Injury 52 (2021) 1002–1010

Contents lists available at ScienceDirect

Injury

journal homepage: www.elsevier.com/locate/injury

Measurement properties of Patient-Reported Outcome Measures in patients with a tibial shaft fracture; validation study alongside the multicenter TRAVEL study

Mandala S. Leliveld, Michael H.J. Verhofstad, Esther M.M. Van Lieshout[1,*], on behalf of TRAVEL Study Investigators

Trauma Research Unit, Department of Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 Rotterdam, the Netherlands

A R T I C L E I N F O

Article history:
Accepted 27 December 2020

Keywords:
Tibial shaft
validation
SMFA
LEFS

A B S T R A C T

The aim of this study was to evaluate the measurement properties of the Short Musculoskeletal Function Assessment (SMFA) and Lower Extremity Functional Scale (LEFS) in patients who sustained a tibial shaft fracture, by comparing them with the scores of a general health-related quality of life instrument scale (i.e., EuroQoL-5D).

Data of 136 patients participating in a multicenter randomized controlled trial comparing incisions for intramedullary nail entry in adults with a tibial shaft fracture were used. Patients completed the SMFA, LEFS, EQ-5D and an anchor question at 2 and 6 weeks, and at 3, 6 and 12 months. Reliability (internal consistency), construct validity, responsiveness (longitudinal validity), floor and ceiling effects, minimal important change (MIC), and smallest detectable change (SDC) were determined.

The SMFA and LEFS (sub)scales showed adequate internal consistency ($0.84<\alpha<0.94$). Construct and longitudinal validity were also adequate (correctly predicted hypotheses between 83%-100%). Floor effects were not present. Ceiling effects were present at

12 months for the SMFA lower extremity dysfunction and bother subscales (22% and 19%, respectively) and the LEFS (19%). MICs could not be determined with the available data. The SDC was 13.84 points for the SMFA and 38.74 points for the LEFS.

This study confirms that the SMFA and LEFS are reliable, valid, and responsive instruments for monitoring functional limitation in patients after sustaining a tibia shaft fracture during at least the first six months post-injury. An anchor-based MIC for the SMFA remains to be determined.

Introduction

Compared with, e.g., hip fractures the incidence rate of lower leg fractures is low, but the average burden of these fractures is

* Corresponding author.
E-mail addresses: m.leliveld@erasmusmc.nl (M.S. Leliveld), m.verhofstad@erasmusmc.nl (M.H.J. Verhofstad), e.vanlieshout@erasmusmc.nl (E.M.M. Van Lieshout).

[1] Membership of the TRAVEL study Investigators:
F.W. Bloemers, Department of Surgery, Amsterdam UMC location VUmc, Amsterdam, The Netherlands
M.M.M. Bruijninckx, Department of Surgery, IJsselland Ziekenhuis, Capelle aan den IJssel, The Netherlands
M.R. De Vries, Department of Surgery, Reinier de Graaf Gasthuis, Delft, The Netherlands
S. Kleinveld, Department of Surgery, Haga Ziekenhuis, Den Haag, The Netherlands
C.M.E.M. Meijs, Department of Surgery, Ziekenhuis Rivierenland, Tiel, The Netherlands
B. Mirck, Department of Surgery, Noordwest Ziekenhuisgroep, The Netherlands
R.M.H.G. Mollen, Department of Surgery, Ziekenhuis Gelderse Vallei, Ede, The Netherlands
E.D. Ritchie, Department of Surgery, Alrijne Ziekenhuis, Leiderdorp, The Netherlands
J.B. Sintenie, Department of Surgery, Elkerliek Ziekenhuis, Helmond, The Netherlands
M.F. Termaat, Department of Surgery, Leids Universitair Medisch Centrum, Leiden, The Netherlands
A. Van Uden, Trauma Research Unit, Department of Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
C.H. Van der Vlies, Department of Surgery, Maasstad Ziekenhuis, Rotterdam, The Netherlands
J. Vermeulen, Department of Surgery, Spaarne Gasthuis, Haarlem, The Netherlands

https://doi.org/10.1016/j.injury.2020.12.030
0020-1383/

high for patients. This is due to the high proportion of patients with lifelong disabilities of some lower leg injuries in combination with the patients' relatively young age. [1,2]

In order to monitor outcome, functional recovery, and quality of life after lower-limb injury or (surgical) treatment, patient-reported outcome measures (PROMs) are becoming increasingly important instruments. For this purpose numerous disease-specific and region-specific PROMs are being used. [3–5]

The SMFA is a patient-reported questionnaire, designed to detect differences in functional status of patients who have a broad range of musculoskeletal disorders. In its original language version the SMFA is proven to be a valid, reliable, and responsive questionnaire. [3] Psychometric properties of the SMFA were tested in a variety of populations of patients: patients with ankle arthritis [6], hip/knee osteoarthritis [7], rheumatoid arthritis [7], severely injured patients (ISS>15) [8], and patients with various musculoskeletal disorders. [3,9,10] Van Son et al. translated and culturally adapted the Dutch version of the SMFA. [11] They additionally adapted some items in order to avoid double-barrelled items. [11]

Whereas the SMFA is developed for patients with any musculoskeletal disorder, the Lower Extremity Functional Scale (LEFS) is a region-specific PROM. The measurement properties of LEFS indicate that it is a reliable, valid, and responsive tool for assessing functional status in several populations with lower extremity musculoskeletal conditions. [12] Furthermore, the LEFS was found to be responsive in patients with total hip or knee replacement [13], general lower extremity dysfunction [14], osteoarthritis [15], and ankle fractures. [16] The LEFS was translated into Dutch from its original language by Hoogeboom et al. [15]

Patient reported outcome measures are useful for measuring the trajectory of functional recovery after sustaining a tibial shaft fracture. For this specific population of patients the measurement properties of the SMFA and LEFS are not fully known. The aim of this study was to evaluate the measurement properties of the SMFA and LEFS (sub)scores in patients that sustained a tibial shaft fracture by comparing them with the scores of a general health-related quality of life instrument scale (*i.e.*, EuroQoL-5D).

This study is registered at the Netherlands Trial Register (NTR5091). The study was approved by the Medical Research Ethics Committees Erasmus MC (Ref.No. MEC-2014-335 and NL49144.078.14) and Local Ethics Boards of all participating centers.

