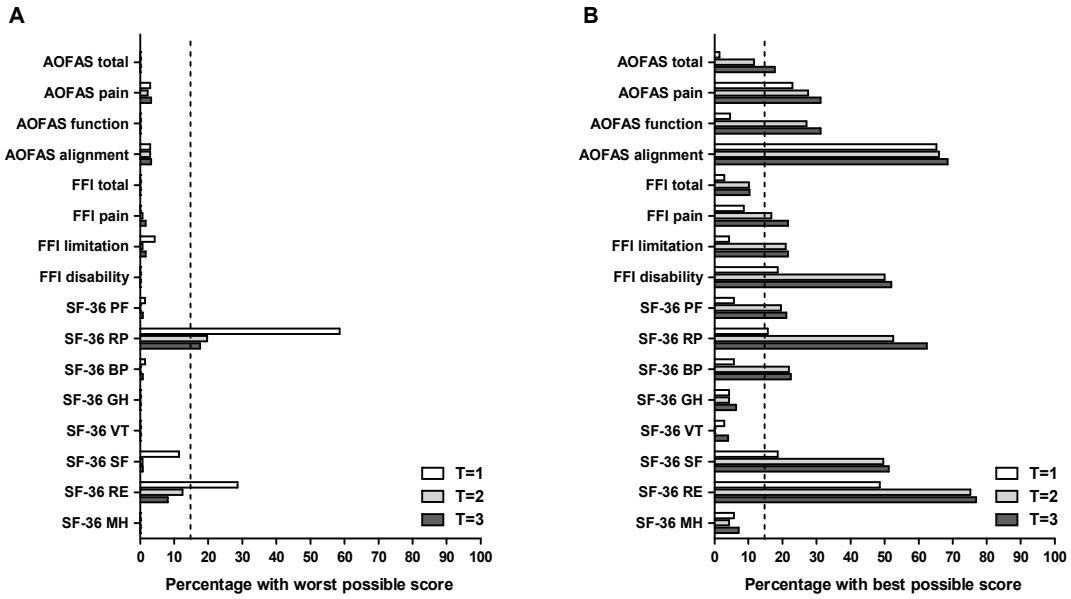
[a] Patients who participated in both groups

Figure 2: AOFAS Ankle-Hindfoot (A), Foot Function Index (B), Short Form-36 PCS (C), and SF-36 MCS (D) scores at each follow-up visit in patients with an ankle fracture



AOFAS, American Orthopaedic Foot and Ankle Society; FFI, Foot Function Index; MCS, Mental Component Summary; PCS, Physical Component Summary; SF-36, Short Form-36.

Figure 3. Floor effects (A) and ceiling effects (B) of the instruments used in patients with an ankle fracture



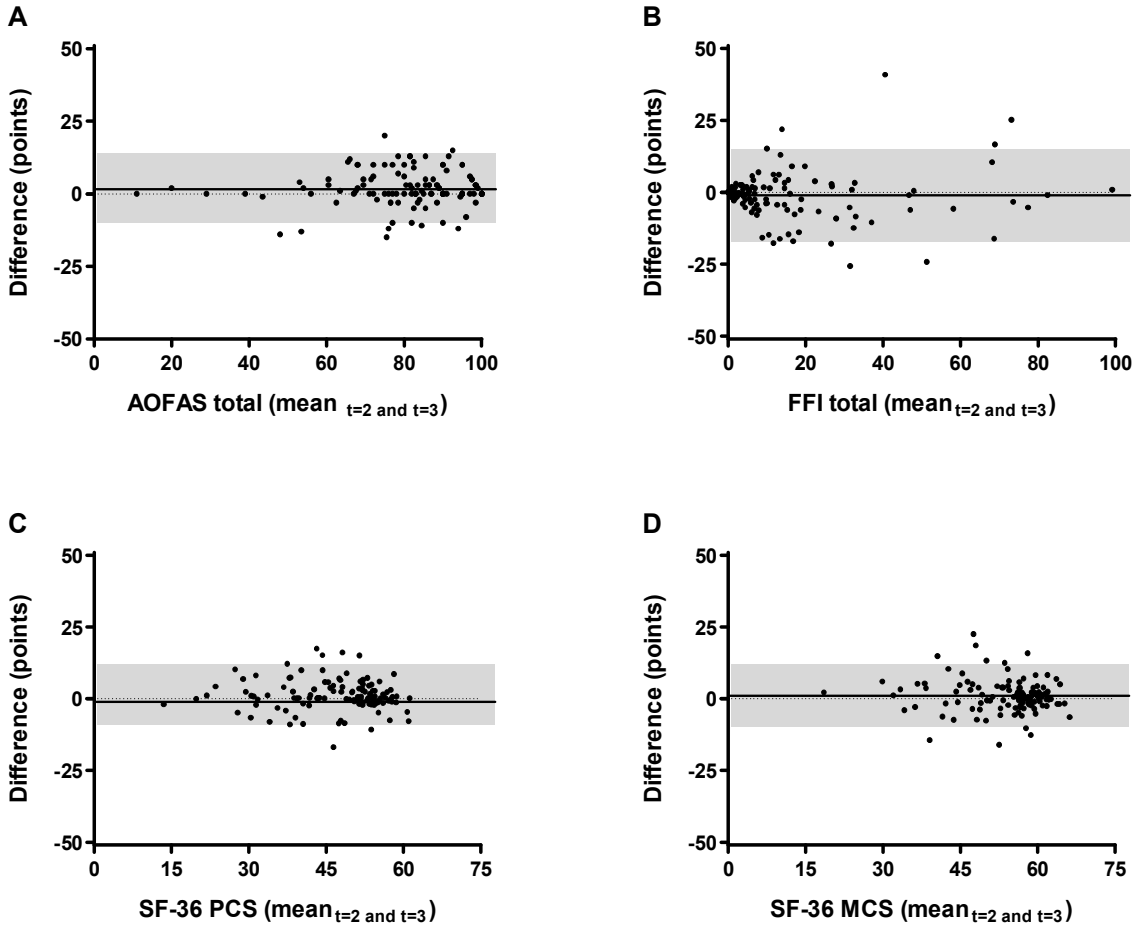
Out of a maximum of 70 at t=1, N=65 for AOFAS function and total, N=69 for AOFAS alignment, and N=70 for AOFAS pain and all (sub)scales of FFI and SF-36.

Out of a maximum of 138 at t=2, N=131 for SF-36 PCS and MCS, N=133 for SF-36 PF, N=136 for SF-36 VT, N=137 for AOFAS function, AOFAS total, and SF-36 RP, BP, SF, and RE, N=138 for AOFAS pain and alignment, all FFI (sub)scales, and SF-36 GH and MH N=138 for AOFAS pain and alignment, 137 for AOFAS function and AOFAS total.

Out of a maximum of 125 at t=3, N=123 for SF-36 PF, PCS, and MCS, N=124 for AOFAS alignment and total, and SF-36 VT, and N=125 for AOFAS pain and function, all FFI (sub)scales, and SF-36 RP, BP, GH, SF, RE, and MH.

The dotted line represents the acceptable 15% of patients with the maximum score. The SF-36 PCS and MCS did not demonstrate a floor or a ceiling effect and are not displayed.

Figure 4. Bland-Altman plots for AOFAS Ankle-Hindfoot (A), Foot Function Index (B), Short Form-36 PCS (C), and SF-36 MCS (D) scores in patients with an ankle fracture



Change scores were calculated from t=2 to t=3.

Each dot represents a single patient. The black line indicates the mean difference. The upper and lower edges of the grey box are the 95% limits of agreement.

AOFAS, American Orthopaedic Foot and Ankle Society; FFI, Foot Function Index; MCS, Mental Component Summary; PCS, Physical Component Summary; SF-36, Short Form-36.

Table 1: Demographic data for the study population

Variable		Outcome
Age (years)		46 (35- 60)
Male gender		75 (52.8%)
Right side affected		58 (40.8%)
Dominant side affected		60 (42.3%)
Malleolar involvement	Unimalleolar	100 (70.4%)
	Bimalleolar	23 (16.2%)
	Trimalleolar	19 (13.4%)
Classification	Weber A	29 (20.4%)
	Weber B	56 (39.4%)
	Weber C	13 (9.2%)
	Unknown	44 (31.0%)
Open fracture		6 (4.2%)
Treatment	Nonoperative	58 (40.8%)
	Operative	84 (59.2%)

Data are shown as median (P_{25}-P_{75}) or as N (%), as applicable.

Table 2: Internal consistency of the instruments used in patients with an ankle fracture

(Sub)scale		N	Number of items	Cronbach's alpha
AOFAS	Total	70	9	0.947[a]
	Pain	70	1	N.A.[b]
	Function	70	7	0.927
	Alignment	70	1	N.A.[b]
FFI	Total	70	23	**0.649**[a]
	Pain	70	9	**0.687**
	Limitation	70	9	0.707
	Disability	70	5	0.854
SF-36	Total	70	35	0.882[a]
	PF	70	10	0.932
	RP	70	4	0.885
	BP	70	2	0.733
	GH	70	5	**0.621**
	VT	70	4	**0.648**
	SF	70	2	0.832
	RE	70	3	0.870
	MH	70	5	0.799
	PCS	70	21	0.846[a]
	MCS	70	14	0.861[a]

Data for t=1 were used.

[a] Values should be interpreted carefully because the total scale is not unidimensional.

[b] Not applicable, as this subscale consists of one item only.

Bold and underlined Cronbach alpha values did not exceed the threshold of 0.70.

AOFAS, American Orthopaedic Foot and Ankle Society; BP, bodily pain; FFI, Foot Function Index; GH, general health perceptions; MCS, mental component summary; MH, general mental health; N.A., not applicable; PCS, physical component summary; PF, physical

functioning; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SF, social functioning; SF-36, Short Form-36; VT, vitality, energy, or fatigue.

Table 3. Construct validity of the instruments in patients with an ankle fracture

(Sub)scale		AOFAS			
		Pain	**Function**	**Alignment**	**Total**
AOFAS	Pain	1	**0.23 [65]**	0.01 [69]	0.66 [65]
	Function	**0.23 [65]**	1	0.28 [65]	0.85 [65]
	Alignment	0.01 [69]	0.28 [65]	1	**0.35 [65]**
	Total	0.66 [65]	0.85 [65]	**0.35 [65]**	1
FFI	Pain	-0.81 [70]	-0.41 [65]	-0.14 [69]	-0.70 [65]
	Limitation	-0.41 [70]	-0.75 [65]	-0.19 [69]	-0.74 [65]
	Disability	-0.34 [70]	-0.80 [65]	-0.23 [69]	-0.77 [65]
	Total	-055 [70]	-0.73 [65]	-0.21 [69]	-0.80 [65]
SF-36	PF	**0.21 [70]**	0.64 [65]	0.21 [69]	0.60 [65]
	RP	0.32 [70]	0.50 [65]	0.19 [69]	**0.58 [65]**
	BP	**0.59 [70]**	0.53 [65]	0.03 [69]	0.67 [65]
	GH	0.15 [70]	-0.01 [65]	-0.09 [69]	0.04 [65]
	VT	0.28 [70]	0.19 [65]	-0.02 [69]	**0.27 [65]**
	SF	0.14 [70]	**0.65 [65]**	0.18 [69]	0.56 [65]
	RE	**0.10 [70]**	0.32 [65]	0.22 [69]	0.33 [65]
	MH	0.24 [70]	0.20 [65]	0.02 [69]	0.24 [65]
	PCS	0.40 [70]	0.62 [65]	0.11 [69]	0.65 [65]
	MCS	0.11 [70]	0.24 [65]	0.13 [69]	0.24 [65]

Data for t=1 were used. Spearman's rank correlation coefficients are given for all possible combinations of (sub)scales, with the N between square brackets. The maximum possible number of patients was 70.

r > 0.6 indicates high correlation, 0.3 < r > 0.6 moderate correlation, and r < 0.3 low correlation. Bold and underlined correlations were not hypothesized correctly.

AOFAS, American Orthopaedic Foot and Ankle Society; BP, bodily pain; FFI, Foot Function Index; GH, general health perceptions; MCS, mental component summary; MH, general

mental health; PCS, physical component summary; PF, physical functioning; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SF, social functioning; SF-36, Short Form-36; VT, vitality, energy, or fatigue.

Table 4. Intraclass correlation coefficient (ICC) and Bland-Altman analysis of the instruments in ankle fracture patients

(Sub)scale		N	ICC(2,1) (95% CI)	SEM	SDC $_{patient}$	Max score	RCI (%)	Mean$_{difference}$ (SD)	95% Limits of agreement
AOFAS	Pain	125	0.85 (0.78-0.89)	3.5	9.7	40	24.3	1.1 (5.0)	-8.6 to 10.8
	Function	124	0.92 (0.89-0.95)	2.1	5.9	50	11.9	0.6 (3.0)	-5.4 to 6.5
	Alignment	124	0.89 (0.85-0.92)	0.9	2.5	10	24.8	0.2 (1.3)	-2.3 to 2.6
	Total	123	0.93 (0.89-0.95)	4.3	12.0	100	12.0	1.8 (6.1)	-10.2 to 13.9
FFI	Pain	125	0.83 (0.76-0.87)	9.4	26.1	100	26.1	-1.5 (13.3)	-27.6 to 24.6
	Limitation	125	0.90 (0.86-0.93)	7.4	20.5	100	20.5	-1.5 (10.5)	-22.0 to 19.0
	Disability	125	0.81 (0.74-0.86)	7.9	22.0	100	22.0	-0.2 (11.2)	-22.2 to 21.8
	Total	125	0.92 (0.89-0.94)	5.9	16.4	100	16.4	-1.2 (8.4)	-17.5 to 15.2
SF-36	PF	120	0.90 (0.87-0.93)	3.18	8.83	56.76	15.6	1.40 (4.50)	-8.43 to 9.23
	RP	124	0.71 (0.59-0.79)	6.36	17.64	55.56	31.7	2.56 (9.00)	-15.07 to 20.20
	BP	124	0.78 (0.70-0.85)	4.07	11.29	60.40	18.7	1.48 (5.76)	-9.80 to 12.77
	GH	125	**0.64 (0.52-0.73)**	5.12	14.20	63.78	22.3	-0.27 (7.24)	-14.47 to 13.93
	VT	123	0.77 (0.68-0.83)	4.06	11.25	68.66	16.4	0.74 (5.74)	-10.51 to 11.99
	SF	124	0.70 (0.60-0.78)	4.89	13.56	57.33	23.7	0.77 (6.92)	-12.79 to 14.32
	RE	124	0.72 (0.63-0.80)	5.31	14.71	55.66	26.4	0.90 (7.50)	-13.81 to 15.60
	MH	125	0.79 (0.70-0.85)	3.86	10.70	63.97	16.7	-1.21 (5.46)	-9.49 to 11.91
	PCS	118	0.85 (0.79-0.89)	3.87	10.72	70.30	15.3	1.10 (5.47)	-9.62 to 11.83
	MCS	118	0.78 (0.70-0.84)	4.10	11.36	77.92	14.6	0.96 (5.80)	-10.42 to 12.30

Change scores were calculated from t=2 to t=3. The maximum possible number of patients was 125. The ICC is shown as correlation coefficient with the 95% CI between brackets. The difference in score from t=2 to t=3 is shown as mean change with SD.

Table 5. Longitudinal validity of the instruments in patients with an ankle fracture

(Sub)scale		AOFAS			
		Pain	**Function**	**Alignment**	**Total**
AOFAS	Pain	1	0.21 [61]	0.12 [65]	0.70 [61]
	Function	0.21 [61]	1	0.05 [61]	0.81 [61]
	Alignment	0.12 [65]	0.05 [61]	1	0.22 [61]
	Total	0.70 [61]	0.81 [61]	0.22 [61]	1
FFI	Pain	**-0.56 [66]**	**-0.19 [61]**	-0.17 [65]	-0.43 [61]
	Limitation	-0.24 [66]	-0.66 [61]	-0.07 [65]	-0.60 [61]
	Disability	-0.06 [66]	**-0.59 [61]**	0.09 [65]	-0.50 [61]
	Total	**-0.33 [66]**	-0.61 [61]	-0.03 [65]	**-0.65 [61]**
SF-36	PF	0.25 [66]	**0.44 [61]**	-0.12 [65]	0.48 [61]
	RP	0.26 [65]	**0.34 [60]**	0.01 [64]	0.37 [60]
	BP	0.39 [65]	0.36 [60]	0.06 [64]	0.46 [60]
	GH	-0.02 [66]	-0.13 [61]	0.13 [65]	-0.05 [61]
	VT	**0.38 [66]**	0.26 [61]	0.10 [65]	0.38 [61]
	SF	0.20 [65]	0.54 [60]	0.03 [64]	0.47 [60]
	RE	-0.08 [65]	0.19 [60]	0.15 [64]	**0.14 [60]**
	MH	0.13 [66]	0.09 [61]	0.08 [65]	0.11 [61]
	PCS	0.34 [65]	0.39 [60]	-0.06 [64]	0.45 [60]
	MCS	-0.07 [65]	0.15 [60]	0.14 [64]	0.06 [60]

Change in scores between t=1 and t=2 were used. The maximum possible number of patients was 70. Spearman's rank correlation coefficients are given for all possible combinations of (sub)scales, with the N between square brackets.

