Author(s): Uimonen, Mikko M.; Ponkilainen, Ville T.; Toom, Alar; Miettinen, Mikko; Häkkinen, Arja H.; Sandelin, Henrik; Latvala, Antti O.; Sirola, Timo; Sampo, Mika; Roine, Risto P.; Lindahl, Jan; Ilves, Outi; Sandbacka, Anna; Repo, Jussi P.

Title: Validity of five foot and ankle specific electronic patient-reported outcome (ePRO) instruments in patients undergoing elective orthopedic foot or ankle surgery

Year: 2021

Version: Accepted version (Final draft)

Please cite the original version:

Uimonen, M. M., Ponkilainen, V. T., Toom, A., Miettinen, M., Häkkinen, A. H., Sandelin, H., Latvala, A. O., Sirola, T., Sampo, M., Roine, R. P., Lindahl, J., Ilves, O., Sandbacka, A., & Repo, J. P. (2021). Validity of five foot and ankle specific electronic patient-reported outcome (ePRO) instruments in patients undergoing elective orthopedic foot or ankle surgery. Foot and Ankle Surgery, 27(1), 52-59. https://doi.org/10.1016/j.fas.2020.02.003

Validity of five foot and ankle specific electronic patient-reported outcome (ePRO) instruments in patients undergoing elective orthopedic foot or ankle surgery

Mikko M. Uimonen, Ville T. Ponkilainen, Alar Toom, Mikko Miettinen, Arja H. Häkkinen, Henrik Sandelin, Antti O. Latvala, Timo Sirola, Mika Sampo, Risto P. Roine, Jan Lindahl, Outi Ilves, Anna Sandbacka, Jussi P. Repo

PII: S1268-7731(20)30029-1

DOI: https://doi.org/10.1016/j.fas.2020.02.003

Reference: FAS 1415

To appear in: *Foot and Ankle Surgery*

Received Date: 12 August 2019

Revised Date: 18 January 2020

Accepted Date: 6 February 2020

Please cite this article as: Uimonen MM, Ponkilainen VT, Toom A, Miettinen M, Häkkinen AH, Sandelin H, Latvala AO, Sirola T, Sampo M, Roine RP, Lindahl J, Ilves O, Sandbacka A, Repo JP, Validity of five foot and ankle specific electronic patient-reported outcome (ePRO) instruments in patients undergoing elective orthopedic foot or ankle surgery, *Foot and Ankle Surgery* (2020), doi: https://doi.org/10.1016/j.fas.2020.02.003

Validity of five foot and ankle specific electronic patient-reported outcome (ePRO) instruments in patients undergoing elective orthopedic foot or ankle surgery

Mikko M. Uimonen[1], Ville T. Ponkilainen[1], Alar Toom[1], Mikko Miettinen[4], Arja H. Häkkinen[2,3], Henrik Sandelin[4], Antti O. Latvala[5], Timo Sirola[4], Mika Sampo[6], Risto P. Roine[7,8], Jan Lindahl[4], Outi Ilves[2], Anna Sandbacka[5], Jussi P. Repo[1]

[1] *Department of Surgery, Central Finland Central Hospital, Jyväskylä, Finland*

[2] *Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland*

[3] *Department of Physical Medicine, Central Finland Healthcare District, Jyväskylä, Finland*

[4] *Department of Orthopedics and Traumatology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland*

[5] *Division of Orthopedic and Trauma Surgery, Department of Surgery, Medical Research Center Oulu, Oulu University Hospital, University of Oulu, Oulu, Finland*

[6] *Comprehensive Cancer Center, Helsinki University Hospital and University of Helsinki, Helsinki, Finland*

[7] *Group Administration, University of Helsinki and Helsinki University Hospital, HUS, Helsinki, Finland*

[8] *Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland*

Address for correspondence: Jussi P. Repo, Department of Surgery, Central Finland Central Hospital, Keskussairaalantie 19, FI-40620 Jyväskylä, Finland; Email: *mrjussirepo@gmail.com*; Phone: +358 14 2691811

Highlights

- Patient-reported outcomes (PROs) are widely accepted measures for evaluating outcomes of surgical interventions. However, there are only few valid electronic PRO (ePRO) instruments available for clinicians and researchers. The aim of this study was to evaluate the validity of electronic versions of five widely used foot and ankle specific PRO instruments.
- The ePRO versions of the Foot and Ankle Ability Measure (FAAM), the Foot and Ankle Outcome Score (FAOS), the Manchester-Oxford Foot Questionnaire (MOXFQ), and the Visual Analogue Scale Foot and Ankle (VAS-FA) provide valid scores for foot and ankle patients.
- Our findings do not support the use of the modified Lower Extremity Function Scale (LEFS) as an electronic outcome measure for patients with orthopedic foot and/or ankle pathologies.

Abstract

Background: Patient-reported outcomes (PROs) are widely accepted measures for evaluating outcomes of surgical interventions. As patient-reported information is stored in electronic health records, it is essential that there are valid electronic PRO (ePRO) instruments available for clinicians and researchers. The aim of this study was to evaluate the validity of electronic versions of five widely used foot and ankle specific PRO instruments.

Methods: Altogether 111 consecutive elective foot/ankle surgery patients were invited face-to-face to participate in this study. Patients completed electronic versions of the Foot and Ankle Ability Measure (FAAM), the Foot and Ankle Outcome Score (FAOS), the modified Lower Extremity Function Scale (LEFS), the Manchester-Oxford Foot Questionnaire (MOXFQ), and the Visual Analogue Scale Foot and Ankle (VAS-FA) on the day of elective foot and/or ankle surgery. Construct validity, coverage, and targeting of the scales were assessed.

Results: Based on general and predefined thresholds, construct validity, coverage, and targeting of the ePRO versions of the FAAM, the FAOS, the MOXFQ, and the VAS-FA were acceptable. Major issues arose with score distribution and convergent validity of the modified LEFS instrument.

Conclusions: The ePRO versions of the FAAM, the FAOS, the MOXFQ, and the VAS-FA provide valid scores for foot and ankle patients. However, our findings do not support the use of the modified LEFS as an electronic outcome measure for patients with orthopedic foot and/or ankle pathologies.