Materials and methods

Study data

All 136 patients who were included in a multicenter randomized controlled trial comparing two incisions for intramedullary nail entry in adult patients with a tibial shaft fracture were used.

Study population

Patients were recruited from September 6, 2015 until June 12, 2018. Patients aged between 16 years and 65 years presenting with a tibial shaft fracture (AO type 42) to the Emergency Department of one of 13 participating hospitals in the Netherlands were eligible for inclusion. Exclusion criteria were 1) polytraumatized patients; 2) concurring injury affecting treatment and recovery; 3) bilateral tibial fractures; 4) pathological or recurrent fracture of the tibia; 5) Gustilo Anderson type IIIC open fractures or open wound on knee; 6) pre-existing knee pathology (e.g., menisci or cruciate ligament); 7) pre-existing functional impairment influencing rehabilitation (e.g., wheelchair-bound); 8) rheumatoid arthritis; 9) bone disease resulting in delayed union (except osteoporosis); 10) problems ensuring follow-up (e.g., no fixed address or cognitive impairment); and 11) insufficient comprehension of the Dutch language.

Questionnaires

Patients were asked to complete Dutch versions of the SMFA [11], LEFS [15], and EuroQoL-5D (EQ-5D-3L) [17] questionnaires at two and six weeks, and at three, six, and 12 months after tibial nailing.

The Dutch SMFA consists of 53 items, each scored using a 5 points Likert scale. It is divided in three subscales: a 23-item Dysfunction lower extremity (LE) scale, a 9-item Dysfunction upper extremity (UE) scale and a 15-item Bother scale. The SMFA score is calculated for each subscale using the formula: ((sum of all items/number of items)-1)*maximum score). The overall score as well as the subscale scores range from 0 to 100 points. Higher scores refer to greater disability.

The LEFS is developed to measure function in patients with a wide range of lower-extremity orthopedic conditions [4]. It is a self-reported measure and consists of 20 items, each with a maximum score of 4. The total possible score of 80 indicates a high functional level. The total score is calculated using the formula: ((sum of all items/number of items)/80).

The EQ-5D-3L is a validated instrument for measuring health-related quality of life [5]. The EQ-5D utility score (EQ-US) ranges from 0 to 1 and is determined from five 1-item domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each item has three possible answers. In addition, the individual's rating of his/her quality of life state is recorded by means of a standard Visual Analog Scale (EQ-VAS), which ranges from 0 to 100. Higher scores represent better health-related quality of life.

Baseline characteristics collected were age, gender, Body Mass Index (BMI), American Society of Anesthesiologists (ASA) classification, and smoking. Fracture details collected were affected side, AO classification, and presence of a fibula fracture and additional injuries.

Statistical analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 25 (SPSS, Chicago, Ill., USA). The Receiver Operating Characteristic (ROC) curve and Youden Index were analyzed using MedCalc version 14.10.2 (MedCalc Software, Ostend, Belgium). Measurement properties of the (sub)scales were determined in compliance with the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines. [18] Data are reported following the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE). [19] Since raw data for individual items were analyzed, missing responses to the questionnaires were not imputed.

Descriptive statistics were used in order to describe the main characteristics of the study participants. Measurement properties of the SMFA, LEFS, and EQ-5D (sub)scores were determined by comparing them with those of the general health-related quality of life instrument EQ-5D.

Reliability was determined by evaluating internal consistency. Internal consistency is a measure of the extent to which items in a (sub)scale are correlated (homogeneous), thus measuring the same concept. [20] For each (sub)scale, correlation between the items was calculated using Cronbach's alpha. Internal consistency can be considered sufficient if the Cronbach's alpha value is between 0.70 and 0.95, provided that the scale is unidimensional. [20] This analysis requires a sample size of 10 per item in the instrument, with a minimum of 100 patients. [20] The six week data were used, since the largest heterogeneity in the degree of recovery and consequently the largest variability in scores were expected at that time.

Validity is the degree to which a questionnaire measures the construct it is supposed to measure. As there was no gold standard in the current study, the validity of the instruments was expressed in terms of the construct validity. Construct validity represents the extent to which scores on a specific questionnaire relate to other measures in a way that is in agreement with prior theoretically derived hypotheses concerning the concepts that are being measured. [20] The six weeks data were used. Continuous data were tested for normality using the Shapiro-Wilk test. As not all of the (sub)scales were normally distributed, Spearman rank correlations between the (sub)scales of the SMFA, LEFS, and EQ-5D were determined in order to assess the construct validity. Strengths of correlation were categorized as high ($r>0.6$), moderate ($0.3<r<0.6$), or low ($r<0.3$). [21] Construct validity was considered sufficient if at least 75% of the results were in accordance with predefined hypotheses in a (sub)sample of at least 50 patients. [20] The hypothesized correlations between the (sub)scale scores are shown in Supplemental Table S1A and were made in consensus between two authors (MSL and EMMVL).

Responsiveness refers to the ability of a questionnaire to detect clinically important changes over time. [20] In addition, the effect size (ES) and standardized response mean (SRM) were determined as measures of the magnitude of change over time.

Longitudinal validity can be considered as a measure of responsiveness. Longitudinal validity refers to the extent to which change in one measurement instrument relates to corresponding change in a reference measure. [22] Analogous to construct validity, longitudinal validity was assessed by testing predefined hypotheses about expected correlations between changes in SMFA, LEFS, and EQ-5D (sub)scale scores. Change scores were calculated as the difference in score from six weeks follow-up to the final measurement at 12 months. Normality was tested using the Shapiro-Wilk test. Since all change scores deviated from a Normal deviation, Spearman correlation coefficients were calculated. Longitudinal validity was considered sufficient if at least 75% of the results were in accordance with predefined hypotheses in a (sub)sample of at least 50 patients. [20] The hypothesized correlations are shown in Supplemental Table S1B and were made in consensus between two authors (MSL and EMMVL).

The Effect Size (ES) and Standardized Response Mean (SRM) were determined as measures of the magnitude over time. The ES was calculated as the mean change in score between two time points (*i.e.*, score at 12 months – score at six weeks) divided by the standard deviation of the first measurement [23]. The SRM was calculated as the mean change in score between two time points (*i.e.*, score at 12 months – score at six weeks) divided by the standard deviation of this change. [23] A value of 0.2-0.4 is considered a small, 0.5-0.7 a moderate, and $\geq$ 0.8 a large effect. [21] Large effect sizes were expected a priori, since at six weeks most patients were expected to have functional limitations, whereas at 12 months large improvement or even full recovery was expected for most patients.

