

BMJ Open Validation of the American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Scale Dutch language version in patients with hindfoot fractures

A Siebe De Boer,¹ Duncan E Meuffels,² Cornelis H Van der Vlies,³ P Ted Den Hoed,⁴ Wim E Tuinebreijer,¹ Michael H J Verhofstad,¹ Esther M M Van Lieshout,¹ AOFAS Study Group

To cite: De Boer AS, Meuffels DE, Van der Vlies CH, *et al.* Validation of the American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Scale Dutch language version in patients with hindfoot fractures. *BMJ Open* 2017;7:e018314. doi:10.1136/bmjopen-2017-018314

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-018314>).

Received 19 June 2017
Revised 7 September 2017
Accepted 29 September 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, Maastad Hospital, Rotterdam, The Netherlands

⁴Department of Surgery, Ikazia 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 used questionnaires for measuring functional recovery after a hindfoot injury. Recently, this instrument was translated and culturally adapted into a Dutch version. In this study, the measurement properties of the Dutch language version (DLV) were investigated in patients with a unilateral hindfoot fracture.

Design Multicentre, prospective observational study.

Setting This multicentre study was conducted in three Dutch hospitals.

Participants In total, 118 patients with a unilateral hindfoot fracture were included. Three patients were lost to follow-up.

Primary and secondary outcome measures

Patients were asked to complete the AOFAS-DLV, the Foot Function Index and the Short Form-36 on three occasions. Descriptive statistics (including floor and ceiling effects), reliability (ie, internal consistency), construct validity, reproducibility (ie, test-retest reliability, agreement and smallest detectable change (SDC)) and responsiveness were determined.

Results Internal consistency was inadequate for the AOFAS-DLV total scale ($\alpha=0.585$), but adequate for the function subscale ($\alpha=0.863$). The questionnaire had adequate construct validity (82.4% of predefined hypotheses were confirmed), but inadequate longitudinal validity (70.6%). No floor effects were found, but ceiling effects were present in all AOFAS-DLV (sub)scales, most pronounced from 6 to 24 months after trauma onwards. Responsiveness was only adequate for the pain and alignment subscales, with a SDC of 1.7 points.

Conclusions The AOFAS Ankle-Hindfoot Scale DLV has adequate construct validity and is reliable, making it a suitable instrument for cross-sectional studies investigating functional outcome in patients with a hindfoot fracture. The inadequate longitudinal validity and responsiveness, however, hamper the use of the questionnaire in longitudinal studies and for assessing long-term functional outcome.

Trial registration number NTR5613; Post-results.

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

Hindfoot fractures are rare, but invalidating injuries. Since most patients are in their wage-earning age combined with the long-term disabilities, these injuries have a high socioeconomic impact.^{1,2} The incidence rate of calcaneal fractures is 11.5 per 100 000 person-years and these fractures occur 2.4 times more frequently in men than women.³ Fractures of the talus are even more rare with a reported annual incidence of 3.2 per 100 000, and occur 4.5 times more often in men.⁴ Despite the facts that these fractures are relatively rare, they have received considerable attention in recent literature, presumable

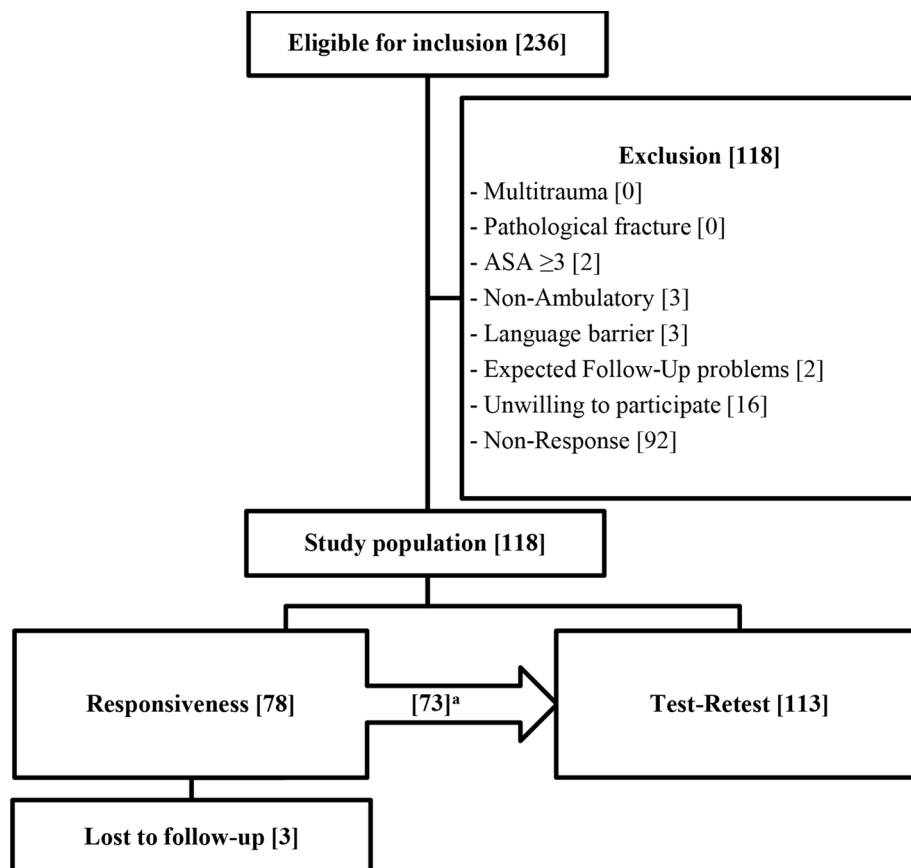


Figure 1 Flowchart. The number of patients in each particular group is shown between square brackets. ^aPatients who participated in both groups. ASA, American Society of Anesthesiologists.

by the long-term recovery and therewith socioeconomic burden.

