The completely patient-reported version of the American Orthopaedic Foot and Ankle Society (AOFAS) score: A valid and reliable measurement for ankle osteoarthritis

Liam D.A. Paget, M.D, Inger N. Sierevelt, MSc, Johannes L. Tol, M.D, Ph.D, Gino M.M.J. Kerkhoffs, M.D, Ph.D, Gustaaf Reurink, M.D, Ph.D

PII: S2059-7754(23)00525-4
DOI: https://doi.org/10.1016/j.jisako.2023.07.003
Reference: JISAKO 137

To appear in: Journal of ISAKOS

Received Date: 24 December 2022
Revised Date: 9 July 2023
Accepted Date: 12 July 2023


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier Inc. on behalf of International Society of Arthroscopy, Knee Surgery and Orthopedic Sports Medicine.
The completely patient-reported version of the American Orthopaedic Foot and Ankle Society
(AOFAS) score: A valid and reliable measurement for ankle osteoarthritis

Liam D.A. PAGET, M.D., 1,2,3 Inger N. SIEREVELT, MSc., 4,5 Johannes L. TOL, M.D.,
Ph.D., 1,2,3,6 Gino M.M.J. KERKHOFFS, M.D., Ph.D., 1,2,3 Gustaaf REURINK, M.D.,
Ph.D., 1,2,3,7

1. Amsterdam UMC, University of Amsterdam, Department of Orthopedic Surgery,
Amsterdam Movement Sciences, Amsterdam, Netherlands

2. Academic Center for Evidence-based Sports medicine (ACES), Amsterdam, the
Netherlands

3. Amsterdam Collaboration for Health and Safety in Sports (ACHSS), AMC/VUmc
IOC Research Center, Amsterdam, the Netherlands

4. Spaarne Gasthuis Academy, Orthopedic department, Spaarne Ziekenhuis, Hoofddorp,
The Netherlands.

5. Specialised Center of Orthopedic Research and Education (stichting SCORE), Xpert
Clinics, Orthopedic department, Amsterdam, the Netherlands

6. Aspetar, Orthopedic and Sports Medicine Hospital, Doha, Qatar

7. The Sport Physician Group, Department of Sports Medicine, OLVG, Amsterdam, The
Netherlands

L.D.A. Paget, M.D.
Department of Orthopaedics
Amsterdam UMC
PO Box 22660
1100 DD Amsterdam, the Netherlands
L.d.paget@amsterdamumc.nl
+31 (0)6 42 14 48 51
Conflicts of Interest statement

Each author certifies that there are no funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article related to the author or any immediate family members. L. D. A. Paget received a research grant from the Marti-Keuning Eckhardt Foundation, a non-profit patient organization.

Ethics

The PRIMA trial is approved by the Medical Ethics Review Committee Amsterdam Medical Center, the Netherlands (ABR 2018-042, approved 23 July 2018) and registered in the Netherlands trial register (NTR7261).

Acknowledgments

We thank the Marti-Keuning Eckhardt foundation for their support in making this study possible.
The completely patient-reported version of the American Orthopaedic Foot and Ankle Society (AOFAS) score: A valid and reliable measurement for ankle osteoarthritis

Abstract

Background: The American Orthopaedic Foot and Ankle score (AOFAS) is an outcome measure for ankle and hindfoot conditions which requires scoring from both the patients and the physician. A completely patient-reported version has been developed and used before, but its measurement properties are unknown. Our goal was to determine the measurement properties and the Minimally Important Change (MIC) of a completely patient-reported AOFAS (PR-AOFAS) in patients with ankle osteoarthritis. Additionally the MIC of both the PR-AOFAS and the AOFAS was estimated, which had not previously been done.

Materials and Methods: The PR-AOFAS of 112 patients was evaluated for reliability, construct validity (using the AOFAS, Foot and Ankle Outcome Score, Ankle Osteoarthritis Score, Visual Analogue Scale, and Short Form-36) and responsiveness. The MIC was estimated using the optimal cut-off point of the ROC curve. This was a sub-study of a randomized clinical trial on the efficacy of platelet-rich plasma injections for ankle OA.

Results: The PR-AOFAS had sufficient construct validity, internal consistency, test-retest reliability and responsiveness. The smallest detectable change at group level was 2.34. The MIC was 6.5 points (95%CI 0.6-14.4).

Conclusions: The measurement properties of the Dutch PR-AOFAS were sufficient in patients with ankle osteoarthritis who are willing to participate in a trial on injection therapy. The minimally important change of the PR-AOFAS is smaller than its Smallest Detectable Change making it more suitable for use in groups of patients, such as a research setting.

Level of Clinical Evidence: 1
Keywords: Ankle, Ankle Osteoarthritis, Clinimetrics, Orthopaedic Surgery, Sports Medicine

What are the new findings

- The Patient reported version of the American Orthopaedic Foot and Ankle Society (AOFAS) score (PR-AOFAS) had sufficient construct validity, internal consistency, test-retest reliability and responsiveness.
- The Minimally Important Change (MIC) was 6.5 points for the PR-AOFAS and is therefore suitable for use in groups such as a research setting.
- The MIC for the original AOFAS was 17.5 points, which has not previously been determined.

