Functional Rating Index

A New Valid and Reliable Instrument to Measure the Magnitude of Clinical Change in Spinal Conditions

Ronald J. Feise, DC,* and J. Michael Menke, MA, DC†

Study Design. A prospective cohort design was used to evaluate the Functional Rating Index in a multicentered setting with 139 participants. The Functional Rating Index is a self-reporting instrument consisting of 10 items, each with 5 possible responses that express graduating degrees of disability.

Objectives. The goal of this study was to evaluate the psychometric qualities of the Functional Rating Index.

Summary of Background Data. The Functional Rating Index combines the concepts of the Oswestry Low Back Disability Questionnaire and the Neck Disability Index and seeks to improve on clinical utility (time required for administration).

Methods. One hundred thirty-nine subjects with spinal complaints participated in four different cohorts to study reliability, validity, responsiveness, and clinical utility.

Results. *Reliability*: Test-retest: Intraclass correlation coefficient was excellent ($ICC_{3,k} = 0.99$); interitem correlation: Item efficiency was good, ranging between 0.54 and 0.82, with a moderate correlation among all items; Cronbach's alpha was excellent (0.92). *Validity*: construct: The Functional Rating Index correlated with the Disability Rating Index (0.76), the Short Form-12 Physical Component Score (0.76), and the Short Form-12 Mental Component Score (0.36). *Responsiveness*: Overall, the size effect was 1.24, which is commendable. *Clinical utility*: Time required by the patient and staff averaged 78 seconds per administration, which is noteworthy. *Effect of Sociodemographics*: Total scores were not affected by education, gender, nor age, suggesting minimal external validity bias.

Conclusions. The Functional Rating Index appears to be psychometrically sound with regard to reliability, validity, and responsiveness and is clearly superior to other instruments with regard to clinical utility. The Functional Rating Index is a promising useful instrument in the assessment of spinal conditions. [**Key words**: back pain, neck pain, outcome, reliability, validity, responsiveness, practicality, activities of daily living] **Spine 2001;26:78–87**

Because the aim of therapy within physical medicine is focused on the reduction of pain and the enhancement of function,[2,8] it is critically important for the physician to understand the everyday functioning and symptoms of the patient and to have a method of quantifying those items. Function is considered the most important measure of severity,[42] and pain is the most common measure.[33] Assessing the patient's perception is considered

vital in judging the end point of therapy. Epstein stated, "A patient's self-evaluation may be more accurate than the clinical, biochemical, or physiologic indexes that we have traditionally relied on."[12] Yet patients can exaggerate responses on questionnaire instruments and distort tests that measure strength, endurance, or range of motion. Moreover, "objective" findings from radiographs and physiologic tests suffer from low reliability and/or validity scores[3,9] and are not well correlated with items that concern patients or society.[8] Even items such as "return to work," which might be considered pertinent to all stakeholders and more objective than a patient's perception, suffer from the confounding of social factors.[42] Patients and payors, in general, do not care whether a patient has a negative Straight Leg Raising Test or a 10% increase in range of motion. Both patients and payors want the patient to be able to perform daily activities and participate in life without pain or restriction.

The need to measure the function of the neck and back (establish clinical effectiveness) has resulted in many instruments being produced in the last 20 years. These instruments can provide reliable and valid methods to quantify a patient's status.[29] Yet, the general problem with these instruments is that they have poor clinical utility; they require too much time for patients to answer and staff members to score. Patients become exhausted because they have too many forms to fill out, and these forms take too much time to complete. Clinicians and researchers have similar reservations about these instruments, which might be a barrier to their use.

The fact that many patients have symptoms in multiple regions (*e.g.*, cervical and lumbar regions) compounds the problem of clinical utility. A recent unpublished cross-sectional study of nine chiropractic clinics by one of the authors (R.J.F.) found that 40% of the sample (n = 90) had simultaneous complaints or diagnoses of both neck and back conditions.[14] Currently, the conscientious physician must use separate instruments to assess low back conditions and neck conditions. Using two instruments per patient requires increased time for patients to complete and for clinicians and their staffs to score and analyze.

■ **Methods**

Objective. The Functional Rating Index (FRI) is an instrument specifically designed to quantitatively measure the subjective perception of function and pain of the spinal musculoskeletal system in a clinical environment. In particular, it evaluates the patient's subjective report of his ability to perform dynamic

Table 1. Psychometric Properties for Established Functional Instruments

	Reliability	Validity	Responsiveness	Administrative Burden*
Low back instruments				
Aberdeen Low Back Scale[34]	√	√	√	10 minutes?
OLBDQ[2,9,21,34]	√	√	√	5 minutes
Waddell Disability Index[2,21,34]	√, Test-retest	Doubtful	Doubtful	5 minutes?
Roland-Morris Scales[2,9,21,37]	√	√	√	5 minutes
Million[2,21]	√, Test-retest	Doubtful	Doubtful	10 minutes
Sickness Impact Profile[9,21]	√	√	√	20 minutes
Pain Disability Index[15,39]	√	√	√	5 minutes?
Backill[40]	√	√	√	10 minutes
Disability Rating Index[35]	√	√	√	5 minutes
Quebec Disability Scale[22]	√	√	√	10 minutes?
Low Back Pain Rating Scale[27,28]	√	√	√	15 minutes
Neck instruments				
Northwick Neck Questionnaire[24]	√	√	√	5 minutes?
NDI[41]	√	√	√	5 minutes?

