

The foot function index is more sensitive to change than the Leeds Foot Impact Scale for evaluating rheumatoid arthritis patients after forefoot or hindfoot reconstruction

Imraan Muradin¹ · Huub J. L. van der Heide¹

Received: 16 July 2015 / Accepted: 31 December 2015 / Published online: 3 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Purpose This study examines the responsiveness of the Foot Functional Index (FFI) and Leeds Foot Impact Scale for Rheumatoid Arthritis (LFIS-RA) in rheumatoid arthritis (RA) patients receiving a forefoot or hindfoot reconstruction.

Methods This was a prospective cohort study including 30 rheumatoid arthritis patients with severe rheumatoid foot deformities in need for surgical correction. Responsiveness was measured using distribution-based methods (standardized effect size, standardized response mean and Guyatt responsiveness ratio) and anchor-based methods (receiver operating characteristics curves and correlation analyses) by making use of an anchor question. To examine the depth of the questionnaires we measured the floor and ceiling effects.

Results The study population consisted of three males and 27 females, with a mean age of 62 years. The mean follow-up time was 38 months. Twenty-two feet received a forefoot reconstruction and eight feet a triple arthrodesis. For the FFI the SES was -0.80 , SRM was -0.85 and the GRR was -1.25 . For the LFIS-RA the SES was 0.58 , SRM was 0.58 and the GRR was 0.88 . The AUC was 0.741 and 0.645 for FFI and LFIS, respectively. Contrary to the LFIS-RA, the FFI showed a significant correlation between change score and the anchor question. Both questionnaires did not show a significant floor or ceiling effect.

Conclusion The FFI showed a large responsiveness and the LFIS-RA showed moderate responsiveness in rheumatoid

arthritis patients receiving forefoot or hindfoot surgery, without floor or ceiling effects in both questionnaires.

Keywords Responsiveness · FFI · LFIS-RA · Rheumatoid arthritis · Triple arthrodesis · Forefoot arthrodesis

Introduction

Despite recent advances in pharmacological management of rheumatoid arthritis, forefoot deformity, with its symptoms, remains a common problem, often requiring operative treatment [1]. The foot is most commonly affected early with a prevalence of up to 90 % for the metatarsophalangeal (MTP) joints and in 15 % the forefoot is the first manifestation of the disease [2]. Deformities are mostly found in the forefoot with symptoms such as atrophy, claw toes, dislocation of the plantar fascia and metatarsophalangeal subluxations [3]. Hindfoot involvement in RA is similarly common, with reported prevalence numbers as high as 50 % through 90 % [4]. The treatment of severe forefoot deformities consists of operative correction through repositioning of the metatarsophalangeal joints of the lesser rays in combination with an arthrodesis of the first metatarsophalangeal joint [3].

In regards to the hindfoot triple arthrodesis is the preferred technique [5].

Concerning the feet reconstruction in RA patients there is a need for structured evaluation of the subjective aspects of the patient's symptoms [6]. Patient-reported outcome measurements (PROMs) are used to evaluate the outcome of operative treatments [2]. To be truly useful PROMs must exhibit good psychometric properties such as reliability, validity and responsiveness [7]. Responsiveness is defined as the ability of a measuring instrument to detect change when it has occurred [8]. The present literature describes 36 different PROMs used

✉ Huub J. L. van der Heide
H.J.L.van_der_Heide@lumc.nl

¹ Department of Orthopaedics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

to measure foot problems in RA patients; solely five instruments were measured for responsiveness [9]. A worldwide (USA, Europe, Australia and Asia) used PROM to measure the impact of foot pathology on function in terms of pain, difficulty and activity restriction is the Foot Function Index [10, 11] as described by Budiman et al. Nelson et al. reported moderate to large responsiveness of the FFI in RA patients receiving foot or ankle surgery; however, they did not specify the operative procedure [12]. The responsiveness of the FFI in end stage ankle arthritis patients, who received either a total ankle replacement or an ankle arthrodesis, showed a high level of responsiveness (both the standardized effect size and standardized response mean above the threshold of 0.8 indicating a high level of responsiveness) [13].

The Leeds Foot Impact Scale for Rheumatoid Arthritis (LFIS-RA) described by Helliwell et al. is a PROM used to measure the outcome (impairment and activity limitations) in RA patients with foot deformities, in mostly eastern Europe countries, including England, Germany, Hungary and the Netherlands [14]. Initial results showed good psychometric properties [15]. The Dutch-translated version of the LFIS-RA showed excellent psychometric properties, but hasn't been tested for responsiveness [16]. This study was conducted to determine the responsiveness of the FFI and LFIS-RA in RA patients with severe feet deformities and to compare them.

