SPINE Volume 27, Number 24, pp 2835–2843

A Clinical Prediction Rule for Classifying Patients with Low Back Pain Who Demonstrate Short-Term Improvement With Spinal Manipulation

Timothy Flynn, PT, PhD,*‡ Julie Fritz, PT, PhD,† Julie Whitman, PT, DSc,‡ Robert Wainner, PT, PhD,*‡ Jake Magel, PT, DSc,‡ Daniel Rendeiro, PT, DSc,‡ Barbara Butler, PT,‡ Matthew Garber, PT, DSc,‡ and Stephen Allison, PT, PhD*

Study Design. A prospective, cohort study of patients with nonradicular low back pain referred to physical therapy.

Objective. Develop a clinical prediction rule for identifying patients with low back pain who improve with spinal manipulation.

Summary of Background Data. Development of clinical prediction rules for classifying patients with low back pain who are likely to respond to a particular intervention, such as manipulation, would improve clinical decision-making and research.

Methods. Patients with nonradicular low back pain underwent a standardized examination and then underwent a standardized spinal manipulation treatment program. Success with treatment was determined using percent change in disability scores over three sessions and served as the reference standard for determining the accuracy of examination variables. Examination variables were first analyzed for univariate accuracy in predicting success and then combined into a multivariate clinical prediction rule.

Results. Seventy-one patients participated. Thirty-two had success with the manipulation intervention. A clinical prediction rule with five variables (symptom duration, fear–avoidance beliefs, lumbar hypomobility, hip internal rotation range of motion, and no symptoms distal to the knee) was identified. The presence of four of five of these variables (positive likelihood ratio = 24.38) increased the probability of success with manipulation from 45% to 95%.

Conclusion. It appears that patients with low back pain likely to respond to manipulation can be accurately identified before treatment. [Key words: manipulation, classification, low back pain, likelihood ratio, clinical prediction rule] **Spine 2002;27:2835–2843**

From the *U.S. Army-Baylor University Graduate Program in Physical Therapy, San Antonio, Texas, the †Department of Physical Therapy, University of Pittsburgh, Pittsburgh, Pennsylvania, and the ‡U.S. Army-Baylor University, Postprofessional Doctoral Program in Orthopaedic & Manual Physical Therapy, Fort Sam Houston, Texas.
Supported in part by a grant from the Foundation for Physical Therapy.
Acknowledgment date: October 31, 2001. First revision date: March 7, 2002. Acceptance date: May 6, 2002.
The manuscript submitted does not contain information about medical device(s)/drug(s).
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.
The opinions and assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.
Address correspondence to Timothy W. Flynn, PT, PhD, U.S. Army-Baylor University Graduate Program in Physical Therapy, 3151 Scott Road, Room 1303, Fort Sam Houston, TX 78234-6138, USA; E-mail: timothy.flynn@cen.amedd.army.mil

Attempts to identify effective interventions for patients with low back pain (LBP) have been largely unsuccessful.[65,66] One explanation offered for the lack of evidence is the inability to define subgroups of patients most likely to respond to a particular intervention.[8,39] Without the ability to match patients to specific interventions, clinicians are left without evidence or guidance for their decision-making. Because of the difficulty in subgrouping patients with LBP based on pathoanatomy,[67] attempts have been made to classify patients based on findings from the history and physical examination.[17,49,60,64] Developing effective, clinically applicable methods for classifying patients with LBP could improve decision-making and outcomes by matching interventions to the patients they are most likely to benefit. Classification methods would also enhance the power of clinical research by permitting researchers to study more homogeneous groups of patients.[56] Identifying methods for classifying patients with LBP has been identified as an important research priority.[1,6,66]

Manipulation is an intervention commonly used in the treatment of individuals with LBP. Several randomized trials have found manipulation to be more effective than placebo[52,54,71] or other interventions.[16,22,36,63] However, other studies have not shown any benefits for manipulation *versus* other interventions.[10,26–28] The disparity may be partly attributable to the admission of a heterogeneous group of patients with LBP without an attempt to identify *a priori* those likely to benefit from the intervention. Other randomized trials have found manipulation to be more beneficial for a subgroup of patients with more acute symptoms[30,42] or more limited straight-leg raise range of motion.[47] No previous studies have sought to develop a multifactorial classification rule that would maximize the prediction of success with manipulation before the intervention.