Floor and ceiling effects are present if more than 15% of the study population rates the worst (floor effect) or best (ceiling effect) possible score on any questionnaire (sub)scale. [24] In the presence of floor and ceiling effects, items might be missing from the upper or lower ends of the scale, reducing content validity. Likewise, patients with the highest or lowest scores cannot be distinguished from one another, indicating limited reliability. [20] Floor and ceiling effect were determined for each follow-up moment separately.

The Minimal Important Change (MIC) is defined as the smallest measurable change in outcome score that is perceived as significant by patients. [25] An anchor-based method was used as this gives a better indication of the importance of the observed change to the patient. [20] In addition to the questionnaires, patients were asked to complete an transition item (anchor question) at six weeks and at three, six, and 12 months evaluating their perception of change in the general condition of the affected leg. The question was: 'How would you judge the condition of your affected leg at this point, compared with the last time you completed this questionnaire?' The item scored from 1 'much better' through 2 'slightly better', 3 'no change', 4 'slightly worse', or 5 'much worse'.

The anchor or transition item was judged as adequate if a Spearman's rank correlation between the anchor and the change score of the questionnaire was $>$ 0.29. [26] The corresponding change score (*i.e.*, score at time of completion of the transition item minus the score at the previous follow-up moment) of patients who answered the transition item as 'slightly better' can be considered the MIC. [27]

As an alternative, the MIC was also calculated for the (sub)scores by plotting the Receiver Operating Characteristics (ROC) curve of the change in score for patients who scored 'slightly better' on the transition item versus patients who scored 'no change'. The area under the ROC curve is provided as measure of discriminatory power. The ROC cutoff point (*i.e.*, the associated criterion of the Youden index) reflects the MIC. This MIC is shown with its 95% confidence interval (CI) after bootstrapping (1000 replicates and 900 random-number seeds).

In addition to the MIC, the Smallest Detectable Change (SDC) was determined. SDC is defined as the smallest intra-personal change in score that represents (with $p<0.05$) a 'real' difference above measurement error. [28] As patients need to be assumed to be stable in the interim period, this was based on the change scores of patients who answered 'no change' on the transition item. First, the standard error of measurement (SEM) was calculated by dividing the standard deviation of the mean difference between both measurements (SD_{change}) by the square root of two [29]. SEM can be considered as a measure of absolute measurement error [20]. For the individual patient, the SDC was calculated as 1.96 x square root of 2 x SEM (herein, SEM = SD_{change} / square root of 2). [20] Ideally, for evaluative purposes, the SDC should be smaller than the MIC. [20]

Results

Patient demographics

A total of 136 patients was enrolled. They had a median age of 35 (P_{25}-P_{75} 24-53) years and a BMI of 24.5 (P_{25}-P_{75} 21.8-26.9) kg/m^2. The majority of the patients was male (n=99; 73%) and had an American Society of Anesthesiologists (ASA) classification of I (*n*=108; 81%). A total of 40 (29%) patients smoked at the age of trauma. The right side was affected in 74 (54%) patients. Most fractures were closed (*n*=109; 80%), AO type 42A (*n*=92; 68%) and also had a fibula fracture (*n*=114; 89%). A minority of 27 (20%) patients had additional injuries (not affecting their recovery from the sustained tibial shaft fracture).

Changes in outcomes scores over time

During the one-year follow-up, all outcome scores consistently improved over time. The median SMFA total score decreased from 43 (P_{25}-P_{75} 33-51) at two weeks to 6 (P_{25}-P_{75} 1-14) at 12 months (Fig. 1). Decreases were noted for both the SMFA Dysfunction lower extremity subscale (from 63 (P_{25}-P_{75} 19-74)) to 4 (P_{25}-P_{75} 1-14) and the SMFA Bother subscale (from 32 (P_{25}-P_{75} 23-47)) to 8 (P_{25}-P_{75} 2-7)). The SMFA Dysfunction upper extremity subscale showed no change in outcome scores over time. The LEFS score increased from 26 (P_{25}-P_{75} 16-39) at two weeks to 89 (P_{25}-P_{75} 74-98) at 12

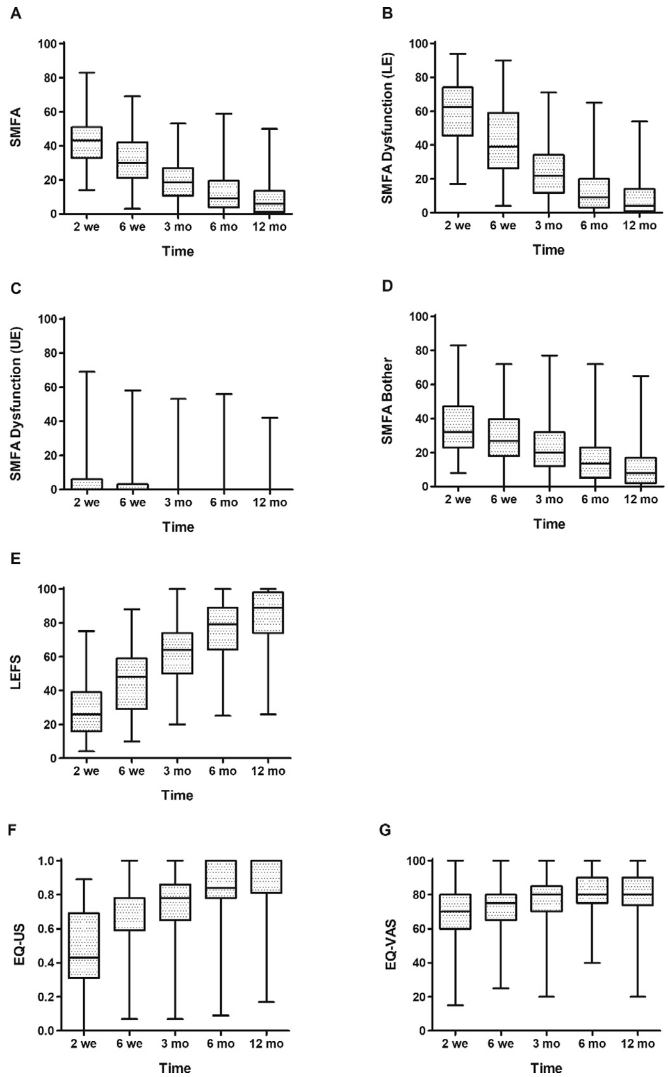

Fig. 1. SMFA (A-D), LEFS (E), and EQ-5D (F-G) (sub)scales at each follow-up visit in patients with a tibial shaft fracture.