Methods

Participants

This prospective cohort study included 30 RA patients with typical RA deformities in need for operative treatment. Data was collected between 2009 and 2013. Regarding the FFI and LFIS, the patients were approached during the regular outpatient clinic appointments and agreed to participate in this study. An assistant who was not involved in the treatment approached the patients to obtain the anchor question. All patients were examined and operated by one orthopaedic surgeon.

Outcome assessment

The outcome was measured using the FFI and LFIS-RA pre-operative and post-operative. Post-operative an anchor question was added, using a seven point Likert-scale; 'How did the situation change regarding your foot since the surgery? (1: *very much deteriorated*, 2: *much deteriorated*, 3: *somewhat deteriorated*, 4: *about the same*, 5: *somewhat improved*, 6: *much improved*, 7: *very much improved*). The FFI consists of 23 items (score range: 0–115) grouped into three subscales; *pain*, *disability* and *activity limitation* related to foot pathology [10]. Each question provides six alternatives. Important to

mention is that a higher score indicates a worse outcome. The LFIS-RA test consist of 51 dichotomous items (score range: 51–102) grouped into four subscales; *impairment*, *activities*, *participation*, and *footwear* [15]. In contrary to the FFI, a higher LFIS-RA score means a better outcome.

Responsiveness

A distinction could be made between internal and external responsiveness [8]. Internal responsiveness is the ability of a measure to change over a particular time frame [17]. External responsiveness is the ability of a measure to change over a particular time frame related to an external measure [8]. The internal responsiveness could be measured using the *standardized effect size* (SES), *standardized response mean* (SRM) and *Guyatt responsiveness ratio* (GRR) [8]. The SES was calculated by dividing the mean change score by the standard deviation of the mean baseline score. The SRM was calculated by dividing the mean change score by the standard deviation of the mean change score. To determine the GRR the study population was divided into two groups based on the anchor question: patients that improved after the procedure and patient that showed little to no change after the procedure. The "*improved*" group consisted of people who answered "*much improved*" and "*very much improved*" to the anchor question; the "*not convincible changed*" group consisted of patients that answered "*somewhat deteriorated*" or "*somewhat improved*". To calculate the GRR the mean change score in the "*improved*" group was divided by the standard deviation of the mean change score in the "*not convincible changed*" group [7]. For the above mentioned approaches a value of 0.5 or less represents an inadequate responsiveness, values between 0.50 and 0.80 a moderate internal responsiveness, and values of 0.80 or greater a large internal responsiveness [18].

External responsiveness

External responsiveness was measured using the area under the curve (AUC) within the ROC curve [19]. An AUC below 0.5 reflects random distribution, AUC between 0.5 and 0.7 is considered to have limited discrimination accuracy, AUC between 0.7 through 0.8 acceptable accuracy, AUC between 0.8 through 0.9 excellent accuracy and an AUC above 0.9 was considered to have outstanding accuracy [19]. In order to perform this analysis it was necessary to dichotomize the anchor question. The "*much improved*" and "*very much improved*" were categorized as the "*improved*" group, and the "*somewhat improved*", "*about the same*", "*somewhat deteriorated*", "*much deteriorated*" and "*very much deteriorated*" as the "*not improved*" group. A practical cut-off point with optimal balance of sensitivity and specificity was chosen. A correlation analysis (Spearman correlation coefficient) provided an

examination of the relationship between the questionnaire and anchor question [20].

Floor and ceiling effects

Floor and ceiling effects in the questionnaires could make it difficult to measure changes after interventions such as a surgery due to distortion of the score distribution. Floor or ceiling effects are considered to be present if more than 15 % of respondents achieved the lowest or highest possible score, respectively [20].

Results

Patient characteristics

The study population consisted of three males (10 %) and 27 females (90 %). The mean age of the group was 62 years (range 44–76; SD 8.7). All patients were reviewed after a mean follow-up of 38 months (range 5–61; SD 16). On 22 feet (73 %) a forefoot reconstruction was performed and on eight feet (27 %) a triple arthrodesis. Approximately 70 % of the patients reported they improved after the surgery (Table 1). During the first measurement, 25 patients had a positive RA-factor (Table 2). Roughly half of the patients used DMARDs during the first measuring point (Table 3).

