Comparison of the Manchester–Oxford Foot Questionnaire (MOXFQ) and the Self-Reported Foot and Ankle Outcome Score (SEFAS) in patients with foot or ankle surgery

Dariusch Arbab,*, Katharina Kuhlmann, Christoph Schnurr, Christian Lüring, Dietmar König, Bertil Bouillon

*Department of Orthopedic Surgery, Klinikum Dortmund, Member Faculty of Health Witten/Herdecke University, Beurhausstraße 40, 44137 Dortmund, Germany
†Medical University Düsseldorf, Moorstraße 5, 40225 Düsseldorf, Germany
‡Department of Orthopedic Surgery, St. Vinzenz Krankenhaus, Schlosstraße 85, 40447 Düsseldorf, Germany
§Department of Orthopedic Surgery, Klinikum Dortmund, Beurhausstraße 40, 44137 Dortmund, Germany
¶IRV Clinics of Orthopedic Surgery, Horstenstraße 2, 41479 Viersen, Germany
#Department of Traumatology and Orthopedic Surgery, Cologne-Merheim Medical Center (CMMC), University of Witten/Herdecke, Ostheimer Str. 200, 51109 Cologne, Germany

A R T I C L E   I N F O

Article history:
Received 12 October 2017
Revised in revised form 22 January 2018
Accepted 30 January 2018

Keywords:
MOXFQ
SEFAS
Patient reported outcome measures
Comparison
Validation

A B S T R A C T

Background: Patient-reported outcome measures are a critical tool in evaluating the efficacy of orthopedic procedures. The intention of this study was to compare the psychometric properties of the Manchester–Oxford Foot Questionnaire (MOXFQ) and the Self-Reported Foot and Ankle Outcome Score (SEFAS) in patients with foot or ankle surgery.

Methods: 177 patients completed both scores and FAOS, Foot and Ankle Outcome Score (FAOS), Short Form 36 (SF-36) and numeric scales for pain and disability (NRS) before and after surgery. Test-Retest reliability, internal consistency, floor and ceiling effects, construct validity, responsiveness and minimal important change were analyzed.

Results: The MOXFQ and SEFAS demonstrated excellent test-retest reliability with interclass correlation coefficient values >0.9 Cronbach’s alpha (α) values demonstrated strong internal consistency. No floor or ceiling effects were observed for both questionnaires. As hypothesized MOXFQ subscales correlated strongly with corresponding SEFAS, FAOS and SF-36 domains. MOXFQ subscales showed excellent responsiveness between preoperative assessment and postoperative follow-up, whereas SEFAS demonstrated moderate responsiveness.

Conclusions: The MOXFQ and SEFAS demonstrated good psychometric properties and proved to be valid and reliable instruments for use in foot and ankle patients. MOXFQ showed better outcomes in responsiveness.

© 2018 European Foot and Ankle Society. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Patient reported outcome measurements (PROMs) can provide reliable and valid measures of patient’s degree of pain, impairment, disability, and quality of life. They are a critical tool in evaluating the efficacy of orthopedic procedures and are increasingly used in clinical trials to assess outcomes of health care [1]. The Manchester–Oxford Foot Questionnaire (MOXFQ) is a PROM for surgery of the foot and ankle, initially assessed in patients undergoing hallux valgus surgery [2,3]. Further evaluation provided evidence in support of the reliability and validity using data from a large sample of patients undergoing surgery of a wide range of foot or ankle conditions [4,5]. The MOXFQ contains three sub-scores for pain, walking/standing and social interaction dimensions as well as a summary index score (MOXFQ-Index). The Self-reported Foot and Ankle Score (SEFAS) is a PROM for surgery of the foot and ankle, initially assessed in patients...
undergoing total ankle replacement due to osteoarthritis or inflammatory arthritis [6]. It proved to be a valid score for different foot, hindfoot and ankle disorders [7,8]. The SEFAS is based on the New Zealand total ankle questionnaire [9], and was translated and culturally adapted by Coster et al. in 2012 [6]. It contains 12 items, with 5 response options. The questionnaire covers different constructs, which are not reported separately in subscales. Pain, limitation of function and other symptoms are the main constructs.

