Taylor & Francis
Taylor & Francis Group

Disability and Rehabilitation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/idre20

Responsiveness of specific and generic patient-reported outcome measures in patients with plantar fasciopathy

Marte Heide, Marianne Mørk, Aasne Fenne Hoksrud, Jens Ivar Brox & Cecilie Røe

To cite this article: Marte Heide, Marianne Mørk, Aasne Fenne Hoksrud, Jens Ivar Brox & Cecilie Røe (19 Oct 2023): Responsiveness of specific and generic patient-reported outcome measures in patients with plantar fasciopathy, Disability and Rehabilitation, DOI: 10.1080/09638288.2023.2267438

To link to this article: https://doi.org/10.1080/09638288.2023.2267438

DISABILITY AND REHABILITATION
https://doi.org/10.1080/09638288.2023.2267438

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Responsiveness of specific and generic patient-reported outcome measures in patients with plantar fasciopathy

Marte Heide[a,b], Marianne Mørk[a,b,c], Aasne Fenne Hoksrud[d], Jens Ivar Brox[a,b] and Cecilie Røe[a,b]

[a]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; [b]Department of Physical Medicine and Rehabilitation, Oslo University Hospital, Nydalen, Norway; [c]Research and Communication Unit for Musculoskeletal Health (FORMI), Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway; [d]Norwegian Olympic and Paralympics Committee and Confederation of Sports, Oslo, Norway

ABSTRACT

Purpose: To evaluate and compare responsiveness characteristics for the Foot Function Index revised short form (FFI-RS), RAND-12 Health Status Inventory (RAND-12), and Numeric Rating Scale (NRS), in patients with plantar fasciopathy receiving non-surgical treatment.

Materials and methods: This study was conducted on a sub-group of patients from an ongoing randomised controlled trial. One-hundred fifteen patients were included. The patient-reported outcome measures (PROMs) were applied at baseline and after 6 months. Responsiveness was calculated using standardised response mean and area under the receiver operating characteristic (ROC) curve. ROC curves were used to compute the minimal important change (MIC) for the outcome measures.

Results: The region specific FFI-RS had best responsiveness and the NRS at rest had lowest responsiveness.

Conclusion: FFI-RS were marginally more responsive than the other PROMs. Responsiveness and MIC estimates should be regarded as indicative rather than fixed estimates.

IMPLICATIONS FOR REHABILITATION

- The region-specific Foot Function Index Revised Short Form could, based on responsiveness perspectives, be recommended as an outcome measurement for patients with plantar fasciopathy.
- Responsiveness and minimal important change estimates are indicative and should be interpreted with caution.

ARTICLE HISTORY
Received 31 March 2023
Revised 28 September 2023
Accepted 3 October 2023

KEYWORDS
Plantar fasciopathy; patient-reported outcome measures; responsiveness; minimal important change; foot function index revised short form; numeric rating scale; RAND-12 health status inventory

Introduction

Patient-reported outcome measures (PROMs) are used to assess patients' subjective ratings of pain, function and overall health and are important tools in clinical trials when evaluating treatment outcomes [1]. Consensus on the preferred PROM in research on plantar fasciopathy is lacking. There is ongoing debate regarding the superiority of questionnaires focusing on the features of foot disorders or generic questionnaires capturing general function and health. Generic PROMs allow for generalisation across different groups and conditions [2]. Condition- or region-specific instruments, however, are thought to be more responsive and therefore more likely to detect clinical changes in targeted patient populations [3,4]. Regardless, the measurements must be valid, reliable, and responsive to change. Validity refers to the degree to which a PROM reflects the constructs that it is intended to measure, while reliability refers to the consistency of a measure and the absence of measurement error [5]. Responsiveness is defined as the ability of a PROM to detect change over time in the construct being measured [6]. This is important when evaluating changes over time and the effects of interventions. To interpret change scores, one benchmark to consider is the minimal important change (MIC), which reflects the minimal within-person change in a score at which patients regard themselves as importantly changed [7,8]. MIC values should be viewed with an awareness of the uncertainty and measurement error of the PROM and in relation to the population under study and follow-up time examined [9,10].

Previous reviews have presented a wide range of specific and generic PROMs employed in research on foot and ankle disorders [11–16]. Jia et al. identified 115 studies investigating 50 non-generic PROMs used in patients with foot or ankle conditions. Limited evidence was found for the psychometric properties of the PROMs, which were evaluated using COSMIN guidelines, and many studies did not assess responsiveness [15]. Sierevelt et al. found that the three most frequently used specific PROMs were the Foot Function Index (FFI), the Foot and Ankle Outcome Score and the Foot and Ankle Activity Measure, though the poor quality of most of the studies included in this systematic review led to inconclusive results regarding the measures' psychometric properties [11].