One subgroup of patients for whom manipulation is proposed to be effective comprises individuals with sacroiliac (SI) region dysfunction.[16,19,22,29,62,71] Numerous clinical findings have been promoted as capable of identifying patients with SI dysfunction; however, previous studies have questioned the reliability and validity of many of these tests.[4,9,20,21,37,40,41,44,45,50,53,61] Previous studies have generally examined individual tests for SI dysfunction in isolation and typically have not included other potentially important factors from the history or

examination. An exception is the work published by Cibulka *et al*, who described a cluster of movement and provocative tests purported to identify SI joint dysfunction.[12–14] However, this test cluster was chosen based on the authors' clinical experience, and the predictive validity is unknown.[12] Furthermore, most studies have used short-term pain relief with an injection of an anesthetic agent into the SI joint as the reference standard against which validity has been judged.[9,21,32,44] The validity of using anesthetic joint blocks as a reference standard for diagnostic studies has not been demonstrated.[3,32,59] The reference standard should represent the condition of interest that the diagnostic test is attempting to identify.[55] If the goal of a test is to identify patients likely to respond to an intervention directed toward the SI region (*e.g.*, manipulation), then the use of a criterion standard of short-term pain relief with an injection is questionable. A more viable alternative may be a longer duration response to the intervention in question. A study using a more "therapeutic" reference standard, which considers a broader variety of clinical findings and attempts to combine clinical findings, will be more useful in identifying a subgroup of patients with SI dysfunction who are likely to respond to spinal manipulation. The purpose of this study was to develop a clinical prediction rule for identifying patients with LBP likely to respond favorably to a specific manipulation technique. A variety of findings were considered against a reference standard of change in disability.

■ Methods

A prospective cohort study of patients with LBP was conducted at two outpatient facilities: Brooke Army Medical Center and Wilford Hall Air Force Medical Center. Patients were between the ages of 18 and 60 years, referred to physical therapy with a diagnosis related to the lumbosacral spine, and had a chief complaint of pain and/or numbness in the lumbar spine, buttock, and/or lower extremity. The baseline Oswestry disability score had to be at least 30%. Exclusion criteria were current pregnancy, signs consistent with nerve root compression (positive straight-leg raise at $<45°$, or diminished lower extremity strength, sensation, or reflexes), prior lumbar spine surgery, or a history of osteoporosis or spinal fracture. All patients were briefed on the purpose of the study and signed an informed consent approved by the Brooke Army Medical Center and Wilford Hall Air Force Medical Center Institutional Review Boards.

Therapists. Eight licensed physical therapists participated. Four were residents in the U.S. Army-Baylor Post-Professional Doctoral Program in Orthopedic and Manual Physical Therapy, and four were instructors in the program. This program is designed to provide physical therapists serving in the U.S. Military with advanced training in orthopedic and manual physical therapy. A 1-day training session was conducted for participating therapists to standardize examination and treatment techniques.

Examination Procedures. Patients completed a baseline examination including demographic information and a pain rating using an 11-point scale.[34] A pain diagram[46] was used to categorize symptoms as low back, buttock/thigh, or distal to the knee based on the distal-most extent of symptoms.[70] The Modified Oswestry Disability Questionnaire (OSW) assessed disability related to LBP.[25] The Fear-Avoidance Beliefs Questionnaire (FABQ) was used to assess the patient's beliefs about the influence of activity on LBP.[69] The FABQ contains two subscales: one is related to general physical activity and the other to work.[69]

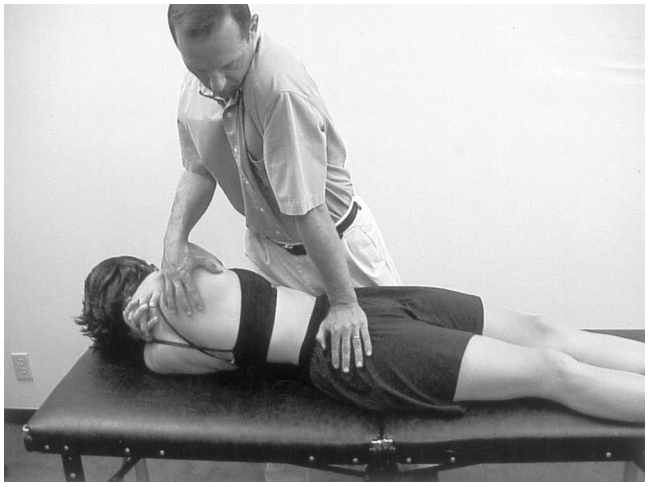

Figure 1. Manipulation technique used in this study.