Internal responsiveness

Pre-operatively, the mean FFI total score was 55.5 (range 28–82; SD 14), and post-operatively it was 44.2 (range 24–66; SD 12). The mean change was a decrease of 11.3 points (range –42 to 13; SD 13). For the FFI the SES was –0.80, SRM was –0.85 and GRR was –1.25 (Table 4). As mentioned before, negative values indicate improvement. Pre-operatively, the mean LFIS-RA score was 76.8 (range 62–98; SD 8), and post-operative it was 81.2 (range 65–102; SD 10). The mean

Table 1 Frequency table for the outcome of the anchor question

Operative procedure	Response answer	Frequency	Percent (%)
Triple arthrodesis	Slightly improved	1	12.5
	Much improved	5	62.5
	Very much improved	2	25.0
	Total	8	100.0
Forefoot reconstruction	Much deteriorated	2	9.1
	Slightly deteriorated	7	31.8
	Slightly improved	4	18.2
	Much improved	6	27.3
	Very much improved	3	13.6
	Total	22	100.0

Table 2 Diagnosis at first measuring moment

Diagnosis	Frequency	Percent (%)
RAF-positive	25	83.3
RAF negative	2	6.7
JIA	1	3.3
Psoriatic arthritis	2	6.7

change was an increase of 4.4 points (range –12 to 23; SD 8). For the LFIS-RA the SES was 0.58, SRM was 0.58 and the GRR was 0.90 (Table 4).

External responsiveness

For the FFI questionnaire the AUC was 0.741 (CI 95 %: 0.558–0.924, SE 0.094, $P=0.025$) (Fig. 1). For the LFIS-RA questionnaire we obtained an AUC of 0.645 (CI 95 %: 0.440–0.850, SE 0.104, $P=0.177$) (Fig. 1). The optimal cut-off point for the FFI was 6 points, with a sensitivity of 81 % and a specificity of 57 %. For the LFIS-RA the cut-off point was 1.5 points, with a sensitivity of 75 % and a specificity of 57 %. There was a significant negative correlation (Spearman correlation coefficient, 0.396; $P=0.030$) between the FFI change scores and anchor questions. No significant correlation between the LFIS-RA changes scores and anchor questions was found (Spearman correlation coefficient = 0.210; $P=0.266$).

Floor and ceiling effect

Neither the LFIS or the FFI showed a significant floor or ceiling effect pre-operative and post-operative (Table 5), due to the fact that less than 15 % of the respondents achieved the highest or lowest possible score.

Discussion

Multiple instruments are developed to measure foot function, foot pain and foot related disability in RA patients. A recent study reviewed the measurement properties of 36 different instruments, and concluded that solely five instruments were measured for responsiveness [9]. Our study is the first to evaluate the responsiveness and floor/ceiling effects of the FFI and

Table 3 Medication use at first measuring moment

Medication	Frequency	Percent (%)
NSAID	4	13.3
DMARD	17	56.7
Biological	8	26.7
Steroids	1	3.3

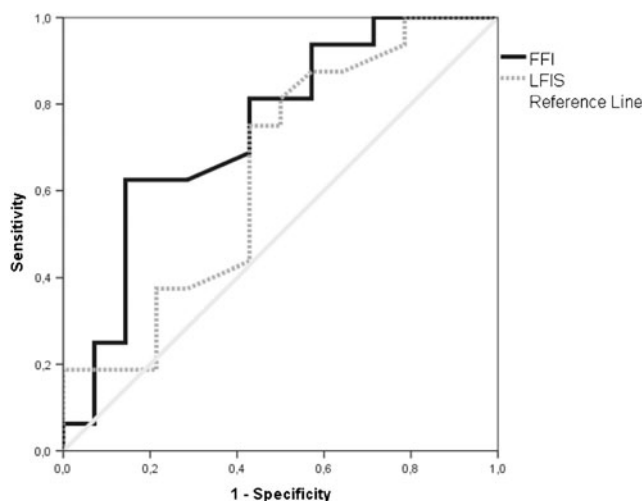
Table 4 Descriptive statistics and internal responsiveness characteristics for the FFI and LFIS-RA

Measure	FFI	LFIS-RA
Mean change (SD)		
Overall group ($n = 30$)	11.3 (13.4)	4.4 (7.7)
Improved group ($n = 16$)	16.42 (11.7)	6.6 (7.5)
Non-improved group ($n = 14$)	5.51 (13.1)	2.0 (7.5)
SES	-0.80	0.58
SRM	-0.85	0.58
GRR	-1.25	0.90

LFIS questionnaire in RA patients who received forefoot and hindfoot correction, and boosts robustness by applying both anchor-based as well as distribution-based approaches to measure responsiveness.