A literature search (Appendix) was conducted in PubMed from January 2017 to December 2021 to identify PROMs used in research on plantar fasciopathy and their reported responsiveness. Prospective and randomised controlled trials (RCTs) that employed PROMs at multiple time points were included. The present study cohort was not included. One hundred and thirty-six studies and 36 different PROMs were identified, with several studies assessing multiple PROMs. The measurement of pain intensity using the

CONTACT Marte Heide marte.heide@medisin.uio.no
Supplemental data for this article can be accessed online at https://doi.org/10.1080/09638288.2023.2267438.

Numeric Rating Scale (NRS), or Visual Analogue Scale was by far the most frequently applied PROM, identified 113 times. This was unsurprising, as these measures are considered valid and responsive in both research and clinical settings [17,18]. The FFI was the second most frequently applied PROM, identified 33 times, while the revised FFI (FFI-R) and its short version (FFI-RS) were identified five times and one time, respectively. The Short Form Health Survey (SF) was the most frequently used generic PROM, with the 36-item version (SF-36) identified 11 times and the 12-item version (SF-12) identified four times. None of the identified studies reported on responsiveness.

The region-specific FFI-RS was developed in 2006 and is the latest version of the FFI. The total score of the FFI-RS is considered a reliable assessment [19]. However, during its development, the FFI-RS was not evaluated for responsiveness or MIC [20]. The FFI-RS was recently translated into Norwegian, and its psychometric properties were evaluated in a population of patients, the majority of whom had plantar fasciopathy [21]. Most patients were participating in an ongoing RCT on effectiveness of conservative interventions for plantar fasciopathy [22]. The FFI-RS was found to be reliable and the total score responsive, with an area under the curve (AUC) of 0.78 (95% confidence interval (CI) 0.69 to 0.87). The smallest detectable change was 6.5 and the MIC was 8.4 at 3-month follow-up [21].

The RAND-36 Health Status Inventory (RAND-36) and its brief version (RAND-12) include the same items as the SF-36 and SF-12, respectively. The RAND-12 is widely used, freely available and considered a valid and reliable generic measure of health-related quality of life [23]. To the best of our knowledge, while the responsiveness of the RAND-12/SF-12 and RAND-36/SF-36 have not been estimated in the plantar fasciopathy population, they have been reported to be responsive in patients with back pain [24,25] and patients with knee osteoarthritis who undergo knee replacement surgery [26].

To our knowledge, no previous studies have compared the responsiveness characteristics of pain measurement, foot-specific and generic health-related quality of life PROMs in the plantar fasciopathy population.

The aim of this study was to employ both anchor- based and distribution-based methods to evaluate and compare the responsiveness characteristics of the NRS, FFI-RS and RAND-12 in patients with plantar fasciopathy receiving non-surgical treatment. We hypothesised that the region-specific FFI-RS would be more responsive than the generic RAND-12 and NRS at 6-month follow-up.

Methods

Study design and settings

In the present study, a sub-population of patients with plantar fasciopathy was used, originally enrolled in a double-blind, randomised, sham-controlled, four-armed trial. The patients were recruited from the outpatient clinic at Oslo University Hospital between March 2018 and January 2022. The aim of the trial was to compare the effectiveness of radial extracorporeal shock wave therapy (rESWT), sham-rESWT, a standardised exercise program and usual care. The trial received approval from the Regional Committee for Medical and Health Research Ethics (2017/1325), and written informed consent was obtained from each patient before enrolment [22].

The rESWT and sham-rESWT were administered once a week, with a maximum of three treatments. The standardised exercise programme was a high-load strength training programme comprising two exercises, the unilateral heel raise and unilateral leg squat, performed three times a week for 12 weeks. Usual care consisted of standardised information on pathogenesis, aetiology, and prognosis as well as advice on physical activity and footwear and customised foot orthoses. All participants, regardless of treatment group, received the standardised information and customised foot orthoses. All treatments concluded approximately 3 months after inclusion except the foot orthoses, which were used throughout the trial in all groups. The primary outcome was heel pain intensity during activity over the previous week, measured using the NRS at 6-month follow-up. The secondary outcomes were measured at 6- and 12-month follow-ups and included the FFI-RS, RAND-12, Patient Global Impression of Change (PGIC) scale, NRS at rest and NRS during activity (12 months) [22]. At the time of this study, 165 patients had completed the 6-month follow-up.