* Times with a question mark indicate an estimated time where no time was reported in the literature.
OLBDQ = Oswestry Low Back Disability Questionnaire; NDI = Neck Disability Index.

movements of the neck and back and/or withstand static positions.

The FRI emphasizes function (realizing that all functions may be influenced by other variables) while concurrently measuring the patient's opinion, attitude, and self-rating of disability.[35] It was developed to provide an assessment instrument that has clinical utility (*i.e.,* easy and fast for both patients and health care teams) yet quantifies the patient's current state of pain and dysfunction in a reliable and valid manner for spinal conditions. The purpose of this study is to determine whether FRI is reliable, valid, and responsive and will reduce administrative burden better than other self-reporting instruments.

Procedures to Build the Instrument. A search of the medical literature from 1970 through 1997 was conducted using electronic data bases (MEDLINE; Index Medicus) to discover which existing instrument(s) might have the most appropriate scale items. Selected for further study were the instruments listed in Table 1, with varying degrees of reliability, validity, responsiveness, and clinical utility.

Overall, no instruments are known to be significantly more advantageous than the Oswestry Low Back Disability Questionnaire (OLBDQ) for the low back, or Neck Disability Index (NDI) for the neck, both of which have been widely researched and validated.[13,41] (The Neck Disability Index is a modification of OLBDQ for neck conditions.) The scale items for OLBDQ and NDI seemed comprehensive without being redundant or irrelevant, but the response scales were cumbersome. Thus, FRI's scale items are modeled from a pool of the two instruments based on the assumption that the selected items measure both cervical and lumbar conditions with equal responsiveness without affecting other psychometric qualities. Nine of the scale items for FRI originated from the NDI and OLBDQ instruments. (Seven were from the NDI instrument, and eight were from the OLBDQ instrument.) (See Table 2.) Selected as the tenth item for this instrument was frequency of pain because it is associated with predicting recovery from musculoskeletal pain,[25] medication use, impairment as rated by an independent board-certified physiatrist,[6] and time to return to work.[31]

In an effort to improve on the response scales of the existing instruments, another search of the literature was performed to discover which response scale might reduce administrative burden without sacrificing responsiveness or reliability. Although the visual analog scale offers a wide range of choice, it yields a clustering of responses, sometimes requires teaching, and can be difficult to score and communicate.[11] Overall, it poses an administrative burden. Hayes concluded that scales with five response options are more sensitive than those with two response options (nominal scales [yes/no] may make an instrument insensitive to small but clinically significant changes), but sensitivity seems to level off after five points, suggesting only incremental utility of a scale with more than five points.[17] In a subsequent study, Guyatt et al compared a 7-point Likert scale and a visual analog scale and concluded that the responsiveness between the scales was similar and recommended that the Likert scale be the response scale of choice because of its ease of administration and interpretation.[16] Nagata et al also found that there was no difference in responsiveness or reliability among the 4, 5, and 7-point scales and the visual analog scales. The researchers concluded that the 5-point scale was the easiest to complete, had the least frequent item omission, and was the most useful when measuring health status.[30]

Table 2. A Comparison of Scale Items From OLBDQ, NDI, and FRI

Scale Item	OLBDQ, Revised	NDI	FRI
Pain intensity	√	√	√
Personal care	√	√	√
Lifting	√	√	√
Sleeping	√	√	√
Social life, recreation	√	√	√
Traveling, driving	√	√	√
Walking	√		√
Standing	√		√
Sitting	√		
Change in pain pattern	√		
Work		√	√
Reading		√	
Headaches		√	
Concentration		√	
Frequency of pain			√

OLBDQ = Oswestry Low Back Disability Questionnaire; NDI = Neck Disability Index; FRI = Functional Rating Index.

General Description of The Functional Rating Index. The FRI instrument contains 10 items that measure both pain and function of the spinal musculoskeletal system. See Appendix. Of these 10 items, 8 refer to activities of daily living that might be adversely affected by a spinal condition, and 2 refer to two different attributes of pain. Because many spinal disabilities are most likely a combination of loss of function and pain and/or the fear of pain, using both pain and function allows for a wider view of a patient's disability.

Using a 5-point scale for each item, the patient ranks his or her perceived ability to perform a function and/or the quantity of pain at the present time ("right now") by selecting one of the five response points that are anchored by bipolar statements (0 = no pain or full ability to function; 4 = worst possible pain and/or unable to perform this function at all).