The rest of Table caption is identical to Table 3.

Table 6. Responsiveness: standardized response mean (SRM) and Effect Size (ES) of the instruments in patients with an ankle fracture

(Sub)scale		N	Mean change	SD_{change}	SRM	$SD_{t=1}$	ES
AOFAS	Pain	66	2.3	8.4	0.27	8.9	0.26
	Function	61	12.3	9.5	1.29	11.5	1.06
	Alignment	65	-0.2	1.8	-0.09	2.7	-0.06
	Total	61	15.1	14.1	1.07	16.9	0.89
FFI	Pain	66	-9.1	18.7	-0.49	21.9	-0.42
	Limitation	66	-23.3	25.3	-0.92	29.9	-0.78
	Disability	66	-17.9	22.9	-0.78	27.1	-0.66
	Total	66	-17.6	18.9	-0.93	23.9	-0.74
SF-36	PF	66	9.04	10.94	0.83	12.98	0.70
	RP	65	11.95	13.25	0.90	10.94	1.09
	BP	65	7.85	10.33	0.76	9.50	0.83
	GH	66	-0.83	8.56	-0.10	8.42	-0.10
	VT	66	1.74	8.89	0.20	8.06	0.22
	SF	65	13.49	13.53	1.00	14.67	0.92
	RE	65	5.28	12.11	0.44	13.36	0.40
	MH	66	1.31	8.40	0.16	9.10	0.14
	PCS	65	8.88	10.03	0.89	9.65	0.92
	MCS	65	2.68	11.21	0.24	11.61	0.23

Change scores were calculated from t=1 to t=2. The maximum possible number of patients was 70.

AOFAS, American Orthopaedic Foot and Ankle Society; BP, bodily pain; ES, effect size; FFI, Foot Function Index; GH, general health perceptions; MCS, mental component summary; MH, general mental health; PCS, physical component summary; PF, physical functioning; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SF, social

functioning; SF-36, Short Form-36; SRM, standardized response mean; VT, vitality, energy, or fatigue.

SUPLEMENTAL MATERIAL

Supplemental Table 1A. Hypothesized correlations between the instruments for construct validity in patients with an ankle fracture.

(Sub)scale		AOFAS			
		Pain	**Function**	**Alignment**	**Total**
AOFAS	Pain	N.A.	moderate	low	high
	Function	moderate	N.A.	low	high
	Alignment	low	low	N.A.	low
	Total	high	high	low	N.A.
FFI	Pain	high	moderate	low	high
	Limitation	moderate	high	low	high
	Disability	moderate	high	low	high
	Total	moderate	high	low	high
SF-36	PF	moderate	high	low	high
	RP	moderate	moderate	low	high
	BP	high	moderate	low	high
	GH	low	low	low	low
	VT	low	low	low	moderate
	SF	low	moderate	low	moderate
	RE	moderate	moderate	low	moderate
	MH	low	low	low	low
	PCS	moderate	high	low	high
	MCS	low	low	low	low

Supplemental Table 1B. Hypothesized correlations between the instruments for longitudinal validity in patients with an ankle fracture.

(Sub)scale		AOFAS			
		Pain	**Function**	**Alignment**	**Total**
AOFAS	Pain	N.A.	low	low	high
	Function	low	N.A.	low	high
	Alignment	low	low	N.A.	low
	Total	high	high	low	N.A.
FFI	Pain	high	moderate	low	moderate
	Limitation	low	high	low	moderate
	Disability	low	high	low	moderate
	Total	low	high	low	moderate
SF-36	PF	low	high	low	moderate
	RP	low	low	low	moderate
	BP	moderate	moderate	low	moderate
	GH	low	low	low	low
	VT	low	low	low	moderate
	SF	low	moderate	low	moderate
	RE	low	low	low	moderate
	MH	low	low	low	low
	PCS	moderate	moderate	low	moderate
	MCS	low	low	low	low

Expected strength of correlation for all possible combinations; $r > 0.6$ indicates high correlation, $0.3 < r > 0.6$ moderate correlation, and $r > 0.6$ low correlation.

AOFAS, American Orthopaedic Foot and Ankle Society; BP, bodily pain; FFI, Foot Function Index; GH, general health perceptions; MCS, mental component summary; MH, general mental health; PCS, physical component summary; PF, physical functioning; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SF, social functioning; SF-36, Short Form-36; VT, vitality, energy, or fatigue.

Supplemental Table 2: AOFAS Ankle-Hindfoot Score Dutch Language Version

Pijn

- ☐ Geen
- ☐ Mild, af en toe
- ☐ Matig, dagelijks
- ☐ Ernstig, bijna altijd aanwezig

Functie

Beperkingen in activiteiten, hulpmiddelengebruik

- ☐ Geen beperkingen; geen hulpmiddelen nodig
- ☐ Geen beperkingen bij dagelijkse activiteiten, wel beperkingen bij recreatieve activiteiten; geen hulpmiddelen nodig
- ☐ Beperkingen bij dagelijkse en recreatieve activiteiten; gebruik van een stok
- ☐ Ernstige beperkingen bij dagelijkse en recreatieve activiteiten; gebruik van een brace, krukken, looprek, rollator of rolstoel

Maximale loopafstand

- ☐ Meer dan 600 meter
- ☐ 400 tot 600 meter
- ☐ 100 tot 400 meter
- ☐ Minder dan 100 meter

Loopondergrond

- ☐ Op geen enkele ondergrond problemen
- ☐ Enige moeite met lopen op oneffen terrein, trappen, hellingen of ladders
- ☐ Veel moeite met lopen op oneffen terrein, trappen, hellingen of ladders

Let op: onderstaande vragen worden door de arts ingevuld.

Afwijkende loopgang

- ☐ Geen tot gering
- ☐ Duidelijk
- ☐ Zeer opvallend

Sagittale beweging (dorsoflexie plus plantairflexie)

- ☐ Normaal of geringe beperking (30° of meer)
- ☐ Matige beperking (15-29°)
- ☐ Ernstige beperking (minder dan 15°)

Achtervoetbeweging (inversie plus eversie)

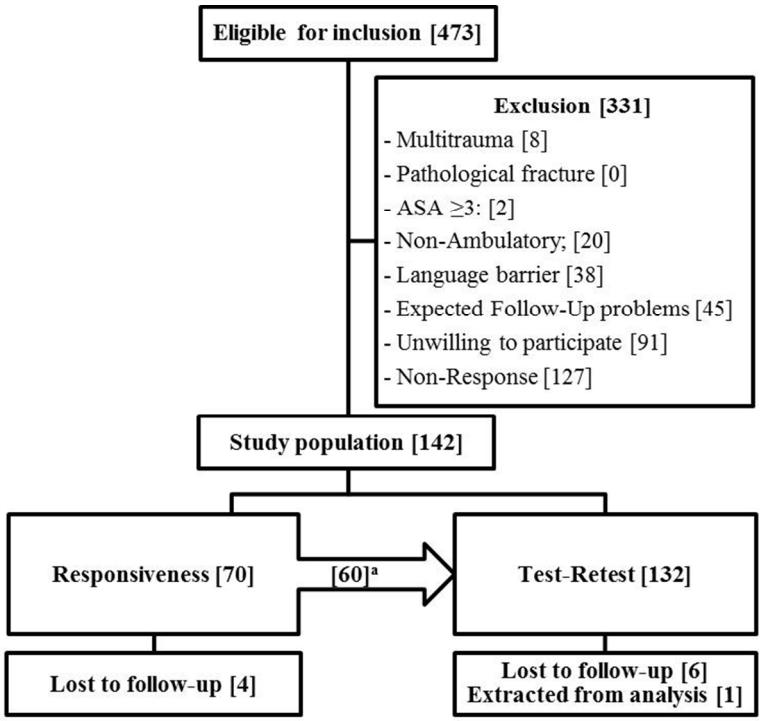
- ☐ Normaal of geringe beperking (75%-100% van normaal)
- ☐ Matige beperking (25-74% van normaal)
- ☐ Opvallende beperking (minder dan 25% van normaal)

Enkel-achtervoet stabiliteit (anteroposterieur, varus-valgus)

☐ Stabiel
☐ Evident instabiel

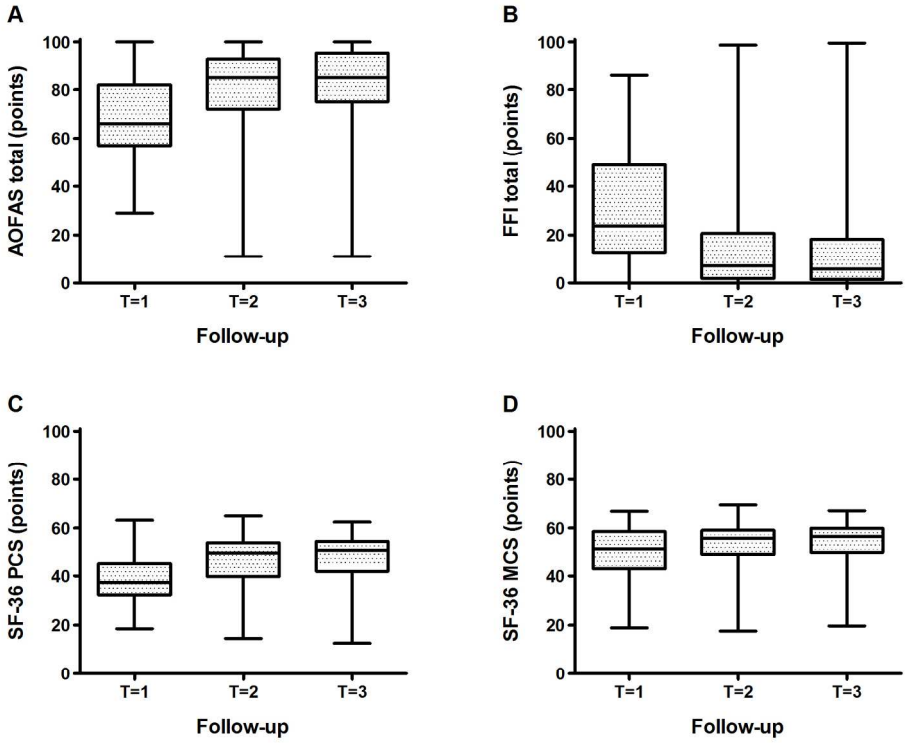
Alignement

☐ Goed, plantigrade voet, enkel-achtervoet fraai gealigneerd
☐ Redelijk, plantigrade voet, enige mate van enkel-achtervoet malalignement, geen klachten of symptomen
☐ Slecht, geen plantigrade voet, ernstige malalignement met klachten of symptomen



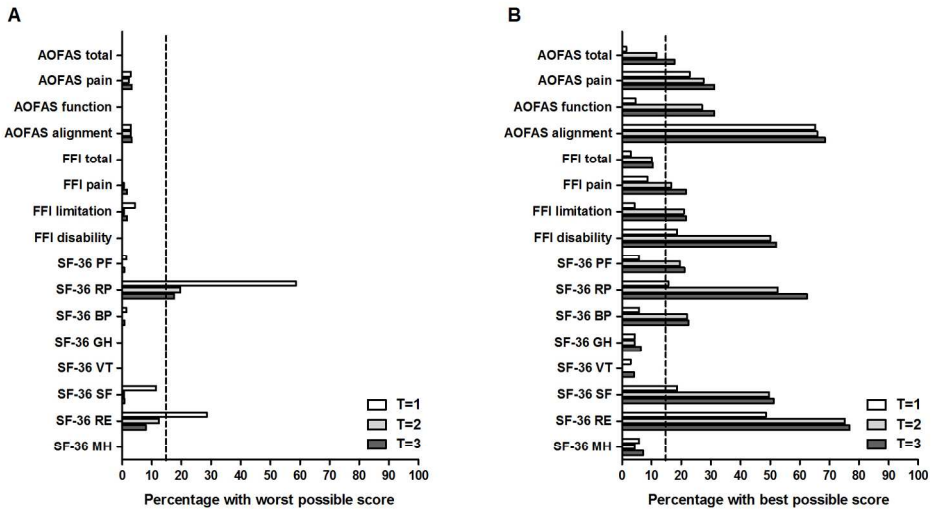
The number of patients in each particular group is shown between square brackets. a Patients who participated in both groups

164x135mm (150 x 150 DPI)



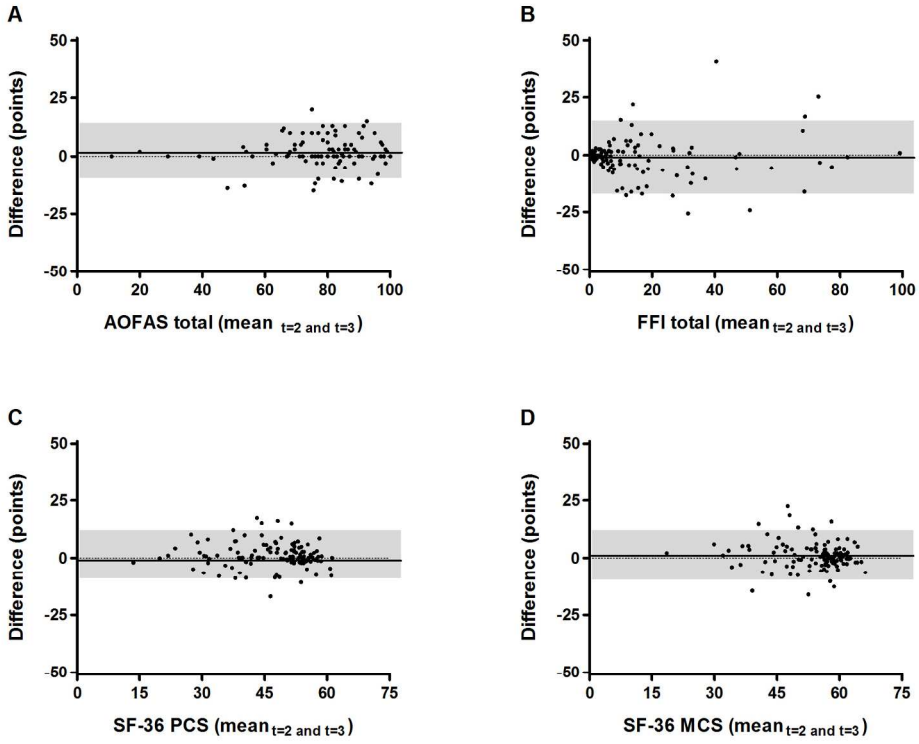
AOFAS, American Orthopaedic Foot and Ankle Society; FFI, Foot Function Index; MCS, Mental Component Summary; PCS, Physical Component Summary; SF-36, Short Form-36.

186x153mm (300 x 300 DPI)



Out of a maximum of 70 at t=1, N=65 for AOFAS function and total, N=69 for AOFAS alignment, and N=70 for AOFAS pain and all (sub)scales of FFI and SF-36.
Out of a maximum of 138 at t=2, N=131 for SF-36 PCS and MCS, N=133 for SF-36 PF, N=136 for SF-36 VT, N=137 for AOFAS function, AOFAS total, and SF-36 RP, BP, SF, and RE, N=138 for AOFAS pain and alignment, all FFI (sub)scales, and SF-36 GH and MH
N=138 for AOFAS pain and alignment, 137 for AOFAS function and AOFAS total.
Out of a maximum of 125 at t=3, N=123 for SF-36 PF, PCS, and MCS, N=124 for AOFAS alignment and total, and SF-36 VT, and N=125 for AOFAS pain and function, all FFI (sub)scales, and SF-36 RP, BP, GH, SF, RE, and MH.
The dotted line represents the acceptable 15% of patients with the maximum score. The SF-36 PCS and MCS did not demonstrate a floor or a ceiling effect and are not displayed.

178x101mm (300 x 300 DPI)



Change scores were calculated from t=2 to t=3.
Each dot represents a single patient. The black line indicates the mean difference. The upper and lower edges of the grey box are the 95% limits of agreement.
AOFAS, American Orthopaedic Foot and Ankle Society; FFI, Foot Function Index; MCS, Mental Component Summary; PCS, Physical Component Summary; SF-36, Short Form-36.