To monitor functional outcome, quality of life and recovery after treatment, patient-reported outcome measures and other instruments are increasingly used in clinical practice and clinical research. The American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Scale is one of the most used assessment tools in foot surgery.⁵ This clinical rating system combines a patient-reported part and a physician-reported part. In its original language version the AOFAS Ankle-Hindfoot Scale, as a complete scale has been shown to be responsive and valid.⁶⁻⁹ The study populations involved non-traumatic diagnoses, such as general ankle-hindfoot complaints,⁸ pending ankle or foot surgery¹⁰ and end-stage ankle osteoarthritis.⁷

Recently, a Dutch version of the AOFAS Ankle-Hindfoot Scale became available.¹¹ It was translated and culturally adapted to the Dutch population according to the guideline for Cross-Cultural Adaptation of Self-Report Measures.^{12 13} The AOFAS Ankle-Hindfoot Scale was shown to be valid, reliable and responsive in patients with an ankle fracture.¹¹ Thus study aimed to determine the measurement properties of the AOFAS-Dutch language version (DLV) in patients who sustained a hindfoot fracture.

METHODS

Study design and ethics statement

This multicentre, prospective, observational study was performed at three hospitals. The study is registered at the Netherlands Trial Register (NTR5613). A detailed study protocol is published elsewhere.¹³ The Medical Research Ethics Committees or Local Ethics Boards of all participating centres approved the study.

Patient recruitment

Patients were recruited from 1 May 2014 to 1 November 2016. Patients were identified from hospital records, based on their International Coding of Diseases, 10th revision (ICD-10) code or Diagnosis-Related Group code. Inclusion criteria were: (1) unilateral hindfoot fracture; (2) age 18 years or older and (3) provision of informed consent. Exclusion criteria were: (1) multiple trauma affecting 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 and (6) expected problems of maintaining follow-up.

A total of 118 individual patients were included; 78 completed $t=1$ and $t=2$, and 113 completed $t=2$ and $t=3$ (figure 1). Three patients were lost to follow-up during the course of the study.

Table 1 Demographic data for the study population

Variable	Outcome
Age (years)	51 (36–58)
Male gender	69 (61.1%)
Right side affected	101 (89.4%)
Dominant side affected	60 (53.1%)
Calcaneal fracture	82 (72.6%)
Talar fracture	36 (31.9%)
Chopart luxation	1 (0.9%)
Closed fracture	113 (100.0%)
Treatment	
Non-operative	72 (73.6%)
Operative	41 (36.3%)

Data are shown as median (P₂₅–P₇₅) or as n (%), as applicable.

The median age was 51 years (P₂₅–P₇₅ 36–58) and the majority of patients (n=69; 61.1%) were men (table 1). The most common injuries were calcaneal fractures (n=82; 72.6%) and talar fractures (n=36; 31.9%). Fractures were mostly treated non-operatively (n=72; 73.6%).

Questionnaires and data collection

Demographic, injury and treatment data were collected from the patient's medical files. To complete the physician-reported part of the AOFAS Ankle-Hindfoot Scale-DLV, a research physician or research assistant performed the physical examination using a standardised protocol. Patients were asked to complete the AOFAS-DLV patient-reported part, Foot Function Index (FFI-DLV) and the Short

Table 2 Overview of measurement properties and definitions used

Measurement property	Definition/calculation	Result	Data
Floor and ceiling effects	Percentage of participants with lowest or highest possible score	>15% ^{13 30 31}	T1–T3
Reliability			
Internal consistency	Cronbach's α value for (sub)scales	0.70–0.95 for unidimensional (sub)scale ³¹	T1
Construct validity	Spearman's rank correlation (r) of scores between (sub)scales ³¹ Strength of correlation categorised as high (r>0.6), moderate (0.3<r< 0.6) or low (r<0.3) ³²	≥75% of correlations in line with predefined hypotheses at n≥50 (online supplementary table 1A,B) ³¹	T1
Test–retest reliability			
ICC _{agreement}	ICC _{agreement} with 95% CI	>0.70 at n≥50 (³¹)	T2, T3
Absolute agreement			
SEM _{agreement} and SDC	SDC _{individual} =1.96×√2×SEM ³¹ SDC _{group} =SDC _{individual} /√n ^{33 34}		T2, T3
RCI	RCI=SDC _{group} /maximum score×100%		T2, T3
Bland-Altman analysis	95% limits of agreement=(mean _{change T3–T2} ±1.96×SD _{change}) ^{31 35}	Zero outside interval indicates measurement bias	T2, T3
Responsiveness			
Longitudinal validity	Spearman's rank correlation of changes in scores (Score _{T2} –Score _{T1}) between (sub)scales. ³¹ Strength of correlation categorised as high (r>0.6), moderate (0.3<r< 0.6) or low (r<0.3) ³²	≥75% of correlations in line with predefined hypotheses at n≥50 (online supplementary table 1A,B) ³¹	T1, T2
Magnitude of change			
ES	ES=(Score _{T2} –Score _{T1})/SD _{T1} ³¹ Effect rated as small (0.2–0.4), moderate (0.5–0.7) or large (≥0.8) ³⁶		T1, T2
SRM	SRM=(Score _{T2} –Score _{T1})/SD _{change} ³¹ Effect rated as small (0.2–0.4), moderate (0.5–0.7) or large (≥0.8) ³⁶		T1, T2

ES, effect size; ICC, intraclass correlation coefficient; RCI, Reliable Change Index, SDC, smallest detectable change; SEM, SE of measurement; SRM, standardised response mean.