Keywords: ePRO, Foot, Ankle, Validation, Psychometrics, Clinimetrics, Patient-reported outcome

Introduction

Since the introduction of evidence-based medicine in 1992, the outcome assessment of surgical treatment using patient reported outcome (PRO) instruments has attracted increasing interest [1, 2]. The main benefits of using PROs are that they allow the detection of changes in patients' functional status, the evaluation of differences between patients with similar conditions, and the evaluation of the effectiveness of different treatments [3].

PROs are widely accepted measures for the evaluation of outcomes after surgical interventions [2, 3]. A wide variety of PRO instruments is available. For example, there are at least 139 different scales used in the field of foot and ankle research [2]. However, a consensus has been lacking on which PROs are valid and reliable for assessing outcomes in foot and ankle surgery [3]. For this reason, among others, the Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) has developed consensus-based checklists to assess and guide the validation of PRO instruments [4, 5]. As the PROs will potentially be stored in electronic patient records, it is essential that there are valid electronic PRO (ePRO) instruments available for clinicians and researchers. These ePROs have been proposed to have comparable validity to the

paper-and-pencil version [6-8]. ePROs will also facilitate data management by investigators and may improve patient compliance [8]. The potential benefits of using ePRO instruments comprise more accurate and complete data, a diminished administrative burden, and lower costs [9-14]. It is essential to evaluate and compare the psychometric properties of ePROs to facilitate selection of the optimal instrument for the population being investigated [7, 17].

The aim of this study was to assess and compare measurement properties of electronic versions of five widely used foot and ankle specific PRO instruments: the Foot and Ankle Ability Measure (FAAM), the Foot and Ankle Outcome Score (FAOS), the modified Lower Extremity Function Scale (LEFS), the Manchester-Oxford Foot Questionnaire (MOXFQ), and the Visual Analogue Scale Foot and Ankle (VAS-FA). A further aim was to examine the relationships between these five ePROs and against sociodemographic and clinical factors and health-related quality of life (HRQoL).

Materials and methods

The Ethics Committee of the Helsinki and Uusimaa Hospital District, Helsinki, Finland, approved the study protocol. Patients were recruited from two large academic centers: Oulu University Hospital (Oulu, Finland) and Peijas Helsinki University Hospital (Vantaa, Finland). Peijas Hospital is currently the largest foot and ankle specialist outpatient clinic in the Nordic countries. The study inclusion criteria were planned foot and ankle surgery, full understanding of written Finnish, and signing an informed consent adhering to the principles of the Declaration of Helsinki [18]. Exclusion criteria were age less than 18 years and severe untreated mental illness (e.g. schizophrenia). In order to acquire sufficient sample size for reliable validity testing, altogether 111 consecutive prospective elective foot and/or ankle surgery patients were invited face-to-face to participate in this study [19].

Patients completed electronic versions of five foot and ankle specific PRO instruments on the day of surgery before the operation using a tablet computer (iPad, Apple). Data were collected using a Webropol survey platform (Webropol Oy, Helsinki, Finland). The patients completed the ePROs in the following order: LEFS, VAS-FA, MOXFQ, FAOS, and FAAM. Furthermore, patients completed a preoperative information form containing clinical and sociodemographic questions as well as an item regarding current general health state of the patients with five response categories from 1 (Excellent) to 5 (Poor). Indication for operation was obtained from the operating surgeons.

Patient-reported outcome measures

Adaptation process for ePRO instruments

All PRO measures have been translated and culturally adapted to Finnish in adherence to the ISPOR guidelines [20] and with the consent or license of the copyright holders. The conversion to electronic versions conformed to the Clinical Outcomes at Oxford University Innovation guidelines for production of electronic clinical outcome assessment (eCOA) measures [7, 21]. Only minor modifications that did not affect content of the questionnaires or items were made. The ePRO versions underwent cognitive testing with 10 patients of the target population, which revealed minor problems with the used software. These issues were corrected before the beginning of the study.

Foot and Ankle Ability Measure (FAAM)

The FAAM is a foot and ankle specific PRO instrument that was developed to evaluate the ability of the ankle, foot, and leg [22]. It contains 28 items. Response options on a five-point scale include “No difficulty”, “Slight difficulty”, “Moderate difficulty”, “Extreme difficulty”, and “Unable to”. Each question is scored from 0 to 4 points, with a low score indicating high ability, and vice versa. The total score is calculated as the sum of all answers, divided by the highest potential score and multiplied by 100. Therefore, the maximum total score is 100, and unanswered items do not affect

the total score. The FAAM can be divided into two subscales: Activities of Daily Living (ADL) (21 items) and Sports (7 items). Minimal clinically important difference (MCID) for the subscales are 5.7 points for ADL and 12.3 points for Sports. FAAM have shown strong relationships with the SF-36 physical function and physical component domains. [22]

Foot and Ankle Outcome Score (FAOS)

The FAOS is a foot and ankle specific PRO instrument that was developed to evaluate pain, function, symptoms, and quality of life [23]. It contains 42 items scored on a five-point scale from 0 to 4 points. The items are scored either by the frequency ("Never", "Rarely", "Sometimes", "Often", "Always") or severity ("None", "Mild", "Moderate", "Severe", "Extreme") of the symptoms. The FAOS can be divided into five subscales: Pain (9 items), Symptoms (7 items), ADL (17 items), Sport and recreation function (5 items), and Quality of life (QoL) (4 items). The total score is calculated as the sum of all items of the subscales, divided by the highest possible score, and multiplied by 100. Therefore, the highest total score is 100 and the lowest total score 0. A higher score means more symptoms. The Cronbach alphas for internal consistency have been high: 0.94, 0.88, 0.97, 0.94, 0.92 for Pain, Symptoms, ADL, Sport and recreation function and QoL, respectively. Test-retest reliability have been high, as Spearman correlation coefficients have been 0.96, 0.89, 0.85, 0.92 and 0.92 for the subscales. [23]

Modified Lower Extremity Function Scale (LEFS)