months. Finally, the health-related quality of life improved from 2 weeks to 12 months, with EQ-US scores increasing from 0.43 (P_{25}-P_{75} 0.31-0.69) to 0.93 (P_{25}-P_{75} 0.81-0.93), EQ-VAS score increasing from 70 (P_{25}-P_{75} 60-80) to 80 (P_{25}-P_{75} 74-90).

Table 1
Internal consistency of the instruments in patients with a tibial shaft fracture.

(Sub)scale	*N*	N items	Cronbach's alpha
SMFA (total)	99	53	0.94[a]
Lower extremity dysfunction	112	23	0.94
Upper extremity dysfunction	112	9	0.84
Bother	108	15	0.90
LEFS (total)	113	20	0.93
EQ-5D EQ-US	116	5	0.66
EQ-VAS	115	1	N.D.[b]

EQ-5D, EuroQoL-5D; EQ-US, EuroQoL-5D Utility Score; EQ-VAS, EuroQoL-5D Visual Analog Scale; LEFS, Lower Extremity Functional Scale; N.D., not determined; SMFA, Short Musculoskeletal Functional Assessment.
Data are shown for the six weeks follow-up. The maximum number of patients was 136.
[a] Value should be interpreted carefully, since the total scale is not unidimensional.
[b] The EQ-VAS consists of a single item. Internal consistency does not apply to a single-item domain.

Reliability

The internal consistency for the SMFA and all subscales was adequate (Cronbach's alpha between 0.84 and 0.94). Internal consistency was also adequate for the LEFS (alpha = 0.93) but inadequate for the EQ-5D (alpha = 0.66). Cronbach's alpha could not be determined for the EQ-5D-VAS because internal consistency does not apply to a single-item domain.

Construct validity

Construct validity is shown in Table 2. Construct validity was adequate for all questionnaires. The Spearman's rank correlation coefficients of the SMFA were consistent with 16 of the 18 (89%) theoretically derived hypotheses. The hypotheses of the SMFA lower and upper extremity dysfunction scales were confirmed in 5 of the 6 (83%) values. All correlations were correctly predicted for the SMFA bother scale (6/6), as were the correlations of the LEFS (6/6). For the EQ-US and EQ-VAS 5 out of 6 (83%) correlations were as hypothesized.

Responsiveness

Longitudinal validity

Spearman's rank correlation coefficients for longitudinal validity are shown in Table 3. For the SMFA the correlations were in line with the predefined hypotheses in 16 out of the 18 (89%) values, indicating sufficient longitudinal validity. For all three SMFA subscales 83% (5/6) hypotheses were predicted correctly. Longitudinal validity was also sufficient for the LEFS, EQ-US and EQ-VAS with

Table 2
Construct validity of the instruments in patients with a tibial shaft fracture.

(Sub)scale	SMFA				LEFS	EQ-5D	
	Total	UE dysfunction	LE dysfunction	Bother	Total	EQ-US	EQ-VAS
SMFA (total)	1	0.95 [99]	0.61 [99]	0.82 [99]	−0.89 [98]	−0.64 [96]	−0.59 [95]
LE dysfunction	0.95 [99]	1	0.48 [112]	0.64 [108]	-0.89 [111]	−0.56 [108]	−0.53 [107]
UE dysfunction	0.61 [99]	0.48 [112]	1	0.43[108]	−0.49 [111]	−0.37 [108]	−0.27 [107]
Bother	0.82 [99]	0.64 [108]	0.43 [108]	1	−0.64 [107]	−0.67 [105]	−0.54 [103]
LEFS (total)	−0.89 [98]	−0.89 [111]	−0.49 [111]	−0.64 [107]	1	0.59 [109]	0.43 [108]
EQ-5D EQ-US	−0.64 [96]	−0.56 [108]	−0.37 [108]	−0.67 [105]	0.59 [109]	1	0.48 [111]
EQ-VAS	−0.59 [95]	−0.53 [107]	−0.27 [107]	-0.54 [103]	0.43 [108]	0.48 [111]	1

Data are shown for the six weeks follow-up. The maximum number of patients was 136. Construct validity is shown as Spearman's rank correlation coefficients (r) with the number of patients included in the correlation between square brackets. r > 0.6 represents high correlation, 0.3 < r < 0.6 moderate correlation, and r < 0.3 low correlation. Bold and underlined correlations were not hypothesized correctly.
EQ-5D, EuroQoL-5D; EQ-USUS, EuroQoL-5D Utility Score; EQ-VAS, EuroQoL-5D Visual Analog Scale; LE, Lower Extremity; LEFS, Lower Extremity Functional Scale; N.D., not determined; SMFA, Short Musculoskeletal Functional Assessment; UE, Upper Extremity.

Table 3
Longitudinal validity of the instruments in patients with a tibial shaft fracture.

