BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012884
Article Type:	Protocol
Date Submitted by the Author:	30-May-2016
Complete List of Authors:	Van Lieshout, Esther M.M.; Erasmus MC, University Medical Center Rotterdam De Boer, A.; 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
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Patient-centred medicine, Emergency medicine, Rehabilitation medicine
Keywords:	Foot & ankle < ORTHOPAEDIC & TRAUMA SURGERY, Dislocation, Fracture, Reliability, Responsiveness, Validity

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: 4361

ABSTRACT

Introduction

The AOFAS Ankle-Hindfoot Score is 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 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 or (fracture) dislocation.

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. Patients will be asked to complete the AOFAS Ankle-Hindfoot Score-DLV, as well as the Foot Function Index (FFI) and the SF-36 (Short Form-36). Patient and injury characteristics will be collected retrospectively. Measurement properties of the AOFAS Ankle-Hindfoot Score-DLV will be determined. Primary outcome measure is the construct validity. Secondary outcome measures include the reliability (i.e., internal consistency), reproducibility (i.e., test-retest reliability, agreement, and smallest detectable change), floor and ceiling effect, and responsiveness.

Discussion

Successful completion of this study will reveal whether or not the AOFAS Ankle-Hindfoot score-DLV is a valid and reliable instrument for studying outcome in patients with a fracture or (fracture) dislocation at the ankle or hindfoot. If proven valid, reliable, and responsive to change in outcome over time, it can be a valuable instrument for comparing treatment modalities or for comparing treatment results across hospitals.

Registration details

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

Strengths of this study

- It is a prospective, observational study with a strong methodologic design
- Statistical analyses 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.

Limitations of this study

- This study will be mostly relevant for the Dutch-speaking regions, but it is also informative for other regions.

BACKGROUND

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]

The AOFAS Ankle-Hindfoot Score, as a complete scale, has been shown to be valid in its original version.[7-10] Good responsiveness has also been shown.[8, 9] The physician-reported part of the scale has been shown to be valid and reliable.[11] The objective portion of the scale has not been evaluated for reliability. 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,[10] and end-stage ankle arthritis.[8]

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 the construct validity, reliability (*i.e.*, internal consistency, test-retest reliability, and measurement error), floor and ceiling effect, and responsiveness, and by calculating the smallest detectable change. Measurement properties will be calculated of 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 patients will be asked to 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. Two hospitals in Rotterdam (The Netherlands) will participate: Erasmus MC, University Medical Center Rotterdam and Ikazia 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 ones.

In order to reduce bias as much as possible, a research physician or research assistant will perform the physical examination that is part of the physician-reported part of the AOFAS Ankle-Hindfoot Score-DLV using a standardized protocol.

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. Patients in group 1 (test of pre-final version) will be asked to complete the pre-final version of the AOFAS Ankle-Hindfoot Score-DLV. Patients in group 2 (responsiveness) and group 3 (test-retest) will be asked to complete the final version of the Dutch AOFAS Ankle-Hindfoot-DLV questionnaire 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 additional injury gives functional limitations at time of enrolment)
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) Patients 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 construct validity of the AOFAS Ankle-Hindfoot Score-DLV will serve as primary outcome measure of the validation study.

The measurement properties of the AOFAS Ankle-Hindfoot Score-DLV will serve as secondary outcome measures in the validation study. The following parameters will be determined as secondary outcome measures:

- 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.[12] 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.*[12] This guideline is based on the review of Guillemin *et al.*[13] 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 complete the questionnaire. 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).

All patients in groups 2 and 3 will be asked to complete three questionnaires during their visit to the outpatient department; the AOFAS Ankle-Hindfoot Score-DLV, the Foot Function Index (FFI-DLV), and the Short Form Health Survey (SF-36-DLV). 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 physician-reported part of 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 5-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]

The Short Form Health Survey (SF-36) is a generic health status questionnaire that gives an indication of health-related quality of life.[14-18] 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). The SF-36 is the most widely evaluated patient-reported outcome measure for assessing general health.[19] It is reliable and easy to complete. A validated Dutch version is available.[20]

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 physician-reported part of 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 patients remain stable and sufficiently long to prevent recall. We consider a time interval of 2-3 weeks to be appropriate.