Scoring. The 10 items of the FRI may be used to profile the nature of the dysfunction and pain, or they may be totaled. The index score is achieved by simply summing up the equally weighted scores, dividing by the total number of possible points, and multiplying by one hundred percent. The range of scores is zero percent (no disability) to 100% (severe disability). The higher the number, the higher the perceived dysfunction and pain; the lower the number, the lower the perceived dysfunction and pain. Following is a calculation example:

1) When all 10 items are completed, the FRI score is calculated as follows:

$$(\text{total score}/40) \times 100\%$$

Little and MacDonald concluded that a relative value (percentage) change in score is a better indicator of outcome than the absolute value change.[26] To calculate pre- and post-test changes, the following formula is recommended:

$$\Delta \text{ FRI } = (\text{pre-FRI\%} - \text{post-FRI\%})/\text{pre-FRI\%}$$

To accommodate for scoring irregularities, the following decision rules were established: 1) When a subject marked two responses on the same item, the responses were averaged; 2) when a subject marked in between two response numbers, the responses were averaged; and 3) when a subject wrote N/A or did not score an item, that subject was excluded from the study.

Statistical Analysis. Descriptive statistics were calculated at baseline (mean, standard error, range), and reliability was established by testing the instrument's stability and internal consistency. The intraclass correlation coefficient ($ICC_{3,k}$) was used to test instrument stability.[32] This statistical test evaluates both the strength of correlation and the concordance (sameness).[23] Two measures were used to test instrument internal consistency: 1) Cronbach's alpha, to measure the correlation between items, and 2) the Spearman rho correlation coefficient, to measure interitem correlation by comparing individual scale items with the index score. Pearson's product moment correlation coefficient was used to test FRI's construct validity. The instrument's responsiveness was tested with the standardized response mean for index score. The differences in baseline score due to age and education were assessed using one-way analysis of variance, whereas gender, marital status, occupation, ethnicity, and days of symptoms were assessed with chi-square.

Note: All data were entered into a database that was analyzed with Minitab 10.51 Xtra (State College, PA).

Table 3. Overview

Setting	Population*	Psychometric Quality Tested	Sample Size
Colorado Springs, CO	400,000	Reliability; test-retest	
Tacoma, WA	590,000	Reliability; test-retest	
Portland, OR	1,240,000	Reliability; test-retest	
San Francisco, CA	1,600,000	Reliability; test-retest	N = 51
Spokane, WA	360,000	Validity	N = 24
Orofino, ID	9,000	Responsiveness	
Lawton, OK	110,000	Responsiveness	N = 36
Seattle, WA	1,920,000	Clinical utility	N = 28
Total sample size			N = 139

* Populations of U.S. Metro areas, 1990.

Overview—Patient Population, Setting and Cohorts. Eight chiropractic clinics in six states with communities ranging in population from 9000 to 1,920,000 participated in the study. Subjects from these clinics were assigned to one of four cohorts. See Table 3.

Subjects were ambulatory and self-selected, and all were seeking professional care for spinal pain and dysfunction (cervical, thoracic, lumbar) at chiropractic practices. Consecutive patients in the clinics were offered participation in this study, and appropriate steps were taken to ensure protection of human subjects by providing written information and obtaining documented consent from each. The following inclusion criteria were used: 1) able to read and write English, 2) signed an informed consent, 3) intact cognitive status, and 4) spinal pain/dysfunction as the chief compliant. Excluded from clinical utility testing (the time required for administration) were those subjects with reading limitations or language barriers that would require translations by a second party.

Personnel Training. The clinics participating in this study had no previous research experience, and each physician-participant signed an informed consent. Standardized training was achieved by providing each field research physician with training manuals that included detailed instructions and follow-up regarding proper protocols, data entry templates and instructions for data collection, and directives and scripts detailing the receptionist's and physician's duties and communications.

New patients were asked to complete a new patient research packet, which included an informed consent, patient case history, and FRI. One cohort received FRI, a 12-item short-form health survey (SF-12),[43] and a Disability Rating Index (DRI)[35] (order of materials were counter-balanced). The following demographic variables were also collected: age, gender, marital status, race, education, occupation, primary International Classification of Diseases code, and time with primary complaint. The completed forms were photocopied with case codes (all patient personal identifiers were removed) and mailed for entry into the computer. All entries were double-checked for accuracy.

Hypotheses. The purpose of this study was to test FRI on the following qualities. First, *reliability:* There should be no statistical variation in FRI results when the instrument is administered to the same participant at two different times. Furthermore, FRI should have high internal consistency among the different questions of the instrument and should have high item analysis consistency. Second, *validity:* FRI should correlate

strongly with the SF-12 Physical Component Score (PCS) and the DRI (the disability rating instrument designed to measure physical function). Ideally, it should have a weak correlation with the SF-12 Mental Component Score (MCS). Third, *responsiveness:* FRI should measure clinically significant physical improvement after chiropractic treatment (according to Cohen, a large size effect is greater than 0.8).[5] Fourth, *clinical utility:* The administrative burden for both patients and health care teams should be less than 2 minutes, and FRI should have a very high completion rate. Additionally, with regard to *sociodemographics*: Gender, age and education should not affect the index score.

Reliability. The two most important properties of reliability are stability and consistency. The *coefficient of stability* is primarily measured by test-retest.[4] Patients at participating clinics were tested on the day of admission and retested on the next scheduled visit. To reduce the memory effect, the patients completed one of the following randomly selected FRI instruments on the initial test and a different version on the retest: 1) normal order, 2) scramble order of test items, or 3) reverse order of the scale. Intraclass correlation coefficient was used to measure stability (a score above 0.80 is considered acceptable).[1]

The *coefficient of internal consistency* is mainly assessed with Cronbach's alpha[7] and item-scale correlation. Item-scale correlation should be greater than 0.40,[20] and interitem and item-total correlations need to be examined for item redundancy. If items have an excessively high interitem correlation (> 0.80), one item may be eliminated. Cronbach's alpha should be greater than 0.80.[1]

Validity.