180x145mm (300 x 300 DPI)

The American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Score; Study Protocol for the Translation and Validation of the Dutch Language Version

Esther M.M. Van Lieshout MSC PhD[1A*], A. Siebe De Boer MD[1A], Duncan E. Meuffels MD PhD[2], P. Ted Den Hoed MD PhD[3], Cornelis H. Van der Vlies, MD PhD[4], Wim E. Tuinebreijer MD PhD[1], Michael H.J. Verhofstad MD PhD[1]

[1]Trauma Research Unit Department of Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

[2]Department of Orthopaedic Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

[3]Department of Surgery, Ikazia Hospital, P.O. Box 5009, 3008 AA Rotterdam, The Netherlands

[4]Department of Surgery, Maasstad Hospital, P.O. Box 9100, 3007 AC Rotterdam, The Netherlands

[A] Both authors contributed equally

[*] Address for correspondence:

Esther M.M. Van Lieshout, MSc PhD

Trauma Research Unit Department of Surgery

Erasmus MC, University Medical Center Rotterdam

P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Email: e.vanlieshout@erasmusmc.nl

Telephone: + 31 10 7031050

Email addresses:

EMMVL: e.vanlieshout@erasmusmc.nl

ASDB: a.deboer@erasmusmc.nl

DEM: d.meuffels@erasmusmc.nl

PTDH: pt.hoed@ikazia.nl

CHVDV: VliesC@maasstadziekenhuis.nl

WET: w.tuinebreijer@erasmusmc.nl

MHJV: m.verhofstad@erasmusmc.nl

Keywords: Ankle; Dislocation; Fracture; Hindfoot; Reliability; Responsiveness; Validity.

Word count: 5007

ABSTRACT

Introduction

The AOFAS Ankle-Hindfoot Score is among the most commonly used instrument for measuring outcome of treatment in patients who sustained a complex ankle or hindfoot injury. It combines a clinician-reported and a patient-reported part. A valid, Dutch version of this instrument is currently not available. Such a translated and validated instrument would allow objective comparison across hospitals or between patient groups, and with shown validity and reliability it may become a quality of care indicator in future. The main aims of this study are to translate and culturally adapt the AOFAS Ankle-Hindfoot Score questionnaire into Dutch according to international guidelines, and to evaluate the measurement properties of the AOFAS Ankle-Hindfoot Score-Dutch Language Version (DLV) in patients with a unilateral ankle or hindfoot fracture.

Methods and analysis

The design of the study will be a multicenter, prospective, observational study (case series) in patients who presented to the Emergency Department with a unilateral ankle or hindfoot fracture or (fracture) dislocation. A research physician or research assistant will complete the AOFAS Ankle-Hindfoot Score-DLV based upon interview for the subjective part and physical examination for the objective part. In addition, patients will be asked to complete the Foot Function Index (FFI) and the Short Form-36 (SF-36). Descriptive statistics (including floor and ceiling effects), internal consistency, construct validity, reproducibility (i.e., test-retest reliability, agreement, and smallest detectable change), and responsiveness will be assessed for the AOFAS DLV.

Ethics and dissemination

This study has been exempted by the medical research ethics committee (MREC) Erasmus MC (Rotterdam, The Netherlands). Each participant will provide written consent to participate and remain anonymized during the study. The results of the study are planned to be published in an international, peer-reviewed journal.

Registration details

The study is registered at the Netherlands Trial Register (NTR5613; 05-jan-2016).

ARTICLE SUMMARY

Strengths and limitations of this study:

- This study involves translation and validation of the AOFAS Ankle-Hindfoot Score into Dutch.
- It is a prospective, multicenter, observational study with a strong methodologic design.
- Statistical analyses will comply with the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines.
- The study is limited to adults (aged 18 years or older) who have adequate comprehension of the Dutch language.
- Although the study will be mostly relevant for the Dutch-speaking regions, it is also informative for other regions.

INTRODUCTION

Complex foot and ankle injuries cause a, usually temporary, loss of function and quality of life. Patient-Reported Outcome Measures (PROMs) are essential in both clinical practice and clinical research; they enable detailed evaluation of (functional) outcome or quality of life after (non-)operative treatment of musculoskeletal (traumatic) injuries from a patient's perspective. Generic instruments such as quality of life questionnaires allow comparison across populations with different injuries or medical conditions. Region-specific instruments, on the other hand, may give more detailed insight into the disabilities, pain, and problems caused by a specific injury. Some instruments are solely PROMs, and others combine a patient-reported with a physician-reported part. Numerous generic and region-specific instruments are available.[1-6]

A frequently used instrument for assessing outcome after ankle and hindfoot injuries is the American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Score. This clinical rating system, developed by Kitaoka *et al.*, combines subjective scores of pain and function provided by the patient with objective scores based on the surgeon's physical examination of the patient (to assess sagittal motion, hindfoot motion, ankle-hindfoot stability, and alignment of the ankle-hindfoot).[7] The scale includes nine items that can be divided into three subscales (pain, function, and alignment). Pain consists of one item with a maximal score of 40 points, indicating no pain. Function consists of seven items with a maximal score of 50 points, indicating full function. Alignment consists of one item with a maximal score of 10 points, indicating good alignment. The maximal score is 100 points, indicating no symptoms or impairments. In the original publication, the AOFAS Ankle-Hindfoot Score was described to be used for ankle replacement, ankle arthrodesis, ankle instability operations, subtalar arthrodesis, subtalar instability operations, talonavicular arthrodesis, calcaneocuboid arthrodesis, calcaneal osteotomy, calcaneus fracture, talus fracture, and ankle fractures.[7]

Evidence that the AOFAS Ankle-Hindfoot Score (as a complete scale) is valid in its original version, is limited.[7-9] Poor to moderate correlation of the AOFAS scores to the SF-36 subscales may also suggest poor construct validity.[10] Adequate responsiveness has been shown.[8, 9] The physician-reported part of the scale has been shown to be valid and reliable.[11] Westphal *et al.* showed correlations between SF-36 and the AOFAS Ankle-Hindfoot Score were strong regarding function and pain subscales, but moderate for all other subscales.[12] Previous studies involved a wide spectrum of diagnoses, such as general ankle-hindfoot complaints,[9] pending ankle or foot surgery,[11] surgically treated calcaneal fractures,[12] and end-stage ankle arthritis.[8] Some of these studies have included mixed populations.

Despite some favorable results, there is also criticism to the use of the AOFAS Clinical Rating Systems, which includes the AOFAS Ankle-Hindfoot Score.[13] Criticism, which includes the limited number of answers per item as well as linguistic issues, may negatively affect reliability and validity, and makes it more prone to ceiling effects.[13, 14] Despite these concerns, the AOFAS Ankle-Hindfoot Score remains among the most commonly used instruments, especially for patients with hindfoot fractures. It is especially an interesting instrument because it asks for hindfoot-specific complaints or deviations, which are not included in other lower extremity-specific instruments.

Currently, a validated Dutch translations of the AOFAS Ankle-Hindfoot Score is not available. Therefore, the aim of the first part of the study is to translate and culturally adapt the AOFAS Ankle-Hindfoot Score questionnaire into Dutch. The aim of the second part is to evaluate the measurement properties of the AOFAS Ankle-Hindfoot Score-Dutch language version (DLV) in patients who sustained a unilateral ankle or hindfoot fracture or (fracture) dislocation by assessing descriptive statistics (including floor and ceiling effects), internal consistency, construct validity, reproducibility (*i.e.*, test-retest reliability, agreement, and smallest detectable change), and responsiveness. Measurement properties will be calculated for

the ankle and hindfoot separately.

METHODS AND ANALYSIS

Study design

This study (protocol version 1.0, date March 24, 2014) will follow a multicenter, prospective, observational study design (*i.e.*, case series). As the research physician and patients will complete questionnaires starting at variable time points during treatment, this study will have a prospective study design with retrospective data collection with regards to the injury and treatment. Three hospitals in Rotterdam (The Netherlands) will participate: Erasmus MC, University Medical Center Rotterdam, Ikazia Hospital, and Maasstad Hospital. The study is registered at the Netherlands Trial Register (NTR5613), registration date January 05, 2016.

Recruitment and consent

All consecutive patients meeting the eligibility criteria (and none of the exclusion criteria) will be included. Participation in this study will not have any influence on treatment. Prior to their outpatient department visit, eligible patients will be invited to participate. Verbal and written information will be given by the principal investigator, research physician, or a research assistant. Written materials will include an information letter, informed consent form, and return envelope. A reminder will be sent to those patients who did not respond within two weeks, in order to ensure a high response rate. If no response is received within three weeks, the patient will be contacted by telephone.

In order to reduce bias as much as possible, a research physician (MD with clinical experience) or research assistant (with a BSc in Medicine) will perform the physical examination that is part of the physician-reported part of the AOFAS Ankle-Hindfoot Score-DLV using a standardized protocol. Both assessors received elaborate training on the administration and physical examination of the AOFAS Ankle-Hindfoot Score by an experienced trauma surgeon.

Study population

All adult patients who visited the Emergency Department of any of the participating hospitals and were diagnosed with a unilateral ankle or hindfoot fracture or (fracture) dislocation will be considered eligible for inclusion. Measurement properties will be assessed for the ankle and the hindfoot subgroups separately. Patients will be identified from hospital records based upon their ICD-10 (International Coding of Diseases, 10th revision) code or Diagnosis Related Group (DRG; in Dutch, DBC) code.

Three subgroups of patients will be enrolled. In group 1 (test of pre-final version) the pre-final version of the AOFAS Ankle-Hindfoot Score-DLV will be completed. In group 2 (responsiveness) and group 3 (test-retest) the final version of the Dutch AOFAS Ankle-Hindfoot-DLV questionnaire will be completed on two occasions, with 5-6 months (group 2) or 2-3 weeks (group 3) in between.

In order to be eligible to participate in this part of the study, a patient must meet all of the following criteria:

1) Patients with a unilateral ankle or hindfoot fracture or (fracture) dislocation (*i.e.*, Ankle-Hindfoot: ankle fracture, calcaneal fracture, talar fracture, subtalar dislocation, tibiotalar dislocation, or Chopart's fracture dislocation)
2) Age 18 years or older
3) Group 2 only: Treatment started between six weeks and three months (ankle) or between three and six months (hindfoot) prior to the start of the study
4) Group 3 only: treatment has started between seven and nine months (ankle) or between six and 24 months (hindfoot) prior to the start of the study
5) Provision of informed consent by patient

A potential subject who meets any of the following criteria will be excluded from participation in this study:

1) Multiple trauma patient (only if functional recovery of additional injuries was not achieved at time of enrolment, as that likely affects the outcome scores)
2) Pathological fracture
3) Severe physical comorbidity (*i.e.*, American Society of Anesthesiologists (ASA) ≥3)
4) Patient was non-ambulatory prior to the injury (*i.e.*, bed or wheelchair-bound)
5) Insufficient comprehension of the Dutch language to understand and complete the questionnaires
6) Patient with expected problems of maintaining follow-up (*e.g.*, no fixed address)

For testing the pre-final version of the Dutch AOFAS Ankle-Hindfoot Score-DLV (group 1), only exclusion criteria 5 and 6 will apply.

Patients are allowed to participate in group 2 and 3, and if so, the second questionnaire for responsiveness will also be used as first questionnaire for test-retest reliability. Table 1 shows a summary of the injuries, identifying codes, and measurements times of this study.

Table 1: Overview of injuries, identifying codes, and measurement times

Group	Injury	Identifying code		Responsiveness		Test retest reliability	
		ICD-10	DRG	t=1	t=2	t=1	t=2
Ankle	Ankle fracture	S825, S826	224	1.5-3 mo	+ 5-6 mo	7-9 mo	+ 2-3 we
Hindfoot	Calcaneal fracture	S920	236, 237	3-6 mo	+ 5-6 mo	6-24 mo	+ 2-3 we
	Talar fracture	S921	241				
	Subtalar dislocation						
	Tibiotalar dislocation	S930					
	Chopart's fracture						
	Dislocation						

ICD-10, International Coding of Diseases, 10th revision; DRG, Diagnosis Related Group; mo, months; we, weeks.

Outcome measures

The measurement properties of the AOFAS Ankle-Hindfoot Score-DLV will be evaluated in this validation study. The following parameters will be determined:

- Construct validity
- Reliability / Internal consistency
- Reproducibility: Test-retest reliability, agreement, and Smallest Detectable Change
- Floor and ceiling effects
- Responsiveness

In addition to the outcome variables mentioned above, the following data will be collected from the patients' medical files:

a) Intrinsic variables (baseline data): age, gender, and dominant side.

b) Injury-related variables: affected side, trauma mechanism, type of injury.

c) Intervention- and outcome-related variables: type of treatment (operative or non-operative), time between injury and start of treatment, achievement of anatomic restoration as judged from X-ray or CT-scan (*i.e.*, <2mm articular step-off or gap).

Study procedures

The study will be divided into two stages. First, the American (original) version of the AOFAS Hindfoot-Ankle Score will be translated into Dutch according to a standardized procedure.[15] Second, the translated version will be tested for measurement properties in a prospective study.

Step 1: Translation of the questionnaire

The translation and cultural adaptation of the AOFAS Ankle-Hindfoot Score questionnaire will be done according to the guideline for Cross Cultural Adaptation of Self-Report Measures by Beaton *et al.*[15] This guideline is based on the review of Guillemin *et al.*[16] and is the official guideline of the American Academy of Orthopaedic Surgeons. The guideline consists of five stages: (1) translation; (2) synthesis; (3) back translation; (4) evaluation by a team of experts; and (5) tests.

In stage one, the English version of the questionnaire will be translated into Dutch independently by two Dutch native speakers who are fluent in English. One person will have knowledge of medicine and the questionnaire, the other will not necessarily.

In stage two, both translations will be combined by the two translators and a team of experts; this team will consist of at least two independent observers. The synthesis process will be carefully documented in a written report. Differences will be resolved by consensus.

In stage three, two persons will independently translate the synthesized Dutch questionnaire back into English. Both translators will be bilingual native English speakers.

Neither translator will receive any background information on the study or the questionnaire. They will have no medical background, will be blind to the original version of the questionnaire and will not be aware or informed about the concepts explored in it. With this back-translation process, the content validity of the questionnaire is checked in order to make sure that the translated version is reflecting the same item content as the original version. Unclear wording in the translated version can be discovered in this stage.

In stage four, the investigator, the translators and the same team of experts will review the two back-translations. Equivalence between the original and Dutch versions of the questionnaire shall be reached in four areas: semantic equivalence (ensuring that the words mean the same thing), idiomatic equivalence (ensuring that colloquialisms or idioms are formulated in equivalent expressions), experiential equivalence (ensuring that each item captures the experience of daily life in the target culture), and conceptual equivalence (ensuring that words hold the same conceptual meaning). Discrepancies will be resolved by consensus. This stage will result in the pre-final Dutch versions of the questionnaire.

In stage five, these pre-final Dutch version will be tested in a group of 20 patients (group 1) presenting themselves with various foot/ankle problems to the outpatient clinic of one of the participating hospitals. These patients will be asked if they understand the questions and if they are able to provide answers to the questions. If all patients report that this is the case and if there are no ambiguities, no further changes to the questionnaires will be necessary; at that point the translated questionnaire will be considered final. The measurement properties of this version will be assessed in Dutch patients as described below.

Step 2: Determining measurement properties of the AOFAS Ankle-Hindfoot Score-DLV

Patient groups 2 and 3 will be used for this evaluation.

- Group 2 (responsiveness) will consist of patients who were (surgically) treated at a participating hospital, between six weeks and three months earlier (ankle) or between three and six months earlier (hindfoot).
- Group 3 (test-retest) will consist of patients who were (surgically) treated at a participating hospital, between seven and nine months earlier (ankle) or between six and 24 months earlier (hindfoot).

In groups 2 and 3 three questionnaires will be completed during the patient's outpatient department visit; the AOFAS Ankle-Hindfoot Score-DLV, the Foot Function Index (FFI-DLV), [2] and the Short Form Health Survey (SF-36-DLV). [17] These instruments were chosen since they were also used for the validation of the original language version.[8] The research physician or research assistant will complete the AOFAS Ankle-Hindfoot Score-DLV during the outpatient department visit. If a patient is unable or unwilling to come to the hospital, a home visit may be planned.

The Foot Function Index (FFI) measures the effect of foot pathology on function in terms of pain and disability. The FFI consists of 23 items divided into three subscales: limitation, pain, and disability. The items are scored on a 10-point Likert scale. For each subscale, the raw score is transformed to a 100-point score; the higher the score, the more limitation/pain/ disability is present. The total score on the FFI is the mean of the subscale scores.[2] Adequate internal consistency, reproducibility and reliability as well as strong correlation with SF-36 have been reported for patients with traumatic foot disorders in some languages.[2, 18, 19] The FFI-DLV will be used [2].

