

Physical Impairment Index: Reliability, Validity, and Responsiveness in Patients with Acute Low Back Pain

Julie M. Fritz, PhD, PT, ATC,* and Sara R. Piva, MS, PT†

Study Design. Cohort study of patients with acute low back pain undergoing physical therapy.

Objectives. Examine the reliability and validity of the Physical Impairment Index in a group of patients with acute low back pain and determine the responsiveness and minimum detectable change of the index and its component tests.

Summary of the Background Data. The Physical Impairment Index was originally described as a reliable and valid means of assessing physical impairment in patients with low back pain. The psychometric properties of the index have not been reported in patients with acute low back pain, nor has its responsiveness been examined.

Methods. Seventy-eight patients with acute (<3 weeks duration) low back pain participating in a clinical trial were assessed at baseline and after 4 weeks. Interrater reliability of the index was examined in a subgroup of 20 patients. Validity was examined through correlations with concurrent measures of pain, disability, and psychosocial variables. Changes in the index over 4 weeks were used to assess responsiveness and minimum detectable change.

Results. Interrater reliability of the index was high (intraclass correlation coefficient = 0.89), and its validity was generally supported by the pattern of correlations. The index was more responsive to change than any of its component tests but was less responsive than the Oswestry disability score. The minimum detectable change on the index was approximately 1 point.

Conclusions. The Physical Impairment Index appears to be a reliable and valid measure of physical impairment for patients with acute low back pain and may be useful as an adjunct outcome measure for studies involving these patients. Further research on patients with chronic pain is needed before it can be advocated for outcomes research with this population. [Key words: physical impairment, responsiveness, minimum detectable change, outcome measures] **Spine 2003;28:1189–1194**

most common impairments assessed in patients with LBP have been range of motion (ROM) measurements of the lumbar spine, straight leg raise ROM, and muscle strength.⁶ When examined individually, these impairment measurements have shown questionable convergent validity with measures of disability, pain, and patient satisfaction outcomes, and little discriminant validity in distinguishing between individuals with and without LBP.^{7,8,9,10} Despite these findings, lumbar impairments continue to figure prominently in the American Medical Association's guidelines for the determination of permanent disability related to LBP¹¹ and are frequently monitored by clinicians for determining the progress of patients in their care.^{12,13,14}

Waddell *et al*³ attempted to overcome some of the difficulties with the measurement of physical impairment in patients with LBP by identifying a group of tests that could be measured reliably and would successfully discriminate between individuals with and without LBP. The final Physical Impairment Index (PII) included seven tests, each scored positive or negative based on published cut-off values (Table 1). The final score of the PII therefore ranges between 0 and 7, with higher numbers indicating greater levels of impairment. Waddell *et al*³ reported on the reliability and validity of the PII in a group of 120 patients with chronic LBP (>3 months duration). The authors reported good or excellent reliability for the individual items of the PII (intraclass correlation coefficient [ICC] or κ coefficients ranging between 0.48 and 0.96). Convergent validity of the PII was supported by significant correlations with disability ($r = 0.51$ with the Roland Morris Questionnaire), work loss in the past year ($r = 0.43$), and pain ($r = 0.27$). The authors also reported high correlations of the PII with concurrent measures of depressive symptoms ($r = 0.26$ with Zung Depressive Inventory), somatization ($r = 0.32$ with the Modified Somatic Perception Questionnaire), nonorganic signs ($r = 0.49$), and nonorganic symptoms ($r = 0.35$).³

Since its initial description, the PII has been used as an outcome measure in several studies involving patients with both acute and chronic LBP.^{15,16,17} However, the reliability and validity of the PII has not been reported for patients with acute LBP, nor has the responsiveness of the PII been studied. The purpose of this study was to examine the reliability and validity of the PII in a group of patients with acute LBP. We also examined the responsiveness and minimum detectable change (MDC) of the PII and its component tests.

Several studies have confirmed the generally low correlations among pain, disability, work status, and physical impairment in patients with low back pain (LBP).^{1,2,3} Physical impairment represents a loss or abnormality of physiologic or anatomic structure or function.^{4,5} The

From the *Department of Physical Therapy and the †School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania.