Introduction

Patient Reported outcome measures (PROMs) enable quantification of outcomes such as pain and function, and its impact on Quality of Life (QoL). To adequately interpret treatment effects, an outcome score must be reliable, valid and responsive to change. [1] The American Orthopaedic Foot and Ankle Society (AOFAS) score is one of the most frequently used scores for foot and ankle conditions. [2] It combines five patient-reported items concerning pain and function with four physician-determined items concerning function and alignment, on a 0-100 point scale. [3]

The measurement properties of the AOFAS were assessed in two studies (n=133 and n=117) with patients with end-stage ankle OA [2, 3] and two studies (n=9 and n= 8) that included patients with ankle OA amongst other ankle and hindfoot conditions. [4, 5] The available data on
ankle OA found the patient-reported part of the AOFAS to be valid and responsive.[2–9] However the Minimally Important Change (MIC) of the AOFAS has not been determined for ankle osteoarthritis. Completely patient-reported outcome scores are easier for the patient, less time-consuming and result in more patient compliance, leading to higher response rates and thereby reduce risk of bias and increase generalizability.[2, 10] In a prospective multicenter randomized controlled trial in the Netherlands, including 152 patients on the routine versus on demand removal of syndesmotic screws the physician-determined items were modified to patient-reported items, allowing for a completely patient-reported version of the AOFAS.[11] However the measurement properties of this completely patient-reported AOFAS had not been determined.[11] Our goal was to determine the measurement properties and the Minimally Important Change (MIC) of a Dutch completely patient-reported AOFAS (PR-AOFAS) in patients with ankle osteoarthritis. Additionally the MIC of both the PR-AOFAS and the AOFAS was estimated, which had not previously been done.
Materials and Methods

Study Design

This is a sub-study of the PRIMA trial, a randomized, double-blind, placebo-controlled, multicentre prospective study, designed to determine the efficacy of Platelet-rich Plasma (PRP) injections in the management of ankle OA.[12, 13] The PRIMA trial is approved by the Medical Ethics Review Committee Amsterdam Medical Center, the Netherlands (ABR 2018–042, approved 23 July 2018) and registered in the Netherlands trial register (NTR7261). This sub-study was sponsored by the Marti-Keuning Eckhardt Foundation, a non-profit patient organization.

Study Population

Patients with ankle OA in six hospitals in the Netherlands (2 University Medical Centres, 2 teaching hospitals, a general hospital and a focus clinic) were informed of the study. All participants signed an informed consent form before participating in the study. Patients were eligible for inclusion if they had a severity of ankle OA pain on a Visual Analogue Scale (VAS 0–100 mm) ≥40 mm during daily activities, radiographs (anteroposterior and lateral view) indicating ≥ grade 2 talocrural OA on the van Dijk classification (joint space narrowing, with or without osteophytes)[14] and were ≥18 years of age. Patients were excluded if they had received injection therapy for ankle OA in the previous 6 months, did not want to receive one of the two therapies, had clinical signs of concomitant OA of one or more other major joints of the lower extremities that negatively affects their daily activity level or had had a previous ankle surgery for OA or osteochondral defects <1 year (not including surgery for an ankle fracture in the past).
Study procedures

This study was performed according to the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist as a guideline.[15] Following inclusion, as far as relevant for this study, patients completed questionnaires at baseline, 6, 12, 26 and 52 weeks.[12] The questionnaires completed included Dutch versions of the the PR-AOFAS, FAOS, AOS, VAS and SF-36.[12] In order to assess the test-retest reliability, at a minimum follow-up of 26 weeks under assumption of no change, an additional PR-AOFAS was sent two weeks after a previous one with an additional question to assess change of symptoms. A Dutch version of the AOFAS was taken during outpatient visits at baseline, 6 weeks and 26 weeks. In order to minimise bias, all patients were seen by the same coordinating research physician of the PRIMA study.

The original patient-physician-determined American Orthopaedic Foot and Ankle Society (AOFAS) score

The AOFAS measures pain function and alignment and was translated and validated in Dutch for ankle fractures.[8] The physician-determined items were performed by the same coordinating research physician. It consists of 9 items and 3 subscales (Pain – 1 item, Function- 7 items and Alignment – 1 item) of which 3 items from the function scale and the alignment scale are physician determined. The Maximum scores on all three subscales are 40, 50 and 10 points, respectively.[16]

A completely patient-reported version of the American Orthopaedic Foot and Ankle Society (PR-AOFAS) score
For the PR-AOFAS the questionnaire from the RODEO trial was used.[11] Here the physician-determined items were made more comprehensible for the patients.[11] These changes are presented in the Supplement. Where previously the physician determined sagittal and hindfoot motion, ankle-hindfoot stability and alignment, the patient was now asked in patient-friendly terms how they would rate these items. For instance: “How would you rate the mobility of your ankle compared to the other side, or compared to when you did not have symptoms.” Similar to the AOFAS, the PR-AOFAS consists of 9 items and 3 subscales (Pain – 1 item, Function- 7 items and Alignment – 1 item).