Outcome measures

Numeric Rating Scale (NRS) is an 11-point Likert scale, where 0 is "no pain" and 10 is "worst imaginable pain" [17].

RAND-12 Health Status Inventory (RAND-12) is a generic PROM that includes 12 questions intended to measure perceived physical and mental health. It consists of eight dimensions (physical functioning, role-limitation physical, role-limitation emotional, mental health, bodily pain, vitality, social functioning and general health) that produce two summary scores: the physical component summary score (PCS12) and mental component summary score (MCS12) [27]. The PCS12 and MCS12 scores in the present study are based on the SF-12 version 1 and a method that enables physical and mental health to be correlated and requires no items to be missing. Scores range from 0 to 100, where higher scores indicate better health. The PCS12 and MCS12 scales are norm-based and are transformed to have a mean of 50 and a standard deviation of 10 (t-scores) in the general US population [27,28].

Foot Function Index Revised Short Form (FFI-RS) includes 34 questions regarding pain, stiffness, difficulty, activity limitations and social issues during the previous week. The items are rated on a 4-point Likert scale ranging from 1 to 4, where 1 represents "no pain" and 4 represents "very strong pain". In addition, 5 ("not applicable") can be selected for six of the questions. All item scores (except for 5, "not applicable") were rescaled to a range of 0–3 and summed. The sum was then divided by the highest possible sum (excluding the score of 5) and transformed into a 0%–100% scale [21]. Lower scores indicate better foot health.

Anchor for important change

A patient global impression of change (PGIC) scale was also used as an outcome in the RCT. The wording of the question was "Compared to the start of the study, how is your general health status today?" The patients were asked to rate their general health status at 6-month follow-up compared to the start of the study on a 7-point Likert scale, where 1 was "very much improved," 2 was "much improved," 3 was "minimally improved," 4 was "unchanged," 5 was "minimally worse," 6 was "much worse" and 7 was "very much worse". The PGIC scale in the present study was applied as an anchor of important change, to discriminate patients with important improvement from patients with no important change [29].

Statistics

Statistical analyses were performed with IBM SPSS Statistics for windows version 28 (Armonk, N.Y:IBM Corp) and Stata version

Table 1. Distribution of PGIC responses at 6-month follow-up ($N = 115$).

PGIC response		% of responses
1	Very much improved	15.7
2	Much improved	31.3
3	Minimally improved	25.2
4	No change	20.9
5	Minimally worse	3.5
6	Much worse	3.5
7	Very much worse	0

PGIC: Patient global impression of change.

17(College Station, TX: StataCorp LLC). Total scores for all outcome measures were calculated at baseline and 6 months. Patients with any missing items from the NRS, RAND-12 and PGIC scales were excluded. For the FFI-RS, missing values were imputed based on predictive mean matching if less than 25% of the items were missing [30]. Any patients with more than 25% of FFI-RS items missing were excluded. The total sample was considered as one cohort and the sample was not divided based on the four intervention arms.

Responsiveness

The answers from the PGIC scale/anchor were dichotomised into two ordinal categories, where scores of 1 or 2 was considered "importantly improved" and scores of 3 to 7 as "not improved". Due to the low number of patients reporting deterioration, a "worse" category was not included (Table 1). Responsiveness was calculated using the standardised response mean (SRM) and the AUC for the receiver operating characteristic (ROC) curve (ROC_{AUC}). SRM was estimated by dividing the mean change score by the standard deviation (SD) of the change scores in the importantly improved and not improved patients. Cut-off points employed by Cohen was used, where an SRM> 0.8 indicates large responsiveness, 0.5 to 0.8 indicates moderate responsiveness and 0.2 to <0.5 indicates low responsiveness [31]. The ROC_{AUC} represents the ability of the PROM to separate patients who have improved from patients who have not improved according to the anchor. A perfect accuracy will give a value of 1, while 0.5 represents chance alone. Responsiveness is considered adequate when the AUC is at least 0.7 [6]. The AUC was estimated with a 95% CI.

Minimal important change (MIC)

The anchor-based ROC method was used to estimate the MIC values for each PROM [8]. ROC curves were plotted as a combination of sensitivity and 1 – specificity for each change score. The sensitivity is the proportion of importantly improved patients correctly identified by the PROM, and specificity is the proportion of patients with no important change correctly identified by the PROM. The two groups of patients – importantly improved and not improved – will have overlapping change scores, and the optimal cut-off point is where the sum of proportions of misclassifications is smallest ((1– sensitivity) + (1–specificity)) or, equivalently, where the point is closest to the upper-left corner [6–8]. The 95% CI was estimated using stratified bootstrap replications [32].