The LEFS is a PRO instrument that was developed to evaluate the function of the lower extremity [24]. The modified version has been shown to be valid for foot and ankle patients using Rasch analysis [25]. It contains 15 items and the answers are on a four-point scale. The categories are scored from 1, representing "Extreme difficulty or unable to perform activity", to 4 points, representing "No difficulty". Test-retest reliability for LEFS have been excellent (r=0.94, Spearman

correlation coefficient), and it has strong relationship with SF-36 physical function (r=0.80) and physical component (r=0.64) domains. The MCID is 9 points. [24]

Manchester-Oxford Foot Questionnaire (MOXFQ)

The MOXFQ is a foot and ankle specific PRO instrument that was developed to evaluate outcomes after hallux valgus surgery [26]. Thereafter, it has been validated for foot and ankle surgery [27, 28]. It contains 16 items and the response options are on a five-point scale from 1 to 5 [26]. The MOXFQ consists of three domains: Walking/standing (7 items), Pain (5 items), and Social interaction (4 items) [28]. The scores are converted to a scale from 0 to 100, where 0 represents low symptoms and 100 the most severe symptoms [28]. The original work reported a Cronbach's alpha of 0.93, indicating high internal consistency [29]. MOXFQ have strong relationship with the SF-36 Physical functioning, Role physical, and Pain domains [30].

Visual Analogue Scale Foot and Ankle (VAS-FA)

The VAS-FA has been validated for foot and ankle patients to assess function, pain, and other complaints [31-33]. The scale contains 20 items on a visual analog scale (0-100 mm, worst to best). The total score is calculated by dividing the result by the highest potential score of the completed items, awarding points between 0 and 100. The VAS-FA allows dividing the items into three modules as follows: Pain (4 items), Function (11 items), and Other complaints (5 items). VAS-FA has shown high correlation (r>0.5) with general health related quality of life (HRQoL) instrument SF-36 [34].

EuroQol instrument (EQ-5D-3L)

The EQ-5D is a generic HRQoL instrument [35] consisting of five dimensions: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression. Each dimension item has a three-level response option: no problems (1), some problems (2), or severe problems (3). The EQ-5D index is

calculated from the dimension scores. In the Finnish version of the EQ-5D, the index score varies from -0.011 to 1, with a lower score indicating poorer HRQoL [36].

Statistical methods

Data are presented as means with standard deviations (SD) or 95% confidence intervals (95% CI), medians with interquartile ranges (IQR), or as counts with percentages. To obtain comparable and parallel scoring for all ePROs, the LEFS and the EQ-5D index score were converted into a 0 to 100 scale, and the FAAM, the FAOS, and the MOXFQ scores were inverted by subtracting the score from 100. After the modifications, all scores of the ePROs were from 0 to 100, with higher scores indicating better outcome. Predefined hypotheses on psychometric features examined are presented in Table 1. The distributions of the scores were assessed to test scale coverage and targeting of ePROs. Floor or ceiling effect was considered confirmed if more than 15% of the patients scored a minimum or a maximum score, respectively [37].

Construct validity refers to the degree to which the instrument scores are consistent with hypotheses regarding internal relationships of the items, relationships with scores of other instruments, and differences between relevant groups. The construct validity of the ePROs was evaluated with regard to structural and convergent validity as well as independency of the sociodemographic and clinical factors. Structural validity refers to the extent to which the dimensions or subscales of the PRO instrument reflect the dimensionality of the measured construct. Structural validity was examined by assessing the internal structure of each ePRO using exploratory factor analysis (EFA) and by assessing the internal consistency by calculating Cronbach's alphas. EFA was used to test unidimensionality of the ePROs and their subscales. A factor loading of 0.4 was used as a cut-off value, indicating that the item typifies the factor sufficiently [38]. Internal consistency of the ePROs was evaluated by calculating Cronbach's alphas of the whole scales and also the subscales. An alpha value of over 0.7 indicates acceptable internal

consistency [39]. Bootstrapping method with 1000 sample replications was used to obtain 95% CIs of the alphas.

Calculation of Spearman correlations of ePRO scores with age and BMI, as well as comparison of the mean ePRO scores between subgroups of sexes, indication for operative treatment, and previous operations of the affected foot or ankle with independent samples t-test served for assessing independency of the sociodemographic and clinical factors. The mean ePRO scores of the two largest subgroups of indication for operative treatment, the patients with osteoarthritis and the patients with foot or ankle deformity, were compared. The strength of the correlation was interpreted as follows: 0.00–0.30 negligible, 0.30–0.50 low, 0.50–0.70 moderate, 0.70–0.90 high and 0.90–1.00 very high correlation. We hypothesized non-significant associations between the ePRO scores and tested factors. Non-significant correlations or t-test values represent independency of the sociodemographic and clinical factors. Independency of these factors improves comparability of these scores between patients with different characteristics.

Convergent validity refers to the extent to which the scores of the instruments supposed to measure the same construct are related to each other. To examine convergent validity, relationships between each ePROs scores were assessed using Spearman correlation coefficients. Higher correlation coefficient values represent better convergent validity. In addition, Spearman correlation coefficients between ePRO scores and EQ-5D index were calculated to assess the relationship between ePROs and HRQoL. To obtain the 95% CIs, bootstrapping method with 1000 replications was used.

SPSS 25.0 (IBM® SPSS® Statistics, USA) and R-3.4.2 software was used for statistical analysis.

Results

None of the patients refused to participate, resulting in a participation rate of 100% (Table 2). The effective response rates were 87%, 92%, 88%, 85%, and 95% for the FAAM, the FAOS, the LEFS, the MOXFQ, and the VAS-FA, respectively.

Scale coverage and targeting

The distributions of the FAOS, the MOXFQ, and the VAS-FA, and their subscales' scores followed normal distribution, with exceptions of single subscales (FAOS ADL and FAOS QoL). The FAAM and the LEFS scores were focused towards the high scores, representing good outcomes. No floor or ceiling effects were observed either in the total scores of the ePROs or in the subscale scores.

Construct validity

Structural validity

EFA of the FAAM ADL subscale revealed two factors (eigenvalues for factors 1 and 2 were 12.9 and 1.02, respectively). All items loaded strongest on factor 1 (loading value range 0.43 – 0.90). In EFA of the Sport subscale of the FAAM, one factor was found (eigenvalue 5.08, loading value range 0.77 – 0.89).

