

BMJ Open The American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Scale; translation and validation of the Dutch language version for ankle fractures

A Siebe de Boer,¹ Roderik J C Tjioe,¹ Fleur Van der Sijde,¹ Duncan E Meuffels,² Pieter T den Hoed,³ Cornelis H Van der Vlies,⁴ Wim E Tuinebreijer,¹ Michael H J Verhofstad,¹ Esther M M Van Lieshout,¹ AOFAS study Group

To cite: de Boer AS, Tjioe RJC, Van der Sijde F, *et al.* The American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Scale; translation and validation of the Dutch language version for ankle fractures. *BMJ Open* 2017;**7**:e017040. doi:10.1136/bmjopen-2017-017040

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2017-017040>).

Received 28 March 2017

Revised 10 May 2017

Accepted 13 June 2017



CrossMark

¹Trauma Research Unit Department of Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

²Department of Orthopedic Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

³Department of Surgery, Ikazia Hospital, Rotterdam, The Netherlands

⁴Department of Surgery, Maastad Hospital, Rotterdam, The Netherlands

Correspondence to

Professor Esther M M Van Lieshout; e.vanlieshout@erasmusmc.nl

ABSTRACT

Objectives The American Orthopaedic 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 Multicentre (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 and the Short Form-36 were completed by the patient. Descriptive statistics (including floor and ceiling effects), reliability (ie, internal consistency), construct validity, reproducibility (ie, 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 6 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 The Netherlands Trial Register (NTR5613; 05-jan-2016).

Strengths and limitations of this study

- This prospective, multicentre, observational study shows substantial, previously unknown information about the performance of the American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Scale.
- The topic of the clinical study is relevant for orthopaedic 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 guidelines.
- Although the study is mostly relevant for the Dutch-speaking regions, it is also informative for other regions.
- Implementation of the (AOFAS) 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, for example, 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 industrialised 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,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 Orthopaedic Foot and Ankle Society (AOFAS), 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 (ie, 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 fractures¹⁴ 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 multicentre, prospective, observational study design (ie, 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 centres. 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 prefinal 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 (see online supplementary table 1).

Validation

Patient recruitment

Patients were recruited from 1 May 2014 to 29 March 2016. Patients were identified from hospital records, based on their International Coding of Diseases, 10th revision (ICD-10) code or Diagnosis Related Group (DRG; in Dutch, DBC) code. Inclusion criteria were: 1) unilateral ankle fracture; 2) age 18 years or older and 3) provision of informed consent by the patient. Treatment should have been started between 6 weeks and 3 months and/or between 7 and 9 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 (ie, American Society of Anaesthesiologists ≥ 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, 10 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-reported and physician-reported part could be answered; a low score was expected. At 7 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