We defined responsiveness as a parameter for measurement instruments to measure change over time. We defined internal responsiveness as the ability of a measure to change over time, and external responsiveness as a change over time corresponding with an external measure. We found a large internal responsiveness of the FFI, for the LFIS we found a moderate internal responsiveness. The use of GRR to measure responsiveness is a controversial topic. Although the use of the GRR is seen by some as the superior measurement for responsiveness by some researchers [8, 21], others claim that the GRR does not reflect the validity of the changed score [22, 23]. This difference arises due to the difference in the definition of responsiveness [7].

In regards to the external responsiveness the FFI had stronger discriminative abilities than the LFIS-RA in the present study population. The LFIS had shown below acceptable discriminative abilities. To demonstrate the above-stated, a 60-year-old patient who underwent forefoot surgery answered the anchored question with “much improved”. The FFI showed a

**Fig. 1** Receiver operating characteristic (ROC) curve analysis**Table 5** Range of values and for the FFI and LFIS-RA

Time of measurement	Number	Lower limit	Upper limit	Floor effect (%)	Ceiling effect (%)
FFI pre-operative	30	28	82	0	0
FI post-operative	30	24	63	0	0
LFIS pre-operative	30	65	98	0	0
LFIS post-operative	30	65	102	0	3

change of -20 points (17 %) and the LFIS-RA a change of three points (2 %). Another patient, a 68-year-old woman, who underwent the same forefoot surgery, answered “very much improved” at the anchored question, showed a FFI change of -24 points (20 %) and an LFIS-RA change of three points (2 %). If we would rely solely on the LFIS-RA questionnaire, we wouldn’t be able to perceive the true change after the surgery.

External responsiveness was measured using the AUC within the ROC curve. A disadvantage of the ROC method is that the AUC has little meaning, and is primarily useful for ranking competing scales [21]. Another disadvantage is that the external criterion for change must be dichotomized; by merging the groups, valuable data is lost [19]. In choosing the cut-off points we strived for a high sensitivity and specificity. The FFI reached a higher sensitivity than the LFIS-RA, with an equal specificity in both questionnaires. Correlation analysis showed a significant correlation between the FFI change score and anchor question. In both the FFI and LFIS we did not observe any floor or ceiling effects.

A limitation of this study is the absence of a gold standard in measuring and expressing the responsiveness [22]. Furthermore, the use of an anchor question is prone to bias [24]. It is very difficult for people to remember their past state. They deduce their prior status from their present state and invoke an implicit theory of change to construct their prior state before the surgery [25], thereby creating a high correlation between measure of change and present state, but a low correlation between measure of change and prior state. Following on from the anchor question, it was remarkable that none of the patients reported “no change”. As Guyatt and Deyo remarked, the answer on the anchor question will never solitarily rely on operative outcome, but subsequently measures satisfaction with the program, rehabilitation process, or desire to show gratitude to those who have spent time and effort trying to help the patient [7]. The FFI questionnaire is an attractive measurement for physicians who wish to have a sensitive, reliable and responsive questionnaire for routine clinical practice [26].

This study shows that the FFI reaches better properties regarding the responsiveness, compared to the LFIS-RA

questionnaire. We could conclude that the use of the FFI is preferred for RA patients with the above-mentioned deformities. A possible explanation for the difference could be that the LFIS was validated in a study population lacking RA patients with severe deformities [15].

Conclusion

The FFI showed a large responsiveness and the LFIS- RA showed moderate responsiveness in RA patients receiving forefoot or hindfoot surgery, without floor or ceiling effects in both questionnaires.