**The Foot and Ankle Outcome Score (FAOS)**

The FAOS is used for functional assessment of ankle and hindfoot conditions, consists of five subscales (pain, other symptoms, activities of daily living (ADL), sport and recreation and foot and ankle related quality of life (QoL)) and has been shown to have sufficient reliability and validity in patients with ankle and hindfoot symptoms.[17] There are a total of 42 items, each question is assigned 0 – 4 points based on the answer given.[17] The scales runs from 0 (extreme symptoms) to 100 points (no symptoms).

**Medical Outcomes Study short form 36 (SF-36)**

The SF-36 is a generic outcome measuring quality of life and has 36 items, consisting of 8 subscales (0-100 points). Of these subscales, 4 make out the Physical Component Summary (PCS) score and 4 make out the Mental Component Summary (MCS) score (0-100 points).[18] The higher the patient scores, the higher the quality of life.

**The Visual Analog Score (VAS) for pain**

The VAS score (VAS 0-100 mm) is measured during activities of daily living, with 0 mm being no pain and 100 mm the worst pain imaginable.
The Ankle Osteoarthritis Score (AOS)

The AOS measures pain (9 items) and disability (9 items) with a total of 18 items on a visual analog scale (0 – 100 mm), and has been shown to be valid, reliable and responsive to change in patients with ankle OA.[2] Higher scores indicate worse pain and disability.

Subjective patient satisfaction

The subjective patient satisfaction was asked according to a 4-point Likert scale (poor, fair, good and excellent) and served as an anchor to determine the responsiveness of the completely patient-reported AOFAS and calculate Minimal clinically Important Change values.
Statistical analysis

Both PR-AOFAS and AOFAS were assessed for reliability, construct validity, responsiveness and Interpretability (floor and ceiling effects).[19–21] Due to the absence of a retest of the AOFAS, only internal consistency was evaluated as a reliability measure. Reliability assessment of the PR-AOFAS comprised of internal consistency, test-retest reliability, standard error of measurement (SEM), and Smallest Detectable Change (SDC). Additionally, the Minimal clinically Important Change (MIC) of both scales was calculated.[20, 21]

Reliability

Reliability shows the degree to which the scale is free of measurement error and will be assessed through internal consistency, test-retest reliability and measurement error. Internal consistency demonstrates the interrelatedness among items of an outcome measure or their subscales. A Cronbach's $\alpha$ between 0.70 and 0.95 demonstrates satisfactory internal consistency for both PR-AOFAS and AOFAS as long as the scale is unidimensional.[21, 22] Test-retest reliability determines the ability to measure the same when patients with unchanged complaints take the questionnaire twice. Patients were only included in the test-retest reliability analysis for the PR-AOFAS if they had confirmed that no change had occurred. The test-retest reliability was determined using the intraclass correlation coefficient (ICC) alongside a standard error of measurement (SEM) and was deemed reliable if ICC ≥ 0.70.[23, 24] The SEM was calculated as the square root of the within-subject variance (i.e. the sum of the between-measures variance and the residual variance).[25] The smallest detectable change (SDC=1.96*$\sqrt{2}$*SEM) provides the magnitude of change needed to have confidence that the change is not a consequence of measurement error.[26]

Construct validity
Construct validity measures the degree to which the intended construct is measured.[19] This was assessed by determining the correlation of the PR-AOFAS and the AOFAS (and their subscales) with the FAOS subscales, AOS, VAS and the PCS and MCS scores of the SF-36 at baseline using Spearman’s correlation coefficient ($r_s$).

Correlation was categorized as high ($r>0.5$) for similar, moderate ($0.3 \leq r \leq 0.5$) for related but dissimilar and low ($r<0.3$) for unrelated constructs.[15]

In total 21 hypotheses were determined for the PR-AOFAS and 21 for the AOFAS. At least moderate a priori hypothesized correlations were to be expected between the total scores and the FAOS subscales, AOS total score, the Physical Component Summary score of the SF-36 and the VAS. For the pain subscale of the PR-AOFAS and AOFAS high correlations were expected with the pain subscales of the FAOS, the AOS and the VAS; moderate correlations with the PCS and the symptoms subscale of the FAOS. For the function subscale of the PR-AOFAS and AOFAS high correlations were expected with the ADL and sport and recreation subscales of the FAOS, the AOS disability subscale and the PCS. Finally, all domains of both outcomes were expected to be unrelated ($r<0.3$) to the MCS. Confirmation of at least 75% of the aforementioned hypotheses indicates sufficient construct validity.[21]

A separate evaluation was performed to assess whether the PR-AOFAS and AOFAS measure identical constructs. For this purpose, additional a priori correlation coefficients of $>0.8$ were to be expected between the PR-AOFAS and the AOFAS and their identical subscales (pain, function and alignment).

