

Self-reported Severity Measures as Predictors of Return-to-work Outcomes in Occupational Back Pain

Marjorie L. Baldwin · Richard J. Butler · William G. Johnson · Pierre Côté

Published online: 24 October 2007
© Springer Science+Business Media, LLC 2007

Abstract *Introduction* We test an array of biopsychosocial, cognitive-behavioral, and work-related factors to identify the best predictors of work disability following a back injury. *Methods* We test the validity of alternative severity measures in predicting the likelihood of four distinct, mutually exclusive patterns of post-injury employment in the first year after onset of back pain. The study sample includes 959 participants in the *ASU Healthy Back Study*, a prospective cohort study of workers who filed claims for occupational back pain between 1999 and 2002. Self-reported severity measures include: NRS-101 measures of the intensity of back or leg pain, Roland–Morris scale of functional disability, and mental and physical components of the SF-12. *Results* All the severity measures have significant predictive power on return-to-work outcomes even after 1 year. Baseline physical functioning and overall mental and physical health status are more predictive of specific patterns of post-injury employment than pain intensity measures, possibly because there is considerable idiosyncratic variation in the pain intensity measures. The mental component of the SF-12, in particular, is relatively robust to alternate specifications, consistently statistically significant, and has the lowest probability significance level in explaining patterns of employment 1 year after injury.

Keywords Low back pain · Return to work · Back functionality · Mental well being

M. L. Baldwin (✉)
School of Health Management and Policy, W. P. Carey School of Business, Arizona State University,
P.O. Box 874506, Tempe, AZ 85287-4506, USA
e-mail: Marjorie.Baldwin@asu.edu

R. J. Butler
Department of Economics, Brigham Young University, Provo, UT, USA

W. G. Johnson
School of Computing and Informatics, Ira A. Fulton College of Engineering, Arizona State University,
Tempe, AZ, USA

P. Côté
Toronto Western Research Institute, Toronto Western Hospital, Toronto, Canada

Introduction

In 2004, there were more than 1.2 million nonfatal occupational injuries and illnesses in the US resulting in lost work time, and 22% were back injuries [1]. The indirect costs of work disability associated with back injuries exceeded \$18.5 billion in 1996, with mean costs approximately \$5,000 [2]. The mean is misleading, however, because the great majority of costs are incurred by a small proportion of cases with lengthy spells of work absence. Estimates from one national workers' compensation insurer indicate that less than 5% of back claims, with durations of work absence exceeding 1 year, account for 65% of total compensation costs [3].

To begin to address the disability burden of occupational back pain, it is imperative to be able to explain, and forecast, return to work outcomes. Numerous studies have tested an array of biopsychosocial, cognitive-behavioral, and work-related factors to identify the best predictors of work disability following a back injury [4–9]. The accumulated evidence shows that self-reported severity measures add significant predictive power to the models. Radiating pain and high levels of functional disability are shown consistently to be risk factors for longer duration disability [10, 11]. Evidence on the association between intensity of back pain and subsequent work outcomes is mixed, however, with some studies finding a significant, but small, effect and others finding none [11, 12]. Measures of general health status, including psychosocial stress, are validated predictors of work disability [8, 10, 12]. At least one study, however, finds that psychological stress factors become insignificant when cognitive variables such as perceived improvement in health and expectations of recovery are included in the model [7].

Variations in the predictive validity of severity measures across studies are attributed, in part, to differences in outcome measures (e.g. returns to work, durations of absence, survival curves) and follow-up periods (typically 3–12 months for prospective studies) [8]. Thus, the predictive power of self-reported severity measures must be revisited as measures of work outcomes are refined and new prospective data become available.

The objective of this study is to test the validity of alternative severity measures in predicting the likelihood of four mutually exclusive patterns of post-injury employment. The study extends the literature in several important ways: We analyze outcomes at three points in the first year after onset, so we are able to compare the validity of severity measures in predicting outcomes over different follow-up periods. We define patterns of post-injury employment to describe both returns to work and repeated spells of work absence. Thus, our model explicitly recognizes that a first return to work is not necessarily the end of the period of work disability associated with occupational back pain.

Methods

Data

Data come from the *Arizona State University Healthy Back Study*, a prospective cohort study of injured workers who file compensation claims for back pain [13]. The study population includes nearly 200,000 workers from five US employers operating in 37 states. The employers are: America West Airlines, American Medical Response, The Earthgrains Co. (now part of Sara Lee Corporation Baking Division), Maricopa County, and Marriott International, Inc. The data are compiled from a telephone survey of injured workers, merged with information from workers' compensation claims files and first reports of injury. Survey protocols are approved by Institutional Review Boards at Arizona State and East Carolina universities.

The protocols stipulate that participating employers notify the research team whenever an employee files a workers' compensation claim for back pain. Workers are contacted by telephone and asked to give their informed consent to participate in the study. Workers who agree to participate complete a baseline interview immediately and follow-up interviews at 1 month after baseline, and at 6 and 12 months after onset. There are, however, lags in completing some baseline and follow-up interviews. For example, 30% of baseline interviews were conducted within 14 days of onset, 30% between 15 days and 28 days, and 40% after 28 days. In cases where the initial interview is conducted after 28 days, we administer a combination baseline/1-month interview, meaning that 1-month return to work outcomes and baseline severity measures are contemporaneous.

The difficulties in obtaining responses at a uniform time after initial onset lead to a heterogeneous intake population, which includes workers with both acute and sub-acute low back pain. This may affect the predictive value of baseline severity measures, as initial severity a few days after onset may be less predictive of long-term employment outcomes than pain or functional limitations that have persisted for several weeks. We conduct sensitivity tests to determine whether our results vary across samples with short or long lags from date of onset to date of baseline interview.

To be eligible for inclusion in the study sample, workers must meet the following criteria: age 18 and over, confirmed back pain (with or without leg pain or sciatica), claim filed between January 1, 1999 and June 30, 2002. Exclusion criteria are: back pain associated with a fracture (identified by the 'nature of injury' code on first reports of injury), claim either denied or litigated, subsequent claim from a worker already enrolled in the study.

We received notifications for 3,621 injured workers who filed eligible back pain claims during the inception period. Slightly over half ($N = 1,832$, 51%) agreed to participate and completed a baseline interview. Relative to the baseline sample, follow-up rates were 87% ($N = 1,585$) at 1 month, 62% ($N = 1,143$) at 6 months, and 42% ($N = 761$) at 1 year.