Patients underwent a standardized history and physical examination. History included mechanism of injury, nature of current symptoms, and prior episodes of LBP. Patients were asked to rank sitting, standing, and walking as to which was best and worst with respect to symptoms. The examination included Waddell's nonorganic signs.[68] Range of motion and status change[49] in symptoms with single lumbar movements were recorded. Supine straight-leg raise and prone hip rotation range of motion were measured. Posteroanterior spring testing[43] was performed for pain provocation and mobility at each lumbar level. Mobility was judged as normal, hypomobile, or hypermobile. Numerous special tests proposed to be diagnostic of SI dysfunction were performed (see Appendix). The tests were divided into three categories: position (tests assessing symmetry of bony landmarks), provocation (tests to reproduce symptoms), and mobility (tests assessing symmetry of pelvic motion).[37] The examination was repeated on the first 55 subjects by a second therapist blinded to the results of the first examiner to determine the reliability of the physical examination variables.

Treatment. Because response to treatment served as the reference standard, all patients were treated with the same protocol for two sessions. At the first session, the therapist performed a manipulation technique with the patient supine. The therapist stood opposite the side to be manipulated. The patient was passively side-bent away from the therapist. The therapist passively rotated the patient and then delivered a quick posterior and inferior thrust through the anterior superior iliac spine (Figure 1). The side to be manipulated was determined with the following algorithm: first, the side of the positive standing flexion test; if this test was negative, the side of tenderness during sacral sulcus palpation was manipulated. If neither side was tender, the side reported by the patient to be more symptomatic was manipulated. If the patient was unable to identify a more

symptomatic side, the therapist flipped a coin to determine the side. Although the manipulation was performed on one side only, Cibulka *et al*[12] found changes in innominate tilt on both sides of the pelvis after applying this technique. Therefore, whereas the algorithm provided a consistent approach, it is likely the manipulation affected bilateral SI regions.

After the manipulation, the therapist noted whether a cavitation was heard or felt by the therapist or patient. If a cavitation was experienced, the therapist proceeded to the other treatment components. If no cavitation was produced, the patient was repositioned, and the manipulation was attempted again. If no cavitation was experienced again, the therapist attempted to manipulate the opposite side. A maximum of two attempts per side was permitted. If no cavitation was produced after the fourth attempt, the therapist proceeded to the other treatment components. Two additional treatment components were included: 1) instruction in a supine pelvic tilt range of motion exercise (the patient was instructed to perform 10 repetitions, 3–4 times daily; and 2) instruction to maintain usual activity level within the limits of pain.[5]

The second treatment session occurred 2–4 days after the first. Before the second session, each patient completed an OSW questionnaire. Percentage improvement in OSW was calculated (initial score − final score)/initial score × 100). If improvement was >50%, the patient was categorized as a success, and study participation ended. If the patient showed <50% improvement, the therapist repeated the examination and the manipulation procedures. The third session occurred 2–4 days after the second. The patient again completed the OSW, and the percentage improvement from the initial score was calculated. If >50% improvement was noted, the patient was categorized as a success. If improvement was ≤50%, the patient was categorized as a nonsuccess. At this point, the patient's study participation ended, and further treatment was administered as needed.

Data Analysis. Kappa coefficients were calculated to determine the interrater reliability of the special tests for SI dysfunction. Patients were dichotomized based on success or nonsuccess with respect to the treatment. Success or nonsuccess was then used as the reference standard. Individual variables from the self-reports, history, and physical examination were tested for their univariate association with the reference standard using independent sample *t* tests for continuous variables and χ^2 tests for categorical variables. Variables with a significance level of $P < 0.15$ were retained as potential prediction variables; a more liberal significance level was chosen at this stage to avoid excluding potential predictive variables. For continuous variables with a significant univariate association, sensitivity and specificity values were calculated for all possible cut-off points and then plotted as a receiver operator characteristic (ROC) curve.[31] The point on the curve nearest the upper left-hand corner represents the value with the best diagnostic accuracy, and this point was selected as the cut-off defining a positive test.[18] Sensitivity, specificity, and positive likelihood ratios (PLR) were calculated for all potential prediction variables.[57] The PLR is calculated as sensitivity/(1 − specificity) and indicates the increase in the probability of success given a positive test result.[58] A PLR of 1 indicates the test does nothing to alter the probability of success, whereas PLR values >1 increase the probability of success given a positive test result. According to Jaeschke *et al*,[33] PLR values between 2.0 and 5.0 generate small shifts in probability, values between 5.0 and 10.0 generate moderate shifts, and values >10.0 generate large and often conclusive shifts in probability. We chose to focus on the PLR as opposed to the negative LR because we were attempting to predict success with manipulation based on positive test results. Potential prediction variables were entered into a stepwise logistic regression equation to determine the most parsimonious set of predictors for success using a multivariate model. A significance of 0.05 was required to enter a variable into the model and a significance of 0.10 was required to remove it. Variables retained in the regression model were used to develop a multivariate clinical prediction rule for classifying patients as likely responders to manipulation.