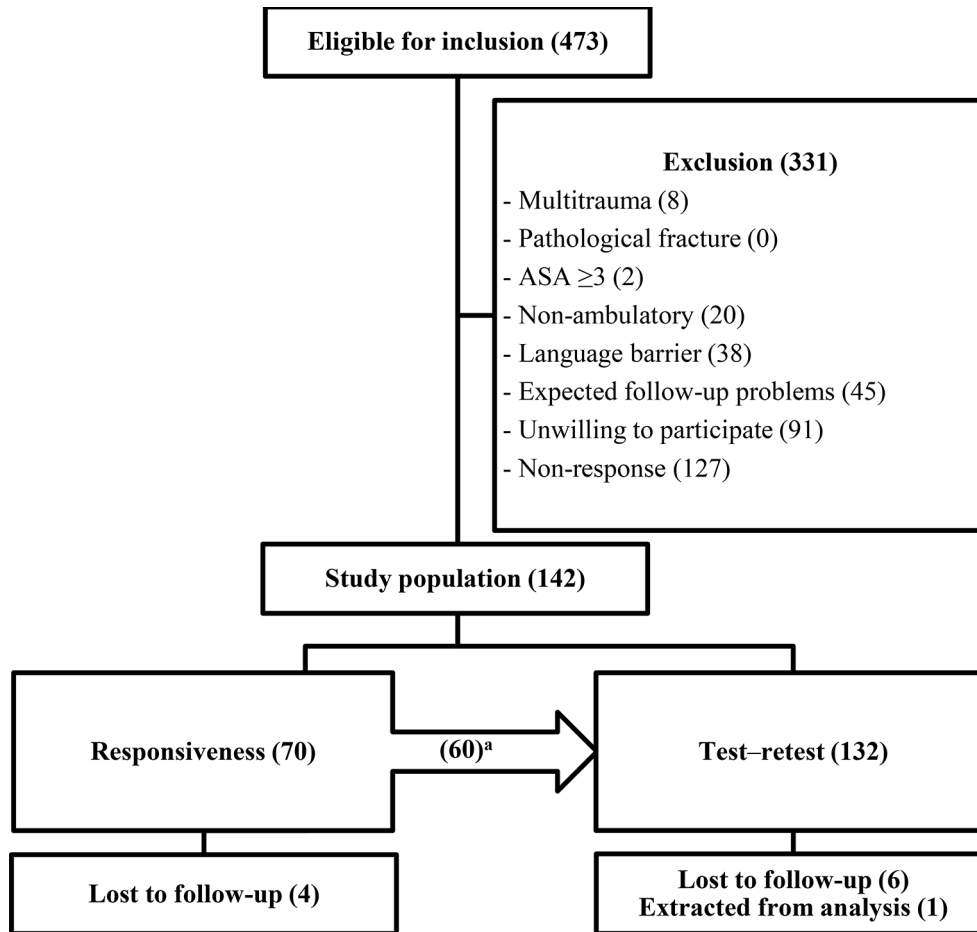


Figure 1 Flowchart. ^aPatients who participated in both groups. ASA, American Society of Anesthesiologists.

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	
Non-operative	58 (40.8%)
Operative	84 (59.2%)

Data are shown as median (P₂₅-P₇₅) or as N (%), as applicable.

(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 0 (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 AOFAS that is part of the physician-reported part of the AOFAS Ankle-Hindfoot Scale-DLV using a standardised 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