In EFAs of the FAOS subscales, only one factor was found for each subscale (eigenvalues for factors of Pain, Symptom, ADL, Sport, and QoL were 4.65, 2.69, 11.3, 3.3, and 2.49, respectively). All items of each subscale, except for the Symptom subscale, loaded strongly on these factors (loading value ranges for Pain, ADL, Sport, and QoL were 0.57 – 0.84, 0.69 – 0.90, 0.67 – 0.95, and 0.65 – 0.86, respectively). For the Symptom subscale, all loading values, except for two items, were over 0.4 (range 0.56 – 0.82). The loadings of items 1 and 2 of the Symptom subscale on factor 1 were 0.30 and 0.36, respectively, indicating insufficient representation of these items on factor 1.

EFA of the LEFS revealed three factors (eigenvalues 7.21, 1.44, and 1.04). All items of the LEFS, except for items 2 and 13, loaded strongest on factor 1 (loading value range 0.51 – 0.86). Although items 2 and 13 loaded stronger on factors 2 and 3, respectively, the loading values on factor 1 were also over 0.4 (0.49 for item 2 and 0.44 for item 13), indicating that these items typify the factor 1 sufficiently.

EFA of the MOXFQ subscales revealed only one factor for each subscale (eigenvalues for factors of the Pain, Walking, and Social subscales were 2.77, 4.09, and 1.55, respectively). The corresponding factor loadings of the items on these subscales ranged between 0.71 and 0.79, 0.69 and 0.82, and 0.54 and 0.72.

EFAs of the VAS-FA subscales revealed one factor for each subscale (eigenvalues 2.07, 5.56, and 1.20 for Pain, Function, and Other subscales, respectively). All items in each subscale loaded strongest on these factors (loading values ranges 0.65 – 0.79, 0.40 – 0.83, and 0.50 – 0.65 for Pain, Function, and Other subscales, respectively).

Cronbach's alphas indicated acceptable internal consistency for all instruments and their subscales, except for the MOXFQ Social and the VAS-FA Other complaints subscales, which had alphas below 0.7 (Table 3).

Independency of the sociodemographic and clinical factors

The examination of associations of the ePRO scores with sociodemographic and clinical factors revealed significant associations in the minority of examined cases (Table 4). There was a low negative correlation between age and FAOS ADL score age (FAOS ADL and age: $r = -0.34$, $p < 0.001$) whereas other significant correlations were negligible strength. Male patients obtained higher scores from VAS-FA Other complaints subscale than female (70 vs. 59, $p = 0.011$). No other significant associations were observed.

Convergent validity

All ePRO scores, except LEFS scores, were correlated (Table 5). Furthermore, there were significant correlations between similar subscale scores. The FAAM ADL and the FAOS ADL subscales were strongly correlated, whereas the LEFS was not correlated with the ADL subscales of the FAAM or the FAOS. The FAAM Sport and the FAOS Sport subscales were highly correlated. The Pain subscales of the FAOS, the MOXFQ, and the VAS-FA were moderately correlated with each other. The Walking subscale of the MOXFQ had a high correlation with the VAS-FA Function subscale. The correlations of all ePRO scores, except for those of the LEFS and the FAOS Symptoms subscale, with the EQ-5D index were of low to moderate strength, albeit significant (Figure 1).

Discussion

The ePRO versions of the FAAM, the FAOS, the MOXFQ, and the VAS-FA provide valid scores in foot and ankle surgery patients. These instruments have acceptable construct validity, structural validity, convergent validity, and independency of the sociodemographic and clinical factors. Furthermore, none of these instruments had problems concerning coverage and targeting (Table 1). However, the LEFS showed major issues with score distribution and convergent validity.

The effective response rate of each ePRO was high. Scale targeting and coverage were appropriate in light of score distributions and the absence of floor and ceiling effects in all ePROs. Nevertheless, the scores of the FAAM and the LEFS were distributed towards the upper end of the scale. Furthermore, structural validity of each ePRO instrument was good, with the exceptions of single subscales with low internal consistency (MOXFQ Social and VAS-FA Other complaints) or non-unidimensional structure (FAOS Symptoms). The findings indicate that the subscales of each ePRO reflect the dimensionality of the measured constructs at least sufficiently. Examination of the

independence from sociodemographic and clinical factors revealed mainly minor violations in all instruments, except for the MOXFQ, which was not associated with the tested demographic and clinical characteristics. The results support the assumption that the variance in the ePRO scores is due to variance in outcomes rather than characteristics of patients. Convergence of all ePROs, except for the LEFS, was high according to the correlations between total scores of the scales. This indicates good convergent validity of ePROs other than the LEFS. ePRO scores seemed to reflect HRQoL-associated construct validity for the FAAM, the FAOS, the MOXFQ, and the VAS-FA and to a lower extent for the LEFS.

The FAAM and the LEFS are outcome measure instruments that focus purely on the functionality of the foot and ankle, whereas the other instruments in this study contain items focusing on other outcome dimensions such as pain, quality of life, and social interaction. The convergence of the FAAM and the LEFS measured by correlation against the generic HRQoL instrument EQ-5D was inferior relative to the other instruments. This may be due to function-related constructs measured by these ePROs, as they do not include HRQoL-related items, unlike the other ePROs. Previous studies have reported both the FAAM and the LEFS scores to be strongly associated with Physical function and Physical component domains of the generic HRQoL instrument SF-36, while there were no associations with Mental component scores [22, 24]. Thus, the LEFS and the FAAM should be considered as specific foot and ankle functionality outcome instruments, rather than comprehensive outcome instruments. Furthermore, the skewness of the distributions of the FAAM and the LEFS scores may be due to patients' foot and ankle issues affecting outcome dimensions other than functionality of the foot. Furthermore, unlike the FAAM scores, the LEFS score did not correlate with other ePROs. As the FAAM questionnaire focuses on the foot condition during last week prior to administration of the questionnaire, the LEFS focuses on the current day of the administration. The longer survey period of the FAAM might capture more general perspective to the functionality and influence of the foot condition in the daily life of

the patients, than that of the LEFS. This is in line with the higher correlations of the FAAM with other ePROs that contains wider scope on the outcome. Thus, the LEFS may be considered even more strictly as a pure foot functionality outcome measure than the FAAM.

