

The Lower Extremity Functional Scale (LEFS): Scale Development, Measurement Properties, and Clinical Application

Background and Purpose. The purpose of this study was to assess the reliability, construct validity, and sensitivity to change of the Lower Extremity Functional Scale (LEFS). **Subjects and Methods.** The LEFS was administered to 107 patients with lower-extremity musculoskeletal dysfunction referred to 12 outpatient physical therapy clinics. **Methods.** The LEFS was administered during the initial assessment, 24 to 48 hours following the initial assessment, and then at weekly intervals for 4 weeks. The SF-36 (acute version) was administered during the initial assessment and at weekly intervals. A type 2,1 intraclass correlation coefficient was used to estimate test-retest reliability. Pearson correlations and one-way analyses of variance were used to examine construct validity. Spearman rank-order correlation coefficients were used to examine the relationship between an independent prognostic rating of change for each patient and change in the LEFS and SF-36 scores. **Results.** Test-retest reliability of the LEFS scores was excellent ($R=.94$ [95% lower limit confidence interval (CI)=.89]). Correlations between the LEFS and the SF-36 physical function subscale and physical component score were $r=.80$ (95% lower limit CI=.73) and $r=.64$ (95% lower limit CI=.54), respectively. There was a higher correlation between the prognostic rating of change and the LEFS than between the prognostic rating of change and the SF-36 physical function score. The potential error associated with a score on the LEFS at a given point in time is ± 5.3 scale points (90% CI), the minimal detectable change is 9 scale points (90% CI), and the minimal clinically important difference is 9 scale points (90% CI). **Conclusion and Discussion.** The LEFS is reliable, and construct validity was supported by comparison with the SF-36. The sensitivity to change of the LEFS was superior to that of the SF-36 in this population. The LEFS is efficient to administer and score and is applicable for research purposes and clinical decision making for individual patients. [Binkley JM, Stratford PW, Lott SA, et al. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther.* 1999;79:371-383.]

Jill M Binkley
Paul W Stratford
Sue Ann Lott
Daniel L Riddle
The North American
Orthopaedic Rehabilitation
Research Network*

Key Words: Disability, Function, Functional scales, Sensitivity to change, Tests and measurements.

Proposed new measures of health status should be viewed with increasing rigor and sophistication with respect to scale development.^{1,2} Numerous generic and disease-specific self-report measures that are suitable for physical therapy clinical practice and that have adequate measurement properties now exist. Generic health status measures assess overall health, including social, emotional, and physical health status, and are intended to be applicable across a broad spectrum of diseases, interventions, and demographic and cultural subgroups.^{1,2} Condition-specific measures, also termed “disease-specific measures,” are designed to assess attributes that are most relevant to the disease or condition of interest.^{1,2} Ideally, disease-specific measures are composed of items that are frequently affected by the condition of interest and that are likely to demonstrate clinically important change.

Several barriers to the widespread adoption of generic and condition-specific measures in clinical practice have been identified, including (1) difficulty of administering the scales and of scoring, (2) difficulty in administering scales for different conditions and anatomical sites, (3) the practitioner’s belief that there is a lack of clinical

meaningfulness for the scores, and (4) inadequate measurement properties for application to individual patients.^{1,3,4} A limitation often cited by clinicians is that many measures are required for a practice that serves a varied caseload. For example, there are numerous condition-specific measures available for people with knee conditions. Measures exist for people with general conditions of the knee,⁵ patellofemoral joint disorders,^{6,7} ligamentous deficiency,^{8–10} and joint replacement.^{11,12} It is conceivable that 4 or 5 measures would be required to accommodate people with knee dysfunction, not to mention the number of measures that may be required to assess people with other lower-extremity, upper-extremity, and spinal problems.

One approach to overcoming the need for multiple measures of health status in clinical practice is to explore whether the measurement properties of condition-specific measures are superior to those of generic measures. Should the measurement properties be similar, a single generic measure or subscale of that measure could be used in place of a number of condition-specific measures. Several generic measures have been applied to a variety of patients with lower-extremity musculoskeletal conditions, including the SF-36,^{13,14} the SF-12,¹⁵ the

JM Binkley, PT, FCAMT, FAAOMPT, is Physical Therapist, Appalachian Physical Therapy, 109A Tipton Dr, Dahlonega, GA 30534 USA (binkley01@sprynet.com), and Assistant Professor (Physical Therapy), School of Rehabilitation Science, McMaster University, Hamilton, Ontario, Canada. Address all correspondence to Ms Binkley at the first address.

PW Stratford, PT, is Associate Professor, School of Rehabilitation Science, and Associate Member, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

SA Lott, PT, is Physical Therapist, Appalachian Physical Therapy, Dahlonega, Ga, and Clinical Professor, Physical Therapy Program, North Georgia College, Dahlonega, Ga.

D Riddle, PhD, PT, is Associate Professor, Department of Physical Therapy, Virginia Commonwealth University, Richmond, Va.

*North American Orthopaedic Rehabilitation Research Network is: Brad Balsora, PT, St Joseph’s Hospital, Hamilton, Ontario, Canada; Paul Beattie, PhD, PT, Department of Physical Therapy, University of Rochester, Rochester, NY; Andrew Berk, PT, Summit Injury Management, Victoria, British Columbia, Canada; Jill Binkley, PT, FCAMT, FAAOMPT, Appalachian Physical Therapy, Dahlonega, Ga; Susan Brennenman, PT, Penn Therapy and Fitness, Philadelphia, Pa; Linda Brett, PT, Kakabeka Physiotherapy, Kakabeka Falls, Ontario, Canada; Jane Burns, PT, Pacific Coast Rehabilitation Center, Vancouver, British Columbia, Canada; Bert Chesworth, PT, FCAMT, University of Western Ontario, London, Ontario, Canada; Doug Conroy, PT, ATC, Conroy Orthopaedic and Sports Physical Therapy, Flossmoor, Ill; Robert Feehley, PT, OCS, Baltimore Sports Rehab, Baltimore, Md; Karen Hayes, PhD, PT, Program in Physical Therapy, Northwestern University Medical School, Chicago, Ill; Scott Hyams, PT, Heartland Healthcare, Sunrise, Fla; Michael Kelo, PT, OCS, Sheltering Arms Physical Rehabilitation Hospital, Chester, Va; Carmen Kirkness, PT, McGill University, Montreal, Quebec, Canada; Kim Kramer, PT, Sartori Hospital, Cedar Falls, Iowa; Jim Krzaczek, PT, OCS, Life Care Medical Center, Glassboro, NJ; Sue Ann Lott, PT, Appalachian Physical Therapy, Dahlonega, Ga; Jane Mennie, PT, Fannin Regional Hospital, Blue Ridge, Ga; Jay Neel, Appalachian Physical Therapy, Dahlonega, Ga; Karen Orlando, PT, Physiotherapy on Bay, Toronto, Ontario, Canada; Beverly Padfield, PT, FCAMT, Four Counties Health Services, Newbury, Ontario, Canada; Corinne Roos, Kettle Creek Physiotherapy and Sports Injuries Clinic, St Thomas, Ontario, Canada; Linda Nolte Smith, PT, MTC, Park at Stony Point Physical Therapy, Richmond, Va; Dan Riddle, PhD, PT, Virginia Commonwealth University, Richmond, Va; Gregory Spadoni, PT, ProActive Physiotherapy, Hamilton, Ontario, Canada; Diane Stratford, PT, West End Physiotherapy Clinic, Hamilton, Ontario, Canada; Paul Stratford, PT, McMaster University, Hamilton, Ontario, Canada; Marcus Walser, PT, Walser Physiotherapy, Thunder Bay, Ontario, Canada; Linda Watts, PT, Algoma Physical Rehabilitation Clinic, Sault Ste Marie, Ontario, Canada; Michael Westaway, PT, FCAMT, Canadian Sport Rehabilitation Institute, Calgary, Alberta, Canada; Myra Westaway, PT, HSC, Lindsay Park Sport Rehab, Calgary, Alberta, Canada.