Attrition is practically inescapable in any long-term study. Participants may relocate without updating contact information (a particular problem for this survey because of the transient nature of the workforce of the largest employer) or may refuse to participate in follow-up interviews for any number of reasons. The 761 participants with a complete set of follow-up interviews are generally comparable to the full sample interviewed at baseline: there are no significant differences in mean baseline severity measures, expectations of recovery, or satisfaction with the pre-injury job. There are small but significant differences in demographic characteristics and work experience, such that workers who complete the interview cycle are, on average, older, have more work experience, and are more likely to be female (Appendix Table AI).

Return to Work Outcomes

The outcome measures are patterns of post-injury employment reported at the 1, 6, and 12 months interviews. At each follow-up date, we classify workers into one of four mutually exclusive employment patterns based on responses to the following questions: (1) "Did you have to take time off from work because of your back injury?" (2) "Have you returned to work?" (3) "Between the time you returned to work and now (date of interview) did you have to take any additional time off work because of your back injury?" The employment patterns are defined as follows:

- Pattern 1 (No absence): The worker takes no time off work following onset of back pain.
- Pattern 2 (Return and stay): The worker returns to work after an initial absence and reports no subsequent spells of absence associated with back pain.

Pattern 3 (Multiple spells): The worker experiences one or more spells of work absence associated with back pain after the initial absence and return to work.

Pattern 4 (Not yet returned): The worker has been absent from work since onset.

Patterns of employment at 6 and 12 months represent a worker's cumulative employment experience to date. A worker who reports a single spell of work absence and return to work before the first follow-up and no additional spells of absence during the year is classified in Pattern 2 (return and stay), at 1, 6, and 12 months. Similarly, a worker who reports, at 6 months, subsequent pain-related absences after the initial absence and return to work is classified in Pattern 3 (multiple spells) at 6 and 12 months.

Patterns 3 and 4 supposedly represent the poorest outcomes for workers and employers, but there are important distinctions between them. Pattern 3, which includes all workers with multiple spells of absence, regardless of the number, timing, or duration of spells, is a heterogeneous category with respect to total time on disability. At 1 year a worker in Pattern 3 could have several weeks of accumulated disability or many months, and we do not know if the worker is currently employed. On the other hand, all workers in Pattern 4 at 1 year have been absent from work for 12 months and are not working.

Self-reported Severity Measures

The primary analysis variables are self-reported measures of pain intensity, functional status, and health-related quality of life. Severity measures reported shortly after onset (baseline interview) are examined at three follow-up points to test their predictive power in the first year after injury. The specific severity measures are: Numeric Rating Scales (NRS) of back and leg pain intensity, the Roland–Morris scale (functional limitations), and the SF-12 questionnaire (health-related quality of life).

Pain Intensity

The intensity/disturbance of back and leg pain are measured with the Numeric Rating Scale (NRS-101). Subjects are asked to select a number from 0 to 100 that best describes their pain over the last week, where '0' indicates the pain is not bothersome at all and '100' indicates the pain is extremely bothersome. The NRS-101 is a valid and reliable instrument that is easy to understand and administer [14].

Functional Limitations

The extent of functional disability associated with back pain is measured with the Roland–Morris Disability Scale [15]. The scale is a low-back-specific functional capacity questionnaire consisting of 24 items from the Sickness Impact Profile-40. The items cover activities of daily living, such as, walking, standing, and climbing stairs, that may be difficult to perform for individuals with low back pain. The Roland–Morris scale has good criterion-based construct and discriminant validity and is the most responsive disability questionnaire for back pain currently available [15–22]. The internal consistency of the questionnaire is well established [15, 19, 23]. The scale has high test-retest reliability when re-administered within a 6-week period: repeated measurements performed on the same day and at 3 weeks are highly correlated [15, 16, 19, 20]. Raw scores range from zero (no disability) to 24 (severe disability), but we transform the raw scores to percentages for our analyses.

Health-related Quality of Life

The SF-12 questionnaire (second revision), a short version of the SF-36, measures physical and mental health-related quality of life [24, 25]. A standard scoring algorithm converts raw scores to two component scores ranging from 0 to 100, where higher scores indicate better physical or mental health. Mean scores for a healthy population equal 50. The physical and mental components of the SF-12 are predictive of the corresponding SF-36 components with r-square values greater than 0.91 [26]. The SF-12 has good test-retest reliability measured over a 2-week period, with correlation coefficients of 0.89 for the physical component and 0.76 for the mental component [26]. Finally, the SF-12 has good internal consistency, validity, and responsiveness in patients with low back pain [27].

Other Variables

Control variables in the return to work models represent worker- and job-related characteristics (from the survey data or administrative claims files) that have been shown to be significant predictors of work absence in prior studies [10, 11]. All control variables are measured either at the baseline interview or the time of claim filing.

Worker-related variables include age, gender (1 = male), expectations of recovery, and employee choice of health care provider. Expectations of recovery are represented as a binary variable where one indicates positive expectations (already recovered, get better soon) and zero represents negative expectations (get better slowly, never get better, get worse). Employee choice equals one if an injured worker is employed in a state where workers have the legal right to choose their initial health provider, and 0 if employed in a state where the employer has the right to choose. Employee choice of provider is a proxy for the potential moral hazard effects of workers' compensation benefits on returns to work, assuming that employee-selected providers are, all else equal, more willing to extend durations of work disability than are employer-selected providers [28].

Job-related variables include work experience (years with pre-injury employer), satisfaction with the pre-injury job, and firm dummies. The job satisfaction variable measures workers' overall satisfaction with their pre-injury jobs on a scale of one (very satisfied) to four (very dissatisfied). Firm dummies control for employers' disability management policies, including the commitment to provide job accommodations.

Preliminary versions of the model include controls for region and prior history of back pain but neither variable is statistically significant in any specification. Estimating the more parsimonious model has no effect on the signs or significance levels of included variables.

Model of Return-to-work Outcomes

We use multinomial logistic models to estimate the relationships between self-reported severity measures (reported at baseline) and post-injury employment patterns at each follow-up period. The categorical dependent variable in the models identifies the employment pattern that describes an injured worker's post-injury job experience to that follow-up point.

Maximizing the likelihood function of the multinomial models yields three sets of parameter estimates: for three employment patterns relative to a reference pattern, in this case Pattern 4 (not yet returned to work). The parameter estimates do not represent the marginal effects of individual variables on observed patterns of employment so, for ease of interpretation, we convert the coefficients to measures of marginal effects and express the results as

semi-elasticity estimates. Semi-elasticity estimates are measures used in the econometrics literature to standardize interpretation of multivariate regression results. A semi-elasticity estimate represents the percentage change in a dependent variable associated with a specified-unit change in an independent variable (1-unit, 10-unit, etc.), holding all other variables in the model constant. In our results on severity, each reported semi-elasticity estimate represents the percentage change in the probability of experiencing a particular employment pattern given a 10-unit change in a severity measure.