The Short Form Health Survey (SF-36) is a generic health status questionnaire that gives an indication of health-related quality of life.[20-27] The SF-36 consists of 36 items (questions)

and provides scores on eight dimensions (subscales): physical functioning (PF), role limitations due to physical health problems (RP), bodily pain (BP), general health perceptions (GH), vitality (VT), social functioning (SF), role limitations due to emotional problems (RE), and general mental health (MH). These eight domains are combined into a Physical Component Summary (PCS) and a Mental Component Summary (MCS). The raw score on each subscale is transferred to a 100-point scale, with a higher score indicating better quality of life. These scores will be converted to a norm-based score and compared with the norms for the general population of the United States (1998), in which each scale was scored to have the same average (50 points) and the same standard deviation (10 points). Dutch norms are available, but will not be used. The Dutch norms were calculated using a smaller sample size than the American study. Moreover, most published studies have used the American norms. On a study population level the means and median values were similar when using the Dutch or American norms, but variance was larger using the Dutch norms than when using the US norms.[28] The SF-36 is the most widely evaluated patient-reported outcome measure for assessing general health.[29] It is reliable and easy to complete. A validated Dutch version will be used.[17]

In order to determine whether the AOFAS Ankle-Hindfoot Score-DLV is able to detect clinical change over time, patients in group 2 will be asked to complete all questionnaires again after five to six months after completing them the first time. A research physician or research assistant will complete the AOFAS Ankle-Hindfoot Score-DLV. For responsiveness, this time interval should be sufficiently long enough for clinical improvement to occur. We consider a time interval of five to six months to be appropriate for all three groups of injuries.

In order to determine the reproducibility (*i.e.*, test-retest reliability) of the AOFAS Ankle-Hindfoot Score-DLV, all questionnaires will be completed again at two to three weeks after completing them the first time (group 3). For test-retest reliability, this time interval needs to be sufficiently short to support the assumption that the patient remains stable and sufficiently

long to prevent recall. We consider a time interval of 2-3 weeks to be appropriate. Patients are asked about presence or absence of change between the two questionnaire administrations. They were asked to complete a transition item (anchor question) evaluating their perception of change in the general condition of their affected ankle. The question was: How would you judge the condition of your ankle, compared with the last time you completed this questionnaire? Patients were given the answer options 'better', 'no change', or 'worse'. Patients reporting a change (either improvement or deterioration) will be excluded from the analysis. Patients who replied 'no change' were considered stable between the two measurements.

Sample size calculation

The pre-final Dutch version of the instrument will be tested in a group of 20 patients (group 1) presenting themselves with various foot/ankle problems to the outpatient clinic of the Erasmus MC (Rotterdam), Ikazia Hospital (Rotterdam), or Maasstad Hospital (Rotterdam).

For groups 2 and 3, recruitment of both the ankle and the hindfoot injury subgroups will continue until complete follow up is ensured for 100 patients. The minimum number of patients needed for determining measurement properties of a PROM depends on the property evaluated. Validity can only be rated positive if at least 75% of the results are in correspondence with prespecified hypotheses, in (sub)groups of at least 50 patients.[30] For calculating the Smallest Detectable Change (SDC) as well as for the assessment of the agreement parameters (reproducibility), a sample size of at least 50 patients is generally considered adequate.[30, 31] The (absence of) floor and ceiling effects also requires a sample size of at least 50 patients. In order to perform a factor analysis (to determine if the AOFAS Ankle-Hindfoot Score-DLV consists of multiple subscales), however, four to ten patients for each item are advised with a minimum of 100 patients.[30, 32] The sample size needed applies both to patients with ankle injuries and hindfoot injuries.

Statistical analysis

Data will be entered into an OpenClinical database. Data will be encoded, and a random sample of entered data will be checked by an independent data monitoring committee. Only the research team, the Medical Research Ethics Committee (MREC), and the health inspection will have legal access to the data.

All statistical analyses will be performed with the Statistical Package for Social Sciences (SPSS, version 21 or higher) and will be reported following the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) and the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines. Descriptive statistics will be used in order to describe the main characteristics of the study participants and the questionnaire scores at the different time points. Data for patients with ankle or hindfoot injuries will be evaluated as two separate groups.

As the raw data for individual items will be analyzed, missing values will not be imputed. Normality of continuous data will be tested with the Shapiro-Wilk test. Descriptive analysis will be performed; continuous data will be reported as mean ± standard deviation (SD) (parametric) or median with percentiles (non-parametric) and categorical data as numbers with percentages.

In order to evaluate if a representative sample participated in this study, the age, gender, and injury location of responders will be compared with that of the non-participants. The categorical variables gender and injury location will be assessed using a Chi-squared test. Age will be compared using a Student's T-test (parametric data) or Mann-Whitney U-test (parametric data).

Construct validity

Validity is the degree to which a patient-reported outcome instrument measures the construct it is supposed to measure. As there is no gold standard in the current study, the validity of the AOFAS Ankle-Hindfoot Score-DLV will be expressed in terms of the construct validity.

Construct validity refers to the extent to which scores on a specific questionnaire relate to other measures in a way that is in agreement with prior theoretically derived hypotheses concerning the concepts that are being measured.[30] In order to evaluate the construct validity of the AOFAS Ankle-Hindfoot Score-DLV, we will formulate a set of hypotheses about the expected magnitude and direction of relationships between the AOFAS (sub)scores and the FFI and the SF-36 (sub)scores. Pearson's product-moment correlation coefficients (parametric data) or Spearman's Rho (rank correlation) coefficients (non-parametric correlation) will be calculated in order to assess construct validity. Correlation coefficients above 0.6, between 0.6 and 0.3 and less than 0.3 will be considered high, moderate, and low correlations, respectively.[33] The AOFAS Ankle-Hindfoot Score is expected to have a high correlation with pain and function (sub)scales (*i.e.*, FFI total score and all three subscales, SF-36 PF, RP, BP, and PCS), a moderate correlation with the SF-36 VT, SF and RE subscales, and a low correlation with SF-36 GH, MH, and MCS. Construct validity will be given a positive rating if at least 75% of the results are in accordance with predefined hypotheses in a (sub)sample of at least 50 patients.[30]

Reliability / internal consistency

Reliability is defined as the degree to which the measurement is free from measurement error.[34] Three elements of reliability will be determined: internal consistency, reproducibility, and measurement error.

Internal consistency is defined as the extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct.[30] The correlation between items on a (sub)scale will be evaluated by calculating Cronbach's alpha for every (sub)scale. Since future use of the AOFAS instrument will be at a group level, internal consistency is considered sufficient if the value for Cronbach's alpha is between 0.70 and 0.95, provided that the scale is unidimensional.[30, 35] If necessary, confirmatory or exploratory factor analysis will be performed, as applicable.

Reproducibility

Reproducibility concerns the degree to which repeated measurements in stable persons (test-retest) provide similar answers.[30] Reproducibility is suggested to consist of two parts: reliability and agreement.[36, 37] The data of group 3 will be used; they will complete all questionnaires twice, with 2-3 weeks in between. Only data for patients reporting 'no change' on the transition item are included as they were considered to be stable between the measurements.

Reliability concerns the degree to which patients can be distinguished from each other, despite measurement error.[30, 38] Evaluation of the test-retest reliability of the AOFAS Ankle-Hindfoot Score-DLV will be performed by calculating the intraclass correlation coefficient ($ICC_{agreement}$) with corresponding 95% confidence interval (CI). An ICC two-way random effects model, type absolute agreement (ICC(2,1)), will be used.[39] Reliability will be given a positive rating when the ICC is at least 0.70 in a sample size of at least 50 patients.[30]

Agreement concerns the absolute measurement error, *i.e.*, how close the scores on repeated measures are, expressed in the unit of the measurement scale at issue.[30] The degree of absolute agreement of the AOFAS Ankle-Hindfoot Score-DLV will be expressed as the standard error of measurement ($SEM_{agreement}$). This SEM equals the square root of the error variance of an analysis of variance (ANOVA) analysis, including the systematic differences ($SEM = \sqrt{(variance_{patient} + variance_{residual})}$).[30, 40, 41]

Based upon the SEM, the Smallest Detectable Change (SDC) will be calculated using the formula; $SDC = 1.96 \times \sqrt{2} \times SEM$.[30] The SDC reflects the smallest within-person change in a score that, with $P < 0.05$, can be interpreted as a ''real'' change, above measurement error, in one individual (SDC_{ind}).[30, 42, 43] The SDC measurable in a group of people (SDC_{group}) will be calculated by dividing the SDC_{ind} by $\sqrt{n}$.[43, 44] Finally, the reliable change index (RCI) will be calculated, representing the SDC as a percentage of the maximum obtainable score.

The degree of absolute agreement of the AOFAS Ankle-Hindfoot Score-DLV will also be determined with a Bland and Altman analysis.[45] The limits of agreement equal the mean change in scores of repeated measurements ($mean_{change}$) ± 1.96 x standard deviation of these changes (SD_{change}).[30] Zero falling outside this interval indicates a bias in the measurements.

Floor and ceiling effects

The validity, reliability and responsiveness of a questionnaire may be jeopardized if floor or ceiling effects are present. It is then likely that extreme items are missing in the lower or upper ends of the questionnaire. As a consequence, respondents with the lowest or highest possible score cannot be distinguished from each other (indicating limited reliability) and changes in these patients cannot be measured (indicating limited responsiveness).[30] Floor and ceiling effects will be determined by calculating the number of individuals that obtained the lowest (0 points; floor) or highest (100 points; ceiling) scores possible and will be considered present if more than 15% of the respondents achieved the lowest or highest score in a sample size of at least 50 patients.[30, 46] Floor and ceiling effects will be determined separately for the different time points.

Responsiveness

Responsiveness is defined as the ability of a questionnaire to detect clinically important changes over time, even if these changes are small.[30, 47] The data of group 2 will be used; they will complete all questionnaires twice, with 5-6 months in between.

The effect size (ES) and standardized response mean (SRM) of the (sub)scales of the AOFAS Ankle-Hindfoot Score-DLV will be determined as measures of the magnitude of change over time. The ES will be calculated by dividing the mean change in score between the two time points by the standard deviation of the first measurement.[48] The SRM will be calculated by dividing the mean change in score between two time points by the standard

deviation of this change.[48] These effect estimates will be interpreted according to Cohen: a SRM of 0.2-0.4 is considered a small effect, 0.5-0.7 a moderate, and 0.8 or higher a large effect.[49]

Responsiveness can be considered to be a measure of longitudinal validity. In analogy to construct validity, this longitudinal validity will be assessed by testing predefined hypotheses about expected correlations between changes in AOFAS Ankle-Hindfoot Score-DLV (sub)scales versus changes in FFI and SF-36 (sub)scales.[30] Change scores of the AOFAS Ankle-Hindfoot Score are expected to have a moderate correlation with changes in the FFI (sub)scales, SF-36 PF, RP, BP, VT, SF, RE, and PCS. A low correlation is expected with changes in the SF-36 GH, MH, and MCS.

ETHICS AND DISSEMINATION

This study will be conducted according to the principles of the Declaration of Helsinki (64th World Medical Association General Assembly, Fortaleza, Brazil, October 2013). This study has been exempted by the medical research ethics committee (MREC) Erasmus MC (Rotterdam, The Netherlands). This MREC acts as central ethics committee for this trial (reference number MEC-2014-215). Approval has been obtained from the local hospital boards in all participating centers. Following review of the protocol, the MREC concluded that this study is not subject to the Medical Research Involving Human Subjects Act (WMO). They concluded that the study is a medical/scientific research, but no patients are subjected to procedures or are required to follow rules of behavior. Consequently, the statutory obligation to provide insurance for subjects participating in medical research (article 7 of the WMO) was also waived. Any important changes in the protocol will be submitted to the accredited MREC. The results of the study are planned to be published in an international, peer reviewed journal. Results of the ankle and hindfoot injury subgroups will be published separately.

DISCUSSION

Modern studies that evaluate treatment efficacy are expected to also take into account the treatment outcome from a patient's perspective. Clinical measures such as mortality, radiographic healing, and rates of complications, re-operation, and readmission are relevant; however, they do not reflect to what extent a patient is able to function in daily living. For that purpose, PROMs and mixed instruments, which combine a patient-reported and a physician-reported part, have been developed. There is a great need for valid instruments in different languages.

The AOFAS Ankle-Hindfoot Score is commonly used in patients with an ankle or hindfoot injury. This instrument combines functional outcome and pain, which are both critical for patients. The AOFAS Ankle-Hindfoot Score is only valid if the score truly reflects function and pain. Completing the questionnaire in duplicate should result in the same score, and during recovery, the change in score should reflect change in functional status of the patient. Both elements of validity of the instrument are determined as part of this study. We expect that the AOFAS Ankle-Hindfoot Score-DLV will prove valid and reliable, giving objective quantitative scores for patients' function and pain after trauma to the ankle or hindfoot. If the data confirm this, the instrument will be available for comparing outcome in future studies, and for comparing treatment outcome across hospitals or between patient groups. Especially the SDC and MIC will reveal important information for sample size calculations in future studies.

Three hospitals in the Netherlands will participate. Inclusion of patients has started May 2014 and the expectation is to include all patients within two years for ankle injuries and three years for hindfoot injuries. With a maximum follow-up of 6.5 months the presentation of data will be expected by end-2016 and end-2017, respectively.

REFERENCES

1. De Groot IB, Favejee MM, Reijman M, et al. The Dutch version of the Knee Injury and Osteoarthritis Outcome Score: a validation study. *Health Qual Life Outcomes* 2008;6:16.
2. Kuyvenhoven MM, Gorter KJ, Zuithoff P, et al. The foot function index with verbal rating scales (FFI-5pt): A clinimetric evaluation and comparison with the original FFI. *J Rheumatol* 2002;29(5):1023-8.
3. De Groot IB, Reijman M, Terwee CB, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthritis Cartilage* 2007;15(1):104-9.
4. Shearer D, Morshed S. Common generic measures of health related quality of life in injured patients. *Injury* 2011;42(3):241-7.
5. Reininga IH, el Moumni M, Bulstra SK, et al. Cross-cultural adaptation of the Dutch Short Musculoskeletal Function Assessment questionnaire (SMFA-NL): internal consistency, validity, repeatability and responsiveness. *Injury* 2012;43(6):726-33.
6. Van Son MA, Den Oudsten BL, Roukema JA, et al. Psychometric properties of the Dutch Short Musculoskeletal Function Assessment (SMFA) questionnaire in patients with a fracture of the upper or lower extremity. *Qual Life Res* 2014;23(3):917-26.
7. Kitaoka HB, Alexander IJ, Adelaar RS, et al. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. *Foot Ankle Int* 1994;15(7):349-53.
8. Madeley NJ, Wing KJ, Topliss C, et al. Responsiveness and validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle Hindfoot Score, and Foot Function Index in end stage ankle arthritis. *Foot Ankle Int* 2012;33(1):57-63.
9. SooHoo NF, Vyas R, Samimi D. Responsiveness of the foot function index, AOFAS clinical rating systems, and SF-36 after foot and ankle surgery. *Foot Ankle Int* 2006;27(11):930-4.

10. SooHoo NF, Shuler M, Fleming LL. Evaluation of the validity of the AOFAS Clinical Rating Systems by correlation to the SF-36. *Foot Ankle Int* 2003;24(1):50-5.

11. Ibrahim T, Beiri A, Azzabi M, et al. Reliability and validity of the subjective component of the American Orthopaedic Foot and Ankle Society clinical rating scales. *J Foot Ankle Surg* 2007;46(2):65-74.

12. Westphal T, Piatek S, Halm JP, et al. Outcome of surgically treated intraarticular calcaneus fractures--SF-36 compared with AOFAS and MFS. *Acta Orthop Scand* 2004;75(6):750-5.

13. Pinsker E, Daniels TR. AOFAS position statement regarding the future of the AOFAS Clinical Rating Systems. *Foot Ankle Int* 2011;32(9):841-2.

14. Guyton GP. Theoretical limitations of the AOFAS scoring systems: an analysis using Monte Carlo modeling. *Foot Ankle Int* 2001;22(10):779-87.

15. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000;25(24):3186-91.

16. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46(12):1417-32.

17. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998;51(11):1055-68.

18. Huh JW, Eun IS, Ko YC, et al. Reliability and Validity of the Korean Version of the Foot Function Index. *J Foot Ankle Surg* 2016;55(4):759-61.