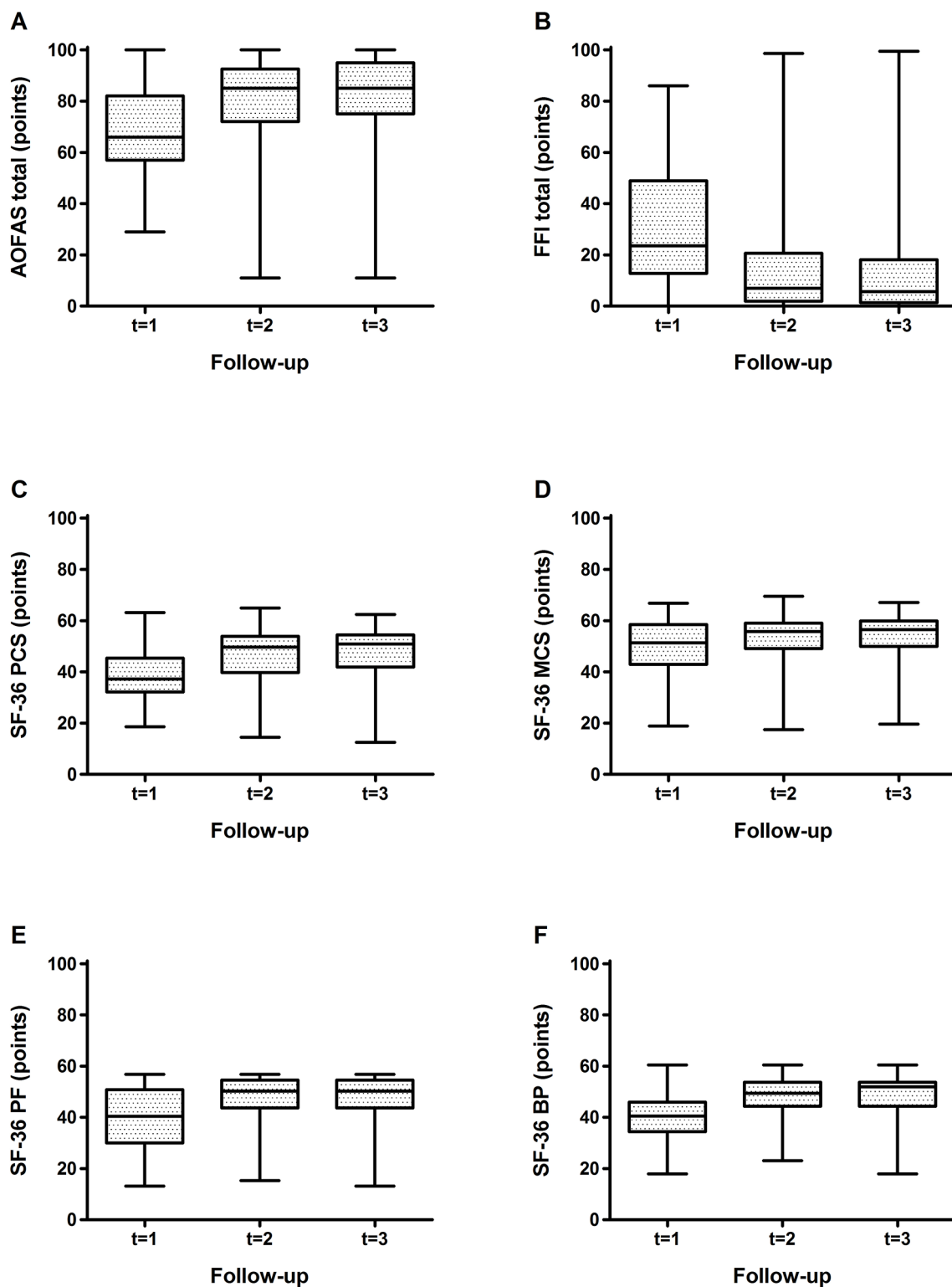


Figure 2 OFAS Ankle-Hindfoot (A), Foot Function Index (B), Short Form-36 PCS (C), SF-36 MCS (D), SF-36 PF (E), and SF-36 BP (F) scores at each follow-up visit in patients with an ankle fracture. AOFAS, American Orthopaedic Foot and Ankle Society; BP, bodily pain; FFI, Foot Function Index; MCS, mental component summary; PCS, physical component summary; PF, physical functioning; SF-36, Short Form-36.