Approval for this study was obtained from the institutional review board associated with the North American Orthopaedic Rehabilitation Research Network based in Dahlonega, Ga. In addition, local institutional review board approval was obtained by clinicians and clinics participating in the study, where necessary.

This project was funded in part by a grant from the Section on Research of the American Physical Therapy Association.

This article was submitted February 24, 1998, and was accepted January 4, 1999.

Functional Status Index,¹⁶ and the Musculoskeletal Functional Assessment Questionnaire.^{4,17,18} The SF-36, SF-12, and Musculoskeletal Functional Assessment Questionnaire have been validated for group decision making only.^{3,4,15,18} The Functional Status Index has been documented to yield reliable and valid measurements in patients with total hip replacement, but sensitivity to change and application to other orthopedic conditions have not been reported for the Functional Status Index.¹⁶

The SF-36 has served as the principal generic measure for comparisons with condition-specific measures.^{4,8,12,19–25} The SF-36 consists of 8 health concept subscales (physical function, role physical, bodily pain, general health, vitality, social function, role emotional, and mental health) and 2 component summary scores. Each subscale score can vary from 0 to 100, with higher scores representing more desirable health states. The physical and mental component summary scores represent weighted composite scores derived from the 8 health concept scales. Each of the component summary scores is scaled to have a mean of 50 and a standard deviation of 10 for the general population of the United States. To date, the responsiveness of several of the SF-36 subscales and the physical component summary score have been shown to be superior or equivalent to condition-specific scales relevant to the lower extremity.^{20,23,25–27} Due to the utilization of designs that are not the most rigorous available²⁸ and because formal statistical comparisons between the observed change indexes were not reported, it is not known whether these observed differences represent true differences in the measures' capacities or whether they are merely a result of sampling variation. There is no strong evidence to suggest that existing condition-specific scales designed for the lower extremity are superior to the SF-36. The SF-36, however, is time-consuming to administer and score in the clinic and was not designed for individual patient decision making.

It is critical that measures of health status be reliable, valid, and responsive to clinical change that occurs over time.^{2,29} The terms "responsiveness" and "sensitivity to change" are often used interchangeably to describe the ability of a measure to detect clinical change.^{2,30–32} *Responsiveness*, as defined by Kirshner and Guyatt,³³ denotes the ability of a scale to detect change. Within Kirshner and Guyatt's taxonomy, responsiveness exists independent of validity. This position has been challenged by Hays and Hadorn,³⁴ who suggested that responsiveness is actually one indication of a measure's validity. An external standard of change was introduced to examine the extent to which a health status measure truly differentiates among patients who have improved, deteriorated, or remained stable and subsequently used in our study and by other authors.^{27,29,34–39} We used the

term "sensitivity to change" to denote the ability of a measure to detect true change in patients' status over time, as we have done in our recent publications.^{35,36} Using this definition, sensitivity to change is a form of validity.^{33,34} Because no criterion standard exists for assessing change in health status, construct validation is used to identify patients or groups of patients who are expected to change by differing amounts. Several methods have been used by us and by other researchers in an attempt to distinguish among patients' levels of change, including other clinical measures (eg, spinal flexion),²⁹ retrospective global rating of change,^{27,29,37–41} the achievement of treatment goals,³⁵ and an external prognostic rating of change.⁴² In a report co-authored by one of the present investigators, the authors contended that a bias can be introduced when the retrospective global rating is performed at the time of the follow-up assessment.⁴²

We believe that there is a need for a functional measure that is easy to administer and score and applicable to a wide range of patients with lower-extremity orthopedic conditions. Our goal was to develop a self-report condition-specific measure that would yield reliable and valid measurements and that would be appropriate for use as a clinical and research tool. Accordingly, the scale development process took into account the barriers identified for clinical implementation of self-report measures. The purpose of this article is to report on the development and initial validation of a newly developed condition-specific measure, the Lower Extremity Functional Scale (LEFS), including the determination of internal consistency, reliability, construct validity, sensitivity to change, and clinical application.

Method

Subjects

Subjects were consecutive patients referred for physical therapy with any lower-extremity musculoskeletal condition (defined as any condition of the joints, muscles, or other soft tissues of the lower extremity). Patients who did not speak English or were unable to read were excluded from the study. Data were collected over a 4-month period. A total of 107 patients were entered into the study. A description of the patients is presented in Table 1.

Data were collected in physical therapy clinics affiliated with the North American Orthopaedic Rehabilitation Research Network (NAORRN). At the time of the study, the NAORRN consisted of 19 physical therapy clinicians and 5 physical therapy researchers in the United States and Canada. The NAORRN was designed to support multicentered research in the orthopedic field. Involvement in the NAORRN is voluntary and unfunded.