Results

Study sample

One hundred and sixty-five patients completing the 6-month follow-up were eligible for analysis. Of the 165 patients, 50 were excluded due to missing items, giving a sample size of 115, which is above the general recommendations for estimating the MIC [7]. Table 2 describes the baseline demographic characteristics of the included and excluded patients. The two groups were not significantly different from each other when compared using the independent sample t-test and chi-square test. Table 3 presents the mean scores for the FFI-RS, NRS and RAND-12 at baseline and 6-month follow-up.

Table 2. Baseline characteristics of included ($N = 115$) and excluded ($N = 50$) patients.

	Included ($N = 115$)	Excluded ($N = 50$)
Sex (female/male)	92/23	45/5
Age in years[a]	45.4 (10.6)	45.1 (11.5)
Duration of symptoms[b]		
<3 months	0 (0)	0 (0)
3–6 months	27 (23.5)	13 (26.0)
6–12 months	34 (29.6)	16 (32.0)
12–24 months	21 (18.3)	7 (14.0)
>24 months	33 (28.7)	14 (28.0)
Unilateral symptoms[b]	72 (62.6)	30 (60.0)
Bilateral symptoms[b]	34 (37.4)	20 (40.0)

N: number of participants.
[a]Values are mean (standard deviation).
[b]Values are number (%).

Table 3. Baseline and 6-month follow-up PROM scores for included patients ($N = 115$).

PROM (score range)	Baseline (SD)	6-month follow-up (SD)
FFI-RS (0–100)	44.8 (19.3)	26.5 (22.8)
NRS a (0–10)	6.1 (2.0)	3.5 (2.6)
NRS r (0–10)	3.6 (2.4)	2.1 (2.2)
PCS12 (0–100)	39.6 (9.7)	46.2(11.4)
MCS12 (0–100)	42.8 (11.1)	46.1 (12.5)

Patient-reported outcome measure (PROM).
N: number of participants.
FFI-RS: Foot Function Index Revised Short Form; NRS a: Numeric Rating Scale during activity; NRS r: Numeric Rating Scale during rest; PCS12: Physical component summary score, RAND-12; MCS12: Mental component summary score, RAND-12.
Values are mean (standard deviation (SD)).

Responsiveness

The ROC_{AUC} and SRM values are reported in Table 4. Four out of five ROC_{AUC} estimates were > 0.7, indicating adequate responsiveness. The FFI-RS had the highest ROC_{AUC} value (0.82; 95% CI 0.75 to 0.90), and the NRS at rest had the lowest ROC_{AUC} value (0.67; 95% CI 0.58 to 0.77). ROC curves for all instruments at 6-month follow-up are reported in Figures 1 and 2. The SRM values in the improved group ranged from 1.06 to 1.82, indicating large responsiveness for all PROMs (for the FFI-RS and NRS, more negative values indicate greater improvement). In the not improved group, the SRM values ranged from 0.01 to 0.70, indicating low to moderate responsiveness.

Minimal important change

The distribution of the responses on the PGIC scale at 6-month follow-up is presented in Table 1; 47% of participants were importantly improved based on the anchor. For the FFI-RS, the MIC was −19 (95% CI −24.2 to −13.8), suggesting that a change in the FFI-RS ≥ 19 represents a meaningful improvement in general health for the patients. The MIC values for the different PROMs are presented in Table 4.

Table 4. Responsiveness and minimal important change.

PROM	ROC_{AUC} (95% CI)	SRM improved	SRM not improved	MIC (95%CI)	Sensitivity/specificity
FFI-RS	0.82 (0.75 to 0.90)	−1.80	−0.56	−19 (-24.2 to −13.8)	0.78/0.77
PCS12	0.81 (0.73 to 0.90)	1.26	0.33	6.4 (3.40 to 9.37)	0.74/0.80
MCS12	0.73 (0.64 to 0.83)	0.82	0.01	4.4 (1.55 to 7.19)	0.65/0.72
NRS a	0.77 (0.68 to 0.85)	−1.82	−0.70	−1.5 (-3.3 to 0.28)	0.89/0.51
NRS r	0.67 (0.58 to 0.77)	−1.06	−0.32	−1.5 (-3.3 to 0.08)	0.63/0.69

PROM: Patient-reported outcome measure; FFI-RS: Foot Function Index Revised Short Form; NRS a: Numeric Rating Scale during activity; NRS r: Numeric Rating Scale during rest; PCS12: Physical component summary score, RAND-12.
Mental component summary score, RAND-12 (MCS12). ROC_{AUC}: area under the ROC curve; SRM improved: mean change score divided by the standard deviation of the mean change score in the group of patients classified as improved; SRM not improved: mean change score divided by the standard deviation of the change score in the group of patients classified as not improved; MIC: minimal important change, the cut-off point on the ROC curve where the sum of misclassification is smallest; sensitivity/specificity: proportion of improved patients correctly classified as improved/proportion of not improved patients correctly classified as not improved.