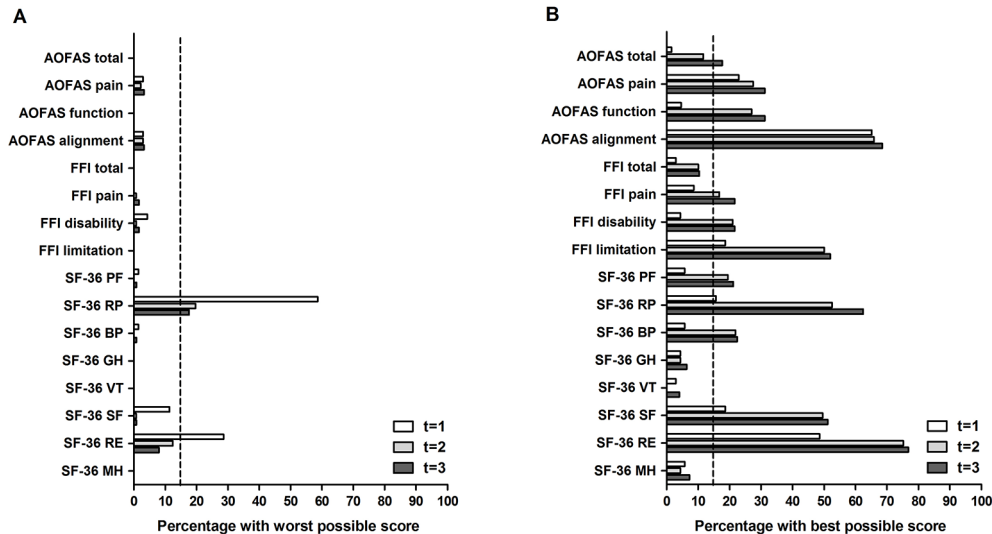


Figure 3 Floor effects (A) and ceiling effects (B) of the instruments used in patients with an ankle fracture. Out of a maximum of 70 at t=1, N=65 for AOFAS function and total, N=69 for AOFAS alignment, and N=70 for AOFAS pain and all (sub)scales of FFI and SF-36. Out of a maximum of 138 at t=2, N=131 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. AOFAS, American Orthopaedic Foot and Ankle Society; FFI, Foot Function Index; SF-36, Short Form-36.

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, V.21). Data are reported following the STrengthening the Reporting of OBservational studies in Epidemiology.²⁶ Since raw data for individual items were analysed, 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 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 α 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 categorised 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 online supplementary table 2A.

Evaluation of the test-retest reliability was performed by calculating the intraclass correlation coefficient ($ICC_{\text{agreement}}$) of (sub)scales administered at t=2 and t=3. ICC is reported with 95% 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 SE of measurement ($SEM_{\text{agreement}}$). For individual patients, the smallest detectable change (SDC) was calculated as $1.96 \times \sqrt{2} \times SEM$.²⁸ The SDC measurable in a group of people (SDC_{group}) was calculated by dividing the SDC in individuals (SDC_{ind}) by \sqrt{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 ($\text{mean}_{\text{change}} \pm 1.96 \times SD$ 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 (online supplementary 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 standardised 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 $(t=2-t=1)/SD_{T_1}$ (28). SRM was calculated as change in score $(t=2-t=1)/SD_{\text{change}}$.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 α 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 α could not be calculated for AOFAS pain and alignment subscales, since these have one item only.

The FFI total scale ($\alpha=0.649$) and pain subscale ($\alpha=0.687$) 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 ($\alpha=0.621$) and vitality ($\alpha=0.648$).

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.

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

(Sub)scale	N	Number of items	Cronbach's α
AOFAS			
Total	70	9	0.947*
Pain	70	1	N.A.†
Function	70	7	0.927
Alignment	70	1	N.A.†
FFI			
Total	70	23	0.649*
Pain	70	9	0.687
Disability	70	9	0.707
Activity limitation	70	5	0.854
SF-36			
Total	70	35	0.882*
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*
MCS	70	14	0.861*

Data for t=1 were used.

Bold and underlined Cronbach's α values did not exceed the threshold of 0.70.

*Values should be interpreted carefully because the total scale is not unidimensional.

†Not applicable, as this subscale consists of one item only.

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]
Disability	-0.41 [70]	-0.75 [65]	-0.19 [69]	-0.74 [65]
Activity limitation	-0.34 [70]	-0.80 [65]	-0.23 [69]	-0.77 [65]
Total	-0.55 [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 hypothesised 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.

Reproducibility

Test-retest reliability

The ICC 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.

For the Bland-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 SRM and the 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 subscale showed a moderate ES for pain (SRM 0.27) and a small ES for alignment (SRM <0.2).

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

(Sub)scale	N	ICC ^{1,2} (95% CI)	SEM	SDC	Max score	RCI (%)	Mean _{difference} (SD)	95% limits of agreement
AOFAS								
Pain	125	0.85 (0.78 to 0.89)	3.5	9.7	40	24.3	1.1 (5.0)	-8.6 to 10.8
Function	124	0.92 (0.89 to 0.95)	2.1	5.9	50	11.9	0.6 (3.0)	-5.4 to 6.5
Alignment	124	0.89 (0.85 to 0.92)	0.9	2.5	10	24.8	0.2 (1.3)	-2.3 to 2.6
Total	123	0.93 (0.89 to 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 to 0.87)	9.4	26.1	100	26.1	-1.5 (13.3)	-27.6 to 24.6
Disability	125	0.90 (0.86 to 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 to 0.86)	7.9	22.0	100	22.0	-0.2 (11.2)	-22.2 to 21.8
Total	125	0.92 (0.89 to 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 to 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 to 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 to 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 to 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 to 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 to 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 to 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 to 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 to 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 to 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 parentheses. The difference in score from t=2 to t=3 is shown as mean change with SD.