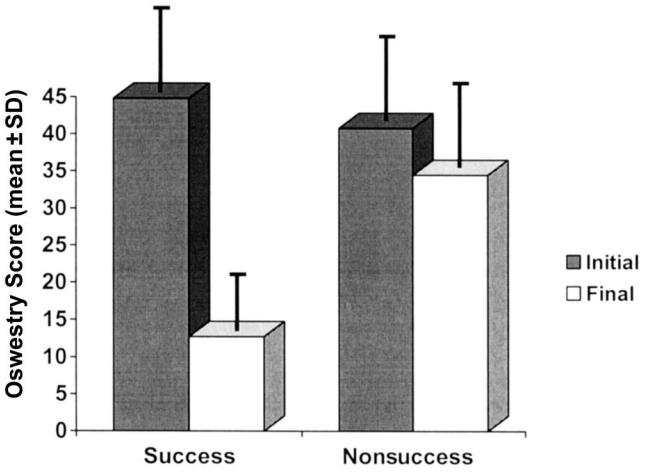

Figure 2. Initial and final Oswestry scores for the success and nonsuccess groups. The mean percent change in the success group was 73.2 ± 15.8%. For the nonsuccess group, the mean percent change was 14.6 ± 18.2%.

■ Results

Seventy-five patients entered the study. Four subjects (5%) did not return after the first session and were not included in the analysis. Two subjects left the study because of personal or work-related circumstances. One subject dropped out because of complications from an ongoing episode of gastrointestinal distress, and one subject failed to return for his final visit. Of the 71 patients completing the study, 29 (41%) were female and 59 (83%) had a prior history of LBP. The mean age was 37.6 ± 10.6 years (range 18–59 years). The mean OSW score at baseline was 42.4 ± 11.7, and at study conclusion 25.1 ± 13.9. The mean percent improvement in OSW over the study period was 41.0 ± 33.9% (range −29.0–100%). Thirty-two patients (45%) were classified as treatment successes, and 39 (55%) were nonsuccesses. Twenty patients were successes after one manipulation session, 12 after two sessions. The mean improvement in OSW in the success group over the study period was 32.5 ± 12.6 points, with a mean percent improvement of 73.2 ± 15.8%. In the nonsuccess group, the mean OSW improvement was 6.2 ± 7.8 points, with a mean percent improvement of 14.6 ± 18.2% (Figure 2).

Among self-reported variables (Table 1), the FABQ work subscale, the presence of symptoms in the back

Table 1. Self-Report Variables Used in the Study

Self-Report Variable	All Subjects (n = 71)	Manipulation Success (n = 32)	Manipulation Nonsuccess (n = 39)	Significance
Pain rating	5.3 (2.0)	5.3 (2.0)	5.2 (1.7)	0.74
Oswestry score	42.4 (11.7)	44.4 (14.4)	40.8 (8.7)	0.21
Fear avoidance beliefs				
Work subscale	12.9 (10.3)	9.1 (8.3)	16.0 (10.8)	0.003
Physical activity subscale	15.1 (5.3)	14.9 (5.0)	15.2 (5.6)	0.85
Pain diagram (%)				
Low back symptoms only	31	41	23	0.11
Buttock/thigh symptoms present	76	69	82	0.19
Symptoms distal to knee	25	13	36	0.024

Values are means (SD) unless otherwise indicated.