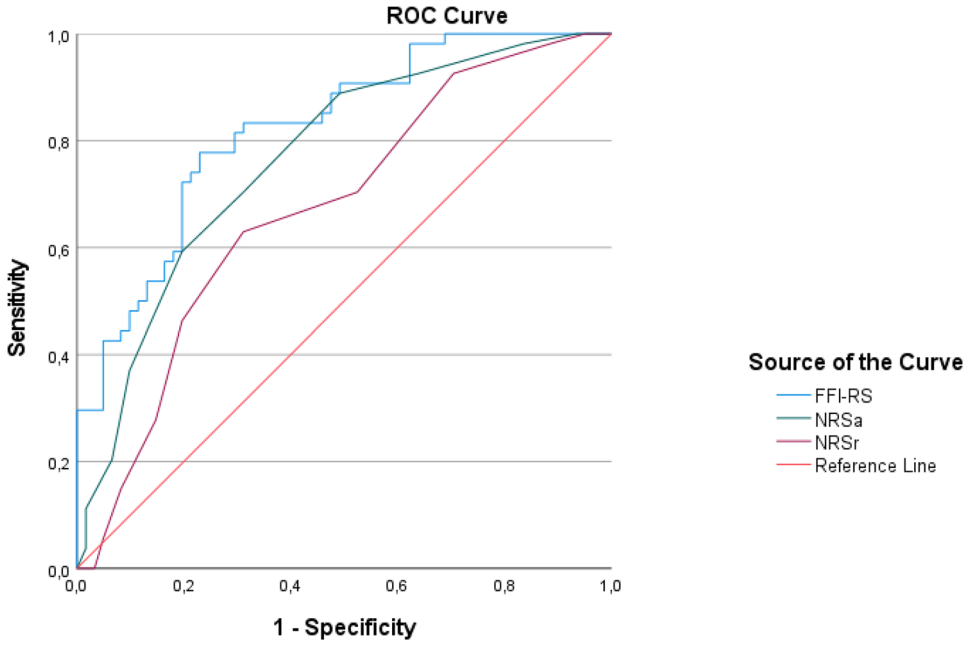

Figure 1. ROC curves of FFI-RS, NRS a and NRS r at 6-month follow-up.

FFI-RS: Foot Function Index Revised Short Form; NRS a: Numeric Rating Scale during activity; NRS r: Numeric Rating Scale during rest.

Discussion

All PROMs except the NRS at rest were responsive in both the ROC_{AUC} and SRM estimates in the importantly improved group. When comparing the estimates for each PROM in Table 4, the FFI-RS comes out marginally on top. These findings are consistent with previous knowledge that specific PROMs tend to be more responsive to intervention-related changes compared to generic measures [3,4,33,34]. The course of plantar fasciopathy can be prolonged before improvement occurs [35] and the primary outcome in the RCT was chosen to be 6 months.

The RAND-12 and its summary scores, the PCS12 and MCS12, were also responsive according to ROC_{AUC} and SRM estimates. The PCS12 was more responsive than the MCS12. In our population, it appears that the items from the PCS12 more closely reflect the aspects of the condition than the items from the MCS12. This is comparable to previous studies on musculoskeletal disorders presenting the responsiveness of summary scores [25,26].

For pain during activity, the AUC value was 0.77, indicating adequate responsiveness, though the 95% CI was < 0.7, which adds some uncertainty to the estimate and reflects that pain is measured by one item, while the other scores are indexes of a range of questions. Pain at rest was not responsive according to the ROC_{AUC} estimate, but pain both during activity and at rest had large responsiveness according to the SRM in the improved group. This discrepancy can be explained by differences in the statistical methods applied to calculate the estimates. A further explanation for the relatively low responsiveness of pain at rest could be that many patients with plantar fasciopathy often experience little or no pain at rest, and therefore the expected change over time is small. Responsiveness as a concept has given rise to debate, particularly regarding statistical approaches and production of conflicting results [36]. Responsiveness indicators should be considered context-specific, as they depend on the population under study, type of intervention, timing of data collection, construct of change being quantified and statistical approach [37].