Content validity. Content validity determines whether the instrument includes relevant items related to what is being measured.[38] To ensure that the items used represent a relevant domain, doctors in 12 chiropractic clinics (located in eight different states) with feedback from their patients supplied content (comprehensive, not redundant or irrelevant) and context (syntactic and scaling) input for the improvement of this instrument over a pilot period of 1 year.

Construct validity. Construct validity is the most important validity test, because criterion validity testing is not possible in the absence of a gold standard against which to test the criterion.[10] Following are the generally accepted rankings for construct validity coefficients: 1.0–0.81 (excellent), 0.80–0.61 (very good), 0.60–0.41 (good), 0.40–0.21 (fair), and 0.20–0 (poor). The Functional Rating Index was tested with DRI, SF-12 PCS, and SF-12 MCS.

Responsiveness. An instrument's responsiveness refers to its ability to detect clinically important change, even if that change is small over time; to discriminate between clinically important and clinically unimportant changes; and to measure true change. Patients were pretested, given 1 month of chiropractic treatment and post-tested. If the initial score of an instrument is low, there is little room for improvement—a floor effect. If the initial score is high, it may overmeasure severity—a ceiling effect.

Effect size. Standardized response mean (SRM) is a statistic that describes the magnitude of responsiveness. Following are the responsiveness standards established by Cohen: small size effect (0.2); medium size effect (0.5); and large size effect (≥0.8).[5]

$$SRM = delta \text{ (mean change)}/sigma \text{ (the standard deviation at baseline)}^{19}$$

Clinical utility. In addition to reliability, validity, and responsiveness, Simmonds recommends the determination of instrument clinical utility.[36] The following are several components of clinical utility: 1) Is it quick and simple to administer? 2) Does it aid in the care of the patient? and 3) Does it present an easy alternative method for gathering data (*e.g.*, by mail, phone)?

Ease of administration was measured with completion rates. If the instrument is simple to use, it is easy to read, coherent, and easy to answer, and it will have high completion ratios. The following are the formulas used to calculate completion ratios for FRI: 1) the number of FRIs with any missing response divided by the total number of FRIs; and 2) the number of items without a response divided by the total number of items.

Time use was measured by carefully timing the following with a calibrated stop watch: 1) time to complete the instrument (measured without the knowledge of the participants); and 2) time to score the instrument (a secretary unfamiliar with FRI scored the instrument and was measured without her knowledge).

■ Results

One hundred percent of the patients (n = 180) who were asked to participate in the research study agreed to do so. Five of those patients were unable to read and write English (no data were gathered from them), and 25 had missing values on the FRI (only their demographic data were used). From the 35 patient cohort that tested validity, 11 patients who completed the FRI failed to complete the SF-12 or DRI. These 11 were included in the demographic data but were not included in the cohort for purposes of comparing FRI, SF-12, and DRI. Overall, 41 patients were excluded and 139 patients participated in this study.

Table 4 compares the subjects who completed the FRI with those who had missing values. Analysis of variance testing for age (P = 0.10) and education (P = 0.33) produced P values that were not significant. Chi-Square testing for gender (P = 0.45), marital status (P = 0.50), occupation (P = 0.10), ethnicity (P = 0.44), and days of symptoms before first visit (P = 0.30) also yielded P values that were not significant.

Table 5 displays the basic statistics for the FRI at baseline. The mean index score was 47.9 with a standard deviation of 22.1. The effect of gender, age, and education on total score was as follows: One-way analysis of variance testing for gender (female n = 75, male n = 47; P = 0.96), age (0–28, n = 32; 29–41, n = 34; 42–51, n = 27; and ≥52, n = 29; P = 0.38), and education (≤12 years, n = 54 and >12 years, n = 68; P = 0.62) yielded P values that were not significant.

Reliability

The time interval between the two tests averaged 1.95 days (1–11). A scatter diagram of the data established a linear relationship between the variables, and intraclass correlation coefficient ($ICC_{3,k}$) was 0.99 (n = 51). Cron-

Table 4. Comparison of Subjects Who Completed the FRI and Those Who Failed to Do So

No Missing Values, n = 150			Missing Values, n = 25		
Gender, %					
Female	60		Female	68	
Male	40		Male	32	
Ethnic, %					
White	89.3		White	84	
African-American	4.7		African-American	4.0	
Hispanic	2.0		Hispanic	8.0	
Other	4.0		Other	4.0	
Occupation, %					
White collar	47.3		White collar	48.0	
Blue collar	26.0		Blue collar	8.0	
Retired	8.7		Retired	24.0	
Student	9.3		Student	8.0	
Other	8.7		Unemployed	4.0	
			Other	8.0	
Marital status, %					
Married	54.0		Married	56.0	
Single	29.3		Single	36.0	
Divorced	11.3		Divorced	0	
Widowed	4.0		Widowed	8.0	
Separated	1.3		Separated	0	
Interval of days between the start of symptoms and first visit, %					
0–42 days	56.0		0–42 days	76.0	
43–365 days	18.0		43–365 days	8.0	
366 days–5 yrs	14.7		366 days–5 yrs	8.0	
>5 yrs	11.3		>5 yrs	8.0	
Age, yrs					
Mean	SD	Range	Mean	SD	Range
41	15.8	12–78	46	19.2	17–82
Education, yrs					
Mean	SD	Range	Mean	SD	Range
13.8	2.4	6–22	13.3	2.5	8–18