In contrast to our results, in the study by Pinsker et al. (2015) [40] comparing psychometric properties of six lower-extremity specific PRO instruments (Foot Function Index FFI, Ankle Osteoarthritis Scale AOS, patient-reported items of the American Orthopaedic Foot and Ankle Society Questionnaire AOFAS, Lower Extremity Functional Scale LEFS, Western Ontario and McMaster Universities Osteoarthritis Index WOMAC, and the Short Musculoskeletal Function Assessment SMFA) in ankle arthroplasty and arthrodesis patients, no clear differences emerged in construct validity between the foot and ankle specific scales (FFI, AOS, AOFAS) and broader lower-extremity scales (LEFS, WOMAC) or general musculoskeletal measure SMFA. The authors concluded that none of the instruments was superior to the others in measurement properties. In addition, a study by Goldstein et al. (2010) [41] compared psychometrics of six foot and ankle specific PRO instruments (Short Form-12 Physical, SMFA, FFI, FAAM, American Academy of Orthopedic Surgeons AAOS) in patients with foot or ankle trauma. They concluded that the region-specific instruments (FFI, FAAM, AAOS, AOFAS) and the more general function instruments (SF-12, SMFA) perform evenly with regard to psychometrics in assessing outcomes. Our results, in turn, suggest differences between the psychometric properties of the instruments in foot and ankle outcomes, which should be considered when selecting the most appropriate instrument for patients with foot and ankle complaints.

In our previous study [42], paper-administered versions of three foot and ankle specific PRO instruments (LEFS, VAS-FA, and WOMAC) were compared in patients who had undergone surgery due to foot or ankle trauma on average four years before completing the PRO instruments. Although a ceiling effect was confirmed only in the LEFS, the scores of each PRO instrument were skewed towards the high end. In addition, the scores were highly correlated with

each other and with the generic HRQoL instrument 15D scores. While the authors concluded that the VAS-FA might be superior to the other measures, there were only minor differences in psychometrics between the PRO instruments. When these results are reflected against the results of our study in which the patients completed the ePROs preoperatively, it seems that after a long follow-up the differences between the PRO instruments diminish. However, when assessing the foot and ankle related issues in a more acute phase, there may be considerable differences between outcome scores of the different PRO instruments.

The findings of this study support the hypothesis that electronic versions of PRO instruments have comparable psychometric properties to paper-administered versions with a number of advantages provided by electronic administration [9-14]. Hence, based on these findings, the use of ePRO instruments should be encouraged since the participation rate has been found to increase markedly when electronic data capture methods are applied instead of paper questionnaires [43].

Strengths of our study include the adequate conversion from the paper PROs to ePROs and testing of their psychometric properties according to international guidelines [4, 39, 44]. To the best of the authors' knowledge, this study is the first to evaluate psychometric properties of electronic versions of widely used foot and ankle specific PRO instruments. As the technology used in research processes advances and use of electronic tools increases, it is important to validate the ePROs properly. A limitation of this study is that it did not assess the responsiveness of the ePROs or the psychometric properties according to the item response theory (IRT). This gap could be addressed in future studies. Furthermore, the patients completed all five comprehensive ePROs concerning the same theme at once. Thus, it is possible that the patients became tired and unfocused in completing the ePRO items, affecting the results.

In conclusion, the ePRO versions of the FAAM, the FAOS, the MOXFQ, and the VAS-FA were shown to provide valid scores in foot and ankle surgery patients. These instruments can be recommended for use in outcome assessment of patients with foot and ankle specific symptoms according to the specific scope of each ePRO. Considerable differences exist between the scopes of these ePROs, which must be taken into account when selecting the most appropriate ePRO for use in each foot and ankle patient population. The ePRO version of the modified LEFS did not fulfill all of the predefined and required psychometric criteria. Based on these results, performance of the modified LEFS is not optimal in this patient population and it should be used with caution.

Funding: This study was funded by the Finnish Cancer Society, the Musculoskeletal and Plastic Surgery Research Center Helsinki, Helsinki University Hospital and University of Helsinki, and competitive Research Funding of Helsinki University Hospital. The funders had no involvement in any phases of the study.

Conflicts of interest

None of the authors declare support from any organization for the submitted work or other conflicts of interest.

Authorship confirmation

The data, models, or methodology used in the research are collected, analyzed and reported completely by our study group. All authors have been a part of planning, analyzing or writing this article. All authors read and approved the final manuscript. Content of the manuscript, figures or tables have not been published or submitted for publication elsewhere.

Acknowledgments: The authors thank Heli Sarpila and Outi Malkavaara for participation in the data collection. The Head of the Orthopedic Department Dr. Timo Juutilainen and Professor Juhana Leppilahti are thanked for their support in conducting the study. Clinical Outcomes at Oxford University Innovation is acknowledged for granting a license to conduct a translation and adaptation of the Finnish version of the MOXFQ instrument and its psychometric testing. Olli Väistö, Janne Jousmäki, Katri Ahonen, Tomi Tillgren, and Miia Aulanko are thanked for ensuring that this study was performed correctly.

References

1. Guyatt, G., et al., *Evidence-based medicine: A new approach to teaching the practice of medicine.* JAMA, 1992. **268**(17): p. 2420-2425.

2. Hunt, K.J.M.D. and D.B.A. Hurwit, *Use of Patient-Reported Outcome Measures in Foot and Ankle Research.* Journal of Bone & Joint Surgery - American Volume, 2013. **95**(16): p. e118.

3. Button, G. and S. Pinney, *A meta-analysis of outcome rating scales in foot and ankle surgery: is there a valid, reliable, and responsive system?* Foot & Ankle International, 2004. **25**(8): p. 521-5.

4. Mokkink, L.B., et al., *The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study.* Quality of Life Research, 2010. **19**(4): p. 539-549.

5. Mokkink, L.B., et al., *COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures.* Quality of Life Research, 2018. **27**(5): p. 1171-1179.