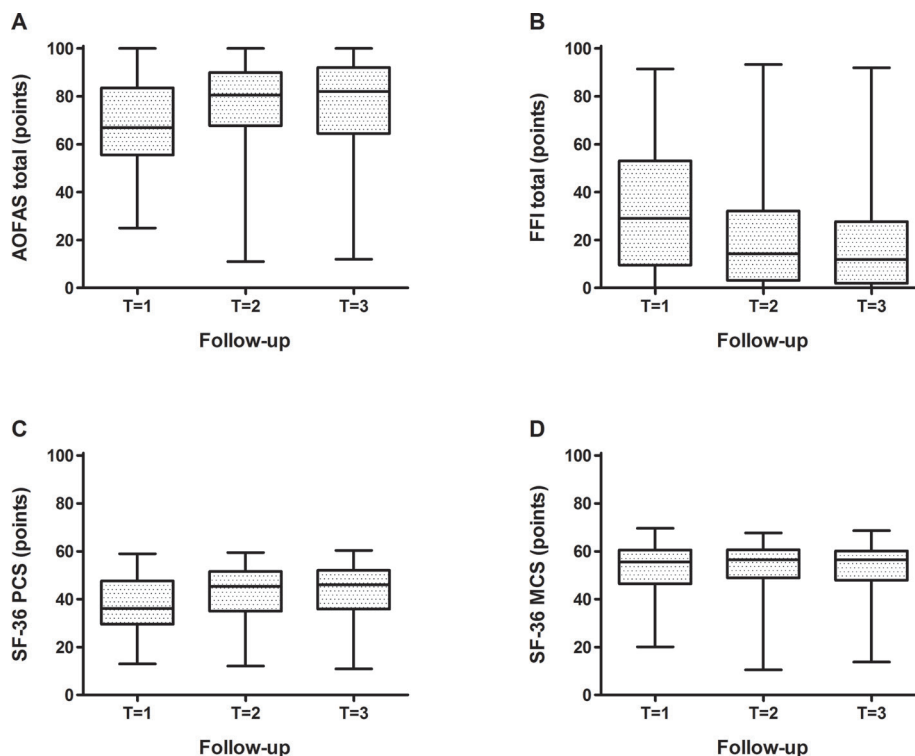


Figure 2 AOFAS Ankle-Hindfoot (A), Foot Function Index (B), SF-36 PCS (C) and SF-36 MCS (D) scores at each follow-up visit in patients with an ankle fracture. AOFAS, American Orthopaedic Foot and Ankle Society; FFI, Foot Function Index; MCS, mental component summary; PCS, physical component summary;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

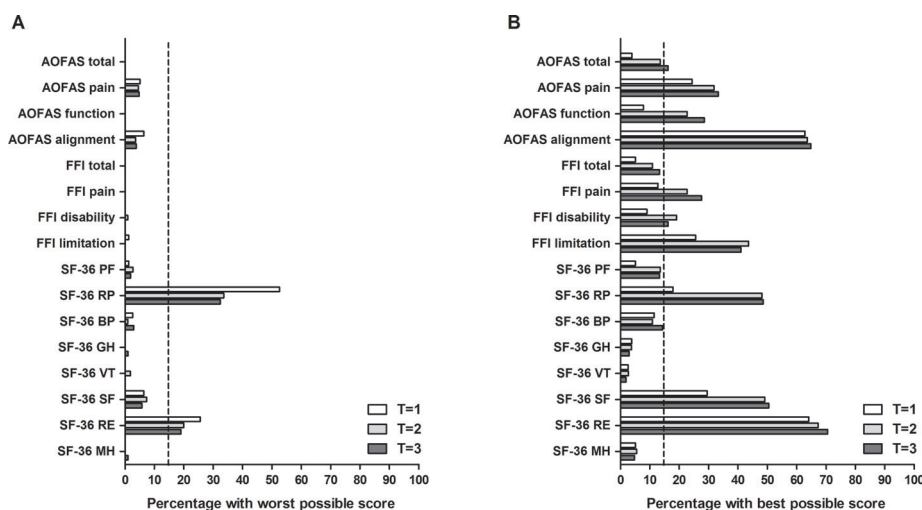


Figure 3 Floor effects (A) and ceiling effects (B) of the instruments used in patients with a hindfoot fracture. Out of a maximum of 78 at t=1, n=77 for AOFAS function and total, n=78 for AOFAS pain and alignment and for all (sub)scales of FFI and SF-36. Out of a maximum of 113 at t=2, n=109 for SF-36 GH, PCS and MCS, and n=110 for all (sub)scales of the AOFAS and FFI, and for all other subscales of the SF-36. Out of a maximum of 113 at t=3, n=105 for all (sub)scales of the AOFAS, FFI and SF-36. The dotted line represents the acceptable 15% of patients with the maximum score. Since for the SF-36 PCS and MCS none of the patients reported the worst or best possible score, they are not shown. 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.

Form Health Survey (SF-36-DLV) questionnaires on three occasions: between 3 and 6 months after trauma (t=1), 5–6 months later (t=2) and 2–3 weeks later (t=3). Patients were allowed to participate in both the responsiveness and test–retest part. A

physician completed the physician-reported part of the AOFAS-DLV.