FRI = Functional Rating Index.

bach's alpha was 0.92 for all variables with the individual items scoring either 0.91 or 0.92 (n = 51). Interitem and item-total correlations (Spearman's rho) indicate that item efficiency was good, ranging between 0.54 and 0.82, with a moderate correlation among all items (n = 51).

Table 5. Basic Statistics for the FRI at Baseline

Variable	n = 111*				
	Mean	SEM	SD	Min	Max
Index score	47.9	2.1	22.1	2.5	97.5
Pain intensity	2.1	0.1			
Sleeping	1.8	0.1			
Personal care	1.4	0.1			
Travel	2.0	0.1			
Work	1.4	0.1			
Recreation	1.8	0.1			
Frequency of pain	2.6	0.1			
Lifting	2.0	0.1			
Walking	1.8	0.2			
Standing	2.3	0.1			
All scale items ranged from 0–4.					

* Note that the 28 patients from the clinical utility cohort were tested only for administrative burden, and their data was not included.
FRI = Functional Rating Index; SEM = standard error of the mean; SD = standard deviation.

Validity
Table 6 demonstrates the relationship between FRI and DRI and the SF-12. Pearson's product moment test produced the following correlations: The Functional Rating Index correlated strongly with SF-12 PCS (0.76) and DRI (0.76) and weakly with the SF-12 MCS (0.36). The correlation for the eight items representing function and the index score was 0.99; for the two items representing pain and the index score, the correlation was 0.80; and for both the pain items and function items, it was 0.71. A scatter diagram of the data established a linear relationship between the variables (n = 51).

Responsiveness
The overall standardized response mean was 1.24 (n = 36); cervical was 1.26 (n = 17); thoracic was 1.61 (n = 5); and lumbar was 1.24 (n = 14).

Clinical Utility
FRI was easy for patients to understand with no training, and no subject asked for or was given more instruction than was printed with the FRI. As a self-administered test, FRI was correctly completed by 88% of the subjects without more instruction than is contained in the questionnaire, and 98% of the scales were completed. The administration of this instrument required a mean of 64 (12–140) seconds for the patient to complete and 14 (5–30) seconds for a staff member to score.

■ Discussion

None of the variables in the two demographic cohorts (those with and without missing FRI data) differed significantly. Implications regarding bias cannot be drawn. The index scores produced a reasonable Gaussian distribution with a mean score of 48%, which minimizes the possibility of a floor effect (not being able to measure those patients with mild problems). This suggests that FRI may be even more responsive in settings with more severe conditions and that the instrument is generally sensitive to a variety of complaints found in the sample. Neither gender, age, nor education predicted any effect on the FRI scores. This suggests the index is acceptable for a wide variety of subjects.

The coefficient of stability (test-retest) using intraclass correlation coefficient was decidedly above the acceptable value.[1] The coefficient of internal consistency, using

Table 6. Construct Validity

Correlations (Pearson's):	N = 24		
	FRI	DRI	SF-12 PCS
DRI	0.76		
SF-12 PCS	−0.76	−0.77	
SF-12 MCS	−0.36	−0.33	0.06

A scatter diagram of the data established a linear relationship between the variables.
FRI = Functional Rating Index; DRI = Disability Rating Index; SF-12 PCS = Short Form Physical Component Score; MCS = Mental Component Score.

Cronbach's alpha, was also above the acceptable value.[1] The item-total correlations were highly correlated with the FRI total score. These correlations establish each item's intercorrelation and effective contribution to overall measurement (efficiency of each item). Using interitem correlations, FRI items correlate moderately with each other, which suggests that redundancy risk is low and that each item focuses on a different perspective of the same attribute. Thus, FRI has been demonstrated to be reliable.

It appeared to experienced clinicians and their patients that FRI measures spinal function and pain, allowing the instrument to perform as designed and to demonstrate good content validity. Although the SF-12 might be considered the "gold standard" for a general measure of health status, there is no irrefutable "gold standard" for assessing functional status. Thus, the authors limited their efforts to construct validity. The Functional Rating Index had a strong association with SF-12 PCS and DRI, which both relate to physical health and function (convergent construct validity), and a much weaker association with SF-12 MCS, which relates to mental health (divergent construct validity), indicating that FRI probably measures some attributes pertaining to mental health but is more relevant to physical health. These findings suggest that a single and distinguishable phenomenon was probably being measured, providing evidence for construct validity.