only, and symptoms distal to the knee were retained as potential prediction variables. From the history (Table 2), duration of symptoms, increasing episode frequency, and ranking standing as the worst position were retained. Five variables were retained from the clinical examination (Table 3): left and right hip internal rotation, hypomobility and pain with lumbar spring testing, and peripheralization with single lumbar movement testing. The special tests for SI dysfunction along with the reliability coefficients are presented in Table 4. As a group, the provocation tests were more reliable with values ranging from fair to substantial agreement. The Gillet test was the only motion test to demonstrate at least moderate agreement. The anterior superior iliac spine and iliac crest in standing and the posterior superior iliac spine symmetry in sitting were the only symmetry tests to have at least fair agreement. Among the special tests for SI dysfunction (Table 4), only the compression–distraction test was retained in the predictive model, although positive findings were more common in the nonsuccess group. Cut-off values for retained continuous variables (duration of symptoms, FABQ work subscale, left and right hip internal rotation) were obtained from receiver operator characteristic curve analyses. Because of their similarity, left and right hip internal rotations were combined into a single variable. Cut-off scores and accuracy statistics for retained variables were calculated (Table 5). Among historical variables, duration of symptoms <16 days was most predictive of success (PLR = 4.39).

The 11 potential prediction variables were entered into the logistic regression. Five were retained in the final model: duration of symptoms <16 days, at least one hip with >35° of internal rotation, hypomobility with lumbar spring testing, FABQ work subscale score <19, and no symptoms distal to the knee (model $\chi^2 = 48.5$, $df = 5$, $P < 0.001$, Nagelkerke $R^2 = 0.67$). These five variables were used to form the clinical prediction rule. Only six subjects (all in the success group) were positive for all five retained prediction variables at baseline (Table 6). Fourteen of 15 subjects with 4 of 5 variables present were in the success group. Of subjects with two or fewer variables present, 25 of 27 were in the nonsuccess group. Accuracy statistics were calculated for each level of the clinical prediction rule (Table 7). Based on the pretest probability of success with manipulation found in this study (45%), and the PLR values calculated, a subject with four or more variables present at baseline increases

Table 2. Variables From the Patient History Used in This Study

Variable	All Subjects	Manipulation Success	Manipulation Nonsuccess	Significance
Age (yrs)	37.6 (10.6)	39.5 (11.7)	36.1 (9.1)	0.19
Gender (% female)	41	47	36	0.35
Duration of symptoms (days)	41.7 (54.7)	23.5 (32.5)	56.7 (64.4)	0.007
Mode of onset (%)				
Gradual	31	26	33	0.64
Sudden (minimal/no perturbation)	39	48	33	0.24
Traumatic (fall, lifting, pulling, *etc.*)	30	26	33	0.44
Prior history of low back pain (%)	83	88	79	0.37
Episodes of low back pain becoming more frequent (%)	35	25	44	0.10
Sitting ranked as best position (%)	31	31	31	0.97
Standing ranked as best position (%)	20	25	15	0.31
Walking ranked as best position (%)	23	16	28	0.21
Sitting ranked as worst position (%)	41	47	36	0.35
Standing ranked as worst position (%)	27	19	36	0.055
Walking ranked as worst position (%)	13	16	10	0.50

Values are means (SD) unless otherwise indicated.