Table 1.
Demographic Characteristics of Patients (N=107)

Characteristic	Summary
Age (y)	
\bar{X}	44.0
SD	16.2
Sex	
Male	46
Female	58
Missing	3
Time since onset (wk)	
\bar{X}	6
Minimum	0
Maximum	250
1st quartile	3
3rd quartile	11
n	98
Site of problem	
Hip	2
Thigh	1
Knee	71
Leg	8
Ankle	14
Foot	8
Missing	3
Surgery	
Yes	62
No	44
Missing	1
Condition	
Hip	
Osteoarthritis	1
Muscle strain	1
Missing	0
Knee/thigh/leg	
Ligament sprain	22
Muscle strain	5
Meniscal injury	10
Osteoarthritis	9
Patellofemoral pain	6
Fracture	3
Total joint replacement	8
Nonspecific sprain/strain	12
Missing	5
Foot/ankle	
Ligament sprain	9
Muscle strain	1
Osteoarthritis	2
Fracture	8
Nonspecific sprain/strain	2
Missing	0
Educational level completed	
Elementary	13
Secondary	46
University	41
Missing	7
Smoker	
Yes	19
No	85
Missing	3

Twelve of the 19 clinicians contributed data to the study (Fig. 1). Informed consent was obtained from all patients.

The LEFS

The conceptual framework that guided the development of the LEFS included that the scale (1) be based on the World Health Organization's model of disability and handicap,⁴³ (2) be efficient to administer, score, and record in the medical record with respect to patient and clinician time, (3) be applicable to a wide variety of patients with lower-extremity orthopedic conditions, including patients with a range of disability levels, conditions, diseases, treatments, and ages, (4) be applicable for documenting function on an individual patient basis as well as in groups, such as for clinical outcomes assessment and clinical research purposes, (5) be developed using a systematic process of item selection and item scaling,² (6) yield reliable measurements (have internal consistency and test-retest reliability), and (7) yield valid measurements (at a single point in time and sensitive to valid change).

Items were generated for the LEFS by a process of reviewing existing questionnaires as well as surveying clinicians and patients. The World Health Organization's model of disability⁴³ served as the basis for the item generation phase of the scale development. The terminology used to define disability and handicap was used as the basis of questions posed about functional limitations to patients. Thirty-five patients with a variety of lower-extremity orthopedic conditions were surveyed to determine important functional limitations associated with their problem. Patients were asked to "identify up to 3 important activities that you are unable to do or are having difficulty with because of your lower-limb problem?" From this survey, we selected 75 items and collapsed them to 22 items by grouping similar activities. "Walking on uneven ground" and "walking on grass," for example, were 2 activities that were grouped together. Three orthopedic physical therapists, each with at least 10 years of experience in orthopedic physical therapy practice, reviewed the 22 items and were given the opportunity to add additional items. We surveyed existing questionnaires. No additional items were identified as important to include in the LEFS by these additional processes.

The initial version of the scale consisted of 22 items. The introductory statement of the questionnaire states: "Today, do you or would you have any difficulty at all with:" followed by a listing of the functional items. Items are rated on a 5-point scale, from 0 (extreme difficulty/unable to perform activity) to 4 (no difficulty). The 5-point difficulty rating scale was selected to maximize the capacity of the scale to measure change (Appendix).

The initial version of the scale was administered to 57 patients who were referred for physical therapy with lower-extremity dysfunction. The anatomical sites represented in this group of patients were: foot and ankle (n=12), knee (n=29), hip (n=8), multiple trauma, (n=5), and missing (n=3). The broad categories of orthopedic conditions in this group were: sprains and strains (n=32), fractures and bone disorders (n=9), osteoarthritis (n=8), articular subluxation or dislocation (n=7), and missing (n=1). Total LEFS scale scores, means, and score distribution were determined for this group. The mean LEFS score, out of a possible score of 88, was 39 (SD=18.0, median score=40.0, range=2-85).

At the individual item level, mean score, median score, standard deviation, range, and frequency of endorsement of each level (0-4) of all items were determined. Inter-item correlations and corrected item-item total correlations were calculated. The corrected item-item total correlation is an estimate of the degree to which a single item score correlates with the total scale score with that item removed. The alpha coefficient, a measure of internal consistency, was determined for the scale and calculated with each of the items removed. The overall goal of this analysis was to ensure that individual item scores were reasonably normally distributed, with mean initial scores of about 50% of the items at approximately the midpoint⁴⁴ of the scale. In order to develop a measure that is applicable to a spectrum of conditions and levels of disabilities, the remaining items were selected to represent different difficulty levels, as indicated by item mean scores that were higher and lower than the midpoint. As a result of the item analysis, 2 items were removed from the original LEFS and 1 item was reworded. A factor analysis performed on the final 20-item questionnaire indicated that all items loaded on a single factor. The factor loadings varied from .44 (walking between rooms) to .86 (performing heavy activities around the house), with 19 of the factor loadings between .58 and .86.

The final version of the LEFS consists 20 items, each with a maximum score of 4. The total possible score of 80 indicates a high functional level (Appendix). The scale is one page, can be filled out by most patients in less than 2 minutes, and is scored by tallying the responses for all of the items. Scoring is performed without the use of a

calculator or computer and requires approximately 20 seconds.

Procedure

Reliability. The LEFS was administered during the initial assessment to patients with lower-extremity musculoskeletal dysfunction referred for physical therapy. The LEFS was readministered to patients 24 to 48 hours following the initial administration in order to examine test-retest reliability. The LEFS was then administered at weekly intervals (within 7 days, ± 1 day) for 4 weeks or until patients were discharged (in cases where discharge occurred prior to 4 weeks). In addition, the SF-36 (acute version) was administered during the initial assessment and at the weekly follow-up assessments. These follow-up intervals allowed examination of the validity of the LEFS measurements as well as comparison of sensitivity to change between the LEFS and SF-36.

Construct validity. In the absence of an accepted measure of function, determination of the validity of functional scales has relied heavily on the concept of construct validity. One or more theories are developed, and the extent to which a measure yields results concordant with the theory provides support for the validity of the measure. In this study, we believed that validity for our measure would be supported if:

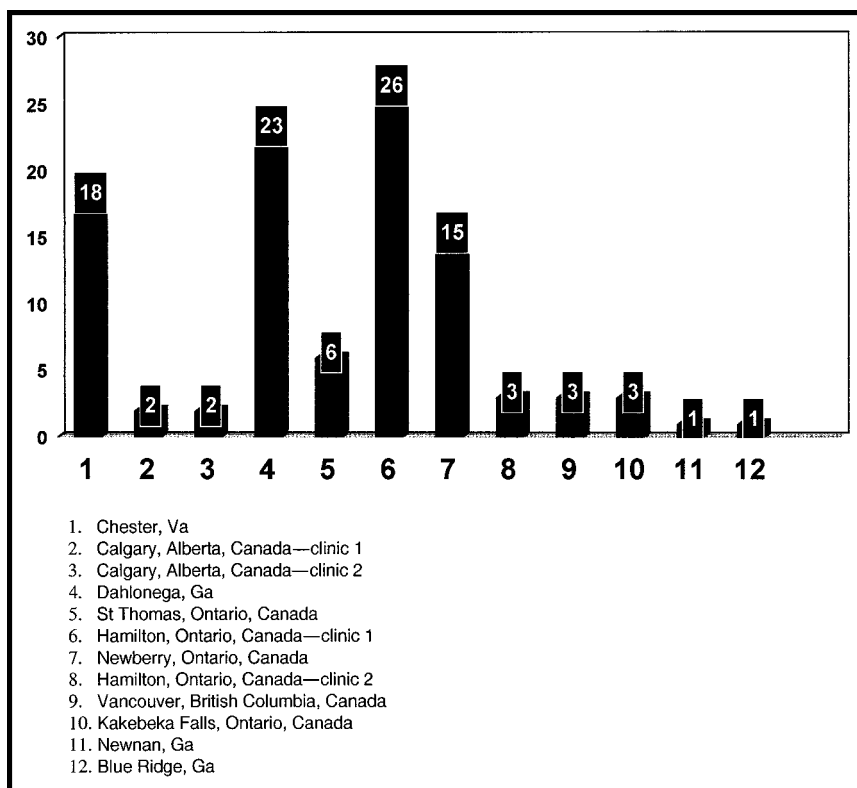


Figure 1. Clinic locations and numbers of patients entered in study.

-2	-1	0	+1	+2	+3	+4
Much Worse	Worse	No Change	Slight Improvement	Moderate Improvement	Large Improvement	Very Large Improvement

Figure 2.
Prognostic rating of change scale.

1. There would be a moderate correlation ($r > .6$) between LEFS scores and SF-36 physical function subscale and SF-36 physical component summary scores at the initial assessment and at the 3-week follow-up assessment.
2. There would be a low correlation ($r < .5$) between LEFS scores and SF-36 mental health subscale and SF-36 mental component summary scores at the initial assessment and at the 3-week follow-up assessment.
3. Patients who had recently undergone surgery (surgery less than 2 weeks prior to initial assessment) would have lower LEFS and SF-36 physical function subscale and physical component summary scores than would patients who did not have recent surgery (no surgery or surgery greater than 2 weeks prior to assessment).
4. Patients with acute conditions would demonstrate lower LEFS scores and SF-36 physical function subscale and physical component summary scores than would patients with chronic conditions.

The SF-36 was selected as the comparison scale for examination of the construct validity of the LEFS scores. The selection of the SF-36 was based on the literature documenting the measurement properties of the SF-36, including its applicability to patients with lower-extremity dysfunction. The reliability, validity, and responsiveness of measurements obtained with the SF-36 have been documented in diverse patient groups. The physical function and pain dimensions appear to be most relevant to orthopedic outpatients.^{20,37} Although several of the SF-36 subscales have the capacity to measure change on outpatients with musculoskeletal conditions, several of the subscales do not change or change minimally in this population.^{20,37} The mental health subscale of the SF-36 demonstrates minimal change in outpatients with musculoskeletal conditions.^{20,37}

In order to examine our argument for validity, which specified that patients with acute conditions would demonstrate more functional limitation than patients with chronic conditions, all patients were assigned a chronicity rating on a 3-point scale by 2 orthopedic physical

therapists blinded to patients' functional scale scores. The ratings were performed allowing discussion of the patient profile, and a single agreed-on score was determined. Patients were placed in one of the following categories based on a review of documentation, which included diagnosis and the time since onset of condition (or the time since surgery or cast removal): (1) acute—less than 4 weeks since onset of condition, surgery, or postfracture immobilization, (2) moderate/unclear, or (3) chronic—more than 4 weeks since onset of condition or having a chronic condition such as osteoarthritis. The basis for the selection of 4 weeks was the judgment of the investigators.

Sensitivity to change. Sensitivity to change was examined using a prognosis rating. Each patient was given a rating of prognosis using a 7-point scale (Fig. 2). Two orthopedic physical therapists who were blind to the patient's functional scale scores performed independent prognostic ratings on each patient, which were subsequently averaged. Prognostic ratings were based on documentation review of patients' diagnoses, age, chronicity, number of comorbid conditions, and type of surgery and time since surgery, where applicable. Raters answered the questions "How much change would you expect in this patient at 1 week following the initial assessment?" and "How much change would you expect in this patient at 3 weeks following the initial assessment?" We believed that, if our assumption about validity was correct, there would be a correlation between (1) the 1-week LEFS and SF-36 scores and the 1-week prognostic ratings and (2) the 3-week LEFS and SF-36 scores and the 3-week prognostic ratings. This approach was based on clinical judgment and previous work by Westaway et al,⁴⁵ whose data suggested that experienced clinicians can make prognoses about patients. The capacity of the LEFS and the SF-36 physical function subscale and physical component summary scores to measure valid change was compared at 1 week and at 3 weeks using this theory for change.

The interrater reliability for the prognostic ratings was determined using a type 3,2 intraclass correlation coefficient (ICC). This class of ICC is appropriate when ratings are averaged and an adjustment has been applied to address a systematic difference between raters.⁴⁶ The interrater reliability of the prognostic rating was $R = .84$

(95% lower limit confidence interval [CI]=.78). Because the goal of the analysis was to examine change, rather than to evaluate intervention, we made no attempt to control interventions.

Data Analysis

Internal consistency, reliability, and minimal detectable change. We used the alpha coefficient to estimate internal consistency, a measure of homogeneity of items.² A type 2,1 intraclass correlation coefficient (ICC) was used to estimate test-retest reliability.⁴⁶ Because many patients with musculoskeletal problems demonstrate true change over a short period of time, test-retest reliability was estimated over a 24- to 48-hour period using the entire patient sample and a subset of patients who were deemed to have more chronic conditions, as determined by the chronicity rating described above, and who were presumably more stable.

Reliability of the LEFS scores was also quantified using the standard error of measurement (SEM), a representation of measurement error expressed in the same units as the original measurement, in this case, LEFS points. Two estimates of the SEM were obtained. The first estimate, based on the alpha coefficient, was used to quantify measurement error at the 90% confidence level about a patient's score at a single point in time. This quantification was achieved by multiplying the SEM by the z value associated with the 90% confidence level (ie, $z=1.65$). The test-retest reliability coefficient obtained for the subset of patients with more chronic conditions was used to estimate the SEM that was used to calculate minimal detectable change (MDC) at the 90% confidence level. To obtain this estimate, the SEM is multiplied by the z value for the confidence level of interest, and this quantity is multiplied by the square root of 2.⁴⁷

Construct validity. Pearson correlation coefficients and 95% one-sided lower limit confidence intervals were calculated to examine the relationship between the LEFS scores and the SF-36 subscale and component summary scores at the initial assessment. One-way analyses of variance were used to examine the hypotheses about validity that specified that there would be a difference in initial LEFS scores and SF-36 physical function subscale and physical component summary scores between: (1) patients with recent surgery and patients without recent surgery and (2) patients with acute conditions and patients with chronic conditions.