Thirty-three percent of the patients in the cohort testing responsiveness had a chronic condition, and all patients received 1 month of chiropractic treatment. The overall standardized response mean for FRI was 1.24, which is greater than Cohen's 0.8 (a size effect of 1.0 is a gain of one standard deviation).[5] Moreover, responsiveness was noted with cervical, thoracic, and lumbar subjects.

The clinics that participated in this research study had no previous research training or culture and, as such, represent typical chiropractic practices. Even so, FRI was easy for patients to understand with no training and required just a little more than one minute (78 seconds) for the patient to complete and a staff member to score. The time required to complete and score the instrument was reasonable, the instrument was easy to administer, score and analyze, and the results were readily interpretable and easy to communicate to the patient. Additionally, with some minor modification, this instrument could be used to gather data by mail or phone.

Study Limitations

There are several study limitations that must be expressed. Demographic variables for this sample are comparable to those found in a survey of chiropractic patients by Hurwitz et al; however, these results may not be extrapolated to other patient populations or settings.[18] Replication of the results in other samples by other research teams is necessary to further investigate this instrument.

The interval between the test and retest was short. A memory effect may have influenced the results,[1] despite attempts to reduce such an effect. The FRI should be retested with a functionally stable cohort with a time interval of 1 or 2 weeks.

The responsiveness of this instrument was gauged with a before-and-after design. As such, the change in scores may have been because of the natural history of the condition, the Hawthorne effect, and/or the placebo effect. Ideally, responsiveness should be tested with a Randomized Clinical Trial design (randomization, a placebo control group and blinding) using prevalidated instruments and a global assessment scheme with patients and experienced clinicians. In this manner, responsiveness and validity could be tested further. Additionally, because of the pattern of FRI's response scales, the instrument should be tested for a response set bias.

The Functional Rating Index is offered as a promising useful instrument in the assessment of spinal conditions. Additional research is needed to compare and improve this instrument. Such research might include a comparison of the sensitivity of FRI with that of OLBDQ for low back patients and with that of NDI for neck patients. Additionally, patients from other sources and with a wide variety of severities should be tested to determine whether FRI is a valuable instrument for the researcher and clinician.

■ Conclusion

The Functional Rating Index is a brief, practical, and inexpensive way to quantify patient physical function and pain related to spinal conditions. The initial evaluation of FRI is encouraging. The psychometric qualities of FRI appear to be stable with regard to reliability, validity, and responsiveness. Moreover, clinicians, researchers, and patients will find this instrument less burdensome than its predecessors.

■ Key Points

● Existing self-reporting instruments measuring spinal pain and dysfunction place an undue administrative burden on patients and health care workers, and, therefore, are underutilized in clinical settings.

● A new instrument, Functional Rating Index (FRI), which combines Oswestry and Neck Disability Index content in a format that reduces administrative burden, establishes itself as a psychometrically robust clinical assessment instrument for spinal pain and dysfunction.

● Based on preliminary research, FRI demonstrates excellent reliability, validity, and responsiveness, and significantly reduces administrative burden.

References

1. Bellamy N. Musculoskeletal Clinical Metrology. Boston: Kluwer Academic Publishers, 1993:11–43.

2. Beurskens AJ, de Vet HC, Koke AJ, et al. Measuring the functional status of patients with low back pain: Assessment of the quality of four disease-specific questionnaires. Spine 1995;20:1017–28.

3. Bogduk N. Musculoskeleletal pain: Towards precision diagnosis. In: Jensen TS, Turner JA, Wiesenfeld-Hallin Z, eds. Progress in Pain Research and Management, vol. 8. Seattle: IASP Press, 1997:507–25.

4. Borg WR, Gall, MD Educational research: An introduction. New York: Longman Publishers, 1992:260.

5. Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1977:1–27.

6. Connally GH, Sanders SH. Predicting low back pain patients' response to lumbar sympathetic nerve blocks and interdisciplinary rehabilitation: The role of pretreatment overt pain behavior and cognitive coping strategies. Pain 1991;44:139–46.

7. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.

8. Deyo RA. Measuring the functional status of patients with low back pain. Arch Phys Med Rehabil 1988;69:1044–53.

9. Deyo RA, Anderson G, Bombardier C, et al. Outcome measures for studying patients with low back pain. Spine 1994;19:2032S–6S.

10. Deyo RA, Battie M, Beurskens AJHM, et al. Outcome measures for low back pain research. Spine 1998;23:2003–13.

11. Dixon JS, Bird HA. Reproducibility along a 10 cm vertical visual analog scale. Ann Rheum Dis 1981;40:87–9.

12. Epstein AM. The outcomes movement—Will it get us where we want to go? N Engl J Med 1990;323:266–70.

13. Fairbank JC, Couper J, Davies JB, et al. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66:271–3.

14. Feise RJ, RJF Consulting, Prescott, Arizona. Cross-sectional study of chiropractic patients: The frequency of multiple conditions. 1999 (unpublished data).

15. Grönblad M, Hupli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. Clin J Pain 1993;9:189–95.

16. Guyatt GH, Townsend M, Berman LB, et al. A comparison of Likert and visual analog scales for measuring change in function. J Chron Dis 1987;40:1129–33.