19. Wu SH, Liang HW, Hou WH. Reliability and validity of the Taiwan Chinese version of the Foot Function Index. *J Formos Med Assoc* 2008;107(2):111-8.

20. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.

21. Mahabier KC, Van Lieshout EMM, Bolhuis HW, et al. HUMeral shaft fractures: measuring recovery after operative versus non-operative treatment (HUMMER): a multicenter comparative observational study. *BMC Musculoskelet Disord* 2014;15:39.

22. Den Hartog D, Van Lieshout EMM, Tuinebreijer WE, et al. Primary hemiarthroplasty versus conservative treatment for comminuted fractures of the proximal humerus in the elderly (ProCon): a multicenter randomized controlled trial. *BMC Musculoskelet Disord* 2010;11:97.

23. De Haan J, Den Hartog D, Tuinebreijer WE, et al. Functional treatment versus plaster for simple elbow dislocations (FuncSiE): a randomized trial. *BMC Musculoskelet Disord* 2010;11:263.

24. Schep NWL, De Haan J, Iordens GIT, et al. A hinged external fixator for complex elbow dislocations: a multicenter prospective cohort study. *BMC Musculoskelet Disord* 2011;12:130.

25. Pan SL, Liang HW, Hou WH, et al. Responsiveness of SF-36 and Lower Extremity Functional Scale for assessing outcomes in traumatic injuries of lower extremities. *Injury* 2014;45(11):1759-63.

26. Obremskey WT, Dirschl DR, Crowther JD, et al. Change over time of SF-36 functional outcomes for operatively treated unstable ankle fractures. *J Orthop Trauma* 2002;16(1):30-3.

27. van Tetering EA, Buckley RE. Functional outcome (SF-36) of patients with displaced calcaneal fractures compared to SF-36 normative data. *Foot Ankle Int* 2004;25(10):733-8.

28. Schep NWL, van Lieshout EMM, Patka P, et al. Long-term functional and quality of live assessment following post-traumatic distraction osteogenesis of the lower limb. *Strategies Trauma Limb Reconstr* 2009;4(3):107-12.

29. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *Bmj* 2002;324(7351):1417.

30. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42.

31. Altman DG. Practical statistics for medical research. London: Chapman and Hall. 1991.

32. De Vet HCW, Terwee CB, Mokkink LB, et al. Measurement in medicine. New York: Cambridge University Press. 2011.

33. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. Boston USA: Houghton MiZin. 1998.

34. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63(7):737-45.

35. Streiner D, Norman G. Health Measurement Scales: A Practical Guide to Their Development and Use. 2nd ed. Oxford, UK: Oxford Medical; 1996. And Nunnally J, Bernstein I. Psychometric Theory. 3rd ed. New York, NY: McGraw-Hill; 1994.

36. De Vet HCW. Observer reliability and agreement. In: Armitage P, Colton T, editors. Encyclopedia of biostatistics. Boston: John Wiley & Sons Ltd. pp. 3123e8. 1998.

37. Stratford P. Reliability: consistency or differentiating among subjects? *Phys Ther* 1989;69(4):299-300.

38. Streiner DL, Norma GR. Health measurement scales. A practical guide to their development and use. New York: Oxford University Press. 2003.

39. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1(1):30-46.

40. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? *Injury* 2011;42(3):236-40.

41. de Vet HC, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59(10):1033-9.

42. Beckerman H, Roebroeck ME, Lankhorst GJ, et al. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res* 2001;10(7):571-8.

43. de Vet HC, Bouter LM, Bezemer PD, et al. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care* 2001;17(4):479-87.

44. De Boer MR, De Vet HC, Terwee CB, et al. Changes to the subscales of two vision-related quality of life questionnaires are proposed. *J Clin Epidemiol* 2005;58(12):1260-8.

45. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10.

46. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4(4):293-307.

47. Guyatt GH, Deyo RA, Charlson M, et al. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989;42(5):403-8.

48. Angst F, Verra ML, Lehmann S, et al. Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. *BMC Med Res Methodol* 2008;8:26.

49. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale (NJ): Lawrence Erlbaum Associates. 1988.

AUTHOR'S CONTRIBUTIONS

EMMVL, ASDB, DEM, CHVDV, PTDH, WET, and MJHV developed the study. ASDB and EMMVL drafted the manuscript. EMMVL will act as trial principal investigator. ASDB, CHVDV, PTDH, DEM, and MHJV will participate in patient inclusion and outcome assessment. ASDB, WET, and EMMVL will perform statistical analysis of the study data. All authors have read and approved the final manuscript.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. There are no statements to declare relating contributorship, data sharing, or ethics approval.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing interests.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3, 4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5, 6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	8, 9
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9, 11
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	8 - 11
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	9-11
		(*b*) Describe any methods used to examine subgroups and interactions	7-11
		(*c*) Explain how missing data were addressed	9
		(*d*) If applicable, explain how loss to follow-up was addressed	7-11
		(*e*) Describe any sensitivity analyses	9-11
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	12
		(b) Give reasons for non-participation at each stage	12
		(c) Consider use of a flow diagram	22 (Figure 1)
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	12
		(b) Indicate number of participants with missing data for each variable of interest	12
		(c) Summarise follow-up time (eg, average and total amount)	12
Outcome data	15*	Report numbers of outcome events or summary measures over time	12 – 14
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12-14
Discussion			
Key results	18	Summarise key results with reference to study objectives	15-17
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-17
Generalisability	21	Discuss the generalisability (external validity) of the study results	15-17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

The American Orthopedic Foot and Ankle Society Ankle-Hindfoot Score; Translation and Validation of the Dutch Language Version for ankle fractures

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017040.R1
Article Type:	Research
Date Submitted by the Author:	10-May-2017
Complete List of Authors:	De Boer, A.; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery Tjioe, Roderik J.C.; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery van der Sijde, Fleur; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery Meuffels, Duncan; Erasmus MC, University Medical Center Rotterdam, Orthopaedic Surgery Den Hoed, P.; Ikazia Ziekenhuis, Surgery Van der Vlies, Cornelis; Maasstad Ziekenhuis Tuinebreijer, Wim; Erasmus MC, University Medical Center Rotterdam Verhofstad, Michael; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery Van Lieshout, Esther M.M.; Erasmus MC, University Medical Center Rotterdam AOFAS Study Group, .; Erasmus MC, University Medical Center Rotterdam, Trauma Surgery
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Emergency medicine, Rehabilitation medicine, Patient-centred medicine, Surgery, Sports and exercise medicine
Keywords:	Ankle, Fracture, Reliability, Responsiveness, Validity

The American Orthopedic Foot and Ankle Society Ankle-Hindfoot Scale; Translation and Validation of the Dutch Language Version for ankle fractures

A. Siebe De Boer MD[1], R.J.C. Tjioe[1], F. Van der Sijde[1], Duncan E. Meuffels MD PhD[2], P. Ted Den Hoed MD PhD[3], Cornelis H. Van der Vlies MD PhD[4], Wim E. Tuinebreijer MD PhD[1], Michael H.J. Verhofstad MD PhD[1], Esther M.M. Van Lieshout MSc PhD[1*], AOFAS study Group*

*D.A. Newhall, E.N. Van der Velden – Macauley, L. Vellekoop, J. Romeo

[1]Trauma Research Unit Department of Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

[2]Department of Orthopedic Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

[3]Department of Surgery, Ikazia Hospital, P.O. Box 5009, 3008 AA Rotterdam, The Netherlands

[4]Department of Surgery, Maasstad Hospital, P.O. Box 9100, 3007 AC Rotterdam, The Netherlands

[*]Corresponding author:

Esther M.M. Van Lieshout, MSc PhD

Trauma Research Unit Department of Surgery

Erasmus MC, University Medical Center Rotterdam

P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Email: e.vanlieshout@erasmusmc.nl

Telephone: + 31 10 7031050

Email addresses:

ASDB: a.deboer@erasmusmc.nl

RJCT: r.tjioe@erasmusmc.nl

FVDS: f.vandersijde@erasmusmc.nl

DEM: d.meuffels@erasmusmc.nl

PTDH: pt.hoed@ikazia.nl

CHVDV: VliesC@maasstadziekenhuis.nl

WET: w.tuinebreijer@erasmusmc.nl

MHJV: m.verhofstad@erasmusmc.nl

EMMVL: e.vanlieshout@erasmusmc.nl

Keywords: Ankle; Fracture; Reliability; Responsiveness; Validity.

ABSTRACT

Objectives: The American Orthopedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Scale is among the most commonly used instruments for measuring outcome of treatment in patients who sustained a complex ankle or hindfoot injury. It consists of a patient-reported and a physician-reported part. A validated, Dutch version of this instrument is currently not available. The aim of this study was to translate the instrument into Dutch and to determine the measurement properties of the AOFAS Ankle-Hindfoot Scale Dutch Language Version (DLV) in patients with a unilateral ankle fracture.

Setting: Multicenter (two Dutch hospitals), prospective observational study.

Participants: In total 142 patients with a unilateral ankle fracture were included. Ten patients were lost to follow up.

Primary and secondary outcome measures: Patients completed the subjective (patient-reported) part of the AOFAS Ankle-Hindfoot Scale-DLV. A physician or trained physician-assistant completed the physician-reported part. For comparison and evaluation of the measuring characteristics, the Foot Function Index (FFI) and the Short Form-36 (SF-36) were completed by the patient. Descriptive statistics (including floor and ceiling effects), reliability (*i.e.*, internal consistency), construct validity, reproducibility (*i.e.*, test-retest reliability, agreement, and smallest detectable change), and responsiveness were determined.

Results: The AOFAS-DLV and its subscales showed good internal consistency (Cronbach's alpha > 0.90). Construct validity and longitudinal validity were proven to be adequate (76.5% of predefined hypotheses were confirmed). Floor effects were not present. Ceiling effects were present from six months onwards, as expected. Responsiveness was adequate, with a smallest detectable change of 12.0 points.

Conclusions: The AOFAS-DLV is a reliable, valid, and responsive measurement instrument for evaluating functional outcome in patients with a unilateral ankle fracture. This implies that the questionnaire is suitable to compare different treatment modalities within this population or to compare outcome across hospitals.

Trial Registration: Netherlands Trial Register (NTR5613; 05-jan-2016).

ARTICLE SUMMARY

Strengths and limitations of this study:

- This prospective, multicenter, observational study shows substantial, previously unknown information about the performance of the American Orthopedic Foot and Ankle Society Ankle-Hindfoot Scale.
- The topic of the clinical study is relevant for orthopedic trauma surgeons, since there is growing need for translated and validated patient reported outcome measures that can be used for determining functional outcome over time.
- The methodological design of the study is strong, and statistical analyses complied with the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines.
- Although the study is mostly relevant for the Dutch-speaking regions, it is also informative for other regions.
- Implementation of the American Orthopedic Foot and Ankle Society Ankle-Hindfoot Scale is limited by the fact that a clinician is required to complete the physician-reported part of the questionnaire. This hampers its use in, *e.g.*, large-scale registers.

BACKGROUND

Ankle fractures are common injuries with a reported incidence rate of 187 fractures per 100,000 people each year (1). Due to an increasing number of people involved in sports and the growing elderly population, this rate is rising significantly in many industrialized countries (1). Ankle fractures can cause a temporary loss of function and quality of life. In order to monitor recovery after treatment, questionnaires regarding functional outcome are increasingly used in clinical practice and clinical research. They enable detailed evaluation of functional outcome and quality of life after (non-)operative treatment of musculoskeletal injuries from a patient's perspective.

Although questionnaires completed by patients alone (so called patient-reported outcome measures; PROMs) may be preferred, many scores combine a patient-reported and a physician-reported part. Examples of PROMs used in foot and ankle research are the Maryland Foot Score (MFS) (2), Foot and Ankle Ability Measure (FAAM) (3), the Foot Function Index (FFI) (4), the Manchester-Oxford Foot Questionnaire (MOXFQ) (5, 6), and the Self-Reported Foot and Ankle Score (SEFAS) (7).

The clinical rating system published by the American Orthopedic Foot and Ankle Society, the AOFAS Ankle-Hindfoot Scale, is one of the mostly used assessment tool in foot surgery (8). This clinical rating system, developed by Kitaoka *et al.*, combines subjective scores of pain and function provided by the patient and objective scores based on the physician's physical examination (*i.e.*, gait, sagittal motion, hindfoot motion, ankle-hindfoot stability, and alignment of the ankle-hindfoot) (9). The questionnaire includes nine items that can be divided into three subscales (pain, function, and alignment). Each of the nine items is scored, accumulating to a total score ranging from 0 points (indicating severe pain and impairment) to 100 points (no symptoms or impairment).

Limitations on the use of the AOFAS Ankle-Hindfoot Scale are the fact that questions have a limited number of answers, some of which can be interpreted differently (10, 11). An advantage is that the physician-reported questions on gait and range of motion provide relevant information that the PROMs do not provide.

The AOFAS Ankle-Hindfoot Scale as a complete scale has been shown to be responsive and valid in its original language version (9, 12-14). The patient-reported part of the scale has been shown to be valid and reliable (15). Reliability of the objective (physician-reported) portion of the scale has not been published. Previous studies involved a wide spectrum of diagnoses, such as general ankle-hindfoot complaints (13), pending ankle or foot surgery (15), surgically treated calcaneal fractures (14), and end-stage ankle osteoarthritis (12).

A validated Dutch version of the AOFAS Ankle-Hindfoot Scale is not available. The aim of this study was to translate the questionnaire into Dutch and to culturally adapt it to the Dutch population. The next aim was to determine the measurement properties of the AOFAS Ankle-Hindfoot Scale Dutch Language Version (AOFAS-DLV) in patients who sustained an ankle fracture.

METHODS

Study design and ethics statement

This study followed a multicenter, prospective, observational study design (*i.e.*, case series) and was performed at two Dutch hospitals. The study is registered at the Netherlands Trial Register (NTR5613). A detailed study protocol is published elsewhere (16). The study was approved by the Medical Research Ethics Committees or Local Ethics Boards of all participating centers. All patients provided informed consent.

Translation

First, the American (original) version of the AOFAS Hindfoot-Ankle Scale was translated and cultural adapted into Dutch according to the guideline for Cross Cultural Adaptation of Self-Report Measures by Beaton *et al*. (17), as described in detail in the published study protocol (16). In the last stage of this guideline the pre-final Dutch version was tested in a group of 20 patients, presenting themselves with various foot/ankle problems in one of the participating hospitals. Since there were no ambiguities or misunderstandings of the questions in this group, the translated questionnaire was considered the final AOFAS Ankle-Hindfoot Scale-DLV (Supplemental Table 1).

Validation

Patient recruitment

Patients were recruited from May 1, 2014 to March 29, 2016. Patients were identified from hospital records, based upon their ICD-10 (International Coding of Diseases, 10th revision) code or Diagnosis Related Group (DRG; in Dutch, DBC) code. Inclusion criteria were; 1)

unilateral ankle fracture; 2) age of 18 years or older; and 3) provision of informed consent by the patient. Treatment should have been started between six weeks and three months and/or between seven and nine months prior to the start of the study. Exclusion criteria were; 1) multiple trauma (only if functional recovery of additional injuries was not achieved at time of enrolment, as that likely affects the outcome scores); 2) pathological fracture; 3) severe physical comorbidity (*i.e.*, American Society of Anaesthesiologists (ASA) ≥3); 4) patient was non-ambulatory prior to the injury; 5) insufficient comprehension of the Dutch language to understand and complete the questionnaires; and 6) expected problems of maintaining follow-up.

In total 142 individual participants were included, 70 completed t=1 and t=2, 132 completed t=2 and t=3 (Figure 1). During the course of the study ten patients were lost to follow up. One patient, who participated in the test-retest part, had to be removed from the analysis; due to removal of osteosynthesis material, the patient reported a change in function between both recordings.

The median age was 46 years (P_{25}-P_{75} 35-60), see Table 1. The majority of patients (N=75; 52.8%) were male. Most ankle fractures were unimalleolar (N=100; 70.4%), and the majority (N=84; 59.2%) were treated operatively.

The AOFAS Ankle-Hindfoot Scale-DLV, the Foot Function Index (FFI-DLV), and the Short Form Health Survey (SF-36-DLV) questionnaires could be completed in total on three occasions: at 2 months (t=1), 7 months (t=2), and 7.5 months (t=3) after trauma. Two months was chosen as first moment after start of weight bearing where both the questions of the patient and physician-reported part could be answered; a low score was expected. At seven months the majority of patients were expected to have reached their maximum recovery, giving the highest possible AOFAS score. That score was also expected at t=3. The time

between the recordings was 5-6 months (responsiveness, t=1 and t=2) and/or 2-3 weeks (test-retest, t=2 and t=3) in between. Patients were allowed to participate in both the responsiveness and test-retest part, and if so, the questionnaires at t=2 were also used as first questionnaire for test-retest reliability.