6. Schulenberg, S.E. and B.A. Yutrzenka, *The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect.* Behavior Research Methods, Instruments, & Computers, 1999. **31**(2): p. 315-321.

7. Coons, S.J., et al., *Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report.* Value in Health, 2009. **12**(4): p. 419-29.

8. Gwaltney, C.J., A.L. Shields, and S. Shiffman, *Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review.* Value in Health, 2008. **11**(2): p. 322-33.

9. Stone, A.A., et al., *Patient non-compliance with paper diaries.* BMJ, 2002. **324**(7347): p. 1193-1194.

10. Tourangeau, R. and T.W. Smith, *Asking sensitive questions: The impact of data collection mode, question format, and question context.* Public Opinion Quarterly, 1996. **60**(2): p. 275-304.

11. Taenzer, P., et al., *Computerized quality-of-life screening in an oncology clinic.* Cancer Practice, 1997. **5**(3): p. 168-175.

12. Bloom, D.E., *Technology, experimentation, and the quality of survey data*. 1998, American Association for the Advancement of Science.

13. Velikova, G., et al., *Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires.* Journal of Clinical Oncology, 1999. **17**(3): p. 998-998.

14. Bushnell, D.M., et al., *Validation of electronic data capture of the irritable bowel syndrome—quality of life measure, the work productivity and activity impairment questionnaire for irritable bowel syndrome and the EuroQol.* Value in Health, 2006. **9**(2): p. 98-105.

15. Muehlhausen, W., et al., *Standards for Instrument Migration When Implementing Paper Patient-Reported Outcome Instruments Electronically: Recommendations from a Qualitative Synthesis of Cognitive Interview and Usability Studies.* Value in Health, 2018. **21**(1): p. 41-48.

16. Rutherford, C., et al., *Mode of administration does not cause bias in patient-reported outcome results: a meta-analysis.* Quality of Life Research, 2016. **25**(3): p. 559-74.

17. Zbrozek, A., et al., *Validation of Electronic Systems to Collect Patient-Reported Outcome (PRO) Data—Recommendations for Clinical Trial Teams: Report of the ISPOR ePRO Systems Validation Good Research Practices Task Force.* Value in Health, 2013. **16**(4): p. 480-489.

18. World Medical, A., *World medical association declaration of helsinki: Ethical principles for medical research involving human subjects.* JAMA, 2013. **310**(20): p. 2191-2194.

19. Mokkink, L.B., et al., *COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs).* User manual, 2018. **78**: p. 1.

20. Wild, D., et al., *Principles of good practice for the translation and cultural adaptation process for patient‐reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation.* Value in Health, 2005. **8**(2): p. 94-104.

21. *Clinical Outcomes at Oxford University Innovation – Guidelines for the production of eCOA Measures*. 2016 1.2.2019]; Available from: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjhydP8-ZngAhXM0KQKHXg5B_oQFjAAegQIAxAC&url=https%3A%2F%2Finnovation.ox.ac.uk%2Fwp-content%2Fuploads%2F2016%2F05%2FClinical-Outcomes-at-Oxford-University-Innovation-Guidelines-for-the-production-of-eCOA-Measures_June-2016.docx&usg=AOvVaw0GxVf9520_vfnyxqSDCbgB.

22. Martin, R.L., et al., *Evidence of Validity for the Foot and Ankle Ability Measure (FAAM).* Foot & Ankle International, 2005. **26**(11): p. 968-983.

23. Roos, E.M., S. Brandsson, and J. Karlsson, *Validation of the Foot and Ankle Outcome Score for Ankle Ligament Reconstruction.* Foot & Ankle International, 2001. **22**(10): p. 788-794.

24. Binkley, J.M., et al., *The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application.* Physical Therapy, 1999. **79**(4): p. 371-383.

25. Repo, J.P., et al., *Rasch analysis of the Lower Extremity Functional Scale for foot and ankle patients.* Disability and Rehabilitation, 2018: p. 1-7.

26. Dawson, J., et al., *A patient-based questionnaire to assess outcomes of foot surgery: validation in the context of surgery for hallux valgus.* Quality of Life Research, 2006. **15**(7): p. 1211-1222.

27. Dawson, J., et al., *The MOXFQ patient-reported questionnaire: assessment of data quality, reliability and validity in relation to foot and ankle surgery.* Foot, 2011. **21**(2): p. 92-102.

28. Morley, D., et al., *The Manchester–Oxford Foot Questionnaire (MOXFQ): Development and validation of a summary index score.* Bone and Joint Research, 2013. **2**(4): p. 66-69.

29. Dawson, J., et al., *The MOXFQ patient-reported questionnaire: assessment of data quality, reliability and validity in relation to foot and ankle surgery.* Foot (Edinb), 2011. **21**(2): p. 92-102.

30. Dawson, J., et al., *Responsiveness of the Manchester-Oxford Foot Questionnaire (MOXFQ) compared with AOFAS, SF-36 and EQ-5D assessments following foot or ankle surgery.* J Bone Joint Surg Br, 2012. **94**(2): p. 215-21.

31. Gur, G., et al., *Validity and Reliability of Visual Analog Scale Foot and Ankle: The Turkish Version.* The Journal of Foot and Ankle Surgery, 2017. **56**(6): p. 1213-1217.

32. Angthong, C., et al., *Visual analogue scale foot and ankle: validity and reliability of Thai version of the new outcome score in subjective form.* Journal of the Medical Association of Thailand, 2011. **94**(8): p. 952-7.

33. Repo, J.P., et al., *Reliability and validity of the Finnish version of the Visual Analogue Scale Foot and Ankle (VAS-FA).* Foot and Ankle Surgery, 2017.

34. Richter, M., et al., *A new foot and ankle outcome score: Questionnaire based, subjective, Visual-Analogue-Scale, validated and computerized.* Foot and Ankle Surgery, 2006. **12**(4): p. 191-199.

35. Brooks, R. and E. Group, *EuroQol: the current state of play.* Health Policy, 1996. **37**(1): p. 53-72.

36. Ohinmaa, A. and H. Sintonen. *Inconsistencies and modelling of the Finnish EuroQol (EQ-5D) preference values*. in *EuroQol Plenary Meeting*. 1998. Health Economics and Health System Research, University of Hannover.