**Responsiveness**

Responsiveness of a PROM is defined as its ability to detect change over time.[19, 27] It was determined by comparing both the PR-AOFAS and AOFAS at baseline and 26 weeks. Change
from baseline at 26 weeks of the 4-point Likert scaled subjective patient satisfaction was used as an anchor. Changes in the anchor were defined as worsened, unchanged, improved, greatly improved.[28] Patients who improved at least 1 category were considered to have had a clinically important improvement.[28] The effect sizes (ES) and Standardized Response Means (SRM) were calculated for each category of changes in the anchor.[29, 30] Based on Cohen’s standardized effect size, the following was hypothesized:[31]

- ES and SRM <0.2 for patients who reported to be unchanged
- ES and SRM ≥0.2 for patients who reported to be improved
- ES and SRM ≥0.5 for patients who reported to be much improved

Responsiveness was further evaluated using the Receiver Operating Characteristic (ROC) curve analyses. ROC curves give information on the true positive rate (sensitivity) and false positive rate (1-specificity) for cut-off points in change score.[32] The Area Under a ROC Curve (AUC) indicates the chance that the patient is indeed correctly improved or unimproved, a value of at least 0.7 was deemed sufficient.[15, 21, 32]

Interpretability

Interpretability is the meaning that can be assigned to a score or change in score.[19] In this study measured interpretability by determining floor and/or ceiling effects. If a large number of patients score the maximum or minimum score, the instrument can fail to detect clinical improvement or deterioration. A floor and/or ceiling effect is present if more than 15% of the study subjects had the lowest or highest possible score.[20, 21]

Minimal clinically Important Change
The Minimal Important Change (MIC) is the minimal amount of change that is perceived by the patient as “important”. [20] Using the previously described anchor the patients were categorized as either improved (or more) and unchanged (patients who were worse were excluded). [33] By use of ROC curve analysis the MIC was calculated as the cut-off point that indicated the least amount of misclassification. [33] A Bootstrapping procedure (with 1000 bootstraps) was performed to estimate the standard error and determine the 95% CI.

Agreement of the PR-AOFAS and the AOFAS

Agreement between the PR-AOFAS and the AOFAS was determined using a Bland-Altman plot with limits of agreement. [20, 34] Additionally, Intraclass correlation coefficients (ICC) were calculated between the PR-AOFAS, the AOFAS as well as between the change from baseline of the both outcome measures.

Results

Inclusion

In total 112 patients were enrolled in the PRIMA study to receive either an intra-articular PRP or placebo (saline) injection and completed all baseline questionnaires and physical examinations. (Table I). In total, 63 (56%) were male and 49 (44%) female. The mean age, duration of symptoms and BMI was 55 years (SD 14 years), 9 years (SD 9 years) and 26.5 kg/m² (SD 3.8 kg/m²), respectively. In total, 95 patients completed the second PR-AOFAS (sent within 2 weeks of a previous one), of which 36 patients (38%) reported a change in complaints. Therefore, 59 patients (62%) were included in the test-retest analysis.

Reliability
The function subscale of the PR-AOFAS had sufficient internal consistency (Cronbach’s $\alpha$ 0.71, n=112). An insufficient internal consistency was found for the total PR-AOFAS (Cronbach’s $\alpha$ 0.68, n=112). The test-retest reliability was good (ICC 0.88; 95% CI 0.81 to 0.93; n=59). The standard error of measurement was 6.48 and gave rise to smallest detectable changes of 17.96 and 2.34 at individual and group (n=59) level, respectively, on a scale that ranged 0-100. The AOFAS had insufficient internal consistency for both the functional subscale (Cronbach’s $\alpha$ 0.40, n=112) and the total score (Cronbach’s $\alpha$ 0.47, n=112).

Construct validity

Spearman’s correlation coefficients were calculated between both PR-AOFAS and AOFAS and the Physical and Mental Component Summary scores of the SF-36, VAS pain, AOS and its subscales and the FAOS scales (Table II). Of the a priori formulated hypotheses, 95% was confirmed for the PR-AOFAS and 71% for the AOFAS. For the PR-AOFAS all correlations were as expected except for the correlations between the pain subscale from the PR-AOFAS and the FAOS symptoms subscale. For the AOFAS sub scales six hypotheses had to be rejected as correlations were not as expected (Table II).

Additionally, all hypotheses concerning the association between the two AOFAS versions were rejected as all correlation coefficient were smaller than expected.

Responsiveness

[Table I]
In total 8 effect sizes and standardized response means for both PR-AOFAS and the AOFAS were calculated to be 100% and 75% in agreement with the pre-defined hypotheses, respectively (Table III). The outcome disagreed with the hypotheses for the AOFAS in the category patients unchanged and exceeded the predefined hypothesis by 0.2. The AUC of the PR-AOFAS and the FAOS were 0.77 (95%CI 0.68 to 0.87) and 0.66 (95%CI 0.54.; 0.76), respectively, implying that only the PR-AOFAS was considered to meet the requirements of proper responsiveness.