Supported by a grant from the Foundation for Physical Therapy. Acknowledgement date: April 30, 2002. First revision date: September 23, 2002. Acceptance date: October 28, 2002.

The submitted manuscript does not contain information about medical devices or drugs.

Foundation funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence to Julie M. Fritz, PhD, PT, ATC, 6035 Forbes Tower, Pittsburgh, PA 15260, USA; E-mail: jfritz@pitt.edu

Table 1. Items Included in the Physical Impairment Index Along with Scoring Criteria

Test	Performance	Scoring	
		Positive	Negative
Total Flexion Range of Motion	The patient stands erect. The inclinometer is held at T12-L1 and the patient is asked to reach down as far as possible towards the toes while keeping the knees straight.	Less than 87°	Greater than or equal to 87°
Total Extension Range of Motion	The patient stands erect. The inclinometer is held at T12-L1. The patient is asked to arch backwards as far as possible. The examiner may support the patient with one hand on the shoulder for balance.	Less than 18°	Greater than or equal to 18°
Average Lateral Flexion Range of Motion	The patient stands erect with the inclinometer aligned vertically in line with the spinous processes of T9 and T12. The patient is asked to lean straight over to one side as far as possible with the fingertips reaching straight down the side of the thigh. The examiner may support the patient's shoulder with one hand. The patient repeats the motion to the opposite side. Average lateral flexion is computed.	Less than 24°	Greater than or equal to 24°
Average Straight Leg Raise Range of Motion	The patient is supine. The inclinometer is positioned on the tibial crest just below the tibial tubercle. The leg is raised passively by the examiner, whose other hand maintains the knee in extension. The leg is raised slowly to the maximum tolerated straight leg raise (not the onset of pain). The opposite leg is tested in the same manner. Average straight leg raise is computed.	Males: less than 66° Females: less than 71°	Males: greater than or equal to 66° Females: greater than or equal to 71°
Spinal Tenderness	The patient is prone with the back muscles relaxed. Palpation is done slowly without sudden pressure. Superficial tenderness to light skin pinch is assessed first. The patient is asked "is that painful?" any response other than "no" is a positive. If superficial tenderness is not present, deep tenderness is assessed by placing firm pressure with the ball of the thumb over the spinous processes and inter-spinous ligaments within 1 cm of the midline from T12 to S2. Again the patient is asked, "is that painful?" Any response other than "no" is a positive test.	Positive superficial or deep spinal tenderness	Negative superficial or deep spinal tenderness
Bilateral Active Straight Leg Raise	The patient is supine with the knees extended. The patient is asked to lift both legs together six inches off the examining surface and hold that position for 5 seconds. The examiner should not count aloud or offer encouragement. The patient may not use the hands to lift the legs. If the patient is unable to maintain the lifted position for 5 seconds, the test is positive.	Positive test	Negative test
Active Sit-Up	The patient is supine with the knees flexed to 90° and the feet flat. The examiner should hold the patient's feet with one hand. The patient is instructed to reach up with the fingertips to touch (not hold) both knees, and hold that position for 5 seconds. The examiner should not count aloud or offer encouragement. If the patient is unable to maintain the position for 5 seconds, the test is positive.	Positive test	Negative test.

■ Methods

Subjects. Seventy-eight subjects (30 women, 48 men; mean age 37.4 years, SD 10.4) participating in a clinical trial of different physical therapy treatments for patients with acute LBP were included. These 78 subjects were the total enrollment in the clinical trial. The duration of LBP for all subjects was less than 3 weeks (mean number of days 5.5, SD 4.6, range 0–19). All subjects had a chief complaint of LBP with or without lower extremity symptoms that was reported to be sustained during work-related activities and was of sufficient magnitude to necessitate modification or cessation of work duties. Subjects underwent a baseline examination during which the PII was assessed along with pain, disability, and various psychosocial measures. All subjects were referred to physical therapy for treatment. A subset of 20 subjects who were evaluated again in physical therapy within 4 days of the baseline assessment was used to test the interrater reliability of the PII. The mean number of days between assessments was 1.2 days (SD 1.0). All examinations were conducted by physical therapists trained in the performance of the PII. Physical therapy treatments received by the subjects were either based on the AHCPR clinical practice guidelines¹⁸ or a classification scheme described by Delitto *et al.*¹⁹ Subjects were reassessed on the PII and the other measurements 4 weeks after the baseline assessment.