We estimate a sequence of models to examine the validity of the severity measures in predicting subsequent patterns of return to work. First, we estimate a ‘basic’ model excluding the severity measures. Second, we estimate a series of models with different severity measures (NRS-101, Roland–Morris, SF-12) added separately to the basic model to determine which severity constructs are significant predictors of post-injury employment patterns. Third, we estimate a model with all the severity measures added to the basic model, to test whether the severity variables are measuring overlapping constructs.

A final question is how well alternate baseline severity measures identify workers who experience the poorest work outcomes, that is, outliers in the cost distribution. To address this issue, we estimate a series of logistic models, in which the dependent variable equals one if a worker experiences Pattern 3 (multiple spells) or 4 (not yet returned), and severity measures are added to the basic model individually. We conduct Hosmer–Lemeshow [29] tests for the goodness of fit of each model in predicting poor work outcomes at each follow-up point.

Results

Study Samples

We restrict our study sample to workers with complete severity and job satisfaction responses on the survey data, and complete demographic and job information from the administrative files. Workers are excluded if the elapsed time between date of injury and date of 1-month survey is more than 90 days. Our final samples include 959 workers (62% of 1,552 interviewed) with complete data at the 1-month follow-up, 585 workers (51% of 1143 interviewed) at 6 months, and 332 (44% of 761 interviewed) at 12 months.

The main reasons for losing observations in the study samples are: missing work experience ($N = 253$); and missing one or more severity measure ($N = 132$). We estimate the probability of missing work experience as a function of all other explanatory variables in the models, and construct Box–Cox plots of predicted probabilities for those with and without reported experience. There is substantial overlap in the distributions, suggesting there are no systematic differences between our study samples and the cases we lose because of missing experience. In a similar manner, we construct plots of predicted values for observations with and without each severity measure, at one, 6 and 12 months, and find substantial overlap in distributions. Although missing patterns of post-injury work experience are not a major cause of lost observations, we re-estimated our multinomial models with a fifth employment pattern (missing) to test the robustness of our results. We find no substantial change in signs or significance levels of the estimated semi-elasticity estimates when missing pattern cases are included in the models. We conclude that our results are representative of the overall population of injured workers from participating firms, but our sample still may not be random. Missing cases may include different subgroups dropping out for different reasons.

Workers in the 1-month cohort are age 18–75 ($M = 38$, $SD = 10.70$); have, on average, less than 10 years job experience at the time of injury ($M = 8.8$, $SD = 13.34$); and are generally satisfied with their jobs (83% report that, overall, they are ‘satisfied’ or ‘very satisfied’). Approximately half are male (48%), and less than one in five have the legal right to choose their initial health provider (18%). Slightly more than one-third of workers have good expectations of recovery from back pain at the baseline interview (36%). Workers in the 6- and 12-month cohorts have virtually identical characteristics, except the proportion of males in the last follow-up is smaller (44%).

Differences in Mean Severity Measures

Table 1 presents means of the severity measures, by employment pattern, for each follow-up sample. Recall that measures of pain intensity and Roland–Morris scores *increase* with greater severity, while SF-12 measures *decrease*. Within each follow-up sample, and almost without exception, the means follow an expected increasing or decreasing pattern as we move from the ‘best’ outcome (Pattern 1—no absence) to the ‘worst’ (Pattern 4—not yet returned). At each follow-up point we observe clinically significant differences in means of the pain intensity measures, Roland–Morris scale, and SF-12 components between workers in Patterns 1 or 2 (‘better’ outcomes) and workers in Patterns 3 or 4 (‘worse’ outcomes). In addition, at each follow-up point we observe clinically significant differences in means of the pain intensity measures and Roland–Morris scale between workers in Pattern 3 (multiple returns) and 4 (not yet returned).

Table 1 Means of severity measures by post-injury employment patterns

Severity measures	Pattern #1: No absence	Pattern #2: Return and stay	Pattern #3: Multiple spells	Pattern #4: Not yet returned
<i>1 month</i>				
Back pain intensity	46.14 (32.13)	48.46 (32.17)	61.62 (31.30)	72.07 (23.78)
Leg pain intensity	26.57 (33.93)	22.61 (31.47)	36.36 (37.00)	48.59 (35.90)
Roland–Morris	38.49 (27.88)	43.13 (29.54)	57.26 (27.31)	67.30 (22.22)
Physical SF-12	40.62 (10.65)	39.17 (10.32)	34.91 (9.66)	32.62 (7.69)
Mental SF-12	49.88 (10.92)	48.60 (11.16)	42.83 (11.86)	41.57 (11.99)
<i>6 months</i>				
Back pain intensity	46.21 (31.64)	47.52 (32.55)	62.59 (30.12)	69.76 (24.69)
Leg pain intensity	27.71 (33.86)	24.97 (33.04)	36.64 (36.74)	54.43 (31.40)
Roland–Morris	38.33 (27.63)	40.31 (28.64)	58.88 (26.41)	71.74 (24.98)
Physical SF-12	40.64 (10.76)	39.48 (10.29)	34.53 (8.98)	31.19 (7.20)
Mental SF-12	49.83 (11.16)	49.74 (10.85)	43.08 (11.83)	40.87 (12.34)
<i>12 months</i>				
Back pain intensity	48.32 (31.62)	47.36 (34.56)	61.28 (29.23)	61.00 (30.57)
Leg pain intensity	27.71 (34.50)	27.88 (34.80)	35.53 (36.41)	54.30 (35.58)
Roland–Morris	37.00 (27.82)	42.39 (30.40)	58.81 (25.18)	65.83 (30.29)
Physical SF-12	40.36 (10.61)	38.60 (10.61)	34.77 (8.86)	30.91 (4.51)
Mental SF-12	50.64 (10.74)	48.32 (11.34)	42.95 (12.55)	40.12 (12.99)

Notes: $N = 959$ at 1 month, 585 at 6 months, and 332 at 12 months. Standard deviations in parentheses

Source: ASU Healthy Back Study, survey data

Table 2 Correlations between severity measures

	Back pain intensity	Leg pain intensity	Roland–Morris	Physical SF-12	Mental SF-12
Back pain intensity	1.00	0.47 (<0.0001)	0.63 (<0.0001)	-0.59 (<0.0001)	-0.36 (<0.0001)
Leg pain intensity	0.47 (<0.0001)	1.00	0.49 (<0.0001)	-0.39 (<0.0001)	-0.37 (<0.0001)
Roland–Morris	0.63 (<0.0001)	0.49 (<0.0001)	1.00	-0.72 (<0.0001)	-0.50 (<0.0001)
Physical SF-12	-0.59 (<0.0001)	-0.39 (<0.0001)	-0.72 (<0.0001)	1.00	0.21 (<0.0001)
Mental SF-12	-0.36 (<0.0001)	-0.37 (<0.0001)	-0.50 (<0.0001)	0.21 (<0.0001)	1.00