A MIC value of −19 appeared for the FFI-RS at 6-month follow-up which contrasts with the previous study by Mørk et al. from virtually the same population, reporting MIC at −8.4 at 3-month follow-up [21]. The relatively wide 95% CI of the MIC value (−24.2 to −13.8) in the present study reflects the uncertainty of the estimate but still does not encompass the value from the previous 3-month follow-up estimate. One possible explanation for the discrepancy between 3- and 6-month follow-ups is that the patients may have changed the way they perceive their health status over that time, a phenomenon known as the response shift effect. A response shift can occur in all PROMs and in particular when patients are asked to make an overall assessment of their condition [8]. In addition,

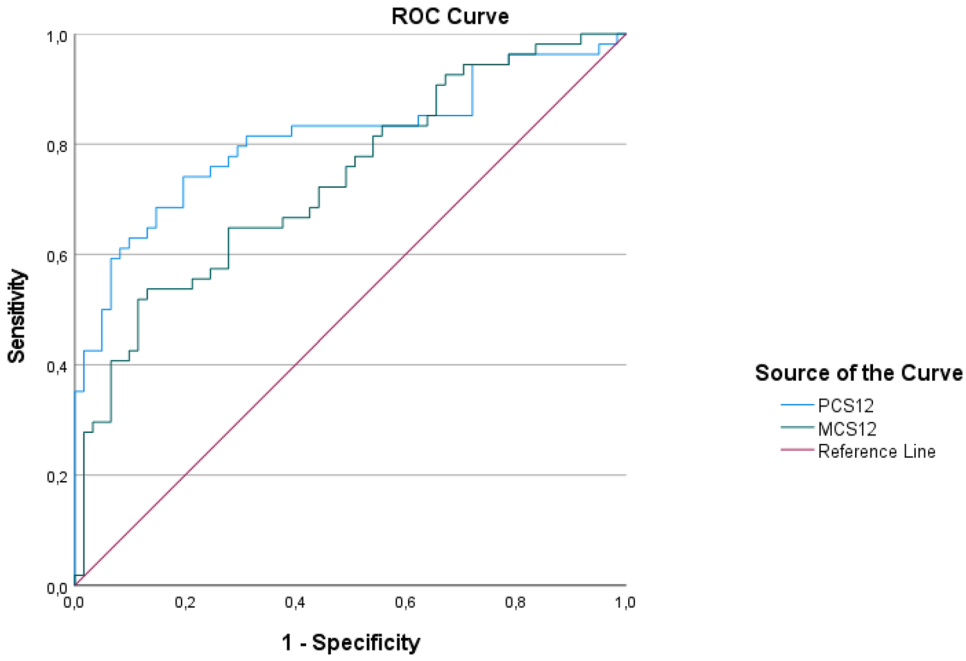

Figure 2. ROC curves of RAND-12 summary scores, PCS12 and MCS12, at 6-month follow-up.
PCS12: Physical component summary score, RAND-12; MCS12: Mental component summary score, RAND-12.

recall bias can influence the answer to the anchor, and a previous study found that global perceived change was strongly influenced by the patient's current status [38].

The MIC was −1.5 for both pain during activity and pain at rest. This is similar to results reported by Farrar et al. comparing clinically important differences in pain intensity measured by the NRS in 10 randomised controlled trials of patients with chronic pain. The same 7-point anchor employed in the present study was used, and improvement was defined similarly as patients reporting to be much or very much improved. They found the MIC to be −1.74 (reported rounded up to 2) based on ROC analysis [39]. The estimates from the analyses are uncertain due to the wide CIs for NRS scores both during activity and at rest. Farrar et al. did not report the 95% CI [39]. SooHoo et al. reported MIC values from the SF-12 version 2 on 391 patients following total knee or hip arthroplasty using a distribution-based method (change in score equal to or greater in magnitude than 0.5 SD for the study population). Similar to the present study, they estimated the PCS to be 4.97 and the MCS to be 5.11 [40].

Strengths and limitations

Strengths of the present study include the novel results on comparative responsiveness of the FFI-RS, NRS and RAND-12 in the plantar fasciopathy population. Both distribution- and anchor-based methods were used, which is a common approach. A limitation of the study is the wording of the anchor question. When not specifically addressing their foot condition, patients may emphasise other constructs related to their health. The interpretation of the anchor may be a limitation of many studies assessing responsiveness and estimating the MIC. For the best possible evaluation of an outcome measure we recommend that the wording of the anchor addresses the specific condition under study. The decision about which items that defined the importantly improved category was made by us and not the patients. To ensure that the defined cut-off point corresponded to the patients' opinions, a second question could have been asked regarding whether the change was meaningful. Another limitation that could have affected the results is that 50 patients were excluded from the analysis due to missing items and that imputation had to be used for the FFI-RS.