Measurements. Subjects completed a series of self-report measures and underwent a physical examination at the time of

the initial and follow-up evaluations. The following measures were collected:

- **Physical Impairment Index:** Each of the seven components of the PII was assessed. The scores on each component were recorded, and the PII was calculated based on the established cut-off scores (Table 1).
- **Modified Oswestry Low Back Disability Questionnaire (OSW):** The OSW is a region-specific health-related quality of life questionnaire that measures disability related to low back pain.²⁰ The questionnaire consists of 10 self-reported items addressing different aspects of function. Each item is scored from 0 to 5, with higher values representing greater disability. The total score is multiplied by 2 and expressed as a percentage. The OSW used in this study was modified by the authors by substituting an item about employment or home-making ability for the item related to sex life in order to improve compliance.^{21,22} This modified OSW has been found to have high levels of test-retest reliability, validity, and responsiveness²¹.
- **Medical Outcomes Survey Short Form (SF-36):** The SF-36 was used to assess general physical and mental health. The SF-36 measures eight dimensions of health (general health, physical function, role—physical, bodily pain, social functioning, mental health, role—emotional, and vitality).²³ The scores of the eight scales can be combined to produce a physical component score (PCS) and a mental component score (MCS). The PCS and MCS are norm-referenced to a

general US population with a mean of 50 and a standard deviation of 10.^{24,25} The SF-36 has well established psychometric properties for the general population²³ and individuals with LBP²⁶.

- Pain rating: Subjects rated their current level of LBP on an 11-point scale ranging from 0 (no pain) to 10 (worst imaginable pain).²⁷ Ten-point numerical pain ratings have been shown to have concurrent and predictive validity as measures of pain intensity^{27,28}.
- Nonorganic signs and symptoms: Waddell *et al*^{29,30} have described seven symptoms²⁹ and five signs³⁰ proposed to indicate the presence of abnormal illness behavior. The seven nonorganic symptoms were assessed through patient self-report, and the signs were determined during the physical examination. High percentage agreement between examiners has been reported for the nonorganic signs.³⁰ Nonorganic signs and symptoms have shown some relation to work loss and disability^{31,32}.
- Center for Epidemiological Studies Depression Scale (CESD): The CESD was used to assess depressive symptoms.³³ The CESD consists of 20 items, each asking the subject to rate the frequency of a depressive symptom on a 0 to 3 scale, for a total possible score from 0 to 60. Higher numbers indicate greater severity of depressive symptoms. The CESD has been tested in general samples and in populations with psychiatric diagnoses and has shown adequate reliability and discriminative validity^{34,35}.
- Fear-Avoidance Beliefs Questionnaire (FABQ): The FABQ seeks to quantify the level of fear of pain and beliefs about the need to change behavior to avoid pain in individuals with LBP.³⁶ Each item of the FABQ is scored 0 to 6, with higher numbers indicating increased levels of fear avoidance beliefs. Two subscales are contained within the FABQ: a seven-item work subscale (score range 0–42) and a four-item physical activity subscale (score range 0–24). The subscales of the FABQ have shown good reliability,³⁷ and previous studies have found the FABQ work subscale to be associated with current and future disability and work loss in patients with chronic^{36,38,39} and acute³¹ LBP.