37. Terwee, C.B., et al., *Quality criteria were proposed for measurement properties of health status questionnaires.* Journal of Clinical Epidemiology, 2007. **60**(1): p. 34-42.

38. Peterson, R.A., *A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis.* Marketing Letters, 2000. **11**(3): p. 261-275.

39. Prinsen, C.A.C., et al., *COSMIN guideline for systematic reviews of patient-reported outcome measures.* Quality of Life Research, 2018. **27**(5): p. 1147-1157.

40. Pinsker, E., et al., *Reliability and validity of 6 measures of pain, function, and disability for ankle arthroplasty and arthrodesis.* Foot & Ankle International, 2015. **36**(6): p. 617-625.

41. Goldstein, C.L., et al., *Comparison of different outcome instruments following foot and ankle trauma.* Foot & Ankle International, 2010. **31**(12): p. 1075-1080.

42. Ponkilainen, V.T., et al., *Assessment of the structural validity of three foot and ankle specific patient-reported outcome measures.* Foot and Ankle Surgery, 2019.

43. Schamber, E.M., et al., *Barriers to completion of patient reported outcome measures.* The Journal of Arthroplasty, 2013. **28**(9): p. 1449-1453.

44. Mokkink, L.B., et al., *The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes.* Journal of Clinical Epidemiology, 2010. **63**(7): p. 737-745.

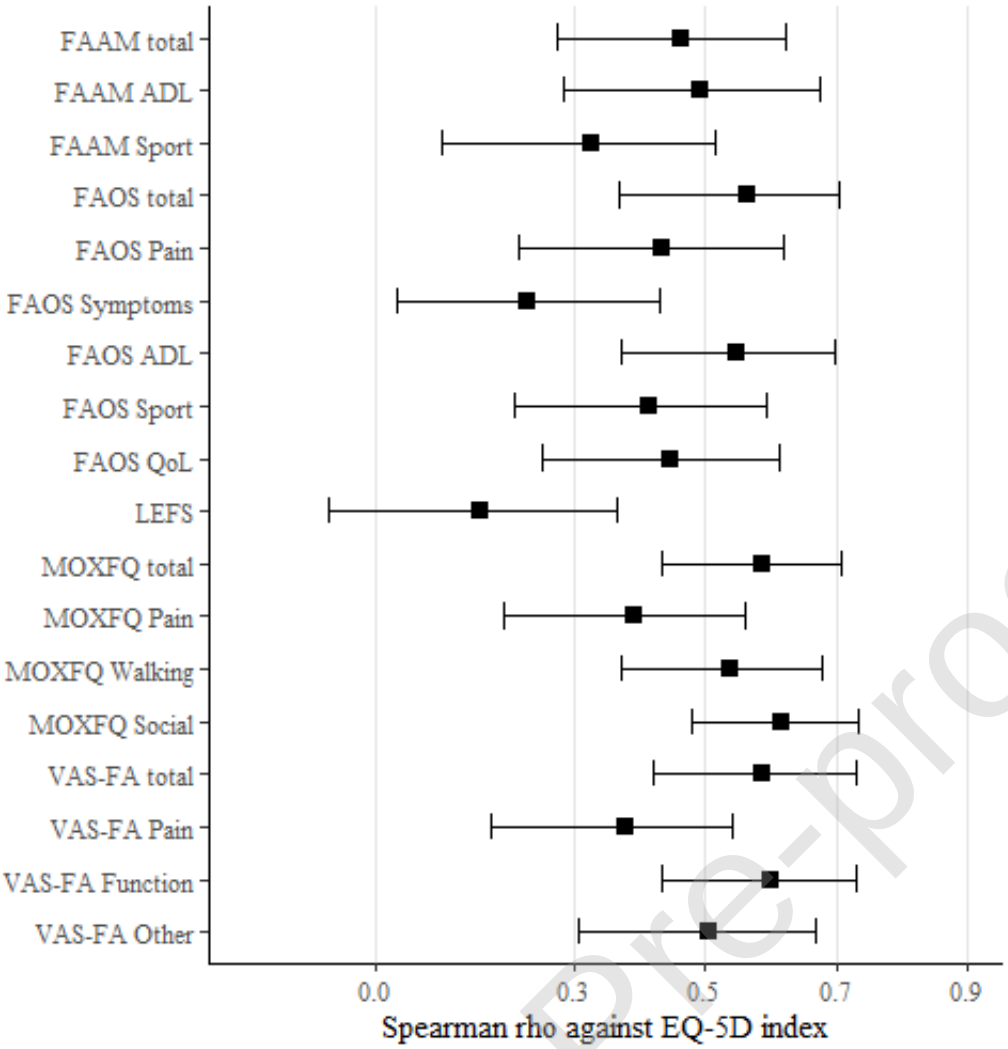

Figure 6. Spearman correlation coefficients and 95% CIs of the ePRO and the subscale scores against the EQ-5D index.

Table 1. Psychometric features, hypotheses, and conclusions for ePROs.

Feature	Hypothesis	FAAM	FAOS	LEFS	MOXFQ	VAS-FA
Scale coverage and targeting						
No floor effect	Min score <15% in the total or the subscale scores	Confirmed	Confirmed	Confirmed	Confirmed	Confirmed
No ceiling effect	Max score <15% in the total or the subscale scores	Confirmed	Confirmed	Confirmed	Confirmed	Confirmed
Construct validity						
Structural validity	Unidimensional subscales in the EFA	Confirmed	Confirmed	Confirmed	Confirmed	Confirmed
	Acceptable internal consistency with Cronbach's alpha > 0.70 for the whole instrument and for the subscales	Confirmed	Confirmed	Confirmed	Confirmed / Rejected	Confirmed / Rejected
Independency of the sociodemographic and clinical factors	Non-significant or negligible associations with age, sex, BMI,	Confirmed	Confirmed / Rejected	Confirmed	Confirmed	Confirmed / Rejected

	indications for operative treatment, and previous operations					
Convergent validity	Significant and at least low correlation with other ePRO scores	Confirmed	Confirmed	Rejected	Confirmed	Confirmed
	Significant and at least low correlation with EQ-5D index	Confirmed	Confirmed	Rejected	Confirmed	Confirmed

Table 2. Clinical and sociodemographic characteristics of patients.