Interpretability

Clear ceiling effects can be seen for both the alignment subscales of the PR-AOFAS and the AOFAS (Table I). A floor effect was observed for the AOFAS pain subscale (Table I).

MIC

The ROC based calculation of the MIC was 6.5 points (95%CI 0.6 to 14.4) for the PR-AOFAS and 17.5 points (95%CI 2.5 to 32.5) for the AOFAS. Accompanying sensitivity and specificity of the PR-AOFAS was 0.42 and 0.80, respectively. The sensitivity and specificity of the AOFAS was 0.36 and 0.89, respectively.

Agreement of the PR-AOFAS and the AOFAS

The ICC of the PR-AOFAS and the AOFAS was 0.70 (95%CI: 0.57 to 0.79) at baseline and 0.56 (95%CI: 0.41 to 0.69) for the change from baseline. Using the Bland-Altman analysis, a systematic difference of 4.6 points (95%CI 2.3 to 6.8) was found with limits of agreement of -18.4 and 27.6 points around this value in favour of the AOFAS (Figure 1).

[Table II]

[Table III]
Figure 1. Bland-Altman Plot for the PR-AOFAS and the AOFAS

AOFAS the original patient-physician-determined American Orthopaedic Foot and Ankle Society (AOFAS) score; PR-AOFAS the completely patient-reported version of AOFAS score; PROM Patient Reported Outcome Measure;
Discussion

Our key finding is that the Dutch version of the PR-AOFAS has sufficient construct validity, internal consistency, test-retest reliability and responsiveness in patients with ankle osteoarthritis who are willing to participate in a trial on injection therapy. The MIC was 6.5 points (95%CI 0.6 to 14.4) for the PR-AOFAS. The AOFAS has insufficient construct validity, internal consistency and responsiveness to change in the same population. The MIC was 17.5 points (95%CI 2.5 to 32.5) for the AOFAS. Although both outcome measures aim to measure identical constructs, correlation coefficients and agreement measures did not support this.

In contrast to the AOFAS, the PR-AOFAS has sufficient construct validity and had a higher correlation with other PROMs (Physical and Mental Component Summary scores of the SF-36, VAS pain, AOS, FAOS pain) than the AOFAS in this study or reported in the literature.[3, 4, 8, 9] Moderate correlations for the AOFAS with all scales of the FAOS and the PCS score of the SF-36 was in line with other studies.[3, 4, 8, 9]

A comparable test-retest (ICC 0.89) capability was found for the PR-AOFAS compared to what is previously reported (ICC 0.89) for the patient-reported items of the AOFAS (evaluating 5 of the 9 items).[3] Although the responsiveness was calculated differently in the literature, the PR-AOFAS was found to be sufficiently responsive and the AOFAS to be insufficiently responsive to change.[4, 5, 8, 9] Similar to previous studies, both the PR-AOFAS and the AOFAS had ceiling effects at baseline in the alignment subscale, but neither scores had ceiling or floor effects in the total score.[8, 9] One study found ceiling effects in the total score after 7.5 months following an ankle fracture, this is likely because a large group of patients fully recovered.[8]

A sufficient internal consistency was found for the function subscale (Cronbach’s α 0.71) and insufficient internal consistency for the total score (Cronbach’s α 0.68) of the PR-AOFAS.
Insufficient internal consistency was found for both the function subscale (Cronbach’s α 0.40) and total score (Cronbach’s α 0.47) of the AOFAS. A Cronbach's α between 0.70 and 0.95 demonstrates satisfactory internal consistency as long as the scale is unidimensional.[21] The internal consistency of the total scores should therefore be interpreted with caution. Currently there is limited comparison in the current literature. Pinsker et al. (2015) reported a Cronbach’s α of 0.84. However, all patient reported items were included in this analysis, including the pain scale (one item). This is incorrect as it is no longer a unidimensional measurement.[3] However, within orthopaedic surgery pain and function may be expected to be correlated, high Cronbach’s α values may therefore be possible for such multidimensional scores. Furthermore the physician determined items of the function scale are left out. De boer et al. (2017) did measure all 7 items of the function scale in two studies and reported a Cronbach’s α of 0.927 and 0.947.[8, 9] However these studies concerned patients with ankle fractures. It must also be taken into consideration that a too high a Cronbach’s α may be an indication that different questions are capturing the same symptoms or limitations.[21]

The minimal important change of the AOFAS had not previously been calculated for patients with ankle osteoarthritis. In this study the minimal clinically important change of the PR-AOFAS and the AOFAS was found to be smaller than its Smallest Detectable Change (individual). As a result this makes it less suitable for follow-up of individual patients in a clinical setting and more suitable for use in groups of patients, such as a research setting. The 95% CI from the MIC values of both outcome measures are large, indicating that this estimator should be interpreted with caution.