At the 4-week reassessment each subject and the subject's treating therapist were asked to rate the overall change in the subject's status since the beginning of physical therapy treatment using a 15-point rating scale described by Jaeschke *et al*.⁴⁰ The global rating of change scale ranges from -7 ("a very great deal worse") to 0 ("about the same") to +7 ("a very great deal better"). Intermittent descriptors of worsening or improving are assigned values from -1 to -6 and from +1 to +6, respectively. Therapists and subjects were blinded to each other's ratings. Ratings of the therapist and the subject were averaged in order to balance the input of the therapist and patient. We considered subjects with an average rating of +3 ("somewhat better") or greater as having undergone clinically meaningful improvement. Subjects with an average rating of +2 ("a little bit better") or less were defined as unimproved.

Data Analysis. Interrater reliability of the PII and its individual components was determined from the subset of subjects undergoing repeat assessments. Reliability for continuous data (total flexion, total extension, average lateral flexion, average SLR, and composite PII) was calculated using the ICC formula.⁴¹ Reliability for categorical data (spinal tenderness, bilat-

eral SLR, and active sit-up) was calculated using the κ coefficient.⁴² A 95% CI was calculated for each reliability coefficient.

The validity of the PII was assessed using correlation coefficients between the baseline PII and the other baseline measurements. Responsiveness of the PII was examined with the calculation of the area under a receiver operating characteristic (ROC) curve. An ROC curve was constructed for the individual components and overall PII score, OSW, and PCS by calculating the sensitivity (true-positive rate) and specificity (true-negative rate) as the change score defining clinically meaningful change varied.⁴³ Sensitivity was calculated by dividing the number of subjects identified by the scale as having improved based on the selected cut-off score by the total number of subjects defined as having improved based on the average global rating. Specificity was calculated by dividing the total number of subjects identified by the scale as not having changed by the total number of subjects defined as unimproved based on the average global rating. The ROC curve was constructed by plotting the sensitivity values on the y-axis and 1 minus the specificity values on the x-axis for different values of the change scores. The area under the curve (AUC) can be used as a quantitative method for assessing a scale's ability to distinguish patients who have improved from those who have not.⁴³

We calculated the MDC for the PII and the individual ROM measurements contained in the index. The MDC represents the amount of change in a measure that needs to be observed before the change can be considered to exceed the bounds of measurement error.^{44,45} The MDC was determined by first calculating the standard error of measurement (SEM). The SEM was calculated as $(SD \times (1-r)^{1/2})$, where r is the reliability coefficient and SD is the square root of the total score variance. We calculated the 95% CI for the SEM using the method described by Stratford and Goldsmith.⁴⁶ There is currently no consensus regarding the number of SEMs an individual's score must change for that change to exceed measurement error confidently.⁴⁷ Previous researchers have reported one SEM as the best measure of meaningful change on health-related quality of life measures.^{45,47} Some have used a $1.96 \times SEM$,^{48,49} while others have used $1.65 \times SEM$.^{50,51} We calculated the MDC using the formula $1.96 \times SEM$ to correspond to a 95% confidence level that true change had occurred.

■ Results

Seventy-eight subjects were enrolled in the study and completed the baseline examination. The 20 subjects used for the reliability analysis did not differ from the entire study population with respect to age, gender, or OSW or PII scores ($P > 0.05$). The interrater reliability coefficients for the seven components and the overall PII score are given in Table 2. The overall PII score demonstrated excellent interrater reliability (ICC 0.89, 95% CI 0.61–0.96). The reliability coefficients of the individual components varied, ranging between 0.91 and 0.35.

Pearson product correlation coefficients between the PII score and other variables assessed at baseline are given in Table 3. The highest correlations were found between the PII and the pain rating ($r = 0.47$), the OSW ($r = 0.43$), nonorganic signs ($r = 0.42$), and nonorganic symptoms ($r = 0.36$). Smaller but statistically significant correlations were found between the PII and the PCS

Table 2. Reliability Coefficients for the Individual Components and Overall Score on the Physical Impairment Index (PII)