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) or Ikazia Hospital (Rotterdam).

The 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.[21] 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.[21, 22] 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.[21, 23] 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.[21] 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.[24] 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.[21]

Reliability / internal consistency

Reliability is defined as the degree to which the measurement is free from measurement error.[25] 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.[21] The correlation between items on a (sub)scale will be evaluated by calculating Cronbach's alpha for every (sub)scale. 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.[21] 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.[21] Reproducibility is suggested to consist of two parts: reliability and agreement.[26, 27] The data of group 3 will be used; they will complete all questionnaires twice, with 2-3 weeks in between.

Reliability concerns the degree to which patients can be distinguished from each other, despite measurement error.[21, 28] 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, will be used.[29] Reliability will be given a positive rating when the ICC is at least 0.70 in a sample size of at least 50 patients.[21]

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.[21] 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$_{residual}$).[21, 30, 31]

Based upon the SEM, the Smallest Detectable Change (SDC) will be calculated using the formula; SDC = 1.96 x $\sqrt{}$2 x SEM.[21] 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$_{ind}$).[21, 32, 33] The SDC measurable in a group of people (SDC$_{group}$) will be calculated by dividing the SDC$_{ind}$ by $\sqrt{}$n.[33, 34] 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.[35] 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 a bias in the measurements.

Agreement will be rated as positive if the SDC (SDC$_{ind}$ for application in individuals and SDC$_{group}$ for use in groups) or the limits of agreement are smaller than the minimally important change (MIC).[21] For the AOFAS Ankle-Hindfoot Score-DLV the MIC has not been published, but often a difference of 10-20 points is considered relevant.

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).[21] 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.[21, 36] 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.[21, 37] The data of group 2 will be used; they will complete all questionnaires twice, with 5-6 months in between. 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.[21]

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, using the data of group 2. 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.[38] The SRM will be calculated by dividing the mean change in score between two time points by the standard deviation of this change.[38] 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.[39]

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.

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. Especially the SDC and MIC will reveal important information for sample size calculations in future studies.

Two 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.

COMPETING INTEREST

The authors declare that they have no competing interests.

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, data sharing, or ethics approval.

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.

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. 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.

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. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

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

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

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

25. 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.

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

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

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

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

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

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

32. 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.

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(4):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(12):1260-8.

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

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

37. 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.

38. 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.

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

SPIRIT
STANDARD PROTOCOL ITEMS: RECOMMENDATIONS FOR INTERVENTIONAL TRIALS

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description	Addressed on page number
Administrative information			
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry	4,7
	2b	All items from the World Health Organization Trial Registration Data Set	See trial register online
Protocol version	3	Date and version identifier	7
Funding	4	Sources and types of financial, material, and other support	22
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors	1,2
	5b	Name and contact information for the trial sponsor	1
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	22
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	22

Introduction			
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	6
	6b	Explanation for choice of comparators	13
Objectives	7	Specific objectives or hypotheses	6
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)	7
Methods: Participants, interventions, and outcomes			
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	7
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	8-9
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	11-14
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)	10-11
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)	
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	7
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	10-14
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	10,13

Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	14-15
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size	8
Methods: Assignment of interventions (for controlled trials)			
Allocation:			
Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	
Methods: Data collection, management, and analysis			
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	11-14
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	

Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	15
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	15-19
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	15-19
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	15
Methods: Monitoring			
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	15
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial	
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	
Ethics and dissemination			
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval	3,19
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)	20

Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	7
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	15
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site	22
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	15
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	20
	31b	Authorship eligibility guidelines and any intended use of professional writers	20
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	
Appendices			
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates	
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable	