Conclusion

The region-specific FFI-RS was marginally more responsive than the other PROMs. Responsiveness and MIC estimates are contextual attributes of a PROM and should be regarded as indicative rather than fixed estimates. Further studies are needed to provide additional insight into the responsiveness characteristics of these PROMs.

Acknowledgments

We would like to thank Sophies Minde Ortopedi AS for their financial support and for providing foot orthoses to the participants in the study.

Disclosure statement

The authors report that there are no competing interests to declare.

Funding

This work was supported by the Sophies Minde Ortopedi, Norway (31768).

Data availability statement

The data for this article stem from a randomised controlled trial registered in the ClinicalTrials.gov database (reference number:

NCT03472989). The datasets used and analysed during the present study are available from the corresponding author upon reasonable request.

References

[1] Nelson EC, Eftimovska E, Lind C, et al. Patient reported outcome measures in practice. BMJ. 2015;350:g7818. doi: 10.1136/bmj.g7818.

[2] Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167–f167. doi: 10.1136/bmj.f167.

[3] Chang EM, Gillespie EF, Shaverdian N. Truthfulness in patient-reported outcomes: factors affecting patients' responses and impact on data quality. Patient Relat Outcome Meas. 2019;10:171–186. doi: 10.2147/PROM.S178344.

[4] Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. Int J Integr Care. 2002;2(4):e15. doi: 10.5334/ijic.65.

[5] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–745. doi: 10.1016/j.jclinepi.2010.02.006.

[6] de Vet HT, Mokkink LB, Knol DL. Responsiveness. Measurement in medicine: a practical guide. Cambridge: cambridge University Press; 2011.

[7] Terwee CB, Peipert JD, Chapman R, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res. 2021;30(10):2729–2754. doi: 10.1007/s11136-021-02925-y.

[8] De Vet HT, Mokkink LB, Knol DL. Interpretability. Measurement in medicine: a practical guide. Cambridge: cambridge University Press; 2011.

[9] Copay AG, Eyberg B, Chung AS, et al. Minimum clinically important difference: current trends in the orthopaedic literature, part II: lower extremity: a systematic review. JBJS Rev. 2018;6(9):e2–e2. doi: 10.2106/JBJS.RVW.17.00160.

[10] Skare Ø, Brox JS, Schrøder CP, et al. Responsiveness of five shoulder outcome measures at follow-ups from 3 to 24 months. BMC Musculoskelet Disord. 2021;22(1):606. doi: 10.1186/s12891-021-04483-3.

[11] Sierevelt IN, Zwiers R, Schats W, et al. Measurement properties of the most commonly used foot- and Ankle-Specific questionnaires: the FFI, FAOS and FAAM. A systematic review. Knee Surg Sports Traumatol Arthrosc. 2018;26(7):2059–2073. doi: 10.1007/s00167-017-4748-7.

[12] Button G, Pinney S. A meta-analysis of outcome rating scales in foot and ankle surgery: is there a valid, reliable, and responsive system? Foot Ankle Int. 2004;25(8):521–525. doi: 10.1177/107110070402500802.

[13] Hunt KJ, Hurwit D. Use of patient-reported outcome measures in foot and ankle research. J Bone Joint Surg Am. 2013;95(16):e118–e1189. doi: 10.2106/JBJS.L.01476.

[14] Shazadeh Safavi P, Janney C, Jupiter D, et al. A systematic review of the outcome evaluation tools for the foot and ankle. Foot Ankle Spec. 2019;12(5):461–470. doi: 10.1177/1938640018803747.

[15] Jia Y, Huang H, Gagnier JJ. A systematic review of measurement properties of patient-reported outcome measures for use in patients with foot or ankle diseases. Qual Life Res. 2017;26(8):1969–2010. doi: 10.1007/s11136-017-1542-4.

[16] Martin RL, Irrgang JJ. A survey of self-reported outcome instruments for the foot and ankle. J Orthop Sports Phys Ther. 2007;37(2):72–84. doi: 10.2519/jospt.2007.2403.

[17] Ferreira-Valente MA, Pais-Ribeiro JL, Jensen MP. Validity of four pain intensity rating scales. Pain. 2011;152(10):2399–2404. doi: 10.1016/j.pain.2011.07.005.

[18] Euasobhon P, Atisook R, Bumrungchatudom K, et al. Reliability and responsivity of pain intensity scales in individuals with chronic pain. Pain. 2022;163(12):e1184–e1191. doi: 10.1097/j.pain.0000000000002692.