Notes: $N = 959$. Probability significance levels in parentheses

Source: ASU Healthy Back Study, survey data

Correlations among the severity measures are reported in Table 2. All correlations are in the expected direction (SF-12 measures being negatively correlated with pain intensity and Roland–Morris scores) and statistically significant at better than the 0.01 level. The highest correlation is between the Roland–Morris and physical SF-12 scales, but there is still substantial independent variation in the two severity measures. The second highest correlation is between the Roland–Morris and back pain intensity scales: this is important because the pain intensity scale is much easier and faster to administer. The mental SF-12 scale has the lowest correlation with other severity measures, suggesting that, as intended, it measures a different dimension of health.

Severity Measures that Work

Our main focus is on the ability of various baseline severity measures to predict return to work patterns at 1, 6, and 12 months after onset. We begin by estimating the multinomial patterns model with the severity measures excluded (the ‘basic’ model) and then add severity measures individually (pain intensity, Roland–Morris, SF-12) to compare their predictive power. Semi-elasticity estimates from the basic model are reported in Appendix Table AII. In the underlying model, all included variables except employee choice have at least one statistically significant coefficient (at the 0.10 level or better) over the three follow-up periods. Variables that have the greatest effect on distributions across employment patterns are the firm dummies and workers’ expectations of recovery at the baseline interview.

Semi-elasticity estimates for the severity measures are reported in Table 3A–C. We find statistically significant associations between reported back pain intensity and return to work outcomes at each follow-up point (Table 3A), but the strength of the association tends to diminish over time. For example, back pain intensity is a good predictor of poor work outcomes at one month (a 10-point increase in reported back pain at baseline is associated with a 19% increase in the probability of ‘not yet returned’ at 1 month) but the sign changes and the effect size diminishes drastically by one year (a 10-point increase in reported back pain at baseline is associated with only a 4% decrease in the probability of ‘not yet returned’ at 1 year). There is a strong association between reported baseline intensity of leg pain and the poorest work outcome (Pattern 4) at 6 and 12 months, but the estimates are imprecise.

The baseline Roland–Morris score is a statistically significant predictor of return to work outcomes across all follow-up periods (Table 3B). Greater functional disability, as measured by a higher Roland–Morris score, has a strong positive association with less favorable work outcomes (Patterns 3 and 4), and a strong negative association with the most favorable outcome (Pattern 1). The strong associations persist over the entire first year: a 10-point increase

Table 3 Semi-elasticity estimates for individual severity measures

	Pattern 1: No absence	Pattern 2: Return and stay	Pattern 3: Multiple spells	Pattern 4: Not yet returned	P-value	-2*log likelihood
A. Pain intensity						
<i>1 month</i>						
Back pain	-0.06 (-0.09, -0.04)	-0.002 (-0.003, 0.002)	0.07 (0.01, 0.13)	0.19 (0.12, 0.27)	<0.0001	2157.51
Leg pain	0.01 (-0.01, 0.03)	-0.04 (-0.07, -0.02)	0.03 (-0.02, 0.07)	0.09 (0.04, 0.14)	0.005	
<i>6 months</i>						
Back pain	-0.05 (-0.09, -0.02)	-0.03 (-0.06, 0.01)	0.10 (0.06, 0.16)	0.13 (0.01, 0.29)	0.002	1302.66
Leg pain	0.004 (-0.03, 0.04)	-0.01 (-0.05, 0.02)	-0.01 (-0.05, 0.03)	0.12 (0.03, 0.20)	0.37	
<i>12 months</i>						
Back pain	-0.03 (-0.08, 0.02)	-0.05 (-0.1, 0.009)	0.09 (0.03, 0.14)	-0.04 (-0.20, 0.10)	0.07	719.51
Leg pain	-0.02 (-0.08, 0.02)	0.03 (-0.02, 0.08)	-0.02 (-0.07, 0.03)	0.17 (0.03, 0.40)	0.28	
B. Roland-Morris						
<i>1 month</i>						
Roland-Morris score	-0.10 (-0.13, -0.08)	-0.01 (-0.04, 0.02)	0.15 (0.01, 0.16)	0.31 (0.24, 0.39)	<0.0001	2134.34
<i>6 months</i>						
Roland-Morris score	-0.11 (-0.16, -0.08)	-0.04 (-0.08, -0.003)	0.17 (0.12, 0.25)	0.41 (0.21, 0.73)	<0.0001	1271.19
<i>12 months</i>						
Roland-Morris score	-0.14 (-0.20, -0.09)	-0.02 (-0.08, 0.06)	0.14 (0.08, 0.20)	0.25 (-0.02, 0.91)	<0.0001	702.45
C. SF-12						
<i>1 month</i>						
Physical SF-12	0.21 (0.14, 0.28)	0.02 (-0.01, 0.09)	-0.31 (-0.49, -0.14)	-0.61 (-0.83, -0.41)	<0.0001	2157.51
Mental SF-12	0.16 (0.10, 0.23)	0.02 (-0.05, 0.09)	-0.31 (-0.44, -0.18)	-0.41 (-0.57, -0.25)	<0.0001	
<i>6 months</i>						
Physical SF-12	0.24 (0.14, 0.37)	0.05 (-0.05, 0.16)	-0.30 (-0.46, -0.15)	-0.87 (-0.16, -0.38)	<0.0001	1302.66
Mental SF-12	0.16 (0.06, 0.27)	0.09 (0.002, 0.19)	-0.31 (-0.44, -0.20)	-0.50 (-0.93, -0.19)	<0.0001	

Table 3 continued

	Pattern 1: No absence	Pattern 2: Return and stay	Pattern 3: Multiple spells	Pattern 4: Not yet returned	P-value	-2*log likelihood
<i>12 months</i>						
Physical SF-12	0.24 (0.10, 0.45)	0.04 (-0.11, 0.22)	-0.20 (-0.37, -0.02)	-1.00 (-3.42, -0.41)	0.011	719.51
Mental SF-12	0.29 (0.17, 0.47)	-0.01 (-0.14, 0.16)	-0.26 (-0.42, -0.13)	-0.49 (-1.62, -0.05)	0.001	

Notes: $N = 959$ at 1 month, 585 at 6 months, 332 at 12 months. Model controls for age, gender, employer, job satisfaction, work experience, expectations of recovery, and provider choice. Results represent % changes in the probability of a particular employment pattern associated with a 10-unit change in the severity measure. 90% confidence interval in parentheses, generated by bootstrapping 1,000 samples

Source: ASU Healthy Back Study, survey and claims data

in the baseline Roland–Morris score, for example, is associated with a 25% increase in the probability of not returning to work within 1 year (Pattern 4).