Sensitivity to change and minimal clinically important difference. Spearman rank-order correlation coefficients were used to examine the relationship between the prognostic rating and change in the following func-

tional status scores at 1 week and 3 weeks: LEFS score, SF-36 physical function subscale score, SF-36 physical component summary score, and SF-36 mental component summary score. The magnitudes of the correlations between the prognostic ratings and the LEFS, SF-36 physical function subscale, and SF-36 physical component summary scores were formally compared using the method of Williams⁴⁸ for dependent data.

The minimal clinically important difference (MCID), defined as the minimal amount of change on the scale required to be considered a clinically important change, was determined using 2 methods. In the first approach, we used the prognostic ratings of change to separate patients into those who were predicted to undergo important change (prognostic ratings of 2, 3, and 4) and those who were predicted to undergo no important change at 3 weeks (prognostic ratings of 0 and 2). The cutpoint of change on the LEFS that maximized the area under a receiver operating characteristic (ROC) curve was determined as the estimate of the amount of change on the LEFS that best classified patients who had changed an important amount from those who had not.^{35,36} Sensitivity and specificity for this cutpoint value were determined. The second approach was a survey of 5 clinicians who reported that they had used the LEFS for a minimum of 4 months and on at least 10 patients as a clinical decision-making tool. Clinicians were asked to estimate the amount of change that they would consider to be clinically important for initial LEFS scores of 10, 25, 40, 55, and 70. Clinicians were asked to identify the minimal amount of change on the LEFS, in scale points, that would suggest that improvement had occurred. The same question was posed to clinicians in terms of deterioration. Clinicians' judgments of MCID were compared with the statistical approach.

Results

Descriptive Statistics

Patients' characteristics are shown in Table 1. Descriptive statistics for the patients by measure are presented in Table 2. None of the patients received the minimum or maximum scores for the LEFS at any of the assessments. Minimum and maximum SF-36 physical function subscale scores were obtained for 1 and 4 patients, respectively, at the initial assessment. Minimum and maximum SF-36 physical function subscale scores were each obtained for 1 patient at the 3-week follow-up assessment.

Reliability

Internal consistency was $\alpha=.96$ ($N=107$). Test-retest reliability estimates were $R=.86$ (95% lower limit CI=.80) for the entire sample ($n=98$) and $R=.94$ (95%

Table 2.

Descriptive Data for Lower Extremity Functional Scale (LEFS) and SF-36 Physical Function Subscale and Physical and Mental Component Summary Scores

	Sample Size	\bar{X}	SD	Median	Minimum	Maximum	Skewness (SE)	Kurtosis (SE)
LEFS (/80)								
Initial	107	34	16	34	5	74	0.34 (0.23)	-0.57 (0.46)
Retest (24-48 h)	98	37	17	36	5	76	0.23 (0.24)	-0.56 (0.48)
Week 1	101	43	18	42	2	77	0.02 (0.24)	-0.84 (0.47)
Week 2	90	45	16	45	9	77	0.16 (0.26)	-0.70 (0.50)
Week 3	58	49	17	50	6	78	-0.21 (0.31)	-0.81 (0.61)
Week 4	35	51	18	50	18	78	-0.33 (0.40)	-1.01 (0.78)
SF-36 physical function subscale scores(/100)								
Initial	101	35	22	35	0.0	100	0.56 (0.24)	-0.01 (0.47)
Week 1	101	46	24	40	10	100	0.46 (0.24)	-0.98 (0.47)
Week 2	88	50	25	45	5	100	0.26 (0.26)	-0.96 (0.51)
Week 3	55	53	24	52.5	0	100	0.04 (0.32)	-0.83 (0.63)
Week 4	35	56	24	60	10	100	-0.27 (0.40)	-0.92 (0.78)
SF-36 physical component summary scores ^a								
Initial	101	31	8	30	19	58	0.78 (0.78)	0.17 (0.47)
Week 1	101	36	9	34	19	58	0.61 (0.24)	-0.37 (0.48)
Week 2	88	37	10	35	20	58	0.33 (0.26)	-0.91 (0.51)
Week 3	55	38	9	36.5	21	58	0.25 (0.32)	-0.94 (0.63)
Week 4	35	40	9	39	26	58	0.11 (0.40)	-1.16 (0.78)
SF-36 mental component summary scores ^a								
Initial	101	52	11	54	20	73	-0.68 (0.24)	0.02 (0.47)
Week 1	101	54	9	56	31	70	-0.48 (0.24)	-0.31 (0.48)
Week 2	87	55	9	56	34	70	-0.52 (0.26)	-0.29 (0.51)
Week 3	55	56	9	59	29	72	-0.91 (0.32)	0.24 (0.63)
Week 4	35	56	9	61	34	69	-0.95 (0.40)	-0.33 (0.78)

^a Component summary scores are scaled to have a mean of 50 and a standard deviation of 10 in the general population of the United States.**Table 3.**Validity Coefficient Estimates^a for Lower Extremity Functional Scale (Single Point in Time) (n=100)

SF-36 Subscale and Component Summary Scores									
Physical Function	Role Physical	Bodily Pain	General Health	Vitality	Social Function	Role Emotional	Mental Health	Physical Component Summary Score	Mental Component Summary Score
.80	.51	.49	.09	.43	.62	.32	.23	.64	.30
(.73)	(.38)	(.35)	(-.08)	(.28)	(.51)	(.16)	(.07)	(.54)	(.14)

^a Lower 1-sided 95% confidence interval shown in parentheses.

lower limit CI=.89) for the subset of patients with more chronic conditions (n=31).

Construct Validity

Table 3 includes the validity coefficient estimates at the initial assessment between the LEFS and SF-36 scores. Correlations between the LEFS scores and the SF-36 physical function subscale and physical component summary scores were $r=.80$ (95% lower limit CI=.73) and $r=.64$ (95% lower limit CI=.54). The correlation between the LEFS scores and the SF-36 mental compo-

nent summary scores was $r=.30$ (95% lower limit CI=.14). There was a difference in LEFS scores between the patients with recent surgery and the patients without recent surgery at the initial assessment ($P=.006$) (Tab. 4). There was a difference in LEFS scores between the patients with acute conditions and the patients with chronic conditions ($P=.027$). There was no difference in SF-36 physical function subscale, physical component summary, and mental component summary scores between the patients with recent surgery and the patients without recent surgery ($P=.117$) or between the

Table 4.