Questionnaires and data collection

The FFI is a scoring system developed to measure the impact of foot pathology. It consists of 23 items, which are grouped into the subscales pain, disability, and activity limitation. Scores for all (sub)scales range from zero (no disability) to 100 (highest level of disability) (4).

The SF-36 Health Survey is a generic measure of health status (18-25). It consists of 36 items, representing eight domains that are grouped into a Physical Component Summary (PCS) and a Mental Component Summary (MCS).

One research physician and one research assistant performed the physical examination that is part of the physician-reported part of the AOFAS Ankle-Hindfoot Scale-DLV using a standardized protocol. Both assessors received elaborate training by an experienced trauma surgeon. Data for each patient was completed by the same assessor. Patients completed the patient-reported part, as well as the FFI and SF-36. Demographic, injury and treatment data were collected from the patient's medical files.

Statistical analysis

Statistical analyses were performed using the Statistical Package for Social Sciences (SPSS, version 21). Data are reported following the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) (26). Since raw data for individual items were analyzed, missing data were not imputed. Descriptive statistics was used in order to describe the main characteristics of the study participants and the questionnaire scores at the different time

points. Measurement properties of the AOFAS-DLV (sub)scales were determined by comparing these (sub)scales with the FFI and SF-36 (sub)scales. They were determined in compliance with the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines (27). A detailed description of the measurement properties and statistical analysis is shown in the published study protocol (16). A summary is given below.

Floor and ceiling effects are present if more than 15% of the study population rates the lowest or highest possible score (16, 28, 29). Data for each time point were evaluated separately.

Internal consistency (measure of reliability) was considered adequate if the Cronbach's alpha value is between 0.70 and 0.95, provided that the scale is unidimensional (28). For reasons of heterogeneity in scores, data for t=1 were used.

Construct validity was assessed by determining the correlation of the AOFAS-DLV (sub)scales with (sub)scales of the FFI and SF-36. Spearman's Rho (rank correlation) coefficients (*r*) were calculated since data were non-parametric. Data of t=1 were used. Strength of correlation was categorized as high ($r > 0.6$), moderate ($0.3 < r < 0.6$), or low ($r < 0.3$) (30). Construct validity was considered adequate if at least 75% of the results were in line with the predefined hypotheses in a (sub)sample of at least 50 patients (28). Expected correlations are given in Supplemental Table 2A.

Evaluation of the test-retest reliability was performed by calculating the intraclass correlation coefficient ($ICC_{agreement}$) of (sub)scales administered at t=2 and t=3. ICC is reported with 95% confidence interval (CI). Reliability was given a positive rating when the ICC is at least 0.70 in a sample size with a minimum of 50 patients (28).

The degree of absolute agreement was expressed as the standard error of measurement ($SEM_{agreement}$). For individual patients, the smallest detectable change (SDC) was calculated

as 1.96 x √2 x SEM (28). The SDC measurable in a group of people (SDC_{group}) was calculated by dividing the SDC in individuals (SDC_{ind}) by √n (31, 32). Finally, the reliable change index (RCI) was calculated, representing the SDC as a percentage of the maximum obtainable score.

The degree of absolute agreement was also determined with a Bland and Altman analysis (33). The limits of agreement equal the mean change in scores of repeated measurements ($mean_{change}$) ± 1.96 x standard deviation of these changes (SD_{change}) (28). Zero falling outside this interval indicates bias in the measurements.

Analogous to construct validity, longitudinal validity (a measure of responsiveness) was assessed by testing predefined hypotheses (Supplemental Table 2B) about expected correlations between changes in AOFAS Ankle-Hindfoot scale-DLV (sub)scales versus changes in FFI and SF-36 (sub)scales (28). Change scores were calculated from t=1 to t=2. Since data were non-parametric, Spearman's rank correlation coefficients were calculated. Longitudinal validity was considered adequate if at least 75% of the results were in line with the predefined hypotheses in a (sub)sample of at least 50 patients (28).

The effect size (ES) and standardized response mean (SRM) were determined as measures of the magnitude of change over time, using the data of t=1 and t=2. ES was calculated as change in score (t=2 - t=1)/SD_{T1} (28). SRM was calculated as change in score (t=2 – t=1)/SD_{change} (28). Values of 0.2-0.4 were considered a small effect, 0.5-0.7 a moderate, and 0.8 or higher a large effect (34). Large effect sizes were expected a priori, since at t=1 patients were expected to have functional limitations, whereas at t=2 full recovery was expected for most patients.

RESULTS

The changes over time in AOFAS-total, FFI-total, SF-36 PCS, SF-36 MCS, SF-36 PF, and SF-36 BP are shown in Figure 2. The AOFAS and SF-36 (all subscales) show an increase in scores in the period from t=1 to t=2. The FFI, focusing on disabilities rather than function, shows a decrease in score. Scores at t=2 and t=3 were similar for all instruments.

Floor and ceiling effects

A floor effect was only present in two SF-36 subscales; namely SF-36 RP subscale at t=1; 58.6% of the patients reported the minimum score, at t=2 (19.7%) and t=3 (17.6%), and the SF-36 RE subscale at t=1 (28.6%); Figure 3a).

A ceiling effect was present in several (sub)scales, and became more evident at longer follow-up (Figure 3b). The AOFAS pain subscale had a ceiling effect from the t=1 onwards, where 22.9% of patients reported the maximum score. From t=2 onwards, ceiling effects were also noted for AOFAS function (27.0%) and alignment (65.9%) subscales, FFI pain (16.7%) and disability (21.0%) subscales, and SF-36 BP (21.9%) and PF (19.5%) subscales. The AOFAS as a total scale only showed a ceiling effect at t=3; 17.7% of patients reported the maximum score.

Reliability

Internal consistency

The Cronbach's alpha for the AOFAS total scale and function subscale were 0.947 and 0.927, respectively, representing adequate internal consistency (Table 2). The value for the total scale should be interpreted carefully as it contains three subscales. Cronbach's alpha could not be calculated for AOFAS pain and alignment subscales, since these have one item only.

The FFI total scale (α = 0.649) and pain subscale (α = 0.687) did not show adequate internal consistency. For the total scale, this may be explained by the fact that it is not unidimensional. All SF-36 (sub)scales showed adequate internal consistency, with the exception of the subscales general health (α = 0.621) and vitality (α = 0.648).

Construct validity

Spearman's rank correlations regarding construct validity are shown in Table 3. Construct validity was adequate for all AOFAS (sub)scales; out of 17 correlations, 14 (82.4%) were in line with predefined hypotheses for the total scale, 13 (76.5%) for the pain subscale, 15 (88.2%) for the function subscale, and 16 (94.1%) for the alignment subscale.

Reproducibility

Test-Retest reliability

The intraclass correlation coefficient indicates the reliability of each (sub)scale (Table 4). The calculated ICC for the total AOFAS (sub)scales ranged from 0.85 to 0.93, indicating adequate test-retest reliability. The ICC was also proven to be adequate (> 0.70) for all FFI and SF-36 (sub)scales, with the exception of SF-36 subscale General Health perceptions (ICC = 0.64).

Agreement and Smallest Detectable Change

The level of agreement is indicated by the SDC and the corresponding RCI, as listed in Table 4. The SDC was 12.0 (RCI: 12.0%) for the AOFAS total scale, 16.4 (RCI: 16.4%) for the FFI total scale, 10.7 (RCI: 15.3%) for the SF-36 PCS subscale, and 11.36 (RCI: 14.6%) for the SF-36 MCS subscale.

The Bland and Altman analysis (Figure 4 and Table 4) there is no bias in measurements, as the 95% Limits of Agreement for the mean change in scores contains zero for every single (sub)scale.

Responsiveness

Spearman's rank correlation coefficients for longitudinal validity are shown in Table 5. Longitudinal validity was adequate for all AOFAS (sub)scales; out of 17 correlations, 15 (88.2%) were in line with predefined hypotheses for the total scale, 14 (82.5%) for the AOFAS pain subscale, 13 (76.5%) for function subscale, and 17 (100%) for alignment subscale.

The Standardized Response Mean (SRM) and the Effect Size (ES) of the instruments are presented in Table 6. The AOFAS total scale (SRM 1.07, ES 0.89) and function subscale (SRM 1.29, ES 1.06) had a large magnitude of change. The one-item subscales showed a moderate effect size for pain (SRM 0.27) and a small effect size for alignment (SRM < 0.2).

DISCUSSION

The results of this study showed that the AOFAS Ankle-Hindfoot scale Dutch Language Version (AOFAS-DLV) is a valid, reliable, and responsive instrument for measuring symptoms and disability in patients who suffered an ankle fracture.

Floor effects were not present for the AOFAS-DLV in this study. Ceiling effects, on the other hand, did occur. The AOFAS total scale showed a ceiling effect at t=3. Ceiling effects were expected to occur at follow-up moments t=2 and t=3, as most patients were expected to have achieved full recovery (and thus the maximum score) at those follow-up moments. Ceiling effects have been reported in another study for the same reason (20). ,Another study found no ceiling effects for the AOFAS Ankle Hindfoot Scale at six months after elective surgery for a variety of chronic ankle and hindfoot disorders (7).

Evaluating the predictions about Spearman's rank correlations between all (sub)scales, the AOFAS scale as a total showed adequate construct validity. This is in correspondence with previous research, conducted by Ibrahim *et al*. (15). Construct validity also showed to be adequate for all AOFAS subscales separately. The correlations between the AOFAS total score and the SF-36 did show to be higher than the correlations found by SooHoo *et al.* (35). Instead of a high correlation, they found the SF-36 subscales bodily pain, and physical functioning to have a moderate correlation with the AOFAS total scale. The difference in correlation was even bigger for the SF-36 PCS, which SooHoo *et al.* found to have a low, instead of a high correlation with the AOFAS total scale in this study (35). A possible explanation for these differences is the difference in study population, as this study only focused on ankle fractures and SooHoo *et al.* included all injuries of the ankle and hindfoot (35).

As far as conclusions can be drawn, the AOFAS Ankle-Hindfoot scale DLV appears to have adequate internal consistency. Cronbach's α for the AOFAS-scale as a total is 0.947. This value however, should be interpreted carefully as this scale is not unidimensional. Pinsker *et al.* also did find Cronbach's α to be adequate (α = 0.84) for the five patient-reported items of the AOFAS Ankle-Hindfoot scale in the original language (10).

The reliability of the AOFAS DLV is proven to be sufficient, as the ICC for the total AOFAS scale was 0.93. Sufficient reliability has been shown before (7, 15).This reflects the instrument as a whole. Being interested in the performance of the AOFAS DLV as a whole, the intraobserver or interobserver reliability of the physician-reported part alone was not analyzed. The separate subscales also showed to be reliable on an independent level, with ICC of > 0.70 for all AOFAS subscales. Validation studies for the Portuguese and Turkish version of the AOFAS Ankle-Hindfoot scale in patients with variable chronic pathologies and joint injuries, respectively, found similar ICC values of 0.92 ($p < 0.001$) and 0.89 ($p= 0.001$), respectively (36, 37).

Responsiveness of the AOFAS-DLV, considered being a product of longitudinal validity and magnitude of change, was adequate in this study. Concerning longitudinal validity, > 75% of all hypothesized correlations for Spearman's Rho were confirmed, indicating adequate longitudinal validity. This confirms previous studies (9, 12-14). Magnitude of change for the outcome measures was high for the AOFAS Ankle-Hindfoot scale DLV as a whole, with an SRM of 1.07 and ES of 0.89. This is comparable to the magnitude of change for the total FFI (SRM -0.93, ES -0.74) and the SF-36 subscales with the highest magnitude of change (PCS, PF, RP and SF) in our study. Values for SRM and ES of the AOFAS-DLV found in this study are in correspondence with the values found in previous research by SooHoo *et al.* (13), regarding the original AOFAS Ankle-Hindfoot scale

(SRM 1.10, ES 1.12). They are also in line with other studies evaluating the AOFAS and the SEFAS (7).

The level of agreement of the AOFAS total scale compared well to the FFI and SF-36 in this study. The SEM for the AOFAS-DLV was 4.3 points. The SDC was 12.0 points. Similar values for SEM and SDC were found in the validation study of the AOFAS Ankle-Hindfoot Scale in Turkish (SEM, 4.8 points and SDC 13.3 points) (36).

The Bland and Altman analysis showed there is no bias in measurements, as the 95% Limits of Agreement for the mean change in scores contained zero for every single (sub)scale. As the AOFAS-DLV shows sufficient reliability and the level of agreement is equivalent to the level of agreement of the SF-36 and FFI (which are both validated patient-reported outcome measures), the reproducibility of the questionnaire is proven to be acceptable.

A limitation could be the arbitrary choice of t=1 and t=2 for calculating longitudinal validity, ES, and SRM. These measurement properties require the largest change scores. Completing the questionnaires early after trauma (*i.e.*, at two months, low scores expected) and at seven months (*i.e.*, maximum recovery expected) was aimed to achieve the largest change score. Despite good measurement properties of the AOFAS-DLV, a limitation of its use is the fact that a physician has to complete a part of the questionnaire. That makes it unsuitable for, *e.g.,* use in large scale registers. For that purpose, PROMs like the FFI, MOXFQ, and SEFAS may be interesting. The last two have sufficient response rates, internal consistency, test-retest reliability, and responsiveness in patients with surgically treated chronic ankle and hindfoot disorders (6, 7). Data for ankle fractures are not yet available. Current data are in support of using the FFI as PROM.

CONCLUSION

This study evaluated the measurement properties of the AOFAS Ankle-Hindfoot scale Dutch Language Version and confirmed it is a reliable, valid, and responsive measurement instrument for evaluating functional outcome in Dutch patients with a unilateral ankle fracture. This makes the questionnaire suitable for comparing outcome in future studies and after different treatment modalities within this study population or for comparing outcome across hospitals or between patient groups.

ETHICS

This study has been exempted by the medical research ethics committee (MREC) Erasmus MC (Rotterdam, The Netherlands). Each participant provided written consent to participate and remained anonymized during the study. The study is registered at the Netherlands Trial Register (NTR5613; 05-jan-2016).

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. There are no statements to declare relating contributorship, data sharing, or ethics approval.

CONTRIBUTORSHIP STATEMENT

EMMVL, ASDB, DEM, CHVDV, PTDH, WET, and MJHV developed the study. ASDB and EMMVL drafted the manuscript. EMMVL acted as trial principal investigator. ASDB, RJCT, FVDS, CHVDV, PTDH, DEM, and MHJV participated in patient inclusion and outcome assessment. ASDB, WET, and EMMVL performed statistical analysis of the study data. All authors have read and approved the final manuscript.

DATA SHARING STATEMENT

All data is processed in this manuscript. There are no further unpublished data from this study available.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing interests.

LIST OF ABBREVIATIONS

AOFAS, American Orthopedic Foot and Ankle Society; BP, bodily pain; ES, effect size; FFI, Foot Function Index; GH, general health perceptions; ICC, intraclass correlation; MCS, mental component summary; MH, general mental health; PCS, physical component summary; PF, physical functioning; RCI, reliable change index; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SDC, smallest detectable change; SEM, standard error of measurement; SF, social functioning; SF-36, Short Form-36; SRM, standardized response mean; VT, vitality, energy, or fatigue.