While Table 3A and B examine severity measures directed specifically at back pain, Table 3C analyzes the predictive power of overall indicators of health status: the physical and mental components of the SF-12. Both measures are statistically significant and clinically important predictors of work outcomes at all follow-up periods. At 1 year, for example, a 10-unit increase in the baseline physical SF-12 is associated with a 24% increase in the probability of experiencing the best work outcome (Pattern 1), while a 10-unit increase in the mental SF-12 is associated with a 29% increase in the probability of the best outcome.

A simple way to determine which severity measures have the best predictive power in the model is to compare log likelihood statistics. The models are not nested, but have the same dependent variables and number of observations, so the log likelihoods are reasonable indicators of the ‘best fit’ model. All the models in Table 3A–C represent improvements on the basic model, but the models with the Roland–Morris score or SF-12 components represent substantial improvements over the model with the pain intensity variables. This is not surprising because the pain intensity measures are single-item scales that are likely to include considerable idiosyncratic response, whereas the other severity measures are composite scales that tend to eliminate the peculiar response patterns inherent in a one-question measure.

What is not clear from the specifications in Table 3A–C is how the severity measures compare when included in the same specification. Is there any one measure that stands out? We address this question in Table 4, which reports semi-elasticity estimates from a model that includes all severity measures. The pain intensity variables are only weakly associated with patterns of employment in this model. The physical SF-12 and Roland–Morris scores are so collinear (see correlations in Table 2) that only the Roland–Morris score is significant at 1 and 6 months, and neither is significant at 1 year. At 1 year, however, both measures have about the same level of statistical significance. This suggests that these two measures (physical functioning and physical health status) are fairly interchangeable for estimating return to work outcomes, at least for these data. The one severity measure that is consistently significant (at better than the 5% level) across all follow-up periods is the mental component of the SF-12, suggesting that mental health is as important as physical health or physical functioning in explaining return to work outcomes after onset of back pain.

Specification Tests

Goodness of Fit

As an alternate test of the relative predictive power of different severity measures, we estimate logistic models of the likelihood of poor employment outcomes and calculate Hosmer–Lemeshow statistics to test goodness of fit [29]. The test partitions predicted observations into 10 uniform cells, then compares the predicted number of individuals with poor outcomes with actual employment outcomes. The null hypothesis is that the fit is good, where a Chi-square value of zero indicates a perfect fit, so the higher the probability significance level, the better the fit of the model to actual outcomes.

Logistic models are estimated for each severity measure, using the same independent variables as the multinomial models. The dependent variable equals one if the worker experiences poor post-injury employment outcomes (Pattern 3 or 4), and zero otherwise. The results, reported in Table 5, indicate that all severity measures are statistically significant predictors of poor outcomes at each follow-up point. In this binary framework, however, the

Table 4 Semi-elasticity estimates for all severity measures

	Pattern 1: No absence	Pattern 2: Return and stay	Pattern 3: Multiple spells	Pattern 4: Not yet returned	P-value	-2*log-likelihood
<i>1 month</i>						
Back pain	-0.01 (-0.04, 0.02)	-0.01 (-0.04, 0.02)	0.01 (-0.05, 0.07)	0.09 (0.002, 0.18)	0.42	2100.97
Leg pain	0.04 (0.01, 0.06)	-0.05 (-0.08, -0.03)	-0.01 (-0.05, 0.07)	0.04 (-0.009, 0.10)	0.01	
Roland-Morris	-0.06 (-0.10, -0.02)	0.01 (-0.03, 0.06)	0.05 (-0.05, 0.14)	0.17 (0.05, 0.28)	0.03	
Physical SF-12	0.12 (0.02, 0.23)	-0.02 (-0.13, 0.09)	-0.22 (-0.47, -0.02)	-0.18 (-0.48, 0.12)	0.14	
Mental SF-12	0.13 (0.05, 0.22)	-0.02 (-0.10, 0.07)	-0.27 (-0.44, -0.12)	-0.18 (0.36, 0.01)	0.003	
<i>6 months</i>						
Back pain	0.01 (-0.04, 0.06)	-0.01 (-0.07, 0.03)	0.02 (-0.04, 0.08)	-0.07 (-0.27, 0.15)	0.84	1254.62
Leg pain	0.03 (-0.007, 0.06)	-0.01 (-0.04, 0.03)	-0.04 (-0.09, 0.002)	0.07 (-0.07, 0.18)	0.26	
Roland-Morris	-0.07 (-0.13, -0.01)	-0.03 (-0.01, 0.03)	0.10 (0.02, 0.18)	0.28 (-0.01, 0.64)	0.24	
Physical SF-12	0.18 (0.03, 0.35)	-0.03 (-0.16, 0.12)	-0.15 (-0.38, 0.06)	-0.46 (-1.39, 0.23)	0.16	
Mental SF-12	0.13 (0.02, 0.25)	0.06 (-0.05, 0.17)	-0.24 (-0.45, -0.11)	-0.25 (-0.77, 0.16)	0.02	
<i>12 months</i>						
Back pain	0.05 (-0.01, 0.29)	-0.04 (-0.11, 0.03)	0.01 (-0.05, 0.07)	-0.22 (-0.87, 0.04)	0.25	684.46
Leg pain	0.02 (-0.05, 0.06)	0.02 (-0.03, 0.08)	-0.05 (-0.06, 0.05)	0.13 (-0.03, 0.49)	0.26	
Roland-Morris	-0.08 (-0.19, 0.04)	-0.02 (-0.12, 0.07)	0.11 (-0.05, 0.14)	-0.03 (-1.0, 0.66)	0.24	
Physical SF-12	0.20 (8E-4, 0.53)	-0.04 (-0.26, 0.24)	-0.04 (-0.47, -0.02)	-1.27 (-6.47, -0.57)	0.16	
Mental SF-12	0.26 (-0.11, 0.51)	-0.03 (-0.19, 0.15)	-0.18 (-0.44, -0.12)	-0.58 (-2.6, -0.10)	0.03	