Construct Validity: Comparison of Lower Extremity Functional Scale (LEFS) and SF-36 Physical Function Subscale Scores for Subgroups of Patients With and Without Surgery and Patients With Acute, Moderate, and Chronic Conditions at Initial Assessment

	Subgroup	N	\bar{X}	SD	F	P
LEFS	No surgery	58	38.2	15.8	7.8	.006
	Surgery	42	29.6	14.2		
SF-36 physical function subscale	No surgery	58	38.8	22.4	2.49	.117
	Surgery	42	31.9	20.2		
LEFS	Acute	29	27.7	13.6	3.8	.027
	Moderate/unclear	37	36.9	15.7		
	Chronic	30	37.4	16.9		
SF-36 physical function subscale	Acute	29	31.6	21.5	0.76	.471
	Moderate/unclear	37	37.7	21.3		
	Chronic	30	37.5	23.7		

Table 5.

Validity Coefficient Estimates of Sensitivity to Change for Lower Extremity Functional Scale (LEFS) and SF-36 Physical Function Subscale and Physical and Mental Component Summary Scores^a

	LEFS Change	SF-36 Physical Function Subscale Change	SF-36 Physical Component Summary Score Change	SF-36 Mental Component Summary Score Change	Prognosis at Week 1
LEFS change		.57 (.44)	.26 (.10)	.13 (-.04)	.36 (.20)
SF-36 physical function subscale change	.77 (.66)		.57 (.44)	-.05 (-.22)	.25 (.09)
SF-36 physical component summary score change	.64 (.48)	.81 (.71)		-.45 (-.57)	.16 (-.01)
SF-36 mental component summary score change	.25 (.02)	.16 (-.07)	-.12 (-.34)		.14 (-.03)
Prognosis at week 3	.64 (.48)	.42 (.21)	.45 (.25)	.28 (-.06)	

^aUpper diagonal represents correlations (95% confidence interval) at week 1 (n=98); lower diagonal represents correlations (95% confidence interval) at week 3 (n=54).

patients with acute conditions and the patients with chronic conditions ($P=.471$) (Tab. 4).

Sensitivity to Change

The correlations relating to change scores are presented in Table 5. There was no difference between the correlations of the prognostic rating with the LEFS and the prognostic rating with the SF-36 physical function subscale score at the initial assessment ($t_{(95)}=1.24$, $P_{(1)}=.106$). There was a difference between the correlations of the prognostic rating with the LEFS and the prognostic rating with the SF-36 physical component summary score at the initial assessment ($t_{(95)}=1.67$, $P_{(1)}=.05$). There was a difference between the correlations of the prognostic rating with the LEFS and the prognostic rating with the SF-36 physical function subscale score at week 3 ($t_{(95)}=3.05$, $P_{(1)}=.002$). There was

also a difference between the correlations of the prognostic rating with the LEFS and the prognostic rating with the SF-36 physical component summary score at week 3 ($t_{(95)}=2.13$, $P_{(1)}=.019$).

Individual Patient Decision Making

With respect to individual patient decision making, the potential error associated with a score on the LEFS at a given point in time is ± 5.3 scale points on the 80-point scale (90% CI) (Tab. 6). The MDC is ± 9 scale points (90% CI). The MCID is approximately 9 scale points. The area under the ROC curve associated with this value is .76, and the sensitivity and specificity are .81 and .70, respectively. The average of the 5 clinician estimates for MCID was 10 scale points, suggesting that the statistical approach has resulted in a reasonable estimate of the MCID.

Table 6.

Measurement Properties of the Lower Extremity Functional Scale (LEFS): Application to Individual Patients

	SEM	Error at Single Point in Time	Minimal Detectable Change	Minimal Clinically Important Difference
LEFS points	±3.9	±5.3	9	9

Discussion and Conclusions

Measurement of functional status in our patients served 2 important and distinct purposes: (1) documentation of physical therapy outcome in *groups* of patients for quality assurance, establishment of clinical standards, or research purposes and (2) documentation of functional level used to set goals and measure functional progress and outcome for *individual* patients. The capacity of the LEFS to detect change in lower-extremity function appears to be superior to that of the SF-36 physical function subscale, as indicated by higher correlations with an external prognostic rating of change. In light of this finding as well as the greater ease of administering and scoring the LEFS, this scale appears to be a good choice for documenting lower-extremity function. Only prospective research, however, validates the use of this measure in clinical decision making. Because the LEFS measures physical function but not overall health, we believe that a generic health status measure such as the SF-36 should be used to supplement the LEFS when the goal is to measure the overall health status of our patients. The LEFS appears to overcome, to some extent, the barriers identified for implementation of a health status measure in clinical practice.

The LEFS is easy to administer and score and is applicable to a wide range of disability levels and conditions and all lower-extremity sites. In our view, the LEFS is more interpretable with respect to understanding error associated measurement and for determining minimally clinically important score changes and is a sufficient measure of reliability, validity, and sensitivity to change, at a level that is commensurate with utilization at an individual patient level.

The LEFS can be used by clinicians as a measure of patients' initial function, ongoing progress, and outcome as well as to set functional goals. For an outpatient orthopedic population, for example, initial and weekly follow-up administration may be considered appropriate. In order to set short- and long-term goals based on a self-report functional scale such as the LEFS, the clinician, in our view, should synthesize the patient's clinical history and findings, as well the measurement properties of the scale (ie, the error associated with a

single-scale measure, MDC, and MCID). The error associated with a given measure on the LEFS is about ± 5 scale points (90% CI). Clinicians, therefore, can be reasonably confident that an observed score is within 5 points of the patient's "true" score. The MDC of the LEFS is ± 9 scale points (90% CI). Clinicians can also be reasonably confident that change on the LEFS of greater than 9 scale points is a true change. This information can be used to base short- and long-term goals for functional change that are at least greater than the MDC. The MCID of the LEFS is about 9 scale points. Clinicians can be reasonably confident that a change of greater than 9 scale points is not only a true change but is also a clinically meaningful functional change. Whether short- or long-term goals are set that are just at or greater than the MDC and MCID for the LEFS will depend on the patient's initial functional level, clinical history and findings, and time frames for the goals.