REFERENCES

1. Daly PJ, Fitzgerald RH, Jr., Melton LJ, Ilstrup DM. Epidemiology of ankle fractures in Rochester, Minnesota. Acta Orthop Scand. 1987;58(5):539-44.
2. Heffernan G, Khan F, Awan N, Riordain CO, Corrigan J. A comparison of outcome scores in os calcis fractures. Ir J Med Sci. 2000;169(2):127-8.
3. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). Foot Ankle Int. 2005;26(11):968-83.
4. Budiman-Mak E, Conrad KJ, Roach KE. The Foot Function Index: a measure of foot pain and disability. J Clin Epidemiol. 1991;44(6):561-70.
5. Dawson J, Coffey J, Doll H, Lavis G, Cooke P, Herron M, et al. A patient-based questionnaire to assess outcomes of foot surgery: validation in the context of surgery for hallux valgus. Qual Life Res. 2006;15(7):1211-22.
6. Dawson J, Boller I, Doll H, Lavis G, Sharp R, Cooke P, et al. The MOXFQ patient-reported questionnaire: assessment of data quality, reliability and validity in relation to foot and ankle surgery. Foot (Edinb). 2011;21(2):92-102.
7. Coster MC, Rosengren BE, Bremander A, Brudin L, Karlsson MK. Comparison of the Self-reported Foot and Ankle Score (SEFAS) and the American Orthopedic Foot and Ankle Society Score (AOFAS). Foot Ankle Int. 2014;35(10):1031-6.
8. Schneider W, Jurenitsch S. Normative data for the American Orthopedic Foot and Ankle Society ankle-hindfoot, midfoot, hallux and lesser toes clinical rating system. Int Orthop. 2016;40(2):301-6.
9. Kitaoka HB, Alexander IJ, Adelaar RS, Nunley JA, Myerson MS, Sanders M. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. Foot Ankle Int. 1994;15(7):349-53.
10. Pinsker E, Inrig T, Daniels TR, Warmington K, Beaton DE. Reliability and validity of 6 measures of pain, function, and disability for ankle arthroplasty and arthrodesis. Foot Ankle Int. 2015;36(6):617-25.
11. Guyton GP. Theoretical limitations of the AOFAS scoring systems: an analysis using Monte Carlo modeling. Foot Ankle Int. 2001;22(10):779-87.
12. Madeley NJ, Wing KJ, Topliss C, Penner MJ, Glazebrook MA, Younger AS. Responsiveness and validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle

Hindfoot Score, and Foot Function Index in end stage ankle arthritis. Foot Ankle Int. 2012;33(1):57-63.

13. SooHoo NF, Vyas R, Samimi D. Responsiveness of the foot function index, AOFAS clinical rating systems, and SF-36 after foot and ankle surgery. Foot Ankle Int. 2006;27(11):930-4.
14. Westphal T, Piatek S, Halm JP, Schubert S, Winckler S. Outcome of surgically treated intraarticular calcaneus fractures--SF-36 compared with AOFAS and MFS. Acta Orthop Scand. 2004;75(6):750-5.
15. Ibrahim T, Beiri A, Azzabi M, Best AJ, Taylor GJ, Menon DK. Reliability and validity of the subjective component of the American Orthopaedic Foot and Ankle Society clinical rating scales. J Foot Ankle Surg. 2007;46(2):65-74.
16. Van Lieshout EMM, De Boer AS, Meuffels DE, Den Hoed PT, Van Der Vlies CH, Tuinebreijer WE, et al. The American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Score; Study Protocol for the Translation and Validation of the Dutch Language Version. BMJ Open 2016:In press.
17. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186-91.
18. de Haan J, den Hartog D, Tuinebreijer WE, Iordens GI, Breederveld RS, Bronkhorst MW, et al. Functional treatment versus plaster for simple elbow dislocations (FuncSiE): a randomized trial. BMC Musculoskelet Disord. 2010;11:263.
19. Den Hartog D, Van Lieshout EM, Tuinebreijer WE, Polinder S, Van Beeck EF, Breederveld RS, et al. Primary hemiarthroplasty versus conservative treatment for comminuted fractures of the proximal humerus in the elderly (ProCon): a multicenter randomized controlled trial. BMC Musculoskelet Disord. 2010;11:97.
20. Mahabier KC, Den Hartog D, Theyskens N, Verhofstad MHJ, Van Lieshout EMM, Hummer Trial Investigators. Reliability, validity, responsiveness, and minimal important change of the Disablities of the Arm, Shoulder and Hand and Constant-Murley scores in patients with a humeral shaft fracture. J Shoulder Elbow Surg. 2016.
21. Obremskey WT, Dirschl DR, Crowther JD, Craig WL, 3rd, Driver RE, LeCroy CM. Change over time of SF-36 functional outcomes for operatively treated unstable ankle fractures. J Orthop Trauma. 2002;16(1):30-3.

22. Pan SL, Liang HW, Hou WH, Yeh TS. Responsiveness of SF-36 and Lower Extremity Functional Scale for assessing outcomes in traumatic injuries of lower extremities. Injury. 2014;45(11):1759-63.
23. Schep NW, De Haan J, Iordens GI, Tuinebreijer WE, Bronkhorst MW, De Vries MR, et al. A hinged external fixator for complex elbow dislocations: a multicenter prospective cohort study. BMC Musculoskelet Disord. 2011;12:130.
24. van Tetering EA, Buckley RE. Functional outcome (SF-36) of patients with displaced calcaneal fractures compared to SF-36 normative data. Foot Ankle Int. 2004;25(10):733-8.
25. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol. 1998;51(11):903-12.
26. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9.
27. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49.
28. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.
29. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4(4):293-307.
30. Hinkle DE WW, Jurs SG. Applied statistics for the behavioral sciences. Boston USA: Houghton MiZin. 1998.
31. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care. 2001;17(4):479-87.
32. de Boer MR, de Vet HC, Terwee CB, Moll AC, Volker-Dieben HJ, van Rens GH. Changes to the subscales of two vision-related quality of life questionnaires are proposed. J Clin Epidemiol. 2005;58(12):1260-8.
33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.

34. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale (NJ): Lawrence Erlbaum Associates. 1998.
35. SooHoo NF, Shuler M, Fleming LL, American Orthopaedic F, Ankle S. Evaluation of the validity of the AOFAS Clinical Rating Systems by correlation to the SF-36. Foot Ankle Int. 2003;24(1):50-5.
36. Analay Akbaba Y, Celik D, Ogut RT. Translation, Cross-Cultural Adaptation, Reliability, and Validity of Turkish Version of the American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Scale. J Foot Ankle Surg. 2016;55(6):1139-42.
37. Rodrigues RC MD, Mizusaki JM, Imoto AM, Peccin MS, Cohen M, Alloza JFM. Translation, cultural adaptation and validity of the American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Scale. Acta Ortop Bras 2008;16:107–11.

Table 1: Demographic data for the study population

Variable		Outcome
Age (years)		46 (35- 60)
Male gender		75 (52.8%)
Right side affected		58 (40.8%)
Dominant side affected		60 (42.3%)
Malleolar involvement	Unimalleolar	100 (70.4%)
	Bimalleolar	23 (16.2%)
	Trimalleolar	19 (13.4%)
Classification	Weber A	29 (20.4%)
	Weber B	56 (39.4%)
	Weber C	13 (9.2%)
	Unknown	44 (31.0%)
Open fracture		6 (4.2%)
Treatment	Nonoperative	58 (40.8%)
	Operative	84 (59.2%)

Data are shown as median (P_{25}-P_{75}) or as N (%), as applicable.

Table 2: Internal consistency of the instruments used in patients with an ankle fracture

(Sub)scale		N	Number of items	Cronbach's alpha
AOFAS	Total	70	9	0.947[a]
	Pain	70	1	N.A.[b]
	Function	70	7	0.927
	Alignment	70	1	N.A.[b]
FFI	Total	70	23	**0.649[a]**
	Pain	70	9	**0.687**
	Disability	70	9	0.707
	Activity limitation	70	5	0.854
SF-36	Total	70	35	0.882[a]
	PF	70	10	0.932
	RP	70	4	0.885
	BP	70	2	0.733
	GH	70	5	**0.621**
	VT	70	4	**0.648**
	SF	70	2	0.832
	RE	70	3	0.870
	MH	70	5	0.799
	PCS	70	21	0.846[a]
	MCS	70	14	0.861[a]

Data for t=1 were used.

[a] Values should be interpreted carefully because the total scale is not unidimensional.

[b] Not applicable, as this subscale consists of one item only.

Bold and underlined Cronbach alpha values did not exceed the threshold of 0.70.

AOFAS, American Orthopedic Foot and Ankle Society; BP, bodily pain; FFI, Foot Function Index; GH, general health perceptions; MCS, mental component summary; MH, general mental health; N.A., not applicable; PCS, physical component summary; PF, physical

functioning; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SF, social functioning; SF-36, Short Form-36; VT, vitality, energy, or fatigue.

Table 3. Construct validity of the instruments in patients with an ankle fracture

(Sub)scale		AOFAS			
		Pain	**Function**	**Alignment**	**Total**
AOFAS	Pain	1	**0.23 [65]**	0.01 [69]	0.66 [65]
	Function	**0.23 [65]**	1	0.28 [65]	0.85 [65]
	Alignment	0.01 [69]	0.28 [65]	1	**0.35 [65]**
	Total	0.66 [65]	0.85 [65]	**0.35 [65]**	1
FFI	Pain	-0.81 [70]	-0.41 [65]	-0.14 [69]	-0.70 [65]
	Disability	-0.41 [70]	-0.75 [65]	-0.19 [69]	-0.74 [65]
	Activity limiation	-0.34 [70]	-0.80 [65]	-0.23 [69]	-0.77 [65]
	Total	-055 [70]	-0.73 [65]	-0.21 [69]	-0.80 [65]
SF-36	PF	**0.21 [70]**	0.64 [65]	0.21 [69]	0.60 [65]
	RP	0.32 [70]	0.50 [65]	0.19 [69]	**0.58 [65]**
	BP	**0.59 [70]**	0.53 [65]	0.03 [69]	0.67 [65]
	GH	0.15 [70]	-0.01 [65]	-0.09 [69]	0.04 [65]
	VT	0.28 [70]	0.19 [65]	-0.02 [69]	**0.27 [65]**
	SF	0.14 [70]	**0.65 [65]**	0.18 [69]	0.56 [65]
	RE	**0.10 [70]**	0.32 [65]	0.22 [69]	0.33 [65]
	MH	0.24 [70]	0.20 [65]	0.02 [69]	0.24 [65]
	PCS	0.40 [70]	0.62 [65]	0.11 [69]	0.65 [65]
	MCS	0.11 [70]	0.24 [65]	0.13 [69]	0.24 [65]

Data for t=1 were used. Spearman's rank correlation coefficients are given for all possible combinations of (sub)scales, with the N between square brackets. The maximum possible number of patients was 70.

$r > 0.6$ indicates high correlation, $0.3 < r > 0.6$ moderate correlation, and $r < 0.3$ low correlation. Bold and underlined correlations were not hypothesized correctly.

AOFAS, American Orthopedic Foot and Ankle Society; BP, bodily pain; FFI, Foot Function Index; GH, general health perceptions; MCS, mental component summary; MH, general

mental health; PCS, physical component summary; PF, physical functioning; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SF, social functioning; SF-36, Short Form-36; VT, vitality, energy, or fatigue.

Table 4. Intraclass correlation coefficient (ICC) and Bland-Altman analysis of the instruments in ankle fracture patients

(Sub)scale		N	ICC(2,1) (95% CI)	SEM	SDC $_{patient}$	Max score	RCI (%)	Mean$_{difference}$ (SD)	95% Limits of agreement
AOFAS	Pain	125	0.85 (0.78-0.89)	3.5	9.7	40	24.3	1.1 (5.0)	-8.6 to 10.8
	Function	124	0.92 (0.89-0.95)	2.1	5.9	50	11.9	0.6 (3.0)	-5.4 to 6.5
	Alignment	124	0.89 (0.85-0.92)	0.9	2.5	10	24.8	0.2 (1.3)	-2.3 to 2.6
	Total	123	0.93 (0.89-0.95)	4.3	12.0	100	12.0	1.8 (6.1)	-10.2 to 13.9
FFI	Pain	125	0.83 (0.76-0.87)	9.4	26.1	100	26.1	-1.5 (13.3)	-27.6 to 24.6
	Disability	125	0.90 (0.86-0.93)	7.4	20.5	100	20.5	-1.5 (10.5)	-22.0 to 19.0
	Activity limitation	125	0.81 (0.74-0.86)	7.9	22.0	100	22.0	-0.2 (11.2)	-22.2 to 21.8
	Total	125	0.92 (0.89-0.94)	5.9	16.4	100	16.4	-1.2 (8.4)	-17.5 to 15.2
SF-36	PF	120	0.90 (0.87-0.93)	3.18	8.83	56.76	15.6	1.40 (4.50)	-8.43 to 9.23
	RP	124	0.71 (0.59-0.79)	6.36	17.64	55.56	31.7	2.56 (9.00)	-15.07 to 20.20
	BP	124	0.78 (0.70-0.85)	4.07	11.29	60.40	18.7	1.48 (5.76)	-9.80 to 12.77
	GH	125	**0.64 (0.52-0.73)**	5.12	14.20	63.78	22.3	-0.27 (7.24)	-14.47 to 13.93
	VT	123	0.77 (0.68-0.83)	4.06	11.25	68.66	16.4	0.74 (5.74)	-10.51 to 11.99
	SF	124	0.70 (0.60-0.78)	4.89	13.56	57.33	23.7	0.77 (6.92)	-12.79 to 14.32
	RE	124	0.72 (0.63-0.80)	5.31	14.71	55.66	26.4	0.90 (7.50)	-13.81 to 15.60
	MH	125	0.79 (0.70-0.85)	3.86	10.70	63.97	16.7	-1.21 (5.46)	-9.49 to 11.91
	PCS	118	0.85 (0.79-0.89)	3.87	10.72	70.30	15.3	1.10 (5.47)	-9.62 to 11.83
	MCS	118	0.78 (0.70-0.84)	4.10	11.36	77.92	14.6	0.96 (5.80)	-10.42 to 12.30

Change scores were calculated from t=2 to t=3. The maximum possible number of patients was 125. The ICC is shown as correlation coefficient with the 95% CI between brackets. The difference in score from t=2 to t=3 is shown as mean change with SD.

Table 5. Longitudinal validity of the instruments in patients with an ankle fracture

(Sub)scale		AOFAS			
		Pain	**Function**	**Alignment**	**Total**
AOFAS	Pain	1	0.21 [61]	0.12 [65]	0.70 [61]
	Function	0.21 [61]	1	0.05 [61]	0.81 [61]
	Alignment	0.12 [65]	0.05 [61]	1	0.22 [61]
	Total	0.70 [61]	0.81 [61]	0.22 [61]	1
FFI	Pain	**-0.56 [66]**	**-0.19 [61]**	-0.17 [65]	-0.43 [61]
	Disability	-0.24 [66]	-0.66 [61]	-0.07 [65]	-0.60 [61]
	Activity limitation	-0.06 [66]	**-0.59 [61]**	0.09 [65]	-0.50 [61]
	Total	**-0.33 [66]**	-0.61 [61]	-0.03 [65]	**-0.65 [61]**
SF-36	PF	0.25 [66]	**0.44 [61]**	-0.12 [65]	0.48 [61]
	RP	0.26 [65]	**0.34 [60]**	0.01 [64]	0.37 [60]
	BP	0.39 [65]	0.36 [60]	0.06 [64]	0.46 [60]
	GH	-0.02 [66]	-0.13 [61]	0.13 [65]	-0.05 [61]
	VT	**0.38 [66]**	0.26 [61]	0.10 [65]	0.38 [61]
	SF	0.20 [65]	0.54 [60]	0.03 [64]	0.47 [60]
	RE	-0.08 [65]	0.19 [60]	0.15 [64]	**0.14 [60]**
	MH	0.13 [66]	0.09 [61]	0.08 [65]	0.11 [61]
	PCS	0.34 [65]	0.39 [60]	-0.06 [64]	0.45 [60]
	MCS	-0.07 [65]	0.15 [60]	0.14 [64]	0.06 [60]

Change in scores between t=1 and t=2 were used. The maximum possible number of patients was 70. Spearman's rank correlation coefficients are given for all possible combinations of (sub)scales, with the N between square brackets.

The rest of Table caption is identical to Table 3.