The aim of this study was to compare the psychometric properties of the Manchester–Oxford Foot Questionnaire (MOXFQ) and the Self-Reported Foot and Ankle Outcome Score (SEFAS) in patients with foot or ankle surgery.

2. Methods

The study was approved by the local research ethics committee (ref 15-252) and performed in accordance with the Declaration of Helsinki. Written informed consent from all participants was obtained.

2.1. Patients and validation procedure

From November 2014 to January 2016 a total of 177 patients undergoing surgery of the foot or ankle were consecutively recruited at a single institution. Eligibility criteria included adult patients undergoing primary foot or/and ankle surgery for osteoarthrits, deformity, rheumatoid arthritis, impingement of the ankle, tendon disorders or bone defect. Patients were asked to complete the MOXFQ, SEFAS, the Foot and Ankle Outcome Score (FAOS) the Short-Form 36 Health Survey (SF-36) and a numeric scale for pain and disability (NRS). MOXFQ, SEFAS, FAOS, SF-36 and NRS were completed 3–14 days before surgery (t1) and again on the morning before surgery (t2) for reliability testing. 6 months after surgery (t3) all participants were asked to complete MOXFQ and SEFAS a last time.

2.2. Instruments

The MOXFQ is a 16-item questionnaire consisting of 3 subscales: walking/standing (w/s) problems (seven items), pain (p) (five items) and issues related to social interaction (si) (four items). Patients score each question on a five-point Likert scale scored from 0 to 4, with 4 representing the worst stage [2]. Raw scores are converted to a metric scale from 0 to 100, where 100 denotes most severe.

The SEFAS is a 12-item questionnaire covering different constructs, which are not separately reported in subscales. We defined 3 subscales consisting of: pain (p) (seven items), limitation of function (lof) (three items) and others (o) (2 items). Patients score each question on a five-point Likert scale scored from 0 to 4, with 0 representing the worst stage and the sum of 48 representing normal function [6]. To facilitate the comparison of both scales, also SEFAS values were converted into a metric scale from 0 to 100, 100 denoting most severe.

The FAOS is a 42-item instrument to evaluate symptoms and functional limitations related to the foot and ankle [10]. It consists of five domains: pain (p), other symptoms (s), activities of daily living (adl), sport and recreational activities (s/r), and foot-and-ankle-related quality of life (qol). Each item is scored on a five-point Likert scale from 0 to 4, with 4 representing the worst stage. Compared to MOXFQ low numbers represent more severe stages.

The SF-36 instrument is a widely used generic patient-reported instrument to measure health related quality of life. It consists of eight domains: physical functioning (pf), role physical (rp), role emotional (re), social functioning (sf), mental health (mh), energy/vitality (e/v), pain (p), general health perception (gh).

The NRS were used to determine pain and disability of the foot and ankle. On a 0–10 scale 10 represents the most severe pain or disability.

2.3. Statistical analysis

The MOXFQ, SEFAS and FAOS subscale scores were entered into a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond WA) and analyzed using SPSS v24 (SPSS Inc. Chicago, Illinois). A p-value <0.05 was considered to indicate statistical significance.

2.4. Reliability

2.4.1. Reproducibility

Reproducibility as test–retest reliability was assessed by calculating intraclass correlation coefficient (ICC, Two-way Random Effect Model Absolute Agreement Definition) between MOXFQ and SEFAS completed at the first visit 3–14 days before surgery (t1) and second time before surgery (t2). An ICC value of 0.7 and above was considered as good [11].

2.4.2. Internal consistency

Reliability also includes internal consistency [11]. Internal consistency is the extent to which items within a scale are homogeneous, thus measuring the same construct [16]. Cronbach’s alpha (α) coefficient calculated to assess internal consistency of the MOXFQ and SEFAS items. Values of alpha of 0.7, 0.8 and 0.9 are considered to represent fair, good and excellent degree of internal consistency, respectively [12].