Measurement (unit)	Reliability Coefficient (95% CI)	SEM	MDC
Total flexion (degrees)	ICC = 0.91 (0.76 to 0.97)	7.71	15.11
Total extension (degrees)	ICC = 0.45 (0.06 to 0.74)	4.03	7.90
Average lateral flexion (degrees)	ICC = 0.39 (-0.01 to 0.69)	3.68	7.21
Average SLR (degrees)	ICC = 0.87 (0.71 to 0.95)	6.41	12.56
Spinal tenderness (yes, no)	Kappa = 0.35 (-0.33 to 1.00)	—	—
Bilateral active SLR (yes, no)	Kappa = 0.80 (0.54 to 1.00)	—	—
Sit-up (yes, no)	Kappa = 0.60 (0.25 to 0.95)	—	—
Overall PII score	ICC = 0.89 (0.61 to 0.96)	0.58	1.14

($r = -0.28$) and FABQ—physical activity subscale ($r = 0.24$).

Sixty-three of the 78 subjects (81%) completed the global rating of change and were used to assess responsiveness. Forty-four of the 63 subjects (70%) were defined as improved based on an average global rating $\geq +3$, and 19 were defined as unimproved. Two additional subjects (one improved, one unimproved) did not undergo 4-week reassessment of the PII or its components. Sixty-one subjects (43 improved, 18 unimproved) were therefore used for the calculation of responsiveness statistics. The subjects included in the responsiveness calculations did not differ from the entire sample with respect to initial or 4-week OSW or PII scores ($P > 0.05$). The AUC values for the individual components and overall PII score, OSW score, and PCS score are given in Table 4. The greatest AUC was found for the OSW, followed by the PII and the PCS. The AUC of the PII exceeded the AUC of any of the individual component measures of the scale.

The SEM and MDC values for the PII and its components are reported in Table 2. The SEM for the overall PII score was 0.58, with an MDC of 1.14. These results indicate that just over 1 point of change in the PII is required to exceed measurement error.

Discussion

The goal of this study was to examine the PII, originally developed using patients with chronic LBP, in a group of

Table 3. Pearson Product Correlation Coefficients Between the Physical Impairment Index (PII) and Other Measures Taken at Baseline

Baseline Measure	Correlation with PII
Oswestry Disability Score	0.43**
SF-36 Physical Health Component Score	-0.28*
SF-36 Mental Health Component Score	0.00
Pain Rating	0.47**
Depressive Symptoms (CES-D)	-0.05
Fear-Avoidance Beliefs about Work	0.22
Fear-Avoidance Beliefs about Physical Activity	0.24*
Non-organic Signs	0.42**
Non-organic Symptoms	0.36**

** $P < 0.01$, * $P < 0.05$.

Table 4. Area Under the ROC Curve and 95% CI for Various Measures Assessed at Baseline and After Four Weeks

Measure	AUC	(95% CI)
Oswestry disability score	0.96	(0.92 to 1.0)
Physical impairment index	0.88	(0.80 to 0.96)
SF-36 physical health component score	0.86	(0.76 to 0.95)
Total flexion	0.81	(0.69 to 0.92)
Total extension	0.70	(0.56 to 0.84)
Average lateral flexion	0.65	(0.50 to 0.79)
Average straight leg raise	0.53	(0.37 to 0.69)

n = 61.

patients with acute LBP (<3 weeks duration). In the original report of the PII, patients with chronic LBP (>3 months duration) were examined, and reliability of the component tests of the PII was reported. Our results found similarly high interrater reliability for measures of total flexion ROM, SLR ROM, and the active sit-up and bilateral SLR tests. Our reliability was lower for total extension and lateral flexion ROM and the spinal tenderness test. Some of the decrease in reliability found in this study is likely due to the more acute population of patients studied. The mean duration of LBP in our sample was 5.5 days. Patients with very acute LBP are expected to manifest changes in clinical presentation more rapidly than patients with more chronic symptoms.⁵² The average time between evaluations in this study was approximately 24 hours, and it is possible that real changes may have occurred in some patients during this time period. The lower reliability in our sample for extension ROM, lateral flexion ROM, and spinal tenderness may also be due to the small sample size on which the reliability was judged in this study. The CI values in our sample are rather wide for these measurements (Table 2), and higher reliability might be found if the study were replicated in another sample.