The AOFAS Ankle-Hindfoot Scale consists of three subscales: pain, function and alignment and includes a total of nine items. The minimum score is 0 points

Table 3 Internal consistency of the instruments used in patients with a hindfoot fracture

(Sub)scale	n	No of items	Cronbach's α
AOFAS			
Total	77	8*	0.585†
Pain	77	1	NA‡
Function	77	6*	0.863
Alignment	77	1	NA‡
FFI			
Total	78	23	0.599†
Pain	78	9	0.635
Disability	78	9	0.558
Activity limitation	78	5	0.841
SF-36			
Total	78	35	0.916†
PF	78	10	0.932
RP	78	4	0.875
BP	78	2	0.769
GH	78	5	0.627
VT	78	4	0.757
SF	78	2	0.841
RE	78	3	0.939
MH	78	5	0.803
PCS	78	21	0.875†
MCS	78	14	0.879†

Data for t=1 were used.

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

*Question about stability has been removed as all patients scored identical answers.

†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; NA, 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.

(indicating severe pain and impairment), the maximum score is 100 points (no impairment).

The FFI is a questionnaire, which focusses on disabilities and measures the impact of foot disorders. The FFI includes three subscales: pain, disability and activity limitations, which are spread over a total of 23 items. In this scoring system, a score of 0 points means 'no disability', 100 points implies the highest level of disability.¹⁴

The SF-36 Health Survey is a generic measure of health status.^{15–22} It consists of 36 items, representing eight domains that are grouped into a physical component summary (PCS) and a mental component summary

(MCS). All (sub)scales are normalised to a mean of 50 points with a SD of 10 points.

Statistical analysis

Statistical Package for Social Sciences (SPSS V.21) was used for analysis. Data are reported following the STrengthening the Reporting of OBServational studies in Epidemiology.²³ Missing data were not imputed. Patient characteristics and questionnaire scores were analysed using descriptive statistics. Measurement properties of the AOFAS Ankle-Hindfoot Scale were determined in compliance with the CONsensus-based Standards for the selection of health Measurement Instruments guidelines.²⁴ The already validated FFI and SF-36 (sub)scales were used to compare the AOFAS-DLV with. A summary of the measurement properties and statistical analysis is given in table 2. A more detailed description is published in the study protocol.¹³

RESULTS

The changes over time in AOFAS-total, FFI-total, SF-36 PCS and SF-36 MCS are shown in figure 2. In the period from t=1 to t=2, the AOFAS, SF-36 PCS and (less pronounced) SF-36 MCS increased in scores. The FFI score decreases as expected, since this questionnaire focusses on disabilities. Scores at t=2 and t=3 were similar for all instruments.

Floor and ceiling effects

A floor effect was only present in the SF-36 RP and RE subscales at all follow-up moments. The percentage of patients reporting the minimum score varied between 52.6% (t=1) and 32.4% (t=3) for SF36 RP and between 25.6% (t=1) and 19.0% (t=3) for the SF36 RE subscale (figure 3A).

Ceiling effects were seen in several (sub)scales, especially at longer follow-up (figure 3B). The AOFAS as a total scale only showed a ceiling effect at t=3; 16.2% of patients reported the maximum score. The AOFAS pain and alignment subscales had a ceiling effect from the t=1 onwards (12.8% and 62.8%, respectively). The AOFAS function subscale showed ceiling effects from t=2 onwards (22.7%). The FFI pain and disability subscales showed ceiling effects from t=2 onwards. The FFI limitation, SF-36 RP, SF and RE subscales showed ceiling effects at all follow-up moments.

Reliability

Internal consistency

For the AOFAS total scale the Cronbach's α was 0.585 (table 3). This may suggest inadequate internal consistency, but as the entire scale contains three subscales, this value should, however, be interpreted carefully. The Cronbach's α for the AOFAS function subscale was 0.863, representing adequate internal consistency. Being single-item domains, Cronbach's

Table 4 Construct validity of the instruments in patients with a hindfoot fracture

(Sub)scale	AOFAS			
	Pain	Function	Alignment	Total
AOFAS				
Pain	1	0.12 [77]	-0.02 [78]	0.54 [77]
Function	0.12 [77]	1	0.31 [77]	0.86 [77]
Alignment	-0.02 [78]	0.31 [77]	1	0.43 [77]
Total	0.54 [77]	0.86 [77]	0.43 [77]	1
FFI				
Pain	-0.70 [78]	-0.38 [77]	-0.18 [78]	-0.63 [77]
Disability	-0.28 [78]	-0.85 [77]	-0.30 [78]	-0.84 [77]
Activity limitation	-0.22 [78]	-0.80 [77]	-0.37 [78]	-0.79 [77]
Total	-0.40 [78]	-0.82 [77]	-0.34 [78]	-0.88 [77]
SF-36				
PF	0.16 [78]	0.79 [77]	0.30 [78]	0.73 [77]
RP	0.30 [78]	0.65 [77]	0.20 [78]	0.66 [77]
BP	0.56 [78]	0.48 [77]	0.16 [78]	0.65 [77]
GH	-0.06 [78]	0.15 [77]	0.22 [78]	0.13 [77]
VT	0.18 [78]	0.23 [77]	0.13 [78]	0.29 [77]
SF	0.20 [78]	0.54 [77]	0.06 [78]	0.53 [77]
RE	0.17 [78]	0.30 [77]	-0.02 [78]	0.31 [77]
MH	0.15 [78]	0.29 [77]	0.09 [78]	0.31 [77]
PCS	0.29 [78]	0.74 [77]	0.33 [78]	0.75 [77]
MCS	0.12 [78]	0.12 [77]	-0.08 [78]	0.15 [77]

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 78.