2.5. Floor and ceiling effects

Floor and ceiling effects were considered to exist if more than 15% of responses reached lowest or highest possible score [13].

2.6. Validity

2.6.1. Construct validity

Describes the extent to which a score relates to other scores [18]. MOXFQ subscales were compared to SEFAS, FOAS and SF-36 and NRS pain and disability using non-parametric correlation coefficients (Spearman’s Rho). Correlation coefficients <0.4 were considered as low, 0.4–0.59 as moderate and 0.6–0.79 as high correlation [5]. For convergent validity high correlation between MOXFQ dimensions w/sand the SEFAS lof, the FAOS domains adl, s/r, and with the SF-36 domains pf were hypothesized. For the MOXFQ dimension p high correlation was expected with SEFAS p, FAOS domain p and SF-36 domain bp and NRS. MOXFQ si domain and the FAOS dimension qol and SF-36 domains sf should show high correlations. Low correlations were expected between subscales on different contents [5,14].

2.7. Responsiveness

Responsiveness is the extent to which a questionnaire is able to detect changes over time or due to an intervention such as surgery [15]. All patients completed MOXFQ and SEFAS before surgery (t1) and 5–6 months after surgery (t3). To test responsiveness effect size (ES) and standardized response means (SRM) were calculated. ES calculated as the difference between the means before and after intervention divided by the standard deviation (SD) of the same measure before treatment [11]. SRM is calculated as the difference between the means before and after treatment divided by the SD of the change. For both, ES and SRM, values of 0.2, 0.5 and 0.8 were regarded as small, moderate and large effects, respectively [11].
2.7.1. Minimal important change (MIC)

Minimal important change is the smallest change in a treatment outcome that a patient or physician would identify as important. MIC describes a threshold above which outcome is experienced as relevant by the patient and avoids the problem of bare statistical significance [15]. One distribution-based approach to calculate MIC is the minimal detectable change (MDC). It is defined as minimum amount of change that can be considered above the threshold of a measurement error. If the change in a score is higher than MDC, it can be considered as a true change [16]. It is calculated from the standard error of measurement (SEM), which is related to the internal consistency/reliability of the score (Cronbach’s alpha). SEM = standard deviation * √(1- Cronbach’s alpha). To allow comparisons with other studies, the MDC was calculated based on the confidence level of 90% (MDC90: MDC = 1.65 * SEM * √(2)). In order to estimate significant change of the scores over time, we performed likelihood ratio tests after applying mixed-effects linear regression with random patient effects to account for repeated (longitudinal) measurements on the same patient [14,17].

3. Results

177 patients, 130 women and 47 men, with a mean age of 57 years (18–92) undergoing surgery of the foot or ankle were consecutively recruited at a single institution and completed the baseline 3–14 days before surgery (t1). On the morning before surgery (t2) 145 patients completed MOXFQ and SEFAS to determine test-retest reliability. 6 months after surgery (t3) 117 patients completed MOXFQ, SEFAS, FAOS, SF-36 and NRS to test responsiveness. 118 patients were undergoing forefoot, 56 patients hindfoot or ankle surgery.

3.1. Reliability

3.1.1. Reproducibility

All three dimension of the MOXFQ demonstrated excellent test-retest reliability with ICC values of 0.97 for walking/standing, 0.94 for pain, 0.92 social interaction and 0.97 for the MOXFQ index. The mean indexes for the baseline and the reliability assessments were 67.7 (standard deviation (SD) 17.7) and 67.4 (SD 18.5), respectively (Table 1). The SF-36 demonstrated excellent test-retest reliability with ICC values of 0.95 for limitation of function, 0.97 for pain, 0.96 other symptoms and 0.97 for the SF-36 total. The mean indexes for the baseline and the reliability assessments were 48.4 (Standard deviation (SD) 19.7) and 49.7 (SD 20.7), respectively (Table 1).