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012884.R1
Article Type:	Protocol
Date Submitted by the Author:	17-Aug-2016
Complete List of Authors:	Van Lieshout, Esther M.M.; Erasmus MC, University Medical Center Rotterdam De Boer, A.; 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
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Patient-centred medicine, Emergency medicine, Rehabilitation medicine
Keywords:	Foot & ankle < ORTHOPAEDIC & TRAUMA SURGERY, Dislocation, Fracture, Reliability, Responsiveness, Validity

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 - 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). Measurement properties of the AOFAS Ankle-Hindfoot Score-DLV will be determined. Outcome measures include the construct validity, reliability (i.e., internal consistency), reproducibility (i.e., test-retest reliability, agreement, and smallest detectable change), floor and ceiling effect, and responsiveness.

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).

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]

The AOFAS Ankle-Hindfoot Score, as a complete scale, has been shown to be valid in its original version.[7-9] However, poor to moderate correlation of the AOFAS scores to the SF-36 subscales may also suggest poor construct validity [10].Adequate responsiveness has also been shown.[8, 9] The physician-reported part of the scale has been shown to be valid and reliable.[11] Westphal *et al.* showed strong 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] These studies included mixed populations. Whether or not the AOFAS Ankle-Hindfoot score would be reliable and valid in homogenous populations consisting of, *e.g.*, only patients with hindfoot fractures, has not been published.

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. Lack of evaluation of measurement properties of the AOFAS Ankle-Hindfoot Score in homogeneous populations, the inclusion of anatomy-specific questions in the instrument, and its continued common use warrant its further evaluation in homogenous populations with either ankle or hindfoot fractures.

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 the construct validity, reliability (*i.e.*, internal consistency, test-retest reliability, and measurement error), floor and ceiling effect, and responsiveness, and by calculating the smallest detectable change. 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 patient 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 ones.

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 additional injury gives functional limitations at time of enrolment)
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) Patients 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 patients' outpatient department visit; the AOFAS Ankle-Hindfoot Score-DLV, the Foot Function Index (FFI-DLV), and the Short Form Health Survey (SF-36-DLV). 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.[17, 18]

The Short Form Health Survey (SF-36) is a generic health status questionnaire that gives an indication of health-related quality of life.[19-23] 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). The SF-36 is the most widely evaluated patient-reported outcome measure for assessing general health.[24] It is reliable and easy to complete. A validated Dutch version is available.[25]

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 patients remain 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. Those reporting a change will be excluded from the analysis.

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.[26] 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.[26, 27] 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.[26, 28] 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.[26] 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.[29] 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.[26]

Reliability / internal consistency

Reliability is defined as the degree to which the measurement is free from measurement error.[30] 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.[26] 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.[26][31] 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.[26] Reproducibility is suggested to consist of two parts: reliability and agreement.[32, 33] The data of group 3 will be used; they will complete all questionnaires twice, with 2-3 weeks in between.

Reliability concerns the degree to which patients can be distinguished from each other, despite measurement error.[26, 34] 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.[35] Reliability will be given a positive rating when the ICC is at least 0.70 in a sample size of at least 50 patients.[26]

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.[26] 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 ($SEM = \sqrt{(variance_{patient} + variance_{residual})}$).[26, 36, 37]

Based upon the SEM, the Smallest Detectable Change (SDC) will be calculated using the formula; $SDC = 1.96 \times \sqrt{2} \times SEM$.[26] 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}).[26, 38, 39] The SDC measurable in a group of people (SDC_{group}) will be calculated by dividing the SDC_{ind} by $\sqrt{n}$.[39, 40] 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.[41] 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}).[26] 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).[26] 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.[26, 42] 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.[26, 43] The data of group 2 will be used; they will complete all questionnaires twice, with 5-6 months in between. 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.[26] 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.