(Sub)scale	SMFA				LEFS	EQ-5D	
	Total	UE dysfunction	LE dysfunction	Bother	Total	EQ-US	EQ-VAS
SMFA (total)	1	0.93 [88]	0.302 [88]	0.65 [88]	−0.77 [87]	−0.30 [85]	−0.15 [81]
LE dysfunction	0.93 [88]	1	0.25 [103]	**0.44 [97]**	−0.78 [102]	−0.30 [99]	−0.15 [94]
UE dysfunction	0.302 [88]	0.25 [103]	1	0.19 [97]	-0.12 [102]	**−0.21 [99]**	0.05 [94]
Bother	0.65 [88]	**0.44 [97]**	0.19 [97]	1	−0.47 [96]	−0.34 [94]	−0.26 [89]
LEFS (total)	−0.77 [87]	−0.78 [102]	−0.12 [102]	−0.47 [96]	1	0.35 [100]	0.24 [95]
EQ-5D EQ-US	−0.30 [85]	−0.30 [99]	**−0.21 [99]**	−0.34 [94]	0.35 [100]	1	0.19 [98]
EQ-VAS	−0.15 [81]	−0.15 [94]	0.05 [94]	−0.26 [89]	0.24 [95]	0.19 [98]	1

Responsiveness is shown as Spearman's rank correlation coefficients (r) of change in scores between six weeks and 12 months with the number of patients included in the correlation between square brackets. The maximum number of patients was 136.
$r > 0.6$ indicates high correlation, $0.3 < r < 0.6$ moderate correlation, and $r < 0.3$ low correlation. Bold and underlined correlations were not hypothesized correctly.
EQ-5D, EuroQoL-5D; EQ-USUS, EuroQoL-5D Utility Score; EQ-VAS, EuroQoL-5D Visual Analog Scale; LE, Lower Extremity; LEFS, Lower Extremity Functional Scale; N.D., not determined; SMFA, Short Musculoskeletal Functional Assessment; UE, Upper Extremity.

Table 4
Responsiveness: standardized response mean (SRM) and Effect Size (ES) of the instruments in patients with a tibial shaft fracture.

(Sub)scale	*N*	Mean change	SD_{change}	SRM	$SD_{6\ weeks}$	ES	
SMFA (total)	88	−23.54	10.45	−2.25	14.89	−1.58	Large
LE dysfunction	103	−33.28	16.88	−1.97	19.83	−1.68	Large
UE dysfunction	103	-2.31	7.31	-0.32	9.12	−0.25	Small
Bother	97	−16.90	11.69	−1.45	16.89	−1.00	Large
LEFS (total)	104	38.25	17.55	2.18	19.00	2.01	Large
EQ-5D US	107	0.20	0.20	1.04	0.22	0.93	Large
VAS	102	6.38	15.51	0.41	14.03	0.45	Small

Change scores were calculated from six weeks to 12 months. The maximum number of patients was 136.
EQ-5D, EuroQoL-5D; EQ-US, EQ-5D Utility Score; EQ-VAS, EQ-5D Visual Analog Scale; LE, Lower Extremity; LEFS, Lower Extremity Functional Scale; N.D., not determined; SD, Standard deviation of mean change; SRM, standardized response mean; SMFA, Short Musculoskeletal Functional Assessment; UE, Upper Extremity.

100% (6/6), 83% (5/6) and 100% (6/6) correlations as hypothesized, respectively.

Standardized Response Mean (SRM) and Effect Size (ES)

The SRM and the ES of the SMFA, LEFS, and EQ-5D and their subscales are reported in Table 4. For the SMFA and the lower extremity dysfunction and bother subscales, the magnitude of change over time was large (SRM between -2,25 and -1.45; ES between -1.58 and -1.00). For the SMFA upper extremity dysfunction subscale the magnitude of change over time was small (SRM -0.32; ES -0.25). The magnitude of the change of the LEFS was large with a SRM of 2.18 and an ES of 2.01. The EQ-US and EQ-VAS showed a large (SRM 1.04; ES 0.93) and small (SRM 0.41; ES 0.45) magnitude of change over time, respectively.

Floor and ceiling effects

Floor and ceiling effects for all instruments are shown in Fig. 2. None of the questionnaires and their subscales showed a floor effect at any point in time. All the measures can therefore be used to accurately interpret the change in status in patients with poor function and health-related quality of life. The SMFA total score showed no ceiling effect. However, for the SMFA lower extremity dysfunction and bother subscales 22% and 19% of all patients, respectively, reported the best score at 12 months follow-up. The SMFA upper extremity dysfunction subscale shows a ceiling effect at each follow-up, increasing from 51% at two weeks to 90% at 12 months follow-up. A ceiling effect of the LEFS is present at 12 months (19%). The EQ-VAS has no ceiling effect, but a ceiling effect of the EQ-US is seen from 3 months (15%) onwards.

Minimal Important Change (MIC) and Smallest Detectable Change (SDC)

The number of patients per transition item for the different time intervals is shown in Supplemental Table S2. Anchor-based MIC and distribution-based SDC values are shown in Table 5. Overall, 146 transition items were reported as 'slightly better' and 103 as 'no change'. None of the transition items were judged as adequate since the Spearman's rank correlation between the anchor and the change scores of the questionnaires were all < 0.3. Therefore the MICs for the evaluated instruments are potentially unreliable and should be interpreted with care (*r* between -0.14 [EQ-US] and 0.24 [SMFA total]; data not shown).

The SDC was 13.84 points (SEM 4.99) for the SMFA total score and 38.74 points (SEM 13.97) for the LEFS, 0.51 (SEM 0.19) for the EQ-US and 22.99 (SEM 8.29) for the EQ-VAS.

Discussion

Clinimetric properties of the SMFA and LEFS have previously been tested in a heterogeneous population of patients with lower-limb conditions and some homogeneous populations of patients with lower-limb conditions. For evaluation of patients with tibial shaft fractures, measurement properties of these questionnaires were not fully known. Data of the current study confirm that the SMFA and LEFS are reliable, valid, and responsive in the study population. The MIC of these questionnaires should be interpreted with care. Both questionnaires are useful for monitoring patients functional limitation after sustaining a tibia shaft fracture for at least the first six months.

Reliability of the SMFA and LEFS was supported by adequate internal consistency. The Cronbach's alpha values of the SMFA and its subscales are comparable with published values of the original

Table 5
Minimal important change and smallest detectable change values of the instruments in patients with a tibial shaft fracture.