	N=111
Age, mean (SD)	48 (14)
Female, n (%)	78 (70)
BMI, mean (SD)	27 (5)
Duration of foot or ankle complaints, years, median (IQR)	3.0 (1.0 – 9.0)
Affected foot or ankle previously operated, n (%)	50 (45)
Indication for operative treatment, n (%)	
Deformity of foot or ankle	45 (41)
Osteoarthritis of foot or ankle	32 (29)
Flat foot or cavoid foot	8 (7.2)
Plantar fasciitis or tendinopathy of foot	8 (7.2)
Soft tissue protrusions or tumors	7 (6.3)
Ankle instability	1 (0.9)
Symptomatic accessory bone (os tibiale)	1 (0.9)
Stress fracture of foot	1 (0.9)
EQ-5D index, mean (SD)	0.65 (0.15)
General health state, mean (SD)	2.8 (0.8)
Education, n (%)	
Basic education	30 (27)
Upper secondary level education	19 (17)
Higher education	62 (56)

Table 3. Scores and Cronbach's alphas of the ePROs and the subscales.

	Score		Min (%)	Max (%)	Cronbach's alpha (95% CI)
	Mean (SD)	SEM			
FAAM	68 (21)	2.1	0	1.0	0.97 (0.96 to 0.98)
ADL	77 (20)	2.0	0	5.3	0.97 (0.96 to 0.98)
Sport	48 (27)	2.7	4.2	1.0	0.94 (0.93 to 0.96)
FAOS	55 (13)	1.3	0	0	0.97 (0.96 to 0.98)
Pain	57 (15)	1.5	1.0	0	0.89 (0.86 to 0.92)
Symptom	47 (11)	1.1	0	0	0.75 (0.68 to 0.82)
ADL	67 (15)	1.5	0	0	0.97 (0.96 to 0.98)
Sport	47 (21)	2.1	4.0	0	0.90 (0.87 to 0.93)
QoL	27 (18)	1.8	8.0	0	0.85 (0.80 to 0.89)
LEFS	82 (15)	1.5	0	6.1	0.92 (0.90 to 0.94)
MOXFQ	57 (15)	1.6	0	1.1	0.93 (0.91 to 0.95)
Pain	44 (20)	2.0	1.1	2.1	0.84 (0.79 to 0.89)
Walking	46 (23)	2.3	2.1	2.1	0.90 (0.87 to 0.93)
Social	49 (20)	2.1	1.1	1.1	0.69 (0.59 to 0.78)
VAS-FA	61 (18)	1.8	1.0	0	0.91 (0.89 to 0.94)
Pain	50 (21)	2.1	1.0	1.9	0.75 (0.67 to 0.83)
Function	66 (19)	1.9	0	1.0	0.90 (0.87 to 0.93)
Other complaints	62 (21)	2.1	1.0	1.9	0.62 (0.51 to 0.74)

SD = Standard deviation

SEM = Standard error of the mean

ADL = Activities of daily living

QoL = Quality of life

Table 4. Associations between basic characteristics and ePRO scores.

	Age		BMI		Sex (mean score)			Indication for operative treatment (mean score)			Previous operation of affected foot or ankle (mean score)		
	r	Sig.	ig.		Iale	Female	Sig.	Osteoarthritis	Deformity	Sig.	Yes	No	Sig.
FAAM	-0.02	-	-0.15	-	67	68	-	67	67	-	66	69	-
Activities of daily liv	-0.15	-	-0.12	-	80	75	-	78	74	-	74	79	-
Sport	0.45	-	-0.21	*	43	50	-	45	51	-	49	47	-
FAOS	-0.23	*	-0.21	*	56	55	-	55	55	-	54	57	-
Pain	-0.15	*	-0.11	-	57	56	-	55	57	-	55	58	-
Symptom	-0.18	-	-0.19	-	46	47	-	43	47	-	45	49	-
Activities of daily liv	-0.34	***	-0.17	-	69	66	-	67	65	-	65	68	-
Sport	-0.11	-	-0.18	-	48	47	-	47	46	-	43	51	-
Quality of life	0.09	-	-0.24	*	23	29	-	29	26	-	28	27	-
LEFS	-0.20	*	-0.04	-	84	80	-	80	83	-	81	82	-
MOXFQ	-0.12	-	-0.13	-	56	58	-	56	57	-	55	59	-
Pain	-0.14	-	-0.09	-	43	45	-	43	44	-	43	45	-
Walking	-0.10	-	-0.16	-	46	47	-	46	47	-	43	50	-

Social	-0.07	-0.07	46	50	-	48	48	-	47	50	-
VAS-FA	-0.21*	-0.13	65	60	-	62	60	-	58	64	-
Pain	0.01	-0.11	49	51	-	50	50	-	50	50	-
Function	-0.27**	-0.23*	70	65	-	67	67	-	62	69	-
Other complaints	-0.19	-0.02	70	59	*	65	58	-	59	65	-

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5. Spearman correlation coefficients between the ePROs and the subscales. All correlations, other than those marked "ns", were significant ($p < 0.05$).

FAAM																			
Activities of daily living																			
Sport		64																	
FAOS	62	79	70																
Pain	72	74	61																
Symptom	63	64	71		65														
Activities of daily living	68	73	61		75	67													
Sport	65	75	63		72	63	72												
Quality of life	73	69	75		67		69	71											
LEFS		63						62											
MOXFQ		64	65	63	74	69	61	72	67	67									
Pain		68	69	66	61	65	60	69	60	68									
Walking		61	63	63	67	69	66	67	63	62			69						
Social		65	66	64	66	60	59	65	69	67			69	76					
VAS-FA		60	68	60	66	68	64	61	65	65	67	79	66	70	72				
Pain		64	67	66	69	68		69	63	62		67	69	62	61				
Function		64	71	65	68	68	64	67	67	62	64	76	67	73	68		67		
Other complaints		67	61		63	65	65	60	64	69	61	67	60	68	67		60	67	

ns = not significant (0.05)