$r > 0.6$ indicates high correlation, $0.3 < r < 0.6$ moderate correlation and $r < 0.3$ low correlation. Bold 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.

α could not be determined for the AOFAS pain and alignment subscales.

The FFI scale only showed adequate internal consistency for the subscale activity limitation ($\alpha=0.841$). The internal consistency was not adequate for the FFI scale as a total ($\alpha=0.599$) and for the subscales pain ($\alpha=0.653$) and disability ($\alpha=0.558$). For the total scale, this may be due to the fact that it is not unidimensional. Except for the subscale GH ($\alpha=0.627$), all SF-36 (sub)scales showed adequate internal consistency.

Construct validity

Spearman's rank correlations regarding construct validity are shown in table 4. Construct validity was only adequate for the AOFAS scale as a total and the function subscale, in both (sub)scales 82.4% of the predefined hypotheses were predicted correctly. For the pain subscale, only 8 out of 17 correlations (47.1%) were in accordance with predefined hypotheses. This was 12 (70.6%) for the alignment subscale. Both percentages were below the 75% threshold.

Reproducibility

Test-retest reliability

The intraclass correlation coefficient (ICC), indicating the reliability, of each (sub)scale is shown in table 5. The ICC for all AOFAS (sub)scales ranged from 0.89 to 0.97, indicating adequate test-retest reliability. For all FFI and SF-36 (sub)scales, the ICC was also adequate (>0.70).

Agreement and smallest detectable change

The level of agreement is indicated by the smallest detectable change (SDC) and the corresponding Reliable Change Index (RCI) (table 5). The SDC was 1.7 (RCI: 1.7%) for the AOFAS total scale, -2.9 (RCI: -2.9%) for the FFI total scale, 0.16 (RCI: 0.2%) for the SF-36 PCS subscale and -0.29 (RCI: -0.4%) for the SF-36 MCS subscale.

The Bland and Altman analysis shows that for each (sub)scale the 95% limits of agreement for the mean change in scores contains zero; this confirms that there is no bias in measurements (figure 4 and table 5).

Table 5 Intraclass correlation coefficient (ICC) and Bland-Altman analysis of the instruments in patients with a hindfoot fracture

(Sub)scale	n	ICC(2,1) (95% CI)	SEM	SDC _{patient}	Max score	RCI (%)	Mean _{difference} (SD)	95% limits of agreement
AOFAS								
Pain	105	0.89 (0.84 to 0.92)	0.4	1.1	40	2.8	0.6 (4.8)	-8.8 to 9.9
Function	105	0.97 (0.95 to 0.98)	0.1	0.4	50	0.8	0.2 (2.6)	-5.0 to 5.4
Alignment	105	0.97 (0.96 to 0.98)	0.1	0.2	10	1.9	0.1 (0.7)	-1.3 to 1.4
Total	105	0.96 (0.93 to 0.97)	0.6	1.7	100	1.7	0.9 (5.7)	-10.3 to 12.1
FFI								
Pain	105	0.92 (0.88 to 0.95)	-1.0	-2.8	100	-2.8	-1.4 (8.8)	-18.8 to 15.9
Disability	105	0.95 (0.92 to 0.96)	-1.1	-2.9	100	-2.9	-1.5 (8.0)	-17.1 to 14.1
Limitation	105	0.95 (0.92 to 0.97)	-1.0	-2.9	100	-2.9	-1.5 (6.1)	-13.5 to 10.5
Total	105	0.95 (0.92 to 0.96)	-1.0	-2.9	100	-2.9	-1.5 (6.6)	-14.5 to 11.5
SF-36								
PF	105	0.91 (0.87 to 0.94)	0.09	0.24	56.76	0.4	0.12 (4.67)	-9.03 to 9.28
RP	105	0.81 (0.73 to 0.86)	0.10	0.28	55.56	0.5	0.14 (8.46)	-16.43 to 16.72
BP	105	0.82 (0.74 to 0.87)	0.08	0.22	60.40	0.4	0.11 (6.14)	-11.90 to 12.14
GH	104	0.92 (0.89 to 0.95)	-0.27	-0.76	63.78	-1.2	-0.39 (4.20)	-8.62 to 7.85
VT	105	0.91 (0.87 to 0.94)	-0.37	-1.03	68.66	-1.5	-0.52 (4.14)	-8.63 to 7.58
SF	105	0.94 (0.91 to 0.96)	0.38	1.04	57.33	1.8	0.53 (4.68)	-8.64 to 9.71
RE	105	0.84 (0.77 to 0.89)	0.27	0.75	55.66	1.4	0.38 (6.85)	-13.03 to 13.80
MH	105	0.93 (0.90 to 0.95)	-0.34	-0.95	63.97	-1.5	-0.49 (4.30)	-8.92 to 7.94
PCS	104	0.85 (0.78 to 0.89)	0.06	0.16	70.30	0.2	0.08 (5.78)	-11.26 to 11.42
MCS	104	0.94 (0.92 to 0.96)	-0.11	-0.29	77.92	-0.4	-0.15 (4.24)	-8.47 to 8.17

Change scores were calculated from t=2 to t=3. The maximum possible number of patients was 113. 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.

AOFAS, American Orthopaedic Foot and Ankle Society; BP, bodily pain; FFI, Foot Function Index; GH, general health perceptions; ICC, Intraclass correlation coefficient; 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, SE of measurement; SF, social functioning; SF-36, Short Form-36; VT, vitality, energy or fatigue.