17. Hayes BE. Measuring customer satisfaction: Development and use of questionnaires. Milwaukee, WI: Quality Press, 1992:58–60.

18. Hurwitz EL, Coulter ID, Adams A, et al. Use of chiropractic services from 1985 through 1991 in the United States and Canada. Am J Public Health 1998;88:771–6.

19. Kaziz LE, Anderson JJ, Meenan RF. Effect size for interpreting changes in health status. Med Care 1989;27:S178–89.

20. Kline P. A Handbook of Test Construction. London: Methuen, 1986:143.

21. Kopec JA, Esdaile JM. Functional disability scales for back pain. Spine 1995;20:1943–9.

22. Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale. Measurement properties. Spine 1995;20:341–52.

23. Kramer MS, Feinstein AR. Clinical biostatistics LIV. The biostatistics of concordance. Clin Pharmacol Ther 1981;29:111–23.

24. Leak AM, Cooper J, Dyer S, et al. The Northwick Park Neck Pain Questionnaire, devised to measure neck pain and disability. Br J Rheumatol 1994;33:469–74.

25. Linton SJ, Hallden K. Risk factors and the natural course of acute and recurrent musculoskeletal pain: Developing a screening instrument. In: Jensen TS, Turner JA, Wiesenfeld-Hallin Z, eds. Progress in Pain Research and Management, vol. 8. Seattle: IASP Press, 1997:527–36.

26. Little DG, MacDonald D. The use of the percentage change in Oswestry Disability Index score as an outcome measure in lumbar spinal surgery. Spine 1994;19:2139–43.

27. Manniche C, Asmussen K, Lauritsen B, et al. Low Back Pain Rating scale: Validation of a tool for assessment of low back pain. Pain 1994;57:317–26.

28. Manniche C, Asmussen KH, Vinterberg H, et al. Back pain, sciatica and disability following first-time conventional haemilaminectomy for lumbar disc herniation. Use of "Low Back Pain Rating Scale" as a postal questionnaire. Dan Med Bull 1994;41:103–6.

29. Millard RW, Jones RH. Construct validity of practical questionnaires for assessing disability of low-back pain. Spine 199;16:835–8.

30. Nagata C, Ido M, Shimizu H, et al. Choice of response scale for health measurement: Comparison of 4, 5, 7-point scales and visual analog scale. J Epidemiol 1996;6:192–7.

31. Ohlund C, Lindstrom I, Areskoug B, et al. Pain behavior in industrial subacute low back pain. Part I. Reliability: Concurrent and predictive validity of pain behavior assessments. Pain 1994;58:201–9.

32. Portney LG, Watkins MP. Foundations of Clinical Research: Application to Practice. East Norwalk, CT: Appleton and Lange, 1993:509–16.

33. Roach KE, Brown, MD, Dunigan KM, et al. Test-retest reliability of patient reports of low back pain. J Orthop Sports Phys Ther 1997;26:253–9.

34. Ruta DA, Garratt AM, Wardlaw D, et al. Developing a valid and reliable measure of health outcome for patients with low back pain. Spine 1994;19:1887–96.

35. Salen BO, Spangfort EV, Nygren AL, et al. The disability rating index: an instrument for the assessment of disability in clinical settings. J Clin Epidemiol 1994;47:1423–34.

36. Simmonds, M. Physical function and physical performance in patients with pain: What are the measures and what do they mean? In: Devor M, chairman, Max M, ed. Pain 1999—An Updated Review: Refresher Course Syllabus. Seattle, WA: IASP Press, 1999:127–36.

37. Stratford PW, Binkley JM. Measurement properties of the RM-18. A modified version of the Roland-Morris Disability Scale. Spine 1997;22:2416–21.

38. Streiner DL, Norman GR. Health Measurement Scales: A Practice Guideline to Their Development and Use. 2nd ed. New York: Oxford University Press, 1995:20–2.

39. Strong J, Ashton R, Large RG. Function and the patient with chronic low back pain. Clin J Pain 1994;10:191–6.

40. Tesio L, Granger CV, Fiedler RC. A unidimensional pain/disability measure for low-back pain syndromes. Pain 1997;69:269–78.

41. Vernon H, Mior S. The neck disability index: A study of reliability and validity. J Manipulative Physiol Ther 1991;14:409–18.

42. Waddell G, Main CJ, Morris EW, et al. Chronic low back pain, psychological distress, and illness behavior. Spine 1984;9:209–13.

43. Ware JE, Kosinski MA, Keller SD. A 12-item short-from health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–33

Address reprint requests to

Ronald J. Feise, DC
1151 E. Timber Ridge Road
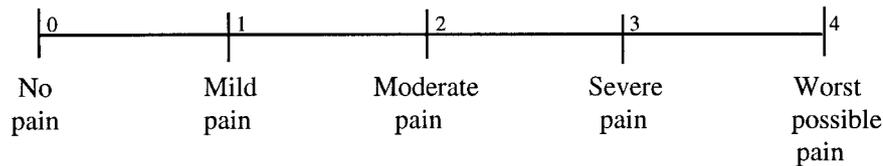*Prescott, AZ 86303*
E-mail: rjf@chiroevidence.com
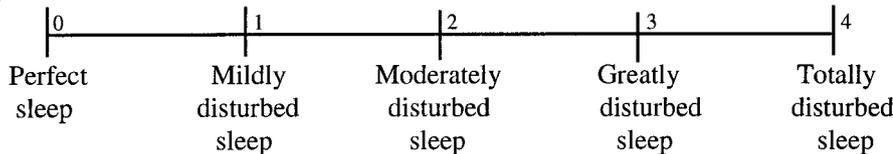
■ **Appendix**

Functional Rating Index

For use with <u>Neck and/or Back Problems</u> only.