Notes: $N = 959$ at 1 month, 585 at 6 months, 332 at 12 months. Model also controls for age, gender, employer, job satisfaction, work experience, and provider choice. Results represent % changes in the probability of a particular employment pattern associated with a 10-unit change in the severity measure 90% confidence interval given in parentheses, generated by bootstrapping 1,000 samples

Source: ASU Healthy Back Study, survey and claims data

Table 5 Chi-square values for Hosmer–Lemeshow tests for goodness of fit: Predicting poor return to work outcomes

	Back/leg pain	Roland–Morris	Physical/mental SF-12
1 Month	8.762 (0.363)	6.685 (0.571)	6.982 (0.539)
6 Months	1.469 (0.993)	9.522 (0.300)	4.607 (0.780)
1 Year	8.491 (0.387)	10.448 (0.235)	10.298 (0.245)

Notes: $N = 959$ at 1 month, 585 at 6 months, 332 at 12 months. Models predict the probability of experiencing patterns 3 (multiple spells) and 4 (not yet returned). Probability significance levels in parentheses

Source: ASU Healthy Back Study, survey and claims data

pain intensity measures dominate other severity measures at six months ($\chi^2 = 1.47$, $P = 0.99$), and retain more of their predictive power at 1 year.

Lag Time to Baseline Interview

As noted above, there is considerable variation within our sample in elapsed time from date of onset of back pain to date of baseline interview. The differences in lag times may substantially affect the predictive power of severity measures, so we re-estimated our models partitioning the sample into three groups: baseline interview within 14 days of onset; baseline interview 15–28 days after onset; and baseline more than 28 days after onset.

Overall the results for partitioned samples support our substantive conclusions. Log-likelihood ratio tests of the equality of coefficients across partitioned samples cannot reject the null hypothesis of no difference in coefficients except for 1 year models which include the pain intensity variables (Table 6). The differences driving these results are: (1) Back pain intensity measured within 14 days of onset has much stronger predictive power on 1-year patterns of employment than back pain measured with longer lag times. (2) Leg pain intensity is a stronger predictor of the poorest outcome (not yet returned) when it is measured closer to onset of back pain. The estimates for partitioned samples tend to have large standard errors, however, so the results should be interpreted with caution.

Claim Management

Negative employer–employee interactions following a work-related injury have been shown to be associated with poorer return to work outcomes. Our survey asks workers how satisfied they are with their employer’s handling of their workers’ compensation claim. When we add a control for workers who are ‘satisfied’ or ‘very satisfied’ (versus workers who are ‘dissatisfied’ or ‘very dissatisfied’) to our models none of our conclusions change. There is no change in sign and little change in magnitude for the vast majority of estimated semi-elasticities; nor does inclusion of the employee satisfaction variable improve the model fit. We exclude the variable from our final specifications because missing values reduce sample sizes by more than 25%.

Duration of First Work Absence

Claim duration is a more common measure of post-injury work outcomes than the long-term employment patterns we describe. We estimated a Weibull duration model for workers with positive durations on indemnity claims replicating our four models with pain intensity, Roland–Morris, SF-12, and all severity measures. The dependent variable is duration of the

Table 6 Semi-elasticity estimates for pain intensity at 1 year: Stratified by time to first interview

Lag time from onset to first interview	Severity measure	Pattern 1: No absence	Pattern 2: Return and stay	Pattern 3: Multiple spells	Pattern 4: Not yet returned
0–14 days	Back pain	0.02 (–0.17, 0.72)	0.21 (0.04, 7.6)	0.36 (0.19, 7.5)	–1.34 (–71.13, –0.11)
	Leg pain	–0.14 (–0.77, 0.17)	–0.12 (–7.74, 0.18)	–0.27 (–7.94, 0.03)	1.45 (–1.33, 75.32)
15–28 days	Back pain	0.02 (–0.29, 0.35)	–0.06 (–0.40, 0.23)	0.08 (–0.21, 0.40)	–0.07 (–8.74, 7.66)
	Leg pain	–0.02 (–0.21, 0.10)	–0.03 (–0.26, 0.08)	0.03 (–0.16, 0.18)	0.29 (–0.44, 4.46)
More than 28 days	Back pain	–0.05 (–0.53, 0.04)	–0.07 (–0.62, 0.02)	0.09 (–0.44, 0.22)	0.27 (–0.60, 17.8)
	Leg pain	–0.004 (–0.19, 0.20)	–0.02 (–0.20, 0.18)	0.01 (–0.16, 0.21)	0.07 (–61, 4.84)

Note: $N = 99$ interviewed at 0–14 days, 103 at 15–28 days, 129 at more than 28 days. One case is excluded because of missing data on lag time. Model controls for age, gender, employer, job satisfaction, work experience, expectations of recovery, and provider choice. Results represent % change in the probability of a particular employment pattern associated with a 10-unit change in the severity measure. 90% confidence interval given in parentheses, generated by bootstrapping 300 samples from each partition

Source: ASU Healthy Back Study, survey and claims data

first spell of work absence. The model controls for right censoring in cases where a worker has not returned to work.

All severity variables have the expected signs, and are jointly significant at better than the 0.0001 level in each duration model. Back pain intensity is never individually significant, consistent with our patterns models showing that back pain at onset is only a weak predictor of eventual work outcomes. The Roland–Morris and SF-12 scores are all individually significant in the models for different severity measures, underscoring the importance of measures of physical functioning and overall health status in predicting work outcomes. In the combined model it is leg pain intensity and physical health status, rather than mental health status, that retain significance. Thus, the relative importance of mental health measures in predicting work outcomes is somewhat ambiguous, depending on the particular outcomes measure used.

Discussion

Occupational back pain imposes a huge disability burden on employers, workers' compensation systems, and injured workers. A large proportion of costs are attributed to a small fraction of workers who are unable to return to work successfully. If workers at high risk of poor return to work outcomes can be identified early, appropriate interventions can be targeted to these potentially high-cost cases. Unfortunately, clinical measures of the causes of back injuries, including radiography and imaging, are not useful predictors for most cases. Diagnoses such as fractures, metastatic cancers, or spinal stenoses are associated with findings of abnormalities in spinal disks or in the spinal cord itself, but these types of conditions represent only approximately 5% of incident cases [30].

Clinically, pain intensity is the severity measure of choice in predicting outcomes following an episode of back pain. Radiating pain, in particular, has been shown to be a significant predictor of outcomes across numerous studies. Our multinomial logistic results suggest that back pain intensity is a weak predictor of patterns of post-injury employment, and would provide little useful information to clinicians or insurers trying to identify potentially high-cost cases (Pattern 4). Measures of physical functioning and health-related quality of life, on the

other hand, are highly significant and clinically important predictors of employment patterns at all follow-up points. When all severity measures are included in a single model, the pain intensity variables become insignificant and their effect size is considerably reduced.