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, using the data of group 2. 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.[44] The SRM will be calculated by dividing the mean change in score between two time points by the standard deviation of this change.[44] 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.[45]

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. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

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

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

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

30. 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.

31. 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.

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

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

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

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

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

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

38. 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.

39. 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.

40. 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.

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

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

43. 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.

44. 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.

45. 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.

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description	Addressed on page number
Administrative information			
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry	4,7
	2b	All items from the World Health Organization Trial Registration Data Set	See trial register online
Protocol version	3	Date and version identifier	7
Funding	4	Sources and types of financial, material, and other support	28
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors	1,2
	5b	Name and contact information for the trial sponsor	1
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	28
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	28

Introduction			
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	5
	6b	Explanation for choice of comparators	12
Objectives	7	Specific objectives or hypotheses	6, 20
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)	8
Methods: Participants, interventions, and outcomes			
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	8
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	9-10
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	12-15
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)	11-12
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)	
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	8
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	11-15
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	11,14

Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	15-16
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size	9
Methods: Assignment of interventions (for controlled trials)			
Allocation:			
Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	
Methods: Data collection, management, and analysis			
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	12-15
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	

Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	16_____
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	16-20_____
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	16-20_____
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	16_____
Methods: Monitoring			
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	16_____
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial	_____
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	_____
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	_____
Ethics and dissemination			
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval	4,20_____
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)	21_____

Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	8
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	16
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site	28
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	16
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	21
	31b	Authorship eligibility guidelines and any intended use of professional writers	21
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	
Appendices			
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates	
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable	

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012884.R2
Article Type:	Protocol
Date Submitted by the Author:	22-Sep-2016
Complete List of Authors:	Van Lieshout, Esther M.M.; Erasmus MC, University Medical Center Rotterdam De Boer, A.; 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
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Patient-centred medicine, Emergency medicine, Rehabilitation medicine
Keywords:	Foot & ankle < ORTHOPAEDIC & TRAUMA SURGERY, Dislocation, Fracture, Reliability, Responsiveness, Validity

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). Measurement properties of the AOFAS Ankle-Hindfoot Score-DLV will be determined. Outcome measures include the construct validity, reliability (i.e., internal consistency), reproducibility (i.e., test-retest reliability, agreement, and smallest detectable change), floor and ceiling effect, and responsiveness.

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 to 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 strong 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 included mixed populations. Whether or not the AOFAS Ankle-Hindfoot score would be reliable and valid in a homogenous population consisting of only patients with ankle fractures, has not been published.

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. Lack of evaluation of measurement properties of the AOFAS Ankle-Hindfoot Score in homogeneous populations, the inclusion of anatomy-specific questions in the instrument, and its continued common use warrant its further evaluation in homogenous populations with either ankle or hindfoot fractures.

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 the construct validity, reliability (*i.e.*, internal consistency, test-retest reliability, and measurement error), floor and ceiling effect, and responsiveness, and by calculating the smallest detectable change. 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 once.

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 additional injury gives functional limitations at time of enrolment)
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). [25] 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.[17, 18] 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.[19-23] 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). The SF-36 is the most widely evaluated patient-reported outcome measure for assessing general health.[24] It is reliable and easy to complete. A validated Dutch version will be used.[25]

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? The item scored 'better','no change', or 'worse'. Patients reporting a change (either improvement or deterioration) will be excluded from the analysis.

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.[26] 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.[26, 27] 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.[26, 28] 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.[26] 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.[29] 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.[26]

Reliability / internal consistency

Reliability is defined as the degree to which the measurement is free from measurement error.[30] 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.[26] 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.[26][31] 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.[26] Reproducibility is suggested to consist of two parts: reliability and agreement.[32, 33] 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.