Compliance with ethical standards

Conflict of interest There are no conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Louwerens JWK, Schrier JCM (2013) Rheumatoid forefoot deformity: pathophysiology, evaluation and operative treatment options. *Int Orthop* 37(9):1719–1729. doi:10.1007/s00264-013-2014-2
- Trieb K et al (2013) The Weil osteotomy for correction of the severe rheumatoid forefoot. *Int Orthop* 37(9):1795–1798. doi:10.1007/s00264-013-2011-5
- Van der Heide HJ, Louwerens JW (2010) Reconstructing the rheumatoid forefoot. *Foot Ankle Surg* 16(3):117–121. doi:10.1016/j.fas.2009.07.001
- Ohly NE, Breusch SJ (2013) Additive hindfoot arthrodesis for rheumatoid hindfoot disease: a clinical study of patient outcomes and satisfaction. *Clin Rheumatol* 32(12):1777–1782. doi:10.1007/s10067-013-2343-y
- Child BJ, Hix J, Catanzariti AR et al (2009) The effect of hindfoot realignment in triple arthrodesis. *J Foot Ankle Surg* 48(3):285–293. doi:10.1053/j.jfas.2009.02.006
- Graves SC, Mann RA, Graves KO (1993) Triple arthrodesis in older adults. results after long-term follow-up. *JBJS J Bone Joint Surg* 75(3):355–362
- Guyatt GH, Deyo RA, Charlson M et al (1989) Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 42(5):403–408. doi:10.1016/0895-4356(89)90128-5
- Husted JA, Cook RJ, Farewell VT et al (2000) Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 53(5):459–468. doi:10.1016/S0895-4356(99)00206-1
- Van der Leeden M, Steultjens MP, Terwee CB et al (2008) A systematic review of instruments measuring foot function, foot pain, and foot-related disability in patients with rheumatoid arthritis. *Arthritis Rheuma* 59(9):1257–1269. doi:10.1002/art.24016
- Budiman Mak E, Conrad KJ, Roach KE (1991) The foot function index: A measure of foot pain and disability. *J Clin Epidemiol* 44(6):561–570. doi:10.1016/0895-4356(91)90220-4
- Budiman Mak E, Conrad KJ, Mazza J et al (2013) A review of the foot function index and the foot function index—revised. *J Foot Ankle Res* 6(1):5. doi:10.1186/1757-1146-6-5
- SooHoo NF, Vyas R, Samimi D (2006) Responsiveness of the foot function index, aofas clinical rating systems, and sf-36 after foot and ankle surgery. *Foot Ankle Int* 27(11):930–934. doi:10.3113/fai.2012.0057
- Madeley NJ, Wing KJ, Topliss C et al (2012) Responsiveness and validity of the sf-36, ankle osteoarthritis scale, AOFAS ankle hindfoot score, and foot function index in end stage ankle arthritis. *Foot Ankle Int* 33(1):57–63. doi:10.3113/fai.2012.0057
- Woodburn J et al (2012) Adaptation and crosscultural validation of the foot impact scale for rheumatoid arthritis using Rasch analysis. *Arthritis Care Res* 64(7):986–992
- Helliwell P, Reay N, Gilworth G et al (2005) Development of a foot impact scale for rheumatoid arthritis. *Arthritis Rheuma* 53(3):418–422. doi:10.1002/art.21176
- Woodburn J, Vliet Vlieland TP, van der Leeden M et al (2011) Rasch analysis of dutch-translated version of the foot impact scale for rheumatoid arthritis. *Rheumatology* 50(7):1315–1319. doi:10.1093/rheumatology/ker003
- De Vet HC, Terwee CB, Bouter LM (2003) Current challenges in clinimetrics. *J Clin Epidemiol* 56(12):1137–1144. doi:10.1016/j.jclinepi.2003.08.012, doi: 0.1017/CBO9780511996214.001
- McGough J, Faraone SV (2009) Estimating the size of treatment effects: moving beyond p values. *Psychiatry* 6(10):21–29
- Deyo RA, Centor RM (1986) Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chron Dis* 39(11):897–906. doi:10.1016/0021-9681(86)90038-X
- Terwee CB, Bot SD, de Boer MR et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42. doi:10.1016/j.jclinepi.2006.03.012
- Deyo RA, Diehr P, Patrick DL (1991) Reproducibility and responsiveness of health status measures. statistics and strategies for evaluation. *Control Clin Trials* 12(4 Suppl):142–158S. doi:10.1016/S0197-2456(05)80019-4
- De Vet HC, Terwee CB, Mokkink LB, Knol DJ (2011) *Measurement in medicine. a practical guide*. Cambridge University Press, Cambridge. doi:10.1007/s11136-012-0123-9
- Hays RD, Hadorn D (1992) Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res* 1(1):73–75. doi:10.1007/BF00435438
- Schwartz CE, Andresen EM, Nosek MA et al (2007) Response shift theory: important implications for measuring quality of life in people with disability. *Arch Phys Med Rehabil* 88(4):529–536. doi:10.1016/j.apmr.2006.12.032
- Ross M, Ross (1989) Relation of implicit theories to the construction of personal histories. *Psychol Rev* 96(2):341–357. doi:10.1037/0033-295x.96.2.341
- Walmsley S, Williams AE, Ravey M, et al (2010) The rheumatoid foot: A systematic literature review of patient-reported outcome measures. *J Foot Ankle Res* 3:12