The results suggest that baseline physical functioning and overall mental and physical health status are more predictive of specific patterns of post-injury employment than pain intensity measures, possibly because there is considerable idiosyncratic variation in the pain intensity measures. This is supported in the literature: When pain intensity is used as an outcome measure following an episode of back pain, clinicians find that NRS-101 pain scores rarely fall to zero. Even patients who report they have improved are not pain free. What appears to matter most in determining patterns of returns to work is not the initial severity of pain, but how well an injured worker is able to function, and how well he or she is able to adapt to the pain.

The one severity measure that retains significance in the patterns models, against all other measures at all follow-up points, is the mental component of the SF-12. Yet other severity measures (leg pain, physical health status) are more predictive in the duration model. There is also conflicting evidence on the predictive power of mental health measures in the literature. Schultz et al. [8] report that the SF-36 mental health score measured at four to 6 weeks after onset is a significant predictor of return to work status at 3 months. Prior research by the same authors, however, reports that psychological distress factors are dominated by cognitive factors (perceptions of health improvement and expectations of recovery) in predicting returns to work [6, 7]. A recent systematic review of the literature finds no high-quality studies supporting the significance of mental health as a predictive factor for duration of work absence among workers with back pain [11]. The role of mental health measures as prognostic factors for outcomes among back pain cases clearly deserves further investigation.

The study has some important limitations. As noted above, we lose a large proportion of our original cases through loss to follow-up and missing data. Comparisons of our study samples with excluded cases show no significant differences in key analysis variables, namely severity measures and expectations of recovery. Nevertheless, our study samples tend to be older and have a higher proportion of females than the original survey sample, and both characteristics are associated with poorer return-to-work outcomes [31]. This may introduce a bias in the distribution of our study sample across employment patterns.

A second limitation is that we were unable to complete interviews at each follow-up point at exactly the same number of days past onset. There were considerable lags in conducting some interviews, particularly baseline interviews, because of late notifications or difficulties contacting injured workers. Thus, the baseline interview should be interpreted as ‘an interview close to onset,’ and follow-up interviews should be interpreted as ‘interviews at approximately one (6, 12) month after onset.’ When the baseline interview was conducted more than 28 days after onset (40% of sample) we administered the baseline and 1-month interviews simultaneously. In these cases ‘baseline’ severity measures and one-month return to work outcomes are collected at the same time, so the severity measures may be endogenous in the patterns model. For this reason we are more confident in our results at 6 and 12 months than at 1 month.

Finally, our models do not include a number of variables that have been shown to be important determinants of return to work outcomes. Some of the variables are unavailable on our data: physical demands of the pre-injury job, diagnostic category (e.g. ruptured disc). Others, such as job accommodations, are not appropriate in a predictive model because they cannot be measured until after a worker returns to work. The results should be interpreted with these exclusions, and the original purpose—to identify severity measures collected at or near onset of back pain that are useful predictors of long-term employment outcomes—in mind.

The useful message for clinicians is that all the self-reported severity measures appear to have significant predictive power on return-to-work outcomes for as long as 1 year, although the pain intensity variables are less useful for predicting specific patterns of employment. The mental component of the SF-12, in particular, is relatively robust to alternate specifications, consistently statistically significant, and has the lowest probability significance level in explaining long-term patterns of employment. This result, which warrants further testing in prospective studies, nevertheless suggests that employment outcomes for workers with occupational back pain could be improved by identifying and treating workers with poor mental health status at the time of onset of back pain.

Acknowledgments This research was supported by a grant, with full freedom to publish, from the National Chiropractic Mutual Insurance Company, a national insurer of doctors of chiropractic. Industry and Institutional funds were also received in support of this work; we are especially grateful to the Office of the Vice President for Research and Economic Affairs, Arizona State University, as well as the College of Family, Home, and Social Sciences at Brigham Young University.

Appendix

Table AI Sample means by survey period

	Baseline (<i>N</i> = 1,832)		1 Month (<i>N</i> = 1,585)		6 Months (<i>N</i> = 1,143)		1 year (<i>N</i> = 761)	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Age	38.41	11.01	38.89	11.00	39.22	10.83	40.18	10.75
Male	0.47	0.49	0.47	0.49	0.47	0.49	0.44	0.49
Work experience	8.13	12.73	8.46	12.94	8.86	13.18	8.99	12.95
Employee choice	0.19	0.39	0.19	0.39	0.19	0.39	0.18	0.39
Job satisfaction	1.89	0.76	1.88	0.76	1.86	0.75	1.87	0.74
Recovery expectations	0.36	0.48	0.35	0.47	0.36	0.48	0.37	0.48
Back pain	51.2	33.04	50.85	32.93	51.8	33.07	52.49	32.91
Leg pain	27.93	34.39	28.09	34.46	29.24	34.56	29.15	34.4
Mental SF-12	48.38	11.54	48.32	11.5	48.44	11.53	48.02	11.85
Physical SF-12	38.73	10.63	38.96	10.65	38.66	10.61	38.22	10.62
Roland–Morris	44.51	29.9	44.49	29.79	45.23	29.61	46.09	29.35

Note: Missing values reduce sample sizes for some variables

Source: ASU Health Back Study, survey and claims data

Table AII Semi-elasticity estimates from the model with no severity measures

	Pattern 1: No absence	Pattern 2: Return and stay	Pattern 3: Multiple spells	Pattern 4: Not yet returned	Probability significance level
<i>1 Month</i>					
Intercept	0.113	0.304	−0.61	−0.916	0.11
Age	−0.001	−0.001	−0.009	0.018	0.26
Male	−0.073	0.186	−0.113	−0.242	0.22
Work experience	0.003	0.0004	0.005	−0.023	0.15