[19] Budiman-Mak E, Conrad K, Stuck R, et al. Theoretical model and rasch analysis to develop a revised foot function index. Foot Ankle Int. 2006;27(7):519–527. doi: 10.1177/107110070602700707.

[20] Budiman-Mak E, Conrad KJ, Mazza J, et al. A review of the foot function index and the foot function index - revised. J Foot Ankle Res. 2013;6(1):5. doi: 10.1186/1757-1146-6-5.

[21] Mørk M, Hoksrud AF, Soberg HL, et al. Psychometric properties of the norwegian foot function index revised short form. BMC Musculoskelet Disord. 2022;23(1):416. doi: 10.1186/s12891-022-05374-x.

[22] Heide M, Mørk M, Røe C, et al. The effectiveness of radial extracorporeal shock wave therapy (rESWT), sham-rESWT, standardised exercise programme or usual care for patients with plantar fasciopathy: study protocol for a double-blind, randomised, sham-controlled trial. Trials. 2020;21(1):589. doi: 10.1186/s13063-020-04510-z.

[23] Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. Ann Med. 2001;33(5):350–357. doi: 10.3109/07853890109002089.

[24] Bronfort G, Bouter LM. Responsiveness of general health status in chronic low back pain: a comparison of the COOP charts and the SF-36. Pain. 1999;83(2):201–209. doi: 10.1016/s0304-3959(99)00103-7.

[25] Luo X, George ML, Kakouras I, et al. Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. Spine (Phila Pa 1976). 2003;28(15):1739–1745. doi: 10.1097/01.BRS.0000083169.58671.96.

[26] Webster KE, Feller JA. Comparison of the short form-12 (SF-12) health status questionnaire with the SF-36 in patients with knee osteoarthritis who have replacement surgery. Knee Surg Sports Traumatol Arthrosc. 2016;24(8):2620–2626. doi: 10.1007/s00167-015-3904-1.

[27] Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 health survey, V.I. Health Qual Life Outcomes. 2007;5(1):54. doi: 10.1186/1477-7525-5-54.

[28] Helse N. RAND 36- Term Short Form Health Survey (SF-36) and RAND 12- kortversjon (SF-12). Skjema for kartlegging av helserelatert livskvalitet 22.april 2021. Available from: https://helse-stavanger.no/seksjon/NKB/Documents/Kartleggingsskjema/Norsk%20RAND-12.pdf.

[29] Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163–170. doi: 10.1179/jmt.2009.17.3.163.

[30] Vink G, Frank LE, Pannekoek J, et al. Predictive mean matching imputation of semicontinuous variables. Stat Neerl. 2014;68(1):61–90. doi: 10.1111/stan.12023.

[31] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J: laurence Erlbaum; 1988.

[32] Clayton P. CUTPT: stata module for empirical estimation of cutpoint for a diagnostic test2013.

[33] Vishwanathan K, Braithwaite I. Construct validity and responsiveness of commonly used patient reported outcome instruments in decompression for lumbar spinal stenosis. J Clin Orthop Trauma. 2021;16:125–131. doi: 10.1016/j.jcot.2021.01.002.

[34] Giesinger K, Hamilton DF, Jost B, et al. Comparative responsiveness of outcome measures for total knee arthroplasty. Osteoarthritis Cartilage. 2014;22(2):184–189. doi: 10.1016/j.joca.2013.11.001.

[35] Hansen L, Krogh TP, Ellingsen T, et al. Long-Term prognosis of plantar fasciitis: a 5- to 15-Year follow-up study of 174 patients With ultrasound examination. Orthop J Sports Med. 2018;6(3):2325967118757983. doi: 10.1177/2325967118757983.

[36] Davidson M, Keating J. Patient-reported outcome measures (PROMs): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. Br J Sports Med. 2014;48(9):792–796. doi: 10.1136/bjsports-2012-091704.

[37] Beaton DE, Bombardier C, Katz JN, et al. A taxonomy for responsiveness. J Clin Epidemiol. 2001;54(12):1204–1217. doi: 10.1016/s0895-4356(01)00407-3.

[38] Grøvle L, Haugen AJ, Hasvik E, et al. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol. 2014;67(5):508–515. doi: 10.1016/j.jclinepi.2013.12.001.

[39] Farrar JT, Young JP, Jr., LaMoreaux L, et al. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001;94(2):149–158. doi: 10.1016/S0304-3959(01)00349-9.

[40] SooHoo NF, Li Z, Chenok KE, et al. Responsiveness of patient reported outcome measures in total joint arthroplasty patients. J Arthroplasty. 2015;30(2):176–191. doi: 10.1016/j.arth.2014.09.026.