(Sub)scale	Scoring	Anchor-based approach						Distribution-based approach			
	Range	N	MIC[A]	AUC	MIC[B]	Sensitivity (%)	Specificity (%)	N	SD_{change}	SEM	SDC
SMFA (total)	0;100	116	8.12 [6.36;9.88]	0.55 [0.47;0.62]	0.92 [-2.89;3.77]	85.3	25.7	74	7.06	4.99	13.84
LE dysfunction	0;100	127	11.62 [9.13;14.10]	0.54 [0.47;0.61]	7.6 [1.09;29.36]	59.8	53.5	86	12.00	8.48	23.52
UE dysfunction	0;100	127	1.29 [0.23;2.35]	0.53 [0.47;0.60]	2.8 [-2.7;11.1]	83.5	22.1	86	7.23	5.11	14.16
Bother	0;100	124	6.08 [4.29;7.87]	0.53 [0.46;0.60]	10.0 [0.3;20.0]	39.5	74.7	83	2.87	6.02	16.68
LEFS (total)	0;100	127	7.32 [3.65;10.99]	0.54 [0.47;0.61]	12.5 [-13.8;31.3]	38.6	76.1	88	19.76	13.97	38.74
EQ-5D US	-0.329;1.000	122	0.09 [0.05;0.14]	0.55 [0.48;0.62]	0.1 [-0.3;0.1]	47.5	65.5	87	0.26	0.19	0.51
VAS	0;100	128	3.64 [1.60;5.68]	0.55 [0.48;0.62]	0.0 [-14.9;10.0]	55.5	57.3	89	11.73	8.29	22.99

Anchor-based and distribution-based methods for Minimal Important Change (MIC) and Smallest Detectable Change (SDC) values, respectively. The MIC and the Area under the Receiver Operating Curve (AUC) are shown with 95% confidence intervals between brackets.
MIC is calculated as mean change score for all respondents reporting "slightly better" on the transition question (MIC [A]) and as criterion in the Receiver Operating Curve (MIC [B]).
AUC, Area under the Receiver Operating Curve; EQ-5D, EuroQoL-5D Assessment; EQ-US, EuroQoL-5D Utility Score; EQ-VAS, EuroQoL-5D Visual Analog Scale; LE, Lower Extremity; LEFS, Lower Extremity Functional Scale; MIC, Minimal Important Change; N.D., not determined; SD_{change}, Standard deviation of mean change; SDC, Smallest Detectable Change; SEM, standard error of measurement; SMFA, Short Musculoskeletal Functional; UE, Upper Extremity.

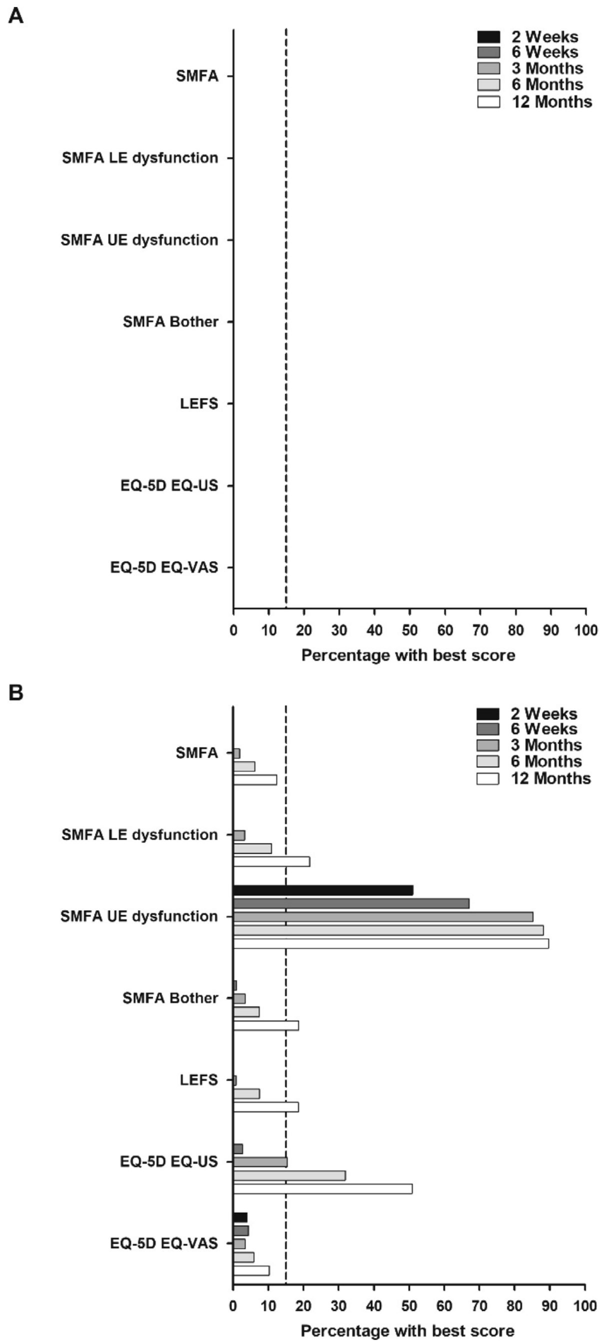

Fig. 2. Floor (A) and ceiling (B) effects of the instruments at each follow-up visit in patients with a tibial shaft fracture.

language version and translated versions. [3,7, 9–11, 30] The Cronbach's alpha value for the LEFS was also high (0.93) and in concordance with literature. [12]

The SMFA and LEFS showed adequate construct validity. The SMFA total score and SMFA bother scale were highly correlated with the EQ-US (*r* -0.64 and *r* -0.67, respectively). The SMFA and its subscales showed a moderate to low correlation with the EQ-VAS ($-0.27 < r < 0.54$). These observations are new and suggest that sustaining a tibial shaft fracture does not necessarily affect a patient's general health perception.

The LEFS scores correlated highly with the SMFA total score (*r* -0.89). This confirms the correlation between these two instruments found by Pinsker et al. in a population of pre- and postoperative ankle arthrodesis and arthroplasty patients. [6] The moderate level of correlation between the LEFS and EQ-US (*r* 0.59) differs from a published study in patients with an ankle fracture (*r* 0.73). [31] In the latter study questionnaires were sent three years after surgery, which may explain the higher correlation between the LEFS and the EQ-US.

Longitudinal validity was adequate for the SMFA total score and LEFS as reflected by the percentage of correctly predicted hypotheses (both 100%). The correlations of the LEFS with the SMFA subscales indicate that changes in lower extremity function have a moderate effect on the bother scale (*r* -0.47) and little effect on upper extremity dysfunction (*r* -0.12). The latter correlation has also been found by De Graaf et al. [32]

The large SRM and ES for the SMFA and LEFS indicate that both instruments can excellently detect clinical change over time. The large SRM value for the SMFA lower extremity dysfunction scale (-1.97) is similar to values found by Busse et al. in tibial shaft fractures [33], tibial plateau fractures [34], and ankle fractures. [10] The SRM of the SMFA bother scale was also comparable to values found in patients who sustained an ankle fracture. [10] In groups with various musculoskeletal disorders and nonoperatively treated patients the responsiveness of the SMFA is much lower. [35] The excellent ability of the LEFS to detect change over time has been shown before in patients with various lower limb injuries. [12] The EQ-US and EQ-VAS showed a moderate and low ES, respectively. This implies that sustaining a tibial shaft fracture has only limited effect on health-related quality over time.