Other authors using similar inclinometric measurement techniques have reported lower interrater reliability coefficients for extension and lateral flexion ROM than for flexion.^{53,54} The magnitude of extension and lateral flexion ROM is much smaller than that of flexion ROM. Because the ICC reliability coefficient expresses a ratio of the total variance among a series of measurements to total variance plus the variance due to error, when the total variance is lower, the ICC will also be diminished.⁵⁵ The impact of the magnitude of the measurement on reliability is also demonstrated by the SEM and MDC values. In our study, even though the ICC value for total flexion (0.91) was better than the ICC for total extension (0.45), the SEM was lower for extension (7.7° vs. 4.0°), as was the MDC (15.1° vs. 7.9°). These results indicate that the measurement of total extension is more precise than that of total flexion, because only about 8° of change is required to exceed measurement error, as opposed to 15° of flexion. We also examined the reliability of the overall PII score and found excellent

reliability (ICC = 0.89) and precision (MDC = 1.1 points). This finding indicates that only about 1 point of change on the PII is needed to exceed measurement error. A change of 2 points clearly will exceed expected measurement error.

The validity of the PII was explored through correlations with concurrent measures of pain, disability, and various psychosocial variables. Similar to the findings of Waddell *et al*³ in a population with chronic LBP, relatively large correlations were found between the PII and the disability questionnaire ($r = 0.43$ with the OSW), nonorganic signs ($r = 0.42$), and nonorganic symptoms ($r = 0.36$). These findings indicate the overlap of impairment with disability and the need for cautious interpretation of the PII in light of the inability to distinguish physical impairment from illness behavior completely.⁵⁶ The correlations between the PII and measures of fear-avoidance beliefs ($r = 0.22$ – 0.24) were lower than those with nonorganic signs and symptoms, indicating that the self-reported fear of work or physical activity was not proportionally reflected in the physical impairment measurement for the patient.

In order to be useful as an outcome measurement, a scale must be responsive to change.⁵⁷ The psychometric property of responsiveness describes a scale's ability to detect change over time that is clinically meaningful.⁵⁸ It can be assessed in several ways. We used the method described by Deyo and Centor,⁴³ which constructs an ROC curve and calculates the AUC as a quantitative method for assessing a scale's ability to distinguish patients who have undergone true change from those who remain stable. The AUC can be interpreted as the probability of correctly identifying the improved patient from randomly selected pairs of improved and unimproved patients⁵⁹ and ranges between 0.5 (no diagnostic accuracy beyond chance) to 1.0 (perfect diagnostic accuracy). As expected, the disease-specific disability measure (OSW) was more responsive than the generic health status measure (SF-36 PCS score).^{60,61} The responsiveness of the PII was slightly better than that of the PCS but was not as good as that of the OSW. This finding likely reflects the fact that measures of strength and motion, while important to the patient, are not as directly linked to the patient's concerns as a disease-specific disability measure that includes questions about specific functional activities such as walking, sitting, and lifting.⁶² The overall PII score was more responsive than any of the individual components, supporting the use of the index score instead of a single measure of impairment such as flexion ROM.

■ Conclusion

The reliability and validity of the PII was generally supported in this group of patients with acute LBP. As was true for patients with chronic LBP, interpretation of physical impairment measures must be conducted cautiously due to their overlap with behavioral and nonorganic findings. The PII was not as responsive as a disease-

specific self-report disability measure but did show adequate responsiveness and measurement precision to permit its use as an adjunct outcome measure for clinical trials involving patients with acute LBP. The reliability and responsiveness of the overall PII score were generally superior to the individual impairment measures contained within the index. Further research into the responsiveness of the PII in patients with chronic LBP is needed before the PII can be advocated for outcomes research with this population.

■ Key Points

- The PII could be measured reliably in patients with acute LBP.
- Impairment was moderately correlated with concurrent measures of illness behavior and therefore needs to be interpreted with caution as a strictly physical phenomenon.
- The PII was found to have high levels of responsiveness to change in patients with acute low back pain, but not as high as the OSW.
- The MDC in the PII Index over 4 weeks of therapy was just over 1 point.