Reliability concerns the degree to which patients can be distinguished from each other, despite measurement error.[26, 34] 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.[35] Reliability will be given a positive rating when the ICC is at least 0.70 in a sample size of at least 50 patients.[26]

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.[26] 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 ($SEM = \sqrt{(variance_{patient} + variance_{residual})}$).[26, 36, 37]

Based upon the SEM, the Smallest Detectable Change (SDC) will be calculated using the formula; $SDC = 1.96 \times \sqrt{2} \times SEM$.[26] 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}).[26, 38, 39] The SDC measurable in a group of people (SDC_{group}) will be calculated by dividing the SDC_{ind} by $\sqrt{n}$.[39, 40] 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.[41] 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}).[26] 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).[26] 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.[26, 42] 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.[26, 43] 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.[44] The SRM will be calculated by dividing the mean change in score between two time points by the standard deviation of this change.[44] 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.[45]

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.[26] 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. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

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

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

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

30. 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.

31. 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.

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

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

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

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

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

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

38. 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.

39. 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.

40. 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.

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

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

43. 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.

44. 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.

45. 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.

SPIRIT

STANDARD PROTOCOL ITEMS: RECOMMENDATIONS FOR INTERVENTIONAL TRIALS

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description	Addressed on page number
Administrative information			
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry	4,7
	2b	All items from the World Health Organization Trial Registration Data Set	See trial register online
Protocol version	3	Date and version identifier	7
Funding	4	Sources and types of financial, material, and other support	28
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors	1,2
	5b	Name and contact information for the trial sponsor	1
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	28
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	28

Introduction			
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	5
	6b	Explanation for choice of comparators	12
Objectives	7	Specific objectives or hypotheses	6, 20
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)	8
Methods: Participants, interventions, and outcomes			
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	8
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	9-10
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	12-15
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)	11-12
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)	
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	8
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	11-15
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	11,14

Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	15-16
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size	9
Methods: Assignment of interventions (for controlled trials)			
Allocation:			
Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	
Methods: Data collection, management, and analysis			
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	12-15
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	

Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	16
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	16-20
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	16-20
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	16
Methods: Monitoring			
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	16
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial	
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	
Ethics and dissemination			
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval	4,20
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)	21

Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	8
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	16
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site	28
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	16
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	21
	31b	Authorship eligibility guidelines and any intended use of professional writers	21
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	
Appendices			
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates	
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable	

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012884.R3
Article Type:	Protocol
Date Submitted by the Author:	10-Nov-2016
Complete List of Authors:	Van Lieshout, Esther M.M.; Erasmus MC, University Medical Center Rotterdam De Boer, A.; 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
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Patient-centred medicine, Emergency medicine, Rehabilitation medicine
Keywords:	Foot & ankle < ORTHOPAEDIC & TRAUMA SURGERY, Dislocation, Fracture, Reliability, Responsiveness, Validity

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.

SPIRIT

STANDARD PROTOCOL ITEMS: RECOMMENDATIONS FOR INTERVENTIONAL TRIALS

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description	Addressed on page number
Administrative information			
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry	4,7
	2b	All items from the World Health Organization Trial Registration Data Set	See trial register online
Protocol version	3	Date and version identifier	7
Funding	4	Sources and types of financial, material, and other support	28
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors	1,2
	5b	Name and contact information for the trial sponsor	1
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	28
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	28

Introduction			
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	5
	6b	Explanation for choice of comparators	12
Objectives	7	Specific objectives or hypotheses	6, 20
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)	8
Methods: Participants, interventions, and outcomes			
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	8
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	9-10
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	12-15
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)	11-12
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)	
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	8
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	11-15
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	11,14

Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	15-16
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size	9
Methods: Assignment of interventions (for controlled trials)			
Allocation:			
Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	
Methods: Data collection, management, and analysis			
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	12-15
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	

Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	16
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	16-20
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	16-20
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	16
Methods: Monitoring			
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	16
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial	
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	
Ethics and dissemination			
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval	4,20
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)	21

Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	8
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	16
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site	28
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	16
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	21
	31b	Authorship eligibility guidelines and any intended use of professional writers	21
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	
Appendices			
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates	
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable	

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license.