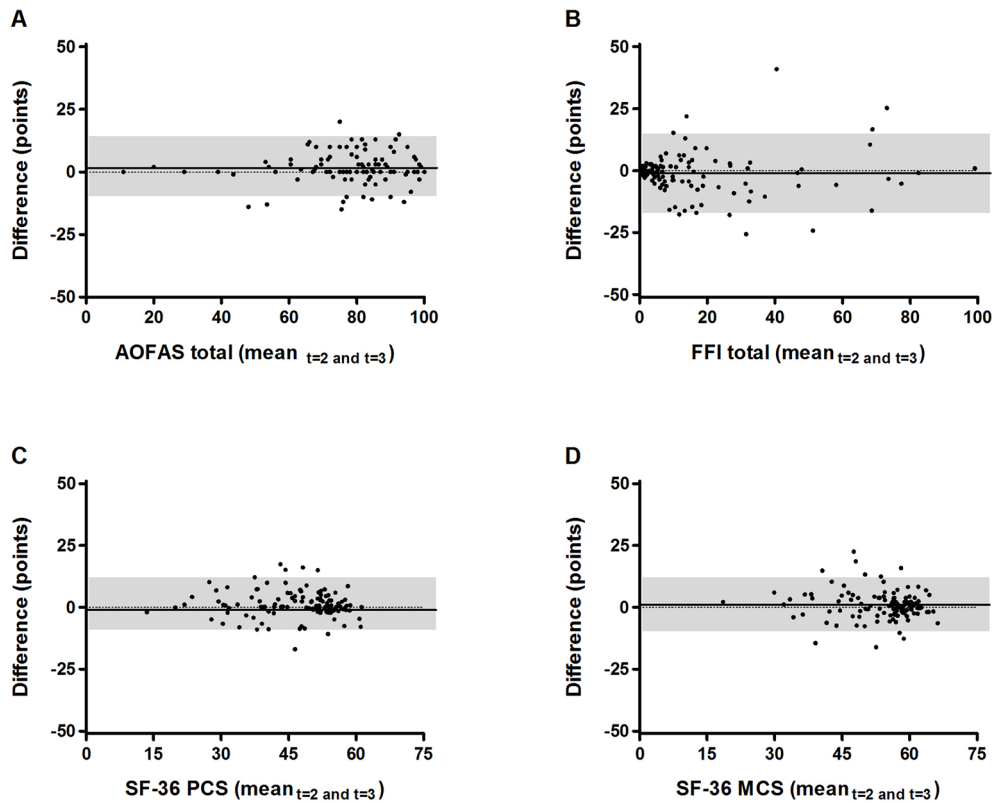


Figure 4 Bland-Altman plots for AOFAS Ankle-Hindfoot (A), Foot Function Index (B), Short Form-36 PCS (C), and SF-36 MCS (D) scores in patients with an ankle fracture. Change scores were calculated from t=2 to t=3. Each dot represents a single patient. The black line indicates the mean difference. The 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.

DISCUSSION

The results of this study showed that the AOFAS-DLV is a valid, reliable and responsive instrument for measuring symptoms and disability in patients who suffered an ankle fracture.

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 6 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 ($\alpha=0.84$) 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 analysed. 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

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](#).

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 hypothesised 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 with 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-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 (ie, at 2 months, low scores expected) and at 7 months (ie, 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, for example, use in large-scale registers. For that purpose, PROMs like the FFI, MOXFQ and

Table 6 Responsiveness: standardised 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 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, standardised response mean; VT, vitality, energy or fatigue.

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 DLV 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.

Correction notice This article has been corrected since it first published. The Research group and Collaborators have been added.