Responsiveness

Spearman's rank correlation coefficients for longitudinal validity are shown in table 6. Longitudinal validity was adequate for the AOFAS pain and alignment subscale; out of 17 correlations, 15 (88.2%) were in line with predefined hypotheses for the pain subscale and 17 (100.0%) for the AOFAS alignment subscale. Longitudinal validity was not sufficient for the function subscale (10/17; 58.8%) and for the total scale (12/17; 70.6%).

The standardised response mean (SRM) and the effect size (ES) of the instruments are shown in table 7. The magnitude of change was large for the AOFAS total scale (SRM 0.79, ES 0.63) and moderate for the function subscale (SRM 0.94, ES 0.61). The ES were small for the one-item subscales pain (SRM 0.26) and alignment (SRM 0.06).

DISCUSSION

The results of this study showed that the AOFAS Ankle-Hindfoot Scale (AOFAS-DLV) has adequate

construct validity and is reliable for measuring functional outcome in patients with a hindfoot fracture. However, longitudinal validity and responsiveness were inadequate in the study population.

Floor effects were not present for the AOFAS-DLV, but all (sub)scales showed an increasing ceiling effect over time. That suggests that an increasing number of patients achieved full recovery over time. This is in line with previous findings.^{11 17} The single-item subscales pain and alignment showed a ceiling effect from t=1 onwards. This could be due to the fact that (minor) extra-articular fractures may not be an issue with alignment. The high rate of operative treatment may also have improved alignment, especially for the intra-articular fractures. Alternatively, the limited answers for the pain and alignment subscales and the choice of administering the AOFAS-DLV at 3–6 months after trauma for the first time, may also have contributed to the ceiling effects.

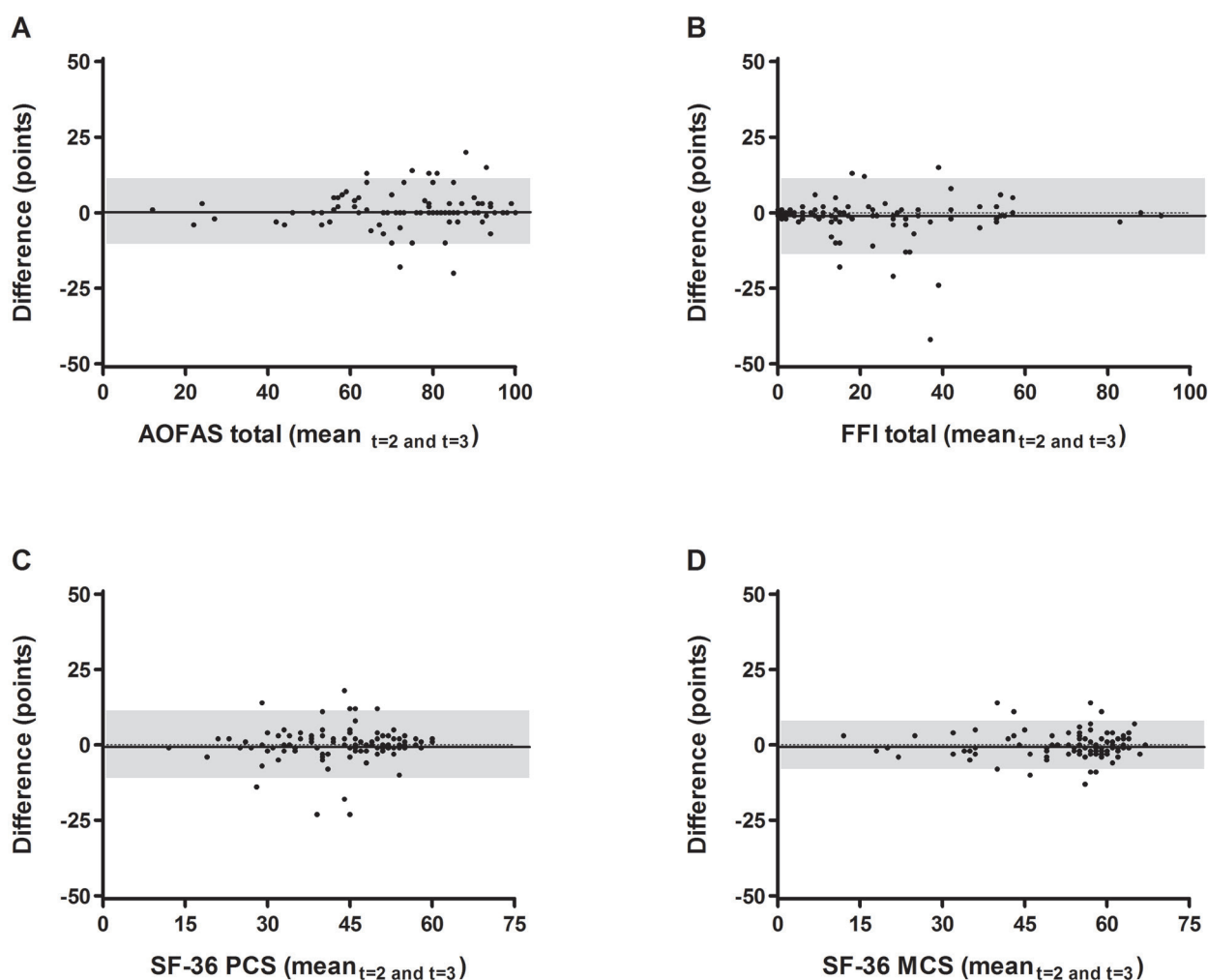


Figure 4 Bland-Altman plots for AOFAS Ankle-Hindfoot (A), Foot Function Index (B), SF-36 PCS (C) and SF-36 MCS (D) scores in patients with a hindfoot fracture. Change scores were calculated from $t=2$ to $t=3$. Each dot represents a single patient. The black line indicates the mean difference. The upper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