References

1. Dionne CE, Von Korff M, Koepsell TD, et al. A comparison of pain, functional limitations, and work status indices as outcome measures in back pain research. *Spine* 1999;24:2339–45.
2. Hazard RG, Haugh LD, Green PA, et al. Chronic low back pain: the relationship between patient satisfaction and pain, impairment, and disability outcomes. *Spine* 1994;19:881–7.
3. Waddell G, Somerville D, Henderson I, et al. Objective clinical evaluation of physical impairment in chronic low back pain. *Spine* 1992;17:617–28.
4. ICIDH. *International Classification of Impairments, Disabilities and Handicaps*. Geneva, Switzerland: World Health Organization; 1980.
5. National Advisory Board on Medical Rehabilitation Research. *Draft V Report and Plan for Medical Rehabilitation Research*. Bethesda, MD: National Institutes of Health; 1992.
6. Deyo RA, Andersson G, Bombardier C, et al. Outcome measures for studying patients with low back pain. *Spine* 1994;19(suppl):2032.s–6s.
7. Bombardier C. Outcome assessment in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000;25:3100–3.
8. Michel A, Kohlmann T, Raspe H. The association between clinical findings on physical examination and self-reported severity in back pain: results of a population-based study. *Spine* 1997;22:296–303.
9. Sullivan MS, Shoaf LD, Riddle DL. The relationship of lumbar flexion to disability in patients with low back pain. *Phys Ther* 2000;80:240–50.
10. Zuberbier OA, Kozlowski AJ, Hunt DG, et al. Analysis of convergent and discriminant validity of published lumbar flexion, extension, and lateral flexion scores. *Spine* 2001;26:472–8.
11. American Medical Association. *The Spine: Guides to the Evaluation of Permanent Impairment*. Chicago, IL: American Medical Association; 1993:94–138.
12. Battie MC, Cherkin DC, Dunn R, et al. Managing low back pain: attitudes and treatment preferences of physical therapists. *Phys Ther* 1994;74:219–26.
13. Flores L, Gatchel RJ, Polatin PB. Objectification of functional improvement after nonoperative care. *Spine* 1997;22:1622–33.
14. Jette AM, Smith K, Haley SM, et al. Physical therapy episodes of care for patients with low back pain. *Phys Ther* 1994;74:101–10.
15. Friedrich M, Gittler G, Halberstadt Y, et al. Combined exercise and motivation program: effect on the compliance and level of disability of patients with chronic low back pain: a randomized controlled trial. *Arch Phys Med Rehabil* 1998;79:475–87.
16. Nattrass CL, Nitschke JE, Disler PB, et al. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. *Clin Rehabil* 1999;13:211–8.