3.1.2. Internal consistency

Cronbach’s alpha (α) value of 0.93 for MOXFQ index, 0.92 for walking/standing and 0.83 for pain demonstrated strong internal consistency. Social interaction values of 0.70 still showed fair internal consistency (Table 1). Cronbach’s alpha (α) value of 0.89 for SEFAS total and 0.87 for pain demonstrated strong internal consistency. Lof values of 0.79 still showed fair internal consistency, whereas os demonstrated low consistency (0.35) (Table 1).

3.2. Floor and ceiling effects

No floor or ceiling effects were observed for any of the three MOXFQ subdomains or SEFAS (Table 1).

3.3. Validity

3.3.1. Construct validity

To examine construct validity the Spearman’s correlation coefficients between MOXFQ, SF-36, FAOS and NRS were examined and shown in Table 2. Convergent validity of the MOXFQ walking/standing subscale was shown with strong correlations (>0.6) with SEFAS lof, FAOS adl, sj and SF-36 domains pf. As hypothesized MOXFQ pain subscale correlated strongly with SEFAS p, FAOS domain p, SF-36 domain bp and NRS. MOXFQ domain si showed strong correlation with FAOS qol and moderate (0.4–0.59) with SF-36 sf. All these findings were statistically significant (p < 0.05).

3.4. Responsiveness

Table 2 shows the responsiveness of the MOXFQ and SEFAS. All MOXFQ subscales demonstrated excellent (ES/SRM > 0.8) responsiveness between preoperative assessment (t2) and postoperative follow-up (t3) indicating that a very large degree of change was detected following surgery. The highest effect size for MOXFQ showed the pain subscale (1.36) representing the best responsiveness, whereas the walking/standing domain showed lowest (1.06), still representing large effects.

With an effect size of 0.73 the SEFAS demonstrated moderate (ES/SRM > 0.5) responsiveness between preoperative assessment (t2) and postoperative follow-up (t3) indicating that a good degree of change was detected following surgery. The highest effect size showed the pain subscale (1.18).

The SEM was 2.75, 2.33, 2.26 and 2.22 for the MOXFQ walking/standing, pain, social interaction and index, respectively. MDC90 (90% confidence level) were 6.42, 5.44, 5.27 and 5.19 for domains standing, pain, social interaction and Index, respectively. The mean

| Table 1 |
|-----------------|-----------------|
| Reliability (n) | 177             |
| Test-retest ICC | 0.97            |
| Test (mean SD)  | 67.7 (17.7)     |
| Retest (mean SD)| 67.4 (18.5)     |
| Internal consistency (n) | 177           |
| Cronbach’s alpha | 0.93           |
| Floor and ceiling effects (n) | 177             |
| Proportions     | 0               |
| Responsiveness (n) | 177             |
| Preoperative (mean SD) |               |
| Postoperative (mean SD) |               |
| Effect size (ES, mean 95% CI) | 1.48, 0.73 |
| Standardized response means (SRM, mean, 95% CI) | 1.17, 0.65 |
| Standard error of measurement (SRM) | 2.22, 2.16 |
| Minimal detectable change (MDC 90, confidence interval 90) | 5.19, 5.04 |
The difference between preoperative and postoperative assessment is shown on Table 2 ranging between 22.23 for the social interaction subscale and 32.75 for pain. All subscales and the index showed higher changes than the MDC which indicates true changes [16]. The SEM for SF-36 was 2.39, 2.13, 3.01 and 2.16 lof, p, os and total, respectively. MDC90 (90% confidence level) were 5.58, 4.97, 7.03 and 5.04 for domains lof, p, os and total, respectively. The mean difference (15.04) between preoperative and postoperative assessment is shown on Table 2. The German SF-36 showed higher changes than the MDC (5.04) which indicates true changes [18].