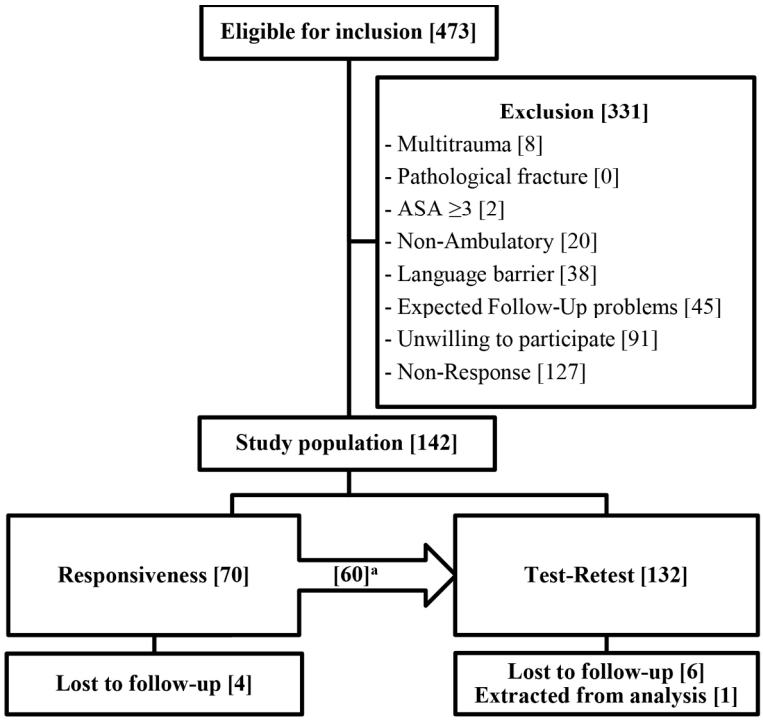
Table 6. Responsiveness: standardized response mean (SRM) and Effect Size (ES) of the instruments in patients with an ankle fracture

(Sub)scale		N	Mean change	SD_{change}	SRM	$SD_{t=1}$	ES
AOFAS	Pain	66	2.3	8.4	0.27	8.9	0.26
	Function	61	12.3	9.5	1.29	11.5	1.06
	Alignment	65	-0.2	1.8	-0.09	2.7	-0.06
	Total	61	15.1	14.1	1.07	16.9	0.89
FFI	Pain	66	-9.1	18.7	-0.49	21.9	-0.42
	Disability	66	-23.3	25.3	-0.92	29.9	-0.78
	Activity limitation	66	-17.9	22.9	-0.78	27.1	-0.66
	Total	66	-17.6	18.9	-0.93	23.9	-0.74
SF-36	PF	66	9.04	10.94	0.83	12.98	0.70
	RP	65	11.95	13.25	0.90	10.94	1.09
	BP	65	7.85	10.33	0.76	9.50	0.83
	GH	66	-0.83	8.56	-0.10	8.42	-0.10
	VT	66	1.74	8.89	0.20	8.06	0.22
	SF	65	13.49	13.53	1.00	14.67	0.92
	RE	65	5.28	12.11	0.44	13.36	0.40
	MH	66	1.31	8.40	0.16	9.10	0.14
	PCS	65	8.88	10.03	0.89	9.65	0.92
	MCS	65	2.68	11.21	0.24	11.61	0.23

Change scores were calculated from t=1 to t=2. The maximum possible number of patients was 70.

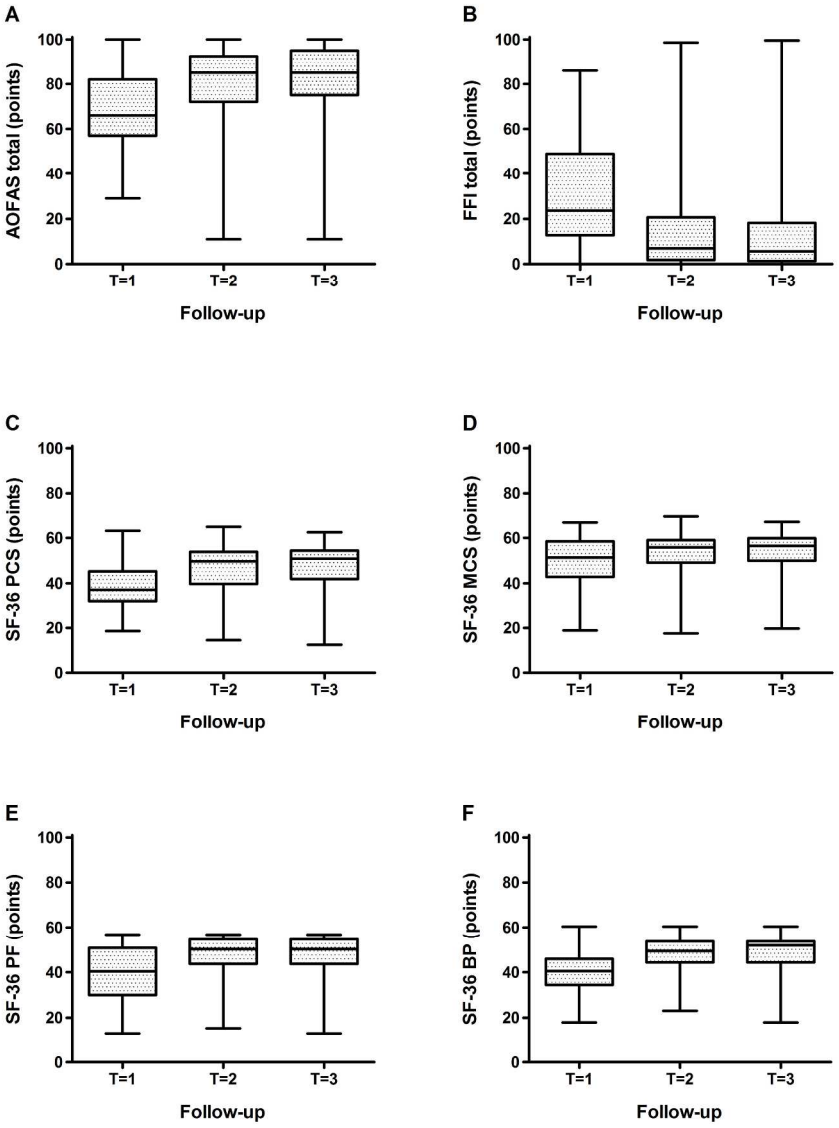
AOFAS, American Orthopedic Foot and Ankle Society; BP, bodily pain; ES, effect size; FFI, Foot Function Index; GH, general health perceptions; MCS, mental component summary; MH, general mental health; PCS, physical component summary; PF, physical functioning; RE, role limitations due to emotional problems; RP, role limitations due to physical health;

SF, social functioning; SF-36, Short Form-36; SRM, standardized response mean; VT, vitality, energy, or fatigue.



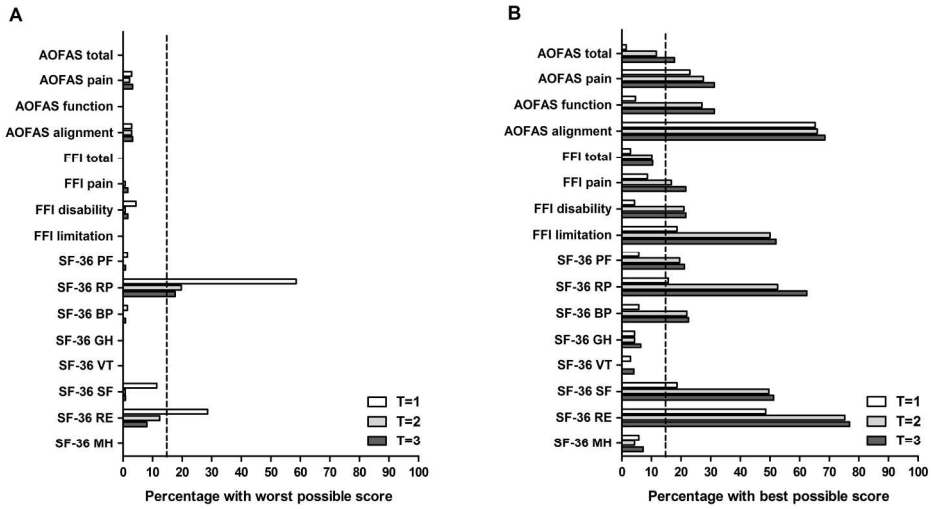
The number of patients in each particular group is shown between square brackets.
a Patients who participated in both groups

170x142mm (300 x 300 DPI)



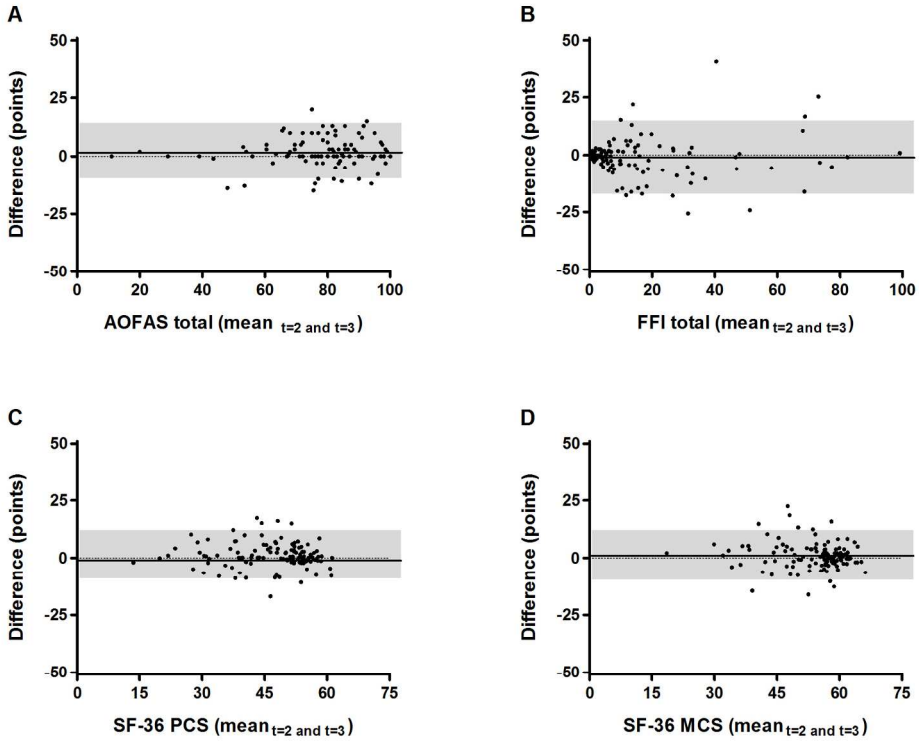
AOFAS, American Orthopedic Foot and Ankle Society; BP, Bodily Pain; FFI, Foot Function Index; MCS, Mental Component Summary; PCS, Physical Component Summary; PF, Physical Functioning; SF-36, Short Form-36.

242x313mm (300 x 300 DPI)



Out of a maximum of 70 at t=1, N=65 for AOFAS function and total, N=69 for AOFAS alignment, and N=70 for AOFAS pain and all (sub)scales of FFI and SF-36.
Out of a maximum of 138 at t=2, N=131 for SF-36 PCS and MCS, N=133 for SF-36 PF, N=136 for SF-36 VT, N=137 for AOFAS function, AOFAS total, and SF-36 RP, BP, SF, and RE, N=138 for AOFAS pain and alignment, all FFI (sub)scales, and SF-36 GH and MH
N=138 for AOFAS pain and alignment, 137 for AOFAS function and AOFAS total.
Out of a maximum of 125 at t=3, N=123 for SF-36 PF, PCS, and MCS, N=124 for AOFAS alignment and total, and SF-36 VT, and N=125 for AOFAS pain and function, all FFI (sub)scales, and SF-36 RP, BP, GH, SF, RE, and MH.
The dotted line represents the acceptable 15% of patients with the maximum score. The SF-36 PCS and MCS did not demonstrate a floor or a ceiling effect and are not displayed.

101x57mm (600 x 600 DPI)



Change scores were calculated from t=2 to t=3.
Each dot represents a single patient. The black line indicates the mean difference. The upper and lower edges of the grey box are the 95% limits of agreement.
AOFAS, American Orthopaedic Foot and Ankle Society; FFI, Foot Function Index; MCS, Mental Component Summary; PCS, Physical Component Summary; SF-36, Short Form-36.

180x145mm (300 x 300 DPI)

SUPPLEMENTAL MATERIAL

Supplemental Table 1: AOFAS Ankle-Hindfoot Scale Dutch Language Version

Pijn
- ☐ Geen
- ☐ Mild, af en toe
- ☐ Matig, dagelijks
- ☐ Ernstig, bijna altijd aanwezig

Functie
Beperkingen in activiteiten, hulpmiddelengebruik
- ☐ Geen beperkingen; geen hulpmiddelen nodig
- ☐ Geen beperkingen bij dagelijkse activiteiten, wel beperkingen bij recreatieve activiteiten; geen hulpmiddelen nodig
- ☐ Beperkingen bij dagelijkse en recreatieve activiteiten; gebruik van een stok
- ☐ Ernstige beperkingen bij dagelijkse en recreatieve activiteiten; gebruik van een brace, krukken, looprek, rollator of rolstoel

Maximale loopafstand
- ☐ Meer dan 600 meter
- ☐ 400 tot 600 meter
- ☐ 100 tot 400 meter
- ☐ Minder dan 100 meter

Loopondergrond
- ☐ Op geen enkele ondergrond problemen
- ☐ Enige moeite met lopen op oneffen terrein, trappen, hellingen of ladders
- ☐ Veel moeite met lopen op oneffen terrein, trappen, hellingen of ladders

Let op: onderstaande vragen worden door de arts ingevuld.

Afwijkende loopgang
- ☐ Geen tot gering
- ☐ Duidelijk
- ☐ Zeer opvallend

Sagittale beweging (dorsoflexie plus plantairflexie)
- ☐ Normaal of geringe beperking (30° of meer)
- ☐ Matige beperking (15-29°)
- ☐ Ernstige beperking (minder dan 15°)

Achtervoetbeweging (inversie plus eversie)
- ☐ Normaal of geringe beperking (75%-100% van normaal)
- ☐ Matige beperking (25-74% van normaal)
- ☐ Opvallende beperking (minder dan 25% van normaal)

Enkel-achtervoet stabiliteit (anteroposterieur, varus-valgus)
- ☐ Stabiel
- ☐ Evident instabiel

Alignement
- ☐ Goed, plantigrade voet, enkel-achtervoet fraai gealigneerd
- ☐ Redelijk, plantigrade voet, enige mate van enkel-achtervoet malalignement, geen klachten of symptomen
- ☐ Slecht, geen plantigrade voet, ernstige malalignement met klachten of symptomen

Supplemental Table 2A. Hypothesized correlations between the instruments for construct validity in patients with an ankle fracture

(Sub)scale		AOFAS			
		Pain	**Function**	**Alignment**	**Total**
AOFAS	Pain	N.A.	moderate	low	high
	Function	moderate	N.A.	low	high
	Alignment	low	low	N.A.	low
	Total	high	high	low	N.A.
FFI	Pain	high	moderate	low	high
	Disability	moderate	high	low	high
	Activity limitation	moderate	high	low	high
	Total	moderate	high	low	high
SF-36	PF	moderate	high	low	high
	RP	moderate	moderate	low	high
	BP	high	moderate	low	high
	GH	low	low	low	low
	VT	low	low	low	moderate
	SF	low	moderate	low	moderate
	RE	moderate	moderate	low	moderate
	MH	low	low	low	low
	PCS	moderate	high	low	high
	MCS	low	low	low	low

Supplemental Table 2B. Hypothesized correlations between the instruments for longitudinal validity in patients with an ankle fracture

(Sub)scale		AOFAS			
		Pain	**Function**	**Alignment**	**Total**
AOFAS	Pain	N.A.	low	low	high
	Function	low	N.A.	low	high
	Alignment	low	low	N.A.	low
	Total	high	high	low	N.A.
FFI	Pain	high	moderate	low	moderate
	Disability	low	high	low	moderate
	Activity limitation	low	high	low	moderate
	Total	low	high	low	moderate
SF-36	PF	low	high	low	moderate
	RP	low	low	low	moderate
	BP	moderate	moderate	low	moderate
	GH	low	low	low	low
	VT	low	low	low	moderate
	SF	low	moderate	low	moderate
	RE	low	low	low	moderate
	MH	low	low	low	low
	PCS	moderate	moderate	low	moderate
	MCS	low	low	low	low

Expected strength of correlation for all possible combinations; $r > 0.6$ indicates high correlation, $0.3 < r > 0.6$ moderate correlation, and $r > 0.6$ low correlation.

AOFAS, American Orthopedic Foot and Ankle Society; BP, bodily pain; FFI, Foot Function Index; GH, general health perceptions; MCS, mental component summary; MH, general mental health; PCS, physical component summary; PF, physical functioning; RE, role limitations due to emotional problems; RP, role limitations due to physical health; SF, social functioning; SF-36, Short Form-36; VT, vitality, energy, or fatigue.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3, 4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5, 6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	8, 9
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9, 11
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	8 - 11
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	9-11
		(*b*) Describe any methods used to examine subgroups and interactions	7-11
		(*c*) Explain how missing data were addressed	9
		(*d*) If applicable, explain how loss to follow-up was addressed	7-11
		(*e*) Describe any sensitivity analyses	9-11
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	12
		(b) Give reasons for non-participation at each stage	12
		(c) Consider use of a flow diagram	22 (Figure 1)
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	12
		(b) Indicate number of participants with missing data for each variable of interest	12
		(c) Summarise follow-up time (eg, average and total amount)	12
Outcome data	15*	Report numbers of outcome events or summary measures over time	12 – 14
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	N/A
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12-14
Discussion			
Key results	18	Summarise key results with reference to study objectives	15-17
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-17
Generalisability	21	Discuss the generalisability (external validity) of the study results	15-17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.