Table 3. Variables From the Baseline Clinical Examination Used in This Study

Variable	All Subjects	Manipulation Success	Manipulation Nonsuccess	Significance
Nonorganic signs*	0.70 (1.0)	0.48 (0.77)	0.74 (1.1)	0.25
Total flexion (°)	75.7 (30.0)	75.3 (31.2)	81.1 (25.7)	0.40
Pelvic flexion (°)	38.0 (20.7)	37.4 (19.9)	41.1 (20.5)	0.45
Lumbar flexion (°)	38.3 (15.4)	37.6 (18.4)	40.0 (12.5)	0.58
Total extension (°)	21.5 (9.5)	21.4 (8.2)	21.4 (9.8)	0.98
Left side bending (°)	27.5 (7.8)	27.6 (7.7)	28.1 (8.0)	0.80
Right side bending (°)	27.3 (7.2)	27.0 (6.4)	28.1 (7.5)	0.53
Left and right side bending discrepancy† (°)	5.2 (4.7)	6.0 (4.4)	4.7 (5.0)	0.26
Left straight leg raise (°)	67.3 (16.8)	69.2 (14.9)	68.5 (16.3)	0.85
Right straight leg raise (°)	67.5 (16.7)	68.6 (14.3)	69.2 (16.7)	0.87
Left and right straight-leg raise discrepancy† (°)	6.0 (6.9)	5.6 (5.6)	6.5 (8.1)	0.59
Left hip internal rotation (°)	29.6 (9.8)	33.0 (9.7)	27.2 (9.3)	0.012
Right hip internal rotation (°)	30.3 (12.5)	29.6 (10.4)	25.9 (10.7)	0.15
Left and right hip internal rotation discrepancy† (°)	6.4 (5.7)	6.8 (6.5)	6.0 (5.3)	0.58
Left hip external rotation (°)	30.1 (11.9)	31.3 (13.3)	29.5 (12.9)	0.48
Right hip external rotation (°)	30.3 (12.5)	31.8 (12.6)	29.3 (11.1)	0.45
Left and right hip external rotation discrepancy† (°)	6.1 (5.3)	5.8 (5.3)	6.6 (5.4)	0.53
Hypomobility at one or more lumbar levels with spring testing (%)	86	97	77	0.016
Pain at one or more lumbar levels with spring testing (%)	92	97	87	0.14
Lateral shift present (%)	13	13	13	0.97
Peripheralizes with lumbar single movement testing (%)	25	16	33	0.088
Centralizes with lumbar single movement testing (%)	6	6	5	0.84

Values represent means (standard deviation) unless otherwise indicated.
* Mann-Whitney test used due to non-normal data distribution.
† Discrepancy values were calculated as the absolute value of the left minus the right.

his or her probability of success with manipulation from 45% to 95%. If the criteria were changed to three or more variables present, the probability of success was only increased to 68%. If two or fewer variables were present, the probability of success was virtually unchanged.

■ Discussion

Clinicians who routinely use spinal manipulation have encountered patients who experience a rapid, even dramatic improvement as a result of one or two treatments, whereas others change very little. The ability to accu-

Table 4. Special Tests for SI Dysfunction Used in This Study

	Reliability	All Subjects (%)	Manipulation Success (%)	Manipulation Nonsuccess (%)	Significance
Provocation tests					
Posterior shear test	0.70	56	56	56	0.99
Sacral sulcus test	0.64	67	75	62	0.23
Patrick test	0.60	44	47	41	0.62
Gaenslen test	0.54	44	47	44	0.78
Resisted hip abduction	0.41	36	34	38	0.72
Sacral thrust test	0.41	54	63	49	0.25
Compression–distraction test	0.26	23	16	31	0.14
Motion tests					
Gillet test	0.59	60	66	56	0.43
Seated flexion test	0.25	66	72	59	0.26
Long-sitting test	0.21	46	39	51	0.35
Prone knee bend test	0.21	56	47	62	0.22
Standing flexion test	−0.08	54	47	62	0.22
Symmetry tests					
ASIS symmetry in standing	0.31	36	41	33	0.53
Iliac crest symmetry in standing	0.23	33	34	33	0.93
PSIS symmetry in sitting	0.23	31	31	31	0.97
PSIS symmetry in standing	0.13	41	44	38	0.65
Ischial tuberosity heights in prone	0.03	41	35	46	0.43
Pubic tubercle symmetry in supine	−0.04	49	41	56	0.18

ASIS = anterior superior iliac spine; PSIS = posterior superior iliac spine.
Reliability measured with kappa and weighted kappa statistics. Percentage values represent the percentage of positive tests in all patients and the success and nonsuccess groups.

Table 5. Accuracy Statistics (With 95% Confidence Intervals) for Individual Variables for Predicting Success