None of the questionnaires showed a floor effect, which is in line with other studies. [12,35] The SMFA total score showed no ceiling effect, however, the SMFA lower extremity dysfunction and bother scale did so at 12 months. The presence of ceiling effects can be expected when instruments are used in time points when participants have much disability and improve over time after an injury or treatment. Treatment effects in this patient population can therefore be measured with the SMFA within the first 6 months after surgery, but effects can be missed later in the follow-up. [10,33,34] Likewise, the LEFS can detect functional improvement during the six months post-surgery (ceiling effect reached at 12 months (19%)). The latter finding is new, since earlier studies on the measurement properties of the LEFS had a follow-up period of 26 weeks or less. [15,16,36] Whereas a ceiling effect of the EQ-US was present at three months and onwards, the EQ-VAS did not have a ceiling effect at all in the current study. The EQ-VAS can thus be used to measure health-related quality of life on the long term in patients that sustained a tibial shaft fracture.

The SDC is a measure of the variation in a scale due to measurement error. Thus, a change score can only be considered to represent a real change if it is larger than the SDC. The SDC values of the SMFA total scale have been reported in three studies. [6, 9,32] In patients with stable ankle arthritis the SDC was 9.60 [6] which is lower than the SDC found in the current study (SDC 13.84). In addition to the stable condition of the patients, the average follow-up time ranged from 7 months to 2.5 years after surgery, which may explain the lower SDC. The SDC's reported by Reininga et al. and De Graaf et al. were higher, but the difference in factor structure of the SMFA hampers a comparison with our results [9,32]. The reported SDC values of the LEFS show much variation (ranging from 2.18 for Spanish LEFS scores in patients diagnosed with lower extremity musculoskeletal conditions [37] to 18.1 in patients with knee osteoarthritis [38], as displayed in the systematic review by Mehta et al. [12] All values are much lower than the SDC found in the current study (SDC 38.74). Explanations for this difference include more heterogeneity of the patient population [36,37], shorter follow-up time [36,37,39], and smaller sample size. [40]

The most important clinimetric property for interpreting change over time is the minimum important change (MIC). For the SMFA no studies have evaluated this measurement property using an anchor-based method. Due to the inadequate correlation between the transition item "slightly better" and change scores, the true MIC remains undetermined. The sample size of 136 patients in this study should have been adequate to determine the MIC. In addition, there were enough (n=146) transition items reported as "slightly better". The time between 2 subsequent follow-up moments varied from 4 weeks to 6 months. The drawback of the anchor-based approach is that it is based on retrospective judgement of change and is susceptible to recall biases. [41]

This study shows that when taking into account the SDC and MIC, the change score should exceed 13.8 points for the SMFA (total score) and 38.7 points for the LEFS to have a clinically relevant change on the questionnaire in patients that sustained a tibial shaft fracture. The MIC for the SMFA remains to be determined. The known anchor-based MIC value of the LEFS (9 [4,42]) only exceeds the SDC found in two studies (2.18 points for the Spanish LEFS [37] and 8.0 points in patients with anterior knee pain [40]. If the SDC is smaller than the MIC, it is possible to distinguish a clinically important change from measurement error with a large amount of certainty. However, this is much more difficult if the SDC is larger than the MIC, since there is a considerable chance that the observed change is caused by measurement error.

This study has some limitations. The relatively long interval between two subsequent follow-up moments (four weeks to six months) hindered an adequate test-retest analysis. It may also have led to recall bias with regard to the transition item and thus affected correct anchor-based MIC and SDC calculations. For future studies that aim to determine the MIC for the SMFA in patients that sustained a tibial shaft fracture we recommend to shorten the intervals between the questionnaires (in combination with the anchor-based questions).

In order to monitor outcome, functional recovery, and quality of life after lower-limb injury or (surgical) treatment, PROMs are becoming increasingly important instruments. By using these instruments different treatment strategies and their outcome can be compared. The SMFA and LEFS are useful instruments for monitoring functional limitation in patients after sustaining a tibia shaft fracture during at least the first 6 months post-injury. Both instruments are reliable, valid, and responsive. The MIC could not be determined reliably.

Declaration of Competing Interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.injury.2020.12.030.