An example of the application of the LEFS to establish functional level, set goals, and track progress and outcome. Consider a patient with an initial LEFS score of 46/80. Based on the error at a given point in time for the LEFS of 5 points, the clinician can be 90% confident that the actual scale score is between 41 and 51. If the patient's condition is deemed to be relatively chronic and is expected to change slowly, the clinician might select a 2-week time frame for a change in score of just at the MDC and MCID of 9 scale points. The short-term goal, therefore, could be: "Increase LEFS score to less than or equal to 54/80." In setting a short-term goal for a patient with a relatively acute condition who is predicted to experience change quickly, a shorter time frame of, for example, 1 week with a greater change than the MDC and MCID may be selected. In this case, the goal may be: "Increase LEFS score to greater than or equal to 60/80." On follow-up, for example, 1 week later, progress is could be determined by the amount of change on the scale. In cases where improvement greater than the MDC and MCID occur, clinicians can be reasonably confident that true (MDC) and important (MCID) change has occurred. In cases where there is no change or change less than the MDC on follow-up, clinicians may be confident that true clinical change has not occurred. In this case, depending on the clinical picture and time frame since the previous assessment, a change in intervention, referral, or discharge of the patient may be considered.

Ceiling and floor effects exist for a health status measure when patients often score at the extremes of normal function or severely restricted function. In the case of a ceiling effect, there is restricted range for improvement because patients begin at the high level of function on the scale. In the case of a floor effect, there is a restricted range for deterioration in functional status. For

example, the existence of both a ceiling effect and a floor effect has been reported for patients with lower-extremity dysfunction for the MFA^{4,18} and for the SF-36 in our study. None of the patients in our study scored at 0 or 80 on the scale at admission or at the 3-week follow-up assessment, indicating that there is no ceiling or floor effect associated with the LEFS in this type of patient population. The implication of ceiling and floor effects is to lower the capacity to detect clinically important change in all patients.

In our study, we also used a rating of expected change as the theory for change. Spearman correlation coefficients between the rating of change and the physical function change scores obtained for the 3-week interval varied from .42 to .64. Our results, coupled with those of Westaway et al,⁴⁵ provide support for using a prognostic rating of change as a theory for evaluating a measure's sensitivity to change. The results of the studies suggest that a correlation coefficient of approximately .50 can be expected. This information may be useful for estimating the sample size for subsequent studies, where a prognostic rating is used as a theory for change.

There are 2 major limitations to this study. The LEFS was conceived as a measure applicable to a broad spectrum of lower-extremity problems. Our sample included only 3 patients with hip and thigh conditions. In addition, all patients in the study were outpatients. Further investigation is needed to document the measurement properties of the LEFS in patients with hip conditions and in other settings, including inpatient orthopedics.

We conceived the LEFS as a measure applicable for people with a broad spectrum of lower-extremity problems. Accordingly, it was necessary that the first study compare the LEFS with a measure of established validity. The SF-36 has been used to assess outcomes in people with hip, knee, and ankle dysfunction.^{4,7,12,19–25} It was for this reason that we chose the SF-36 as the comparison measure. Deyo, when reflecting on the proliferation of outcome measures, stated, “while the development of new instruments would be encouraged where necessary, we may hope that investigators will not reinvent the wheel.”^{49(p1052)} New measures are appropriate when their measurement properties or efficiency—in terms of the burden on both the respondent and those required to score the measure—are superior to existing measures.⁵⁰ It is for this reason that a one-sided research question was posed: Is the LEFS superior to the SF-36? The results of our study, in our opinion, provide evidence supporting the superiority of the LEFS over the SF-36 for assessing lower-extremity function. Subsequent inquiry concerning the LEFS should center on head-to-head comparisons with condition-, disease-, or region-specific measures. Rather than asking whether the LEFS

is superior to existing measures, future research should inquire about the equivalence of the LEFS and the competing measures of interest.

Selection of self-report measures suitable for documenting outcomes in clinical practice and in clinical trials and choosing a condition-specific or generic health status measure should be dependent, in part, on the goals of measurement. Condition-specific measures, such as the LEFS, often do not include measures of psychosocial function and tend to be less influenced by comorbid states.^{8,19} The LEFS, however, is superior to the SF-36 in terms of clinical efficiency and sensitivity to change for the documentation of physical function in patients with lower-extremity dysfunction. Generic measures, such as the SF-36, are not generally practical for application at an individual patient level due to the length of the scale and complexity of scoring. Because the conceptual frameworks for generic and disease-specific measures—such as the LEFS—differ, we believe that they can be viewed as being complementary rather than competing measures. Indeed, there is considerable agreement that a comprehensive assessment should include the administration of both generic and disease-specific measures.^{8,19} In clinical practice, the administration of both a generic measure, such as the SF-36, and a condition-specific measure, such as the LEFS, at admission and discharge, with weekly re-evaluation of using the condition-specific measure, would achieve the benefits offered by both types of measures.

Acknowledgments

We acknowledge the contributions to this project of Gregory Alcock, PT, and Aly Mawani, PT, who were students in the School of Rehabilitation Science at McMaster University, Hamilton, Ontario, Canada, at the time of this study.

References

- 1 Nelson EC, Berwick DM. The measurement of health status in clinical practice. *Med Care*. 1989;27(suppl):S77–S90.
- 2 Streiner DL; Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford, England: Oxford University Press; 1995.
- 3 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Qual Life Res*. 1995;4:293–307.
- 4 Martin DP, Engelberg R, Agel J, Swiontkowski MF. Comparison of the Musculoskeletal Function Assessment Questionnaire with the Short Form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile health-status measures. *J Bone Joint Surg Am*. 1997;79:1323–1335.
- 5 Irrgang JJ, Snyder-Mackler L, Wainner RS, et al. Development of a patient-reported measure of function of the knee. *J Bone Joint Surg Am*. 1998;80:1132–1145.
- 6 Kujala UM, Jaakkola LH, Koskinen SK, et al. Scoring of patellofemoral disorders. *Arthroscopy*. 1993;9:159–163.