Adequate construct validity of the AOFAS total scale and function subscale is also in correspondence with previous research.^{10 11} The AOFAS subscales pain and alignment did not show adequate construct validity, in contrast with earlier data in ankle fractures.¹¹ The AOFAS pain and alignment subscales consist only of one item. In the hindfoot series, the correlations with other (sub)scales were generally overestimated for the pain subscale and underestimated for the alignment subscale. This difference is unlikely due to the (heterogeneity) in (sub)scale scores between the ankle and hindfoot fracture cohorts. There is also no clear pathophysiological explanation for this difference, other than the fact that hindfoot and ankle fractures are different injuries. Another possible explanation may be a difference in follow-up moment used for hindfoot and ankle fractures.

With a Cronbach's α above 0.7, internal consistency of the AOFAS-DLV function subscale was adequate. For the total scale, this remains inconclusive; the Cronbach's α of 0.585 should be interpreted carefully as the total scale is not unidimensional. In ankle fractures,¹¹ ankle sprains²⁵ and ankle arthroplasty and arthrodesis,²⁶ the Cronbach's α for the total scale ranged from 0.92 to 0.95. To our knowledge, no recent literature on this topic is available for hindfoot fractures. Deleting the pain question increases Cronbach's α to 0.843 (data not shown). This may suggest that the pain question is difficult to answer for patients. This could be due to the fact that three out of four answers combine pain severity and frequency. Such linguistic issues have been noted before.^{26 27}

The ICC values between 0.89 and 0.97 confirm adequate test-retest reliability of the AOFAS-DLV

Table 6 Longitudinal validity of the instruments in patients with a hindfoot fracture

(Sub)scale	AOFAS			
	Pain	Function	Alignment	Total
AOFAS				
Pain	1	-0.02 [74]	-0.02 [75]	0.80 [74]
Function	-0.02 [74]	1	-0.07 [74]	0.52 [74]
Alignment	-0.02 [75]	-0.07 [74]	1	0.10 [74]
Total	0.80 [74]	0.52 [74]	0.10 [74]	1
FFI				
Pain	-0.64 [75]	-0.08 [74]	0.08 [75]	-0.57 [74]
Disability	-0.11 [75]	-0.50 [74]	0.04 [75]	-0.39 [74]
Activity limitation	-0.03 [75]	-0.63 [74]	-0.06 [75]	-0.39 [74]
Total	-0.32 [75]	-0.44 [74]	0.02 [75]	-0.54 [74]
SF-36				
PF	0.11 [75]	0.43 [74]	-0.07 [75]	0.32 [74]
RP	0.13 [75]	0.02 [74]	-0.04 [75]	0.15 [74]
BP	0.33 [75]	0.05 [74]	0.06 [75]	0.33 [74]
GH	-0.07 [74]	0.07 [73]	0.29 [74]	0.07 [73]
VT	0.00 [75]	0.21 [74]	0.09 [75]	0.18 [74]
SF	-0.11 [75]	0.31 [74]	-0.10 [75]	0.14 [74]
RE	-0.12 [75]	0.13 [74]	0.00 [75]	-0.04 [74]
MH	0.02 [75]	0.20 [74]	0.09 [75]	0.14 [74]
PCS	0.23 [74]	0.10 [73]	0.09 [74]	0.33 [73]
MCS	-0.17 [74]	0.13 [73]	-0.02 [74]	-0.05 [73]

Change in scores between $t=1$ and $t=2$ were used. The maximum possible number of patients was 75. Spearman's rank correlation coefficients are given for all possible combinations of (sub)scales, with the n between square brackets. $r>0.6$ indicates high correlation, $0.3<r<0.6$ moderate correlation, and $r<0.3$ low correlation. Bold 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.

total scale and all subscales. Similar ICCs (ranging from 0.89 to 0.95) were found for the Turkish and Portuguese version of the AOFAS Ankle-Hindfoot Scale in patients with foot and ankle disorders.^{11 28 29}

Responsiveness is a product of magnitude of change and longitudinal validity. The longitudinal validity of the AOFAS subscales pain and alignment was adequate (ie, $>75\%$ of the hypothesised correlations predicted correctly). However, the AOFAS subscale function and the total scale were not proven adequate, as only 58.8% and 70.6% of the predefined hypothesis were confirmed, respectively. The inadequate longitudinal validity makes the AOFAS-DLV less useful for longitudinal studies measuring recovery over time in patients with a hindfoot fracture. Longitudinal validity was adequate for all (sub)scales of the AOFAS-DLV in patients with ankle fractures in previous research.¹¹ In the hindfoot series, the correlations of the difference in score between $t=1$ and $t=2$ with other (sub) scales were generally overestimated for the AOFAS

function subscale and total scale. Similar as for the construct validity, there is no clear pathophysiological explanation for this difference, other than the difference in (severity of) the injuries and follow-up moments used.

The magnitude of change was moderate for the AOFAS Ankle-Hindfoot scale DLV as a total, with a SRM of 0.79 and an ES of 0.63. This is comparable to the magnitude of change for the total FFI (SRM 0.89, ES 0.60) and the SF-36 subscales PCS, PF and RP as in our recent study on ankle fractures.¹¹ Previous data for hindfoot injuries are not available.