17. Williams RA, Pruitt SD, Doctor JN, et al. The contribution of job satisfaction to the transition from acute to chronic low back pain. *Arch Phys Med Rehabil* 1998;79:366-74.
18. Bigos S, Bowyer O, Braen G, et al. *Acute Low Back Problems in Adults*. AHCPR Publication 95-0642. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services; 1994.
19. Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative management. *Phys Ther* 1995;75:470-89.
20. Fairbank JC, Couper J, Davies JB, et al. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy* 1980;66:271-3.
21. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther* 2001;81:776-88.
22. Hudson-Cook N, Tomes-Nicholson K, Breen A. A revised Oswestry disability questionnaire. In: Roland MO, Jenner JR, eds. *Back Pain: New Approaches to Rehabilitation and Education*. New York, NY: Manchester University Press; 1989:187-204.
23. Ware JE, Sherbourne CD. The MOS 36-item short form health survey (SF-36), I: conceptual framework and item selection. *Med Care* 1992;30:473-83.
24. Ware JE, Kosinski M, Bayless 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 Survey. *Med Care* 1995;33(suppl):264-79.
25. Ware JE, Kosinski M, Keller SD. *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. Boston, MA: The Health Institute; 1994.
26. Patrick DL, Deyo RA, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. *Spine* 1995;20:1899-908.
27. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? *Pain* 1994;58:387-92.
28. Jensen MP, Turner JA, Romano JM, et al. Comparative reliability and validity of chronic pain intensity measures. *Pain* 1999;83:157-62.
29. Waddell G, Main CJ, Morris EW, et al. Chronic low-back pain, psychologic distress, and illness behavior. *Spine* 1984;9:209-13.
30. Waddell G, McCulloch JA, Kummel E, et al. Nonorganic signs in low-back pain. *Spine* 1980;5:117-25.
31. Fritz JM, George SZ, Delitto A. The role of fear avoidance beliefs in acute low back pain: relationships with current and future disability and work status. *Pain* 2001;94:7-15.
32. Kummel BM. Nonorganic signs of significance in low back pain. *Spine* 1996;21:1077-81.
33. Radloff LS. The CES-D scale: a self-report depression scale for research in a general population. *Appl Psychol Meas* 1977;1:385-401.
34. Turk DC, Okifuji A. Detecting depression in chronic pain patients: adequacy of self-reports. *Behav Res Ther* 1994;32:9-16.
35. Weissman MM, Sholomskas D, Pottenger M, et al. Assessing depressive symptoms in five psychiatric populations: a validation study. *Am J Epidemiol* 1977;106:203-14.
36. Waddell G, Newton M, Henderson I, et al. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993;52:157-68.
37. Jacob T, Baras M, Zeev A, et al. Low back pain: reliability of a set of pain measurement tools. *Arch Phys Med Rehabil* 2001;82:735-42.
38. Crombez G, Vlaeyen JW, Heuts PH, et al. Pain-related fear is more disabling than fear itself: evidence on the role of pain-related fear in chronic back pain disability. *Pain* 1999;80:329-39.
39. Hadjiistavropoulos HD, Craig KD. Acute and chronic low back pain: cognitive, affective, and behavioral dimensions. *J Cons Clin Psych* 1994;62:341-9.
40. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
41. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-6.
42. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
43. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 1986;11:897-906.
44. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. *J Rheumatol* 2001;28:400-5.
45. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999;52:861-73.
46. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther* 1997;77:745-50.
47. Wyrwich KW, Nienaber NA, Tierney WM, et al. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469-78.
48. Hébert R, Spiegelhalter DJ, Brayne C. Setting the minimal metrically detectable change on disability rating scales. *Arch Phys Med Rehabil* 1997;78:1305-8.
49. Jacobson NS, Follette WC, Revenstorf D. Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behav Ther* 1984;15:336-52.
50. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 2002;82:8-24.
51. Stratford PW, Binkley J, Solomon P, et al. Defining the minimum level of detectable change for the Roland-Morris Questionnaire. *Phys Ther* 1996;76:359-65.
52. Coste J, Delecoeuillierie G, Cohen de Lara A, et al. Clinical course and prognostic factors in acute low back pain: an inception cohort study in primary care practice. *BMJ* 1994;308:577-80.
53. Chen SP, Samo DG, Chen EH, et al. Reliability of three sagittal motion measurement methods: surface inclinometers. *J Occup Environ Med* 1997;39:217-23.
54. Nitschke JE, Natrass CL, Disler PB, et al. Reliability of the American Medical Association Guides' model for measuring spinal range of motion: its implication for whole-person impairment rating. *Spine* 1999;24:262-8.
55. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York, NY: John Wiley & Sons; 1986.
56. Main CJ, Waddell G. Behavioral responses to examination: a reappraisal of the interpretation of "nonorganic signs." *Spine* 1998;23:2367-71.
57. Kirshner B, Guyatt GH. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27-36.
58. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171-8.
59. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
60. Suarez-Almazor ME, Kendall C, Johnson JA, et al. Use of health status measures in patients with low back pain in clinical settings: comparison of specific, generic and preference-based instruments. *Rheumatology* 2000;39:783-90.
61. Taylor SJ, Taylor AE, Foy MA, et al. Responsiveness of common outcome measures for patients with low back pain. *Spine* 1999;24:1805-12.
62. Deyo RA. Functional status in low back pain. *Arch Phys Med Rehabil* 1988;69:1044-53.