Variable Associated With Success	Sensitivity	Specificity	Positive Likelihood Ratio
Duration of symptoms ≤15 days	0.56 (0.39, 0.72)	0.87 (0.73, 0.94)	4.39 (1.83, 10.51)
Episodes not becoming more frequent	0.75 (0.58, 0.87)	0.44 (0.29, 0.59)	1.33 (0.95, 1.87)
Standing not ranked as worst position	0.84 (0.67, 0.93)	0.36 (0.23, 0.52)	1.31 (1.0, 1.74)
FABQ work subscale ≤18	0.84 (0.68, 0.93)	0.49 (0.34, 0.64)	1.65 (1.17, 2.31)
Symptoms in the low back only	0.41 (0.26, 0.58)	0.77 (0.62, 0.87)	1.76 (0.87, 3.58)
Symptoms not distal to the knee	0.88 (0.72, 0.95)	0.36 (0.23, 0.52)	1.36 (1.04, 1.79)
At least one hip internal rotation range of motion >35°	0.50 (0.34, 0.66)	0.85 (0.70, 0.93)	3.25 (1.44, 7.33)
Hypomobility at one or more lumbar levels with spring testing	0.97 (0.84, 0.99)	0.23 (0.13, 0.38)	1.26 (1.05, 1.51)
Pain at one or more lumbar levels with spring testing	0.97 (0.84, 0.99)	0.13 (0.056, 0.27)	1.11 (0.97, 1.27)
Does not peripheralize with lumbar single movement testing	0.84 (0.68, 0.93)	0.33 (0.21, 0.49)	1.27 (0.97, 1.65)
Negative compression/distraction test	0.84 (0.68, 0.93)	0.31 (0.19, 0.46)	1.22 (0.94, 1.58)

FABQ = Fear-Avoidance Beliefs Questionnaire.

rately predict which patients will have which response *a priori* would be immensely beneficial for clinical decision-making. Similar to other studies,[21,40,41,44] we were unable to show acceptable accuracy for any individual tests proposed to identify SI dysfunction. Furthermore, we found that the reliability of these tests in a population of individuals with LBP is less than optimal. As noted by previous researchers, provocation tests as a whole are more reliable tests than motion or symmetry tests.[9,37] However, by considering other variables and combining findings, we were able to develop a clinical prediction rule that may be useful for assisting clinicians in classifying patients as likely to respond to this manipulation technique.

The developed clinical prediction rule contains five variables: duration of symptoms <16 days, at least one hip with >35° of internal rotation, lumbar hypomobility, no symptoms distal to the knee, and an FABQ work score <19. These findings are generally consistent with previous theories and research. Randomized trials have suggested that patients with more acute symptoms respond better to manipulation.[30,42] Our results support this hypothesis. Hip rotation range of motion discrepancies have been reported in patients with LBP.[2,14,51] Previous studies in patients with "nonspecific" LBP have found greater external rotation than internal rotation.[2,11,51] As a whole, patients in this study had greater external than internal rotation; however, increased internal rotation was associated with manipulation success.

Table 6. Number of Subjects in the Success and Nonsuccess Groups at Each Level of the Clinical Prediction Rule

No. of Predictor Variables Present	No. of Subjects in the Manipulation Success Group	No. of Subjects in the Manipulation Nonsuccess Group
5	6	0
4	14	1
3	10	13
2	2	19
1	0	5
0	0	1

Manipulation is thought to be indicated in the presence of hypomobility. Interestingly, although the technique used in this study is described as affecting the SI region, it was lumbar hypomobility that entered the prediction model. This finding reinforces the idea that the manipulation technique is not specific to the SI region but impacts the lumbar spine as well.[7,17,29] Manipulation is generally thought to be contraindicated in patients with radiculopathy.[5] We excluded patients with signs of nerve root compression. However, some patients with symptoms distal to the knee were included, and these patients tended not to succeed. Finally, the FABQ quantifies a patient's fear of pain and subsequent avoidance of activity.[69] The FABQ work subscale has been previously correlated with work loss and disability in patients with chronic and acute LBP.[24,35,69] Our results suggest that patients with high levels of fear-avoidance beliefs about work activities are unlikely to respond to manipulation. These individuals likely require an alternative treatment approach.[15]

The usefulness of a clinical prediction rule for classifying patients is best expressed using likelihood ratio statistics. The PLR expresses the change in odds favoring the outcome when the patient meets the prediction rule's criteria.[57] In our sample, 45% of subjects were successful without any attempt at prediction. In other words, randomly manipulating individuals with nonradicular LBP may result in success about 45% of the time. Using a criterion of at least 4 of 5 variables present at baseline (PLR = 24.38), the probability of success is raised to 95%; therefore, these individuals should be manipulated. If only three variables are present, the probability increased to 68%, which is likely sufficient to warrant an attempt at manipulation in these patients. When two or fewer variables are present, the probability of success changes little, and clinicians should consider alternative treatments if such can be identified that may have a probability of success >45%.