- 7 Chesworth BM, Culham EG, Tata GE, Peat M. Validation of outcome measures in patients with patellofemoral syndrome. *J Orthop Sports Phys Ther.* 1989;10:302–308.
- 8 Kantz ME, Harris WJ, Levitsky K, et al. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Med Care.* 1992;30:MS240–MS252.
- 9 Lysholm J, Gillquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *Am J Sports Med.* 1982;10:150–154.
- 10 Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop.* 1985;198:43–49.
- 11 Bellamy N, Buchanan WW, Goldsmith C, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip and knee. *J Rheumatol.* 1988;15:1833–1840.
- 12 Hawker G, Melfi CA, Paul J, et al. Comparison of a generic (SF-36) and a disease-specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. *J Rheumatol.* 1995;22:1193–1196.
- 13 Ware JE Jr, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. *Med Care.* 1992;30:473–483.
- 14 Ware JE Jr, Kosinski M, Bayliss MS, et al. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care.* 1995;33:AS264–AS279.
- 15 Ware JE Jr, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220–232.
- 16 Jette AM. The Functional Status Index: reliability and validity of a self-report functional disability measure. *J Rheumatol.* 1987;14:15–21.
- 17 Engelberg R, Martin DP, Agel J, et al. Musculoskeletal Function Assessment Instrument: criterion and construct validity. *J Orthop Res.* 1996;14:182–192.
- 18 Martin DP, Engelberg R, Agel J, et al. Development of a musculoskeletal extremity health status instrument: the Musculoskeletal Function Instrument. *J Orthop Res.* 1996;14:173–181.
- 19 Bombardier C, Melfi CA, Paul J, et al. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med Care.* 1995;4:AS131–AS144.
- 20 Jette DU, Jette AM. Physical therapy and health outcomes in patients with knee impairments. *Phys Ther.* 1996;76:1178–1187.
- 21 Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perception of patients about total hip replacement. *J Bone Joint Surg Br.* 1996;78:185–190.
- 22 Shapiro ET, Richmond JC, Rockett SE, et al. The use of a generic, patient-based health assessment (SF-36) for evaluation of patients with anterior cruciate ligament injuries. *Am J Sports Med.* 1996;24:196–200.
- 23 Katz J, Harris TM, Larson MG, et al. Predictors of functional outcomes after arthroscopic partial meniscectomy. *J Rheumatol.* 1992;19:1938–1942.
- 24 Bennett KJ, Torrance GW, Moran LA, et al. Health state utilities in knee replacement surgery: the development and evaluation of McKnee. *J Rheumatol.* 1997;24:1796–1805.
- 25 Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Quality in Health Care.* 1996;5:581–588.
- 26 Weinberger M, Samsa GP, Tierney WM, et al. Generic versus disease-specific health status measures: comparing the Sickness Impact Profile and the Arthritis Impact Measurement scales. *J Rheumatol.* 1992;19:543–546.
- 27 Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol.* 1997;50:239–246.
- 28 Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther.* 1996;76:1109–1123.
- 29 Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis.* 1986;39:897–906.
- 30 Liang MH. Evaluating measurement responsiveness. *J Rheumatol.* 1995;22:1191–1192.
- 31 Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Control Clin Trials.* 1991;12(suppl):142S–158S.
- 32 Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol.* 1997;50:79–93.
- 33 Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis.* 1985;38:27–36.
- 34 Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res.* 1992;1:73–75.
- 35 Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. *Phys Ther.* 1998;78:1197–1207.
- 36 Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Phys Ther.* 1998;78:1186–1196.
- 37 Chatman AB, Hyams SP, Neel JM, et al. The Patient-Specific Functional Scale: measurement properties in patients with knee dysfunction. *Phys Ther.* 1997;77:820–829.
- 38 Stratford PW, Gill C, Westaway M, Binkley JM. Assessing disability and change on individual patients: a report of a patient-specific measure. *Physiotherapy Canada.* 1995;47:258–263.
- 39 Beurskens AJHM, de Vet HCW, Koke AJA. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain.* 1996;65:71–76.
- 40 Stratford PW, Binkley JM. Measurement properties of the RM-18: a modified version of the Roland-Morris disability scale. *Spine.* 1997;22:2416–2421.
- 41 Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale: measurement properties. *Spine.* 1995;20:341–352.
- 42 Norman GR, Stratford PW, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol.* 1997;50:869–879.
- 43 *Constitution of the World Health Organization.* Geneva, Switzerland: World Health Organization; 1948.
- 44 Jette DU, Jette AM. Health status assessment in the occupational health setting. *Orthop Clin North Am.* 1998;27:891.
- 45 Westaway M, Stratford PW, Binkley JM. The Patient-Specific Functional Scale: validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther.* 1998;27:331–338.
- 46 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.

47 Stratford PW, Binkley J, Solomon P, et al. Defining the minimum level of detectable change for the Roland-Morris Questionnaire. *Phys Ther*. 1996;76:359–365.

48 Williams EJ. The comparison of regression variables. *Journal of the Royal Statistical Society, Series B*. 1959;21:396–399.

49 Deyo RA. Measuring the functional status of patients with low back pain. *Arch Phys Med Rehabil*. 1988;69:1044–1053.

50 Armadio P, Beaton D, Bombardier C, et al. Measuring disability and symptoms of the upper limb: a validation study of the DASH questionnaire [Abstract]. *Arthritis Rheum*. 1996;39:S112.

Appendix.

Lower Extremity Functional Scale

We are interested in knowing whether you are having any difficulty at all with the activities listed below because of your lower limb problem for which you are currently seeking attention. Please provide an answer for **each** activity.

Today, do you or would you have any difficulty at all with:

(Circle one number on each line)

Activities	Extreme Difficulty or Unable to Perform Activity	Quite a Bit of Difficulty	Moderate Difficulty	A Little Bit of Difficulty	No Difficulty
a. Any of your usual work, housework, or school activities.	0	1	2	3	4
b. Your usual hobbies, recreational or sporting activities.	0	1	2	3	4
c. Getting into or out of the bath.	0	1	2	3	4
d. Walking between rooms.	0	1	2	3	4
e. Putting on your shoes or socks.	0	1	2	3	4
f. Squatting.	0	1	2	3	4
g. Lifting an object, like a bag of groceries from the floor.	0	1	2	3	4
h. Performing light activities around your home.	0	1	2	3	4
i. Performing heavy activities around your home.	0	1	2	3	4
j. Getting into or out of a car.	0	1	2	3	4
k. Walking 2 blocks.	0	1	2	3	4
l. Walking a mile.	0	1	2	3	4
m. Going up or down 10 stairs (about 1 flight of stairs).	0	1	2	3	4
n. Standing for 1 hour.	0	1	2	3	4
o. Sitting for 1 hour.	0	1	2	3	4
p. Running on even ground.	0	1	2	3	4
q. Running on uneven ground.	0	1	2	3	4
r. Making sharp turns while running fast.	0	1	2	3	4
s. Hopping.	0	1	2	3	4
t. Rolling over in bed.	0	1	2	3	4
Column Totals:					

SCORE: _____/80

Error (single measure): ±5 scale points
MDC: 9 scale points
MCID: 9 scale points