Collaborators AOFAS study Group: D A Newhall, E N Van derVelden-Macauley, L Vellekoop, J Romeo

Contributors 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 MJHV 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.

Funding 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, datasharing or ethics approval.

Competing interests None declared.

Ethics approval Medical Research Ethics Committees or Local Ethics Boards of all participating centres.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data are processed in this manuscript. There are no further unpublished data from the study available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Daly PJ, Fitzgerald RH, Melton LJ, *et al.* Epidemiology of ankle fractures in Rochester, Minnesota. *Acta Orthop Scand* 1987;58:539–44.
- Heffernan G, Khan F, Awan N, *et al.* A comparison of outcome scores in os calcis fractures. *Ir J Med Sci* 2000;169:127–8.
- Martin RL, Irrgang JJ, Burdett RG, *et al.* Evidence of validity for the foot and ankle ability measure (FAAM). *Foot Ankle Int* 2005;26:968–83.
- Budiman-Mak E, Conrad KJ, Roach KE. The Foot function index: a measure of foot pain and disability. *J Clin Epidemiol* 1991;44:561–70.
- Dawson J, Coffey J, Doll H, *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;15:1211–22.
- Dawson J, Boller I, Doll H, *et al.* The MOXFQ patient-reported questionnaire: assessment of data quality, reliability and validity in relation to foot and ankle surgery. *Foot* 2011;21:92–102.
- Cöster MC, Rosengren BE, Bremander A, *et al.* Comparison of the self-reported foot and ankle score (SEFAS) and the American Orthopedic Foot and Ankle Society score (AOFAS). *Foot Ankle Int* 2014;35:1031–6.
- 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;40:301–6.
- Kitaoka HB, Alexander IJ, Adelaar RS, *et al.* Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. *Foot Ankle Int* 1994;15:349–53.
- Pinsker E, Inrig T, Daniels TR, *et al.* Reliability and validity of 6 measures of pain, function, and disability for ankle arthroplasty and arthrodesis. *Foot Ankle Int* 2015;36:617–25.
- Guyton GP. Theoretical limitations of the AOFAS scoring systems: an analysis using Monte Carlo modeling. *Foot Ankle Int* 2001;22:779–87.
- 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;33:57–63.
- 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;27:930–4.
- 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;75:750–5.
- 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;46:65–74.
- Van Lieshout EMM, De Boer AS, Meuffels DE, *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*. In Press. 2016.
- Beaton DE, Bombardier C, Guillemin F, *et al.* Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
- 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.
- Den Hartog D, Van Lieshout EM, 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.
- Mahabier KC, Den Hartog D, Theyskens N, *et al.* Reliability, validity, responsiveness, and minimal important change of the disabilities of the arms, shoulder and hand and Constant-Murley scores in patients with a humeral shaft fracture. *J Shoulder Elbow Surg* 2017;26:e1–e12.
- Obremsky 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;16:30–3.
- 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;45:1759–63.
- Schep NW, De Haan J, Iordens GI, *et al.* A hinged external fixator for complex elbow dislocations: a multicenter prospective cohort study. *BMC Musculoskelet Disord* 2011;12:130.
- 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;25:733–8.
- Ware JE, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol* 1998;51:903–12.
- von Elm E, Altman DG, Egger M, *et al.* The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344–9.
- Mokkink LB, Terwee CB, Patrick 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;19:539–49.
- Terwee CB, Bot SD, de Boer MR, *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
- Hinkle DE WW, Jurs SG. *Applied statistics for the behavioral sciences*. Boston USA: Houghton Mifflin, 1998.
- 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;17:479–87.
- 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;58:1260–8.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
- Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale (NJ): Lawrence Erlbaum Associates, 1998.
- 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;24:50–5.
- Analay Akbaba Y, Celik D, Ogut RT *et al.* 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;55:1139–42.
- Rodrigues RC MD, Mizusaki JM, Imoto AM, *et al.* Translation, cultural adaptation and validity of the American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Scale. *Acta Ortop Bras* 2008;16:107–11.