Although the PR-AOFAS was expected to be similar in construct it cannot replace the AOFAS
due to the systematic difference between the outcome measures and low level of agreement. There may be several explanations for this poor agreement. Firstly, the lower correlation could be explained by the fact that both instruments are measuring something else. The insufficient construct validity when compared to other PROM’s may be a possible explanation. However, also disability weights assigned by patients have been found to be higher than the weights based on the judgement of the physician in osteoarthritis and other musculoskeletal conditions.[35] This is further supported by the fact that the PR-AOFAS has a higher correlation with the other PROMs, as they all measure the patient’s perspective. One study also found that the AOFAS had a similar responsiveness to other PROMs once the objective items were removed.[2] Secondly, the correlation of identical patient-reported items also differed. Answering the same patient-reported items could therefore have been affected by the presence of a physician or by being in a hospital or clinical setting. Finally, the ceiling effect in the alignment subscale of both the PR-AOFAS and the AOFAS, and the floor effect in the pain subscale of the AOFAS, may result in restricted variation with a subsequent lower correlation.

Strengths of this study include the structural and elaborate evaluation of a new ankle-specific PROM with sufficient measurement properties and a moderate to high correlation with a large range of ankle-specific and generic measuring instruments; a large sample size; and the calculation of the minimal important change in a homogenous population with the same diagnosis. Limitations include a limited anchor with a 4-point Likert scale and small samples for both instruments in the worsened and greatly improved categories for the evaluation of the responsiveness. Also the content validity of neither PR-AOFAS nor the AOFAS was assessed. Content validity includes relevance and comprehensibility of the items. Relevance and comprehensibility, usually conducted through cognitive interviews of a small sample of patients,
measures the patient’s understanding of the item as it is intended (i.e. does the patient understand
the difference between ankle and foot mobility) and its deemed relevance to the construct.
Content validity should be evaluated in future research.

**Conclusion**

The Dutch version of the PR-AOFAS has sufficient construct validity, internal consistency, test-
retest reliability and responsiveness in patients with ankle osteoarthritis who are willing to
participate in a trial on injection therapy. The minimally important change of the PR-AOFAS is
smaller than its Smallest Detectable Change making it suitable for use in groups of patients, such
as a research setting.

**References**

   the foot and ankle. *Journal of Orthopaedic and Sports Physical Therapy.*
   https://doi.org/10.2519/jospt.2007.2403

   S. E. (2012). Responsiveness and validity of the SF-36, Ankle osteoarthritis scale, AOFAS
   ankle hindfoot score, and foot function index in end stage ankle arthritis. *Foot and Ankle
   International, 33*(1), 57–63. https://doi.org/10.3113/FAI.2012.0057

   and validity of 6 measures of pain, function, and disability for ankle arthroplasty and
   arthrodesis. *Foot and Ankle International.* https://doi.org/10.1177/1071100714566624


https://doi.org/10.1053/j.jfas.2016.06.001


35. Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., …

**Tables and Figures**

Table I Descriptives and percentages highest and lowest scores of the completely patient-reported AOFAS, original patient-physician-determined AOFAS, SF-36 subscales, VAS pain, AOS and FAOS at baseline

<table>
<thead>
<tr>
<th>n=112</th>
<th>Mean (SD)</th>
<th>Median (IQR)</th>
<th>Highest score, n (%)</th>
<th>Lowest score, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR-AOFAS</td>
<td>59.4 (17.9)</td>
<td>64.0 (50.0 to 72.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PR-AOFAS Pain</td>
<td>19.7 (8.6)</td>
<td>20 (20.0 to 20.0)</td>
<td>1 (0.9%)</td>
<td>14 (12.5%)</td>
</tr>
<tr>
<td>PR-AOFAS Function</td>
<td>31.2 (10.3)</td>
<td>33.0 (26.0 to 38.0)</td>
<td>1 (0.9%)</td>
<td>0</td>
</tr>
<tr>
<td>PR-AOFAS alignment</td>
<td>8.1 (3.2)</td>
<td>10.0 (5.0 to 10.0)</td>
<td>78 (69.6%)</td>
<td>9 (8%)</td>
</tr>
<tr>
<td>AOFAS</td>
<td>63.6 (14.4)</td>
<td>67.0 (57.0 to 72.0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AOFAS Pain</td>
<td>18.1 (9.0)</td>
<td>20.0 (20 to 20)</td>
<td>0</td>
<td>19 (17%)</td>
</tr>
<tr>
<td>AOFAS Function</td>
<td>36.3 (7.3)</td>
<td>37.0 (32.0 to 42.0)</td>
<td>1 (0.9%)</td>
<td>0</td>
</tr>
<tr>
<td>Measure</td>
<td>Mean (SD)</td>
<td>Range</td>
<td>Good (%)</td>
<td>Poor (%)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------</td>
<td>---------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>AOFAS alignment</strong></td>
<td>9.2 (2.3)</td>
<td>10.0 (10 to 10)</td>
<td>97 (86.6%)</td>
<td>3 (2.7%)</td>
</tr>
<tr>
<td>SF-36 Physical Component</td>
<td>44.3</td>
<td>45.3 (38.9 to 50.2)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Component Summary score</strong></td>
<td></td>
<td>(7.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-36 Mental Component</td>
<td>43.4</td>
<td>43.1 (39.9 to 47.8)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Summary score</strong></td>
<td></td>
<td>(5.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAS pain</td>
<td>53.0</td>
<td>52.5 (40.0 to 69.5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AOS</td>
<td>48.4</td>
<td>47.1 (33.9 to 62.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOS pain</td>
<td>39.9</td>
<td>39.5 (24.5 to 52.5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AOS disability</td>
<td>57.0</td>
<td>58.8 (41.8 to 73.4)</td>
<td>0</td>
<td>2 (1.8%)</td>
</tr>
<tr>
<td>FAOS pain</td>
<td>55.4</td>
<td>52.8 (41.7 to 71.5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FAOS symptoms</td>
<td>46.0</td>
<td>46.4 (32.1 to 57.1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FAOS ADL</td>
<td>70.7</td>
<td>75.7 (52.9 to 87.9)</td>
<td>2 (1.8%)</td>
<td>0</td>
</tr>
<tr>
<td>FAOS sport &amp; R</td>
<td>33.5</td>
<td>30.0 (15.0 to 45.0)</td>
<td>0</td>
<td>6 (5.4%)</td>
</tr>
</tbody>
</table>