The Bland and Altman analysis confirmed absence of systematic bias for repeated recordings of the AOFAS Ankle-Hindfoot Scale-DLV. With an SDC of 1.7 points, the measurement error is very small. This measurement error was lower than reported for a variety of foot and ankle disorders in the Turkish population (SDC 13.3) and for ankle fractures in the Dutch population (SDC 12.0).^{11 28}

Table 7 Responsiveness: standardised response mean (SRM) and effect size (ES) of the instruments in patients with a hindfoot fracture

(Sub)scale	n	Mean change	SD _{change}	SRM	SD _{t=1}	ES
AOFAS						
Pain	75	3.2	12.4	0.26	10.1	0.32
Function	74	7.7	8.2	0.94	12.7	0.61
Alignment	75	0.1	2.2	0.06	3.1	0.04
Total	74	11.3	14.3	0.79	17.9	0.63
FFI						
Pain	75	-9.2	22.9	-0.40	25.1	-0.37
Disability	75	-17.4	20.2	-0.86	29.6	-0.59
Activity limitation	75	-15.4	23.4	-0.66	29.6	-0.52
Total	75	-14.6	16.3	-0.89	24.5	-0.60
SF-36						
PF	75	8.15	9.89	0.82	13.06	0.62
RP	75	7.59	12.46	0.61	11.72	0.65
BP	75	3.53	10.32	0.34	11.16	0.32
GH	74	-0.21	8.18	-0.03	9.23	-0.02
VT	75	0.70	8.73	0.08	9.73	0.07
SF	75	6.18	15.05	0.41	14.00	0.44
RE	75	1.08	15.35	0.07	13.17	0.08
MH	75	0.74	8.89	0.08	10.11	0.07
PCS	74	6.60	7.65	0.86	10.28	0.64
MCS	74	-0.39	11.36	-0.03	11.16	-0.03

Change scores were calculated from t=1 to t=2. The maximum possible number of patients was 75.

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.

CONCLUSION

The AOFAS Ankle-Hindfoot Scale DLV has adequate construct validity and is reliable, making it a suitable instrument for cross-sectional studies investigating functional outcome in patients with a hindfoot fracture. The inadequate longitudinal validity and responsiveness, however, hamper the use of the questionnaire in longitudinal studies and for assessing long-term functional outcome.

Collaborators D A Newhall, J Romeo, R J C Tjioe, F Van der Sijde, E N Van der Velden-Macauley, L Vellekoop

Contributors EMMVL, ASDB, DEM, CHVdV, PTDH, WET and MHJV developed the study. ASDB and EMMVL drafted the manuscript. EMMVL acted as trial principal investigator. ASDB, CHVdV, PTDH, DEM and MHJV participated inpatient inclusion and outcome assessment. ASDB, WET and EMMVL performed statistical analysis of the study data. All authors have read and approved the final manuscript.

Competing interests None declared.

Patient consent Obtained.

Ethics approval 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 anonymised during the study. The study is registered at the Netherlands Trial Register (NTR5613; 05 Jan 2016).

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 this 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

1. Brauer CA, Manns BJ, Ko M, *et al.* An economic evaluation of operative compared with nonoperative management of displaced intra-articular calcaneal fractures. *J Bone Joint Surg Am* 2005;87:2741-9.
2. De Boer AS, Schepers T, Panneman MJ, *et al.* Health care consumption and costs due to foot and ankle injuries in the Netherlands, 1986-2010. *BMC Musculoskelet Disord* 2014;15:128.
3. Mitchell MJ, McKinley JC, Robinson CM. The epidemiology of calcaneal fractures. *Foot* 2009;19:197-200.
4. Court-Brown CM, Caesar B. Epidemiology of adult fractures: a review. *Injury* 2006;37:691-7.
5. 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.

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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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*. In Press. 2017;7:e017040.
12. 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.
13. Van Lieshout EM, De Boer AS, Meuffels DE, *et al.* American Orthopaedic Foot and Ankle Society (AOFAS) ankle-hindfoot score: a study protocol for the translation and validation of the dutch language version. *BMJ Open* 2017;7:e012884.
14. 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.
15. 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.
16. 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.
17. Mahabier KC, Den Hartog D, Theyskens N, *et al.* Reliability, validity, responsiveness, and minimal important change of the disabilities of the arm, shoulder and hand and constant-murley scores in patients with a humeral shaft fracture. *J Shoulder Elbow Surg* 2017;26:e1–e12.
18. Obrebsky 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. von Elm E, Altman DG, Egger M, *et al.* The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344–9.
24. 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.
25. da Cunha RA, Hazime FA, da Silva Martins MC, *et al.* Translation, cross-cultural adaptation, and clinimetric testing of instruments used to assess patients with ankle sprain in the brazilian population. *J Orthop Sports Phys Ther* 2016;46:1042–50.
26. 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.
27. Guyton GP. Theoretical limitations of the AOFAS scoring systems: an analysis using monte carlo modeling. *Foot Ankle Int* 2001;22:779–87.
28. 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;55:1139–42.
29. Rodrigues RC MD, Mizusaki JM, Imoto AM, *et al.* cultural adaptation and validity of the American Orthopaedic Foot and Ankle Society (AOFAS) Ankle-Hindfoot Scale. *Acta Ortop Bras* 2008;16:107–11.
30. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
31. 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.
32. Hinkle DE WW, Jurs SG. *Applied statistics for the behavioral sciences*. Boston USA: Houghton Mifflin, 1998.
33. 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.
34. 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.
35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
36. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale (NJ): Lawrence Erlbaum Associates, 1998.