In order to properly assess your condition, we must understand how much your **neck and/or back problems** have affected your ability to manage everyday activities. For each item below, **please circle the number which most closely describes your condition right now.**
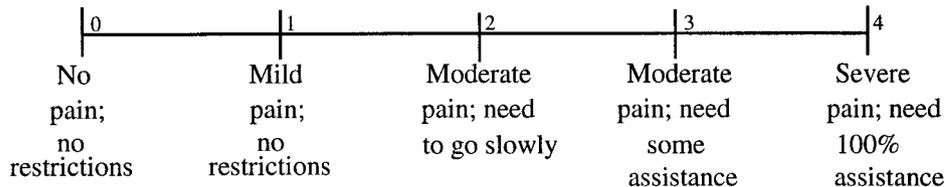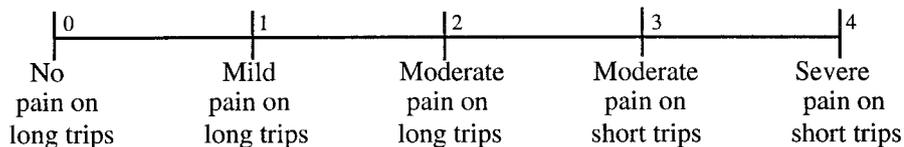
1. Pain Intensity

0	1	2	3	4
No pain	Mild pain	Moderate pain	Severe pain	Worst possible pain

2. Sleeping

0	1	2	3	4
Perfect sleep	Mildly disturbed sleep	Moderately disturbed sleep	Greatly disturbed sleep	Totally disturbed sleep

3. Personal Care (washing, dressing, etc.)

0	1	2	3	4
No pain; no restrictions	Mild pain; no restrictions	Moderate pain; need to go slowly	Moderate pain; need some assistance	Severe pain; need 100% assistance

4. Travel (driving, etc.)

0	1	2	3	4
No pain on long trips	Mild pain on long trips	Moderate pain on long trips	Moderate pain on short trips	Severe pain on short trips

5. Work

0	1	2	3	4
Can do usual work plus unlimited extra work	Can do usual work; no extra work	Can do 50% of usual work	Can do 25% of usual work	Cannot work

6. Recreation

0	1	2	3	4
Can do all activities	Can do most activities	Can do some activities	Can do a few activities	Cannot do any activities

7. Frequency of pain

0	1	2	3	4
No pain	Occasional pain; 25% of the day	Intermittent pain; 50% of the day	Frequent pain; 75% of the day	Constant pain; 100% of the day

8. Lifting

0	1	2	3	4
No pain with heavy weight	Increased pain with heavy weight	Increased pain with moderate weight	Increased pain with light weight	Increased pain with any weight

9. Walking

0	1	2	3	4
No pain; any distance	Increased pain after 1 mile	Increased pain after 1/2 mile	Increased pain after 1/4 mile	Increased pain with all walking

10. Standing

0	1	2	3	4
No pain after several hours	Increased pain after several hours	Increased pain after 1 hour	Increased pain after 1/2 hour	Increased pain with any standing

_____   _____
Patient's Signature   **Date**

■ Point of View

Dan Cherkin, PhD
Group Health Center for Health Studies
Seattle, Washington

Drs. Feise and Menke have done a remarkably thorough job of evaluating the psychometric properties of their Functional Rating Index (FRI), a new clinician-friendly instrument for assessing both back and neck pain. The authors' primary rationale for developing this hybrid instrument was to increase the clinical use of such measures by minimizing administrative burden. The authors clearly succeeded in achieving this goal, as the 10-item FRI required only 78 seconds for the patient to complete and the staff to score!

The value of this instrument for researchers is less clear. Most spine research focuses on a specific region of the spine and uses an assessment tool specific to that region. Thus, studies of low back pain tend to use well-validated measures such as the Roland or the Oswestry disability questionnaires, whereas studies of neck pain often use the Neck Disability Index (NDI). Given that the FRI includes 8 of the 10 Oswestry items and 7 of the 10 NDI items, one might expect these three instruments to have similar properties. Nevertheless, researchers are reluctant to abandon tried and tested measures every time an attractive new measure appears. Compared with clinicians, researchers are more concerned about the validity, reliability, and responsiveness of their measures and less concerned with finding a measure than can easily be incorporated into clinical practice.

Although the potential value of the FRI for clinicians is clear, it will need to be left to future studies to determine whether the FRI has much to offer researchers. Future studies of back and neck pain should consider including the FRI along with back- or neck-specific outcome measures to permit a head-to-head comparison. Until then, researchers may be best advised to continue to use the well-validated region-specific measures.