*Journal Pre-proof*
FAOS QoL  28.2  31.3 (18.8 to 37.5)  0  9 (8.0%) 
(16.6) 

PR-AOFAS The completely patient reported version of the American Orthopaedic Foot and Ankle Society (AOFAS) score; PR Patient Reported; SF-36 Short Form 36; AOFAS original patient-physician determined AOFAS; VAS Visual Analogue Scale; AOS Ankle Osteoarthritis Score; FAOS Foot and Ankle Outcome Score; n number; SD Standard Deviation; IQR Interquartile Range; PCS Physical Component Score; MCS Mental Component Score; ADL Activities of Daily Living; S & R Sport and Recreation; QoL Quality of Life.
Table II. Construct validity using Spearman’s correlation coefficients ($R_s$) between the PR-AOFAS and AOFAS, Physical and Mental Component Summary scores of the SF-36, VAS pain, AOS and the FAOS scales

<table>
<thead>
<tr>
<th></th>
<th>PR-AOFAS</th>
<th>PR-AOFAS</th>
<th>PR-AOFAS</th>
<th>PR-AOFAS</th>
<th>AOFAS</th>
<th>AOFAS</th>
<th>AOFAS</th>
<th>AOFAS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Pain</td>
<td>Function</td>
<td>Alignment</td>
<td>Total</td>
<td>Pain</td>
<td>Function</td>
<td>Alignment</td>
</tr>
<tr>
<td>AOFAS Total</td>
<td>0.70</td>
<td>0.54</td>
<td>0.67</td>
<td>0.36</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AOFAS Pain</td>
<td>0.55</td>
<td><strong>0.63</strong></td>
<td>0.43</td>
<td>0.23</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AOFAS Function</td>
<td>0.62</td>
<td>0.32</td>
<td><strong>0.67</strong></td>
<td>0.34</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AOFAS Alignment</td>
<td>0.30</td>
<td>0.15</td>
<td>0.31</td>
<td><strong>0.28</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td><strong>0.68</strong></td>
<td>0.44</td>
<td><strong>0.68</strong></td>
<td>0.35</td>
<td>0.66</td>
<td>0.47</td>
<td>0.62</td>
<td>0.28</td>
</tr>
<tr>
<td>SF-36 MCS</td>
<td>-0.24</td>
<td>-0.18</td>
<td>-0.23</td>
<td>-0.11</td>
<td>-0.19</td>
<td>-0.18</td>
<td>-0.12</td>
<td>-0.20</td>
</tr>
<tr>
<td>VAS pain</td>
<td>-0.71</td>
<td>-0.57</td>
<td>-0.64</td>
<td>-0.39</td>
<td>-0.61</td>
<td>-0.52</td>
<td>-0.49</td>
<td>-0.19</td>
</tr>
<tr>
<td>AOS</td>
<td>-0.68</td>
<td>-0.51</td>
<td>-0.65</td>
<td>-0.34</td>
<td>-0.55</td>
<td>-0.48</td>
<td>-0.47</td>
<td>-0.14</td>
</tr>
<tr>
<td>AOS pain</td>
<td>-0.66</td>
<td>-0.56</td>
<td>-0.60</td>
<td>-0.34</td>
<td>-0.52</td>
<td>-0.51</td>
<td>-0.42</td>
<td>-0.10</td>
</tr>
<tr>
<td>AOS disability</td>
<td>-0.56</td>
<td>-0.35</td>
<td>-0.57</td>
<td>-0.27</td>
<td>-0.44</td>
<td>-0.34</td>
<td>-0.41</td>
<td>-0.15</td>
</tr>
<tr>
<td>FAOS pain</td>
<td><strong>0.71</strong></td>
<td><strong>0.59</strong></td>
<td>0.63</td>
<td>0.30</td>
<td><strong>0.50</strong></td>
<td><strong>0.53</strong></td>
<td>0.36</td>
<td>0.05</td>
</tr>
<tr>
<td>FAOS symptoms</td>
<td><strong>0.39</strong></td>
<td>0.28</td>
<td>0.36</td>
<td>0.18</td>
<td><strong>0.14</strong></td>
<td><strong>0.25</strong></td>
<td>0.06</td>
<td>-0.09</td>
</tr>
<tr>
<td>FAOS ADL</td>
<td><strong>0.69</strong></td>
<td>0.51</td>
<td><strong>0.67</strong></td>
<td>0.33</td>
<td><strong>0.54</strong></td>
<td>0.54</td>
<td><strong>0.43</strong></td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.59</td>
<td>0.61</td>
<td>0.60</td>
<td>0.40</td>
<td>0.36</td>
<td>0.36</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>FAOS S &amp; R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAOS QoL</td>
<td>0.57</td>
<td>0.57</td>
<td>0.58</td>
<td>0.58</td>
<td>0.46</td>
<td>0.42</td>
<td>0.38</td>
<td>0.17</td>
</tr>
</tbody>
</table>