References

[1] Weijermars W, Bos N, Stipdonk H. Health burden of serious road injuries in the Netherlands. Traffic Inj Prev 2016;17(8):863–9.
[2] Leliveld MS, Polinder S, Panneman MJM, Verhofstad MHJ, Van Lieshout EMM. Epidemiologic trends for isolated tibia shaft fracture admissions in The Netherlands between 1991 and 2012. Eur J Trauma Emerg Surg 2019.
[3] Swiontkowski MF, Engelberg R, Martin DP, Agel J. Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. J Bone Joint Surg Am 1999;81(9):1245–60.
[4] Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther 1999;79(4):371–83.
[5] Brooks R. EuroQol: the current state of play. Health Policy 1996;37(1):53–72.
[6] Pinsker E, Inrig T, Daniels TR, Warmington K, Beaton DE. Reliability and validity of 6 measures of pain, function, and disability for ankle arthroplasty and arthrodesis. Foot Ankle Int 2015;36(6):617–25.
[7] Wollmerstedt N, Kirschner S, Faller H, König A. Reliability, validity and responsiveness of the German Short Musculoskeletal Function Assessment Questionnaire in patients undergoing surgical or conservative inpatient treatment. Qual Life Res 2006;15(7):1233–41.
[8] van Delft-Schreurs CC, van Son MA, de Jongh MA, Gosens T, Verhofstad MH, de Vries J. Psychometric properties of the Dutch Short Musculoskeletal Function Assessment (SMFA) questionnaire in severely injured patients. Injury 2016;47(9):2034–40.
[9] Reininga IH, el Moumni M, Bulstra SK, Olthof MG, Wendt KW, Stevens M. Cross-cultural adaptation of the Dutch Short Musculoskeletal Function Assessment questionnaire (SMFA-NL): internal consistency, validity, repeatability and responsiveness. Injury 2012;43(6):726–33.
[10] Ponzer S, Skoog A, Bergström G. The Short Musculoskeletal Function Assessment Questionnaire (SMFA): cross-cultural adaptation, validity, reliability and responsiveness of the Swedish SMFA (SMFA-Swe). Acta Orthop Scand 2003;74(6):756–63.
[11] Van Son MA, Den Oudsten BL, Roukema JA, Gosens T, Verhofstad MH, De Vries J. Psychometric properties of the Dutch Short Musculoskeletal Function Assessment (SMFA) questionnaire in patients with a fracture of the upper or lower extremity. Qual Life Res 2014;23(3):917–26.
[12] Mehta SP, Fulton A, Quach C, Thistle M, Toledo C, Evans NA. Measurement Properties of the Lower Extremity Functional Scale: A Systematic Review. J Orthop Sports Phys Ther 2016;46(3):200–16.
[13] Yeung TS, Wessel J, Stratford P, Macdermid J. Reliability, validity, and responsiveness of the lower extremity functional scale for inpatients of an orthopaedic rehabilitation ward. J Orthop Sports Phys Ther 2009;39(6):468–77.
[14] Cacchio A, De Blasis E, Necozione S, Rosa F, Riddle DL, di Orio F, et al. The Italian version of the lower extremity functional scale was reliable, valid, and responsive. J Clin Epidemiol 2010;63(5):550–7.
[15] Hoogeboom TJ, de Bie RA, den Broeder AA, van den Ende CH. The Dutch Lower Extremity Functional Scale was highly reliable, valid and responsive in individuals with hip/knee osteoarthritis: a validation study. BMC Musculoskelet Disord 2012;13:117.
[16] Lin CW, Moseley AM, Refshauge KM, Bundy AC. The lower extremity functional scale has good clinimetric properties in people with ankle fracture. Phys Ther 2009;89(6):580–8.
[17] Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. Ned Tijdschr Geneeskd 2005;149(28):1574–8.
[18] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19(4):539–49.
[19] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Epidemiology 2007;18(6):800–4.
[20] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34–42.
[21] Cohen J. Statistical power analysis for the behavioral sciences. New York Academic Press; 1977. p. 474.
[22] 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:e15.
[23] Angst F, Verra ML, Lehmann S, Aeschlimann A. Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. BMC Med Res Methodol 2008;8:26.
[24] McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4(4):293–307.
[25] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10(4):407–15.
[26] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61(2):102–9.
[27] De Vet H, Terwee C, Mokkink L, Knol D. Measurement in Medicine, a Practical Guide. Cambridge University Press; 2011.
[28] 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):792–6.
[29] de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59(10):1033–9.
[30] Taylor MK, Pietrobon R, Menezes A, Olson SA, Pan D, Bathia N, et al. Cross-cultural adaptation and validation of the Brazilian Portuguese version of the short musculoskeletal function assessment questionnaire: the SMFA-BR. J Bone Joint Surg Am 2005;87(4):788–94.
[31] Garratt AM, Naumann MG, Sigurdsen U, Utvåg SE, Stavem K. Evaluation of three patient reported outcome measures following operative fixation of closed ankle fractures. BMC Musculoskelet Disord 2018;19(1):134.
[32] de Graaf MW, Reininga IH, Wendt KW, Heineman E, El Moumni M. The Short Musculoskeletal Function Assessment: a study of the reliability, construct validity and responsiveness in patients sustaining trauma. Clin Rehabil 2019;33(5):923–35.
[33] Busse JW, Bhandari M, Guyatt GH, Heels-Ansdell D, Mandel S, Sanders D,

et al. Use of both Short Musculoskeletal Function Assessment questionnaire and Short Form-36 among tibial-fracture patients was redundant. J Clin Epidemiol. 2009;62(11):1210–17.
[34] Dattani R, Slobogean GP, O'Brien PJ, Broekhuyse HM, Blachut PA, Guy P, et al. Psychometric analysis of measuring functional outcomes in tibial plateau fractures using the Short Form 36 (SF-36), Short Musculoskeletal Function Assessment (SMFA) and the Western Ontario McMaster Osteoarthritis (WOMAC) questionnaires. Injury 2013;44(6):825–9.
[35] Bouffard J, Bertrand-Charette M, Roy JS. Psychometric properties of the Musculoskeletal Function Assessment and the Short Musculoskeletal Function Assessment: a systematic review. Clin Rehabil 2016;30(4):393–409.
[36] Pan SL, Liang HW, Hou WH, Yeh TS. Responsiveness of SF-36 and Lower Extremity Functional Scale for assessing outcomes in traumatic injuries of lower extremities. Injury 2014;45(11):1759–63.
[37] Cruz-Díaz D, Lomas-Vega R, Osuna-Pérez MC, Hita-Contreras F, Fernández Á, Martínez-Amat A. The Spanish lower extremity functional scale: a reliable, valid and responsive questionnaire to assess musculoskeletal disorders in the lower extremity. Disabil Rehabil 2014;36(23):2005–11.
[38] Williams VJ, Piva SR, Irrgang JJ, Crossley C, Fitzgerald GK. Comparison of reliability and responsiveness of patient-reported clinical outcome measures in knee osteoarthritis rehabilitation. J Orthop Sports Phys Ther 2012;42(8):716–23.
[39] Metsavaht L, Leporace G, Riberto M, Sposito MM, Del Castillo LN, Oliveira LP, et al. Translation and cross-cultural adaptation of the lower extremity functional scale into a Brazilian Portuguese version and validation on patients with knee injuries. J Orthop Sports Phys Ther 2012;42(11):932–9.
[40] Watson CJ, Propps M, Ratner J, Zeigler DL, Horton P, Smith SS. Reliability and responsiveness of the lower extremity functional scale and the anterior knee pain scale in patients with anterior knee pain. J Orthop Sports Phys Ther 2005;35(3):136–46.
[41] Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56(5):395–407.
[42] Abbott JH, Schmitt J. Minimum important differences for the patient-specific functional scale, 4 region-specific outcome measures, and the numeric pain rating scale. J Orthop Sports Phys Ther 2014;44(8):560–4.