4. Discussion

Patient related outcome measures have become an important tool in clinical practice and clinical trials to assess outcome of health care [1,19]. Several scores are in use for evaluating the outcome of foot and ankle surgery but none has been accepted as gold standard [7]. The MOXFQ is an increasingly used PROM in foot and ankle surgery which has been extensively tested and translated into German, Italian, Dutch, Turkish, Persian and Spanish [2–5,14,20–26]. The SF-36 has been initially validated for patients with ankle osteoarthritis and later for patients with a wide variety of different disorders of the foot and ankle proving good validity, reliability and responsiveness [6,8]. In a systematic literature review about PROM’s in hallux valgus the SF-36 showed good psychometric properties with good availability and less items compared to the Manchester–Oxford Foot questionnaire [27]. It is available in English, Swedish and German [28]. In this study MOXFQ and SF-36 demonstrate good psychometric properties showing good validity, reliability and responsiveness in patients undergoing foot and ankle surgery.

Test-retest reliability of the MOXFQ and SF-36 showed excellent results with ICC values ranging from 0.92 to 0.97. Strong internal consistency has been demonstrated for the MOXFQ Index (Cronbach’s alpha 0.93), walking/standing (0.92) and pain (0.83) domain. The social interaction (0.70) subscale showed fair internal consistency.

The SF-36 showed strong internal consistency for total index (Cronbach’s alpha 0.89) and SF-36 pain (0.87). The limitation of function (0.79) subscale showed fair internal consistency, whereas the other symptoms subscale demonstrated low consistency (0.35). Low internal consistency for os subscale was expected due to inhomogeneous questions asking for different constructs.

Construct validity was determined by comparing the MOXFQ with the SF-36 and FAOS. Comparison between MOXFQ and SF-36 has not published before.

The MOXFQ walking/standing subscale showed strong correlations (>0.6) with SF-36 lof, FAOS adl, s/r and SF-36 domains pf. As hypothesized MOXFQ pain subscale correlated strongly with SF-36 p, FAOS domain p, SF-36 domain bp and NRS. MOXFQ domain si showed strong correlation with FAOS qol and moderate (0.4–0.59) with SF-36 sf. Divergent validity was shown by low correlation between MOXFQ domains and SF-36 general and mental health subscales, respectively.

Table 2 illustrates the responsiveness of the MOXFQ and SF-36. For MOXFQ all subscales showed excellent (ES/SRM > 0.8) responsiveness between preoperative (t2) and postoperative follow-up (t3). Whereas SF-36 demonstrated moderate (ES/SRM > 0.5) responsiveness in 118 patients with an effect size of 0.73 between preoperative (t2) and postoperative follow-up (t3). The highest effect size for MOXFQ showed the pain subscale (1.36) representing the best responsiveness, whereas the walking/standing domain showed lowest (1.06), still representing large effects.

The mean difference between preoperative and postoperative assessment show higher changes than the MDC for all three MOXFQ dimensions and the MOXFQ index just like the SF-36 so that all changes can be interpreted as real changes [18].

The comparison of the MOXFQ and the SF-36 demonstrated that both scores show good psychometric properties. Validity and reliability were excellent and comparable in both measures.

The MOXFQ showed higher responsiveness as an extent to which a questionnaire is able to detect changes over time or due to an intervention such as surgery. It has been translated and evaluated in many different languages where the SF-36 into German and Swedish only, which might limit international distribution. Last mentioned
points might prove a slight advantage even so the MOXFQ contains of more questions which could take some extra time to complete. As both scores showed good general practicability and good results for validity, reliability and responsiveness both proved to be excellent PROMS for patients undergoing foot and ankle surgery.

5. Conclusion

Comparison of the MOXFQ and SEFAS showed that both scores demonstrate good psychometric properties. Our study demonstrated that both questionnaires are valid and reliable instruments for patients with foot and ankle disorders and can be used as a tool for evaluating the efficacy of surgical procedures and in clinical trials to assess outcomes of health care. The MOXFQ proved slight advantage showing higher responsiveness as an extent to which a questionnaire is able to detect changes over time or due to an intervention such as surgery and availability in more languages.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Source of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fas.2018.01.003.

References