Table AII continued

	Pattern 1: No absence	Pattern 2: Return and stay	Pattern 3: Multiple spells	Pattern 4: Not yet returned	Probability significance level
Employee choice	-0.210	0.117	0.148	0.315	0.21
Job satisfaction	0.0002	-0.057	0.275	-0.129	0.06
Recovery expectations	0.189	0.225	-0.543	-1.044	<0.0001
Firm1	0.781	-0.570	-0.916	-0.173	<0.0001
Firm 2	0.271	-0.067	-0.321	-0.569	0.01
<i>6 months</i>					
Intercept	-0.256	0.991	-0.641	-3.526	0.001
Age	0.002	-0.011	0.01	0.029	0.12
Male	-0.325	0.204	0.056	0.620	0.06
Work experience	-0.001	-0.001	0.005	-0.015	0.77
Employee choice	-0.109	0.168	-0.042	-0.398	0.64
Job satisfaction	0.021	-0.160	0.201	0.167	0.11
Recovery expectations	0.336	0.168	-0.585	-1.175	0.0004
Firm1	0.853	-0.317	-0.718	-0.304	<0.0001
Firm 2	0.447	-0.359	-0.153	0.383	0.01
<i>12 months</i>					
Intercept	-0.342	1.135	-0.292	-5.184	0.01
Age	0.005	-0.019	0.012	0.027	0.15
Male	-0.287	0.062	0.16	0.865	0.31
Work experience	-0.011	0.007	0.004	0.004	0.61
Employee choice	0.2	-0.111	-0.001	-0.715	0.73
Job satisfaction	-0.066	-0.141	0.192	0.102	0.35
Recovery expectations	0.417	0.274	-0.697	-0.291	0.002
Firm1	0.995	-0.382	-0.720	0.472	0.0004
Firm 2	0.453	-0.480	-0.209	2.222	0.02

Notes: $N = 959$ at 1 month, 585 at 6 months, 332 at 12 months. Results represent % changes in the probability of a particular employment pattern associated with a 1-unit change in the independent variable. -2 times log-likelihood = 2217.76 at 1 month, 1327.39 at 6 months, and 730.95 at 12 months

Source: ASU Health Back Study, survey and claims data

References

1. Bureau of Labor Statistics, U.S. Department of Labor. Lost-worktime injuries and illnesses: characteristics and resulting time away from work, 2004. DOL Publication No. USDL 05-2312 December 13, 2005. Retrieved from <http://www.bls.gov/news.release/osh2.nro.htm>.
2. Mychaskiw MA, Thomas J III. Indirect costs due to back pain in the United States [abstract]. Contributed podium presentations of ISPOR Seventh Annual International Meeting. Value Health 2002;5(3):117–137.
3. Hashemi L, Webster BS, Clancy EA. Trends in disability duration and cost of workers' compensation low back pain claims (1988–1996). J Occup Environ Med 1998;40(12):1110–1119.
4. Gallagher RM, Rauh V, Haugh LD, Milhous R, Callas PW, Langelier R, McClallen JM, Frymoyer J. Determinants of return-to-work among low back pain patients. Pain 1989;39(1):55–67.

5. Dozois DJ, Dobson KS, Wong M, Hughes D, Long A. Predictive utility of the CSQ in low back pain: individual vs. composite measures. *Pain* 1996;66(2–3):171–180.
6. Schultz IZ, Crook JM, Berkowitz J, Meloche GR, Milner R, Zuberbier OA, Meloche W. Biopsychosocial multivariate predictive model of occupational low back disability. *Spine* 2002;27(23):2720–2725.
7. Schultz IZ, Crook J, Meloche GR, Berkowitz J, Milner R, Zuberbier OA, Meloche W. Psychosocial factors predictive of occupational low back disability: towards development of a return-to-work model. *Pain* 2004;107(1–2):77–85.
8. Schultz IZ, Crook J, Berkowitz J, Milner R, Meloche GR. Predicting return to work after low back injury using the Psychosocial Risk for Occupational Disability Instrument: a validation study. *J Occup Rehabil* 2005;15(3):365–376.
9. Truchon M, Côté D. Predictive validity of the Chronic Pain Coping Inventory in subacute low back pain. *Pain* 2005;116(3):205–212.
10. Shaw WS, Pransky G, Fitzgerald TE. Early prognosis for low back disability: intervention strategies for health care providers. *Disabil Rehabil* 2001;23(18):815–828.
11. Steenstra IA, Verbeek JH, Heymans MW, Bongers PM. Prognostic factors for duration of sick leave in patients sick listed with acute low back pain: a systematic review of the literature. *Occup Environ Med* 2005;62(12):851–860.
12. Grotle M, Vollestad NK, Veierod MB, Brox JI. Fear-avoidance beliefs and distress in relation to disability in acute and chronic low back pain. *Pain* 2004;112(3):343–352.
13. Côté P, Baldwin ML, Johnson WG. Early patterns of care for occupational back pain. *Spine* 2005;30;5:581–587.
14. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986;27:117–26.
15. Roland M, Morris R. A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;8(2):141–4.
16. Deyo RA. Comparative validity of the Sickness Impact Profile and shorter scales for functional assessment in low-back pain. *Spine* 1986;11(9):951–4.
17. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 1986;39(11):897–906.
18. Stratford PW, Binkley J, Solomon P, Gill C, Finch E. Assessing change over time in patients with low back pain. *Phys Ther* 1994;74(6):528–33.
19. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI. The Quebec Back Pain Disability Scale measurement properties. *Spine* 1995;20(3):341–52.
20. Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris Questionnaire. *Phys Ther* 1996;76(4):359–65; discussion 366–8.
21. Leclaire R, Blier F, Fortin L, Proulx R. A cross-sectional study comparing the Oswestry and Roland-Morris functional disability scales in two populations of patients with low back pain of different levels of severity. *Spine* 1997;22(1):68–71.
22. Beurskens AJ, de Vet HC, Koke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996;65(1):71–6.
23. Hsiegh CY, Phillips RB, Adams AH, Pope MH. Functional outcomes of low back pain: Comparison of four treatment groups in a randomized controlled trial. *J Manipulative Physiol Ther* 1992;15(1):4–9.
24. Ware JE, Kosinski M, Keller SD. SF-12. How to Score the SF-12 Physical, Mental Health summary Scales. 1st ed. Boston MA: The Health Institute, New England Medical Center; 1995.
25. Ware JE, Turner-Bowker DM, Kosinski M, et al. SF-12v2TM: How to Score Version 2 of the SF-12 Health Survey. Lincoln, RI: QualityMetric Incorporated; 2002.
26. Ware J Jr, Kosinski M, Keller KD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34(3):220–33.
27. Luo X, George ML, Kakouras I, Edwards CL, Pietrobon R, Richardson W, Hey L. Reliability, validity, and responsiveness of the Short Form 12-item survey (SF-12) in patients with back pain. *Spine* 2003;28(15):1739–45.
28. Butler RJ. The economics of social insurance, employee benefits. Boston MA: Kluwer Academic Publishers; 1996.
29. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley and Sons; 2000.
30. Jarvik JG, Deyo RA. Diagnostic evaluation of low back pain with emphasis on imaging. *Ann Intern Med* 2002;137:586–597.
31. Johnson WG, Baldwin ML, Butler RJ. Back pain and work disability: the need for a new paradigm. *Industrial Relations* 1998;37(1):9–34.