**Bold** illustrates a-priori defined hypothesized correlations. Underlines confirm the a-priori defined hypotheses. Negative relations are due to the reversed scale for these measures. PR-AOFAS The completely patient-reported version of the American Orthopaedic Foot and Ankle Society (AOFAS) score; PROM Patient Reported Outcome Measure; AOFAS original patient-physician-determined AOFAS; PCS Physical Component Summary score of the Short Form-36, MCS Mental Component Summary score of the Short Form-36, VAS Visual Analogue Scale; AOS Ankle Osteoarthritis Score; FAOS Foot and Ankle Outcome Score; ADL Activities of Daily Living; S & R Sport and Recreation; QoL Quality of Life;
Table III. Descriptives, effect sizes and standardized response means of the completely patient reported AOFAS and the original patient-physician-determined AOFAS to change

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Satisfaction</th>
<th>n</th>
<th>Mean (SD) Baseline</th>
<th>Mean (SD) 26 weeks</th>
<th>Mean Difference (95% CI)</th>
<th>Effect size</th>
<th>Standardized Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR-AOFAS score</td>
<td>Worsened</td>
<td>8</td>
<td>64.5 (19.4)</td>
<td>55.4 (20.1)</td>
<td>-9.13 (-25.5; 7.3)</td>
<td>-0.47</td>
<td>-0.47</td>
</tr>
<tr>
<td></td>
<td>Unchanged</td>
<td>38</td>
<td>59.4 (18.1)</td>
<td>62.3 (23.9)</td>
<td>2.84 (-2.7; 8.3)</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Improved</td>
<td>46</td>
<td>58.8 (17.6)</td>
<td>74.5 (15.4)</td>
<td>15.70 (11.5; 19.9)</td>
<td>0.89</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>Greatly improved</td>
<td>4</td>
<td>53.3 (14.0)</td>
<td>78.5 (12.6)</td>
<td>25.25 (11.8; 38.7)</td>
<td>1.80</td>
<td>2.98</td>
</tr>
<tr>
<td>AOFAS</td>
<td>Worsened</td>
<td>8</td>
<td>67.1 (11.9)</td>
<td>70.4 (15.1)</td>
<td>3.25 (-7.1; 13.6)</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Unchanged</td>
<td>38</td>
<td>62.1 (15.3)</td>
<td>68.9 (15.5)</td>
<td>6.84 (1.9; 11.7)</td>
<td>0.45</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Improved</td>
<td>46</td>
<td>65.2 (14.4)</td>
<td>78.0 (12.9)</td>
<td>12.74 (9.0; 16.5)</td>
<td>0.88</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>Greatly improved</td>
<td>4</td>
<td>55.2 (15.5)</td>
<td>80.8 (4.8)</td>
<td>25.60 (9.5; 41.7)</td>
<td>1.65</td>
<td>1.97</td>
</tr>
</tbody>
</table>

PR-AOFAS the completely patient-reported version of the American Orthopaedic Foot and Ankle Society (AOFAS) score; AOFAS the original patient-physician-determined AOFAS; PROM Patient Reported Outcome Measure; n number; 95% CI Confidence Interval;
Figure 1. Bland-Altman Plot for the completely patient reported AOFAS and the original patient-physician-determined AOFAS

AOFAS the original patient-physician-determined American Orthopaedic Foot and Ankle Society (AOFAS) score; PR-AOFAS the completely patient-reported version of AOFAS score; PR: Patient Reported; CI indicates the confidence interval.