An important consideration in the examination of diagnostic tests is the reference standard against which tests are judged. Previous studies of tests for SI dysfunction have generally used immediate pain relief with SI

Table 7. Clinical Prediction Rule

No. of Predictor Variables Present	Sensitivity	Specificity	Positive Likelihood Ratio	Probability of Success* (%)
5	0.19 (0.09, 0.35)	1.00 (0.91, 1.00)	infinite (2.02, infinite)	—
4+	0.63 (0.45, 0.77)	0.97 (0.87, 1.0)	24.38 (4.63, 139.41)	95
3+	0.94 (0.80, 0.98)	0.64 (0.48, 0.77)	2.61 (1.78, 4.15)	68
2+	1.00 (0.89, 1.0)	0.15 (0.07, 0.30)	1.18 (1.09, 1.42)	49
1+	1.00 (0.89, 1.0)	0.03 (0.005, 0.13)	1.03 (1.01, 1.15)	46

* The probability of success is calculated using the positive likelihood ratio and assumes a pretest probability of success of 45%. Accuracy statistics with 95% confidence intervals for individual variables for predicting success.

joint anesthetic injection. In our opinion, clinicians performing these tests are not as interested in pathoanatomic speculations (*i.e.*, is the SI joint generating the pain?) as they are in determining if the patient will respond to a particular intervention. We therefore used a reference standard representative of the desired outcome of the tests (*i.e.*, responding to manipulation). The use of 50% improvement on the OSW as the reference standard was based on previous research involving the intervention used in this study. In three previous studies, patients thought to be matched to this intervention experienced mean improvements in OSW scores from 57% to 83%, whereas patients receiving unmatched interventions experienced mean improvements ranging from 20% to 38% over a 1–4-week period.[16,22,23] We therefore thought that requiring 50% improvement in the OSW over a 2–4-day period would provide adequate distinction between patients responding to the intervention and those simply benefiting from the favorable natural history of LBP.

The patients participating in this study should be representative of patients seeking physical therapy services in large metropolitan areas. The eight physical therapists involved in the study had varying degrees of skills in spinal manipulative therapy. However, the manipulation technique employed is a standard technique used in physical therapist education programs. Therefore, the results should be generalizable to outpatient clinics treating individuals with LBP.

A three-step process for developing and testing a clinical prediction rule is recommended.[48] The first step is developing the rule, the second step is validation, and the third step is an assessment of the impact of the rule on clinical behavior. The purpose of the present study was to develop a clinical prediction rule that would identify individuals with LBP who respond favorably to a specific spinal manipulation. In the present study, only one manipulation technique was used, and it is unknown whether other techniques would provide similar results. Validation of the proposed clinical prediction rule is the purpose of an ongoing randomized controlled trial where subjects meeting the prediction criteria receive either the spinal manipulation technique or a competing therapy. Ultimately, any clinical prediction rule must be shown to improve outcomes and clinical decision-making before it can be advocated for widespread use.[38,48]

■ Key Points

- Special tests purported to identify patients with low back pain who will respond to manipulation were largely unsuccessful in doing so.
- The best univariate predictor of success with manipulation was the duration of the current symptoms; more acute symptoms were more likely to respond favorably.
- Five variables were identified to form a clinical prediction rule for patients with low back pain likely to respond favorably to spinal manipulation: duration of symptoms <16 days, FABQ work subscale score <19, at least one hip with >35° of internal rotation range of motion, hypomobility in the lumbar spine, and no symptoms distal to the knee.
- The presence of four of five variables in the prediction rule increased the likelihood of success with manipulation from 45% to 95%.

References

1. Clinical research agenda for physical therapy. Phys Ther 2000;80:499–513.
2. Barbee-Ellison JB, Rose SJ, Sahrmann SA. Patterns of hip rotation range of motion: comparisons between healthy subjects and patients with low back pain. Phys Ther 1990;70:537–41.
3. Barnsley L, Lord S, Wallis B, et al. False-positive rates of cervical zygapophysial joint blocks. Clin J Pain 1993;9:124–30.
4. Bemis T, Daniel M. Validation of the long sitting test on subjects with iliosacral dysfunction. J Orthop Sports Phys Ther 1987;8:336–45.
5. Bigos S, Bowyer O, Braen G, et al. Acute low back problems in adults [AHCPR Publication No. 95-0642]. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, 1994.
6. Borkan JM, Koes B, Reis S, et al. A report from the second international forum for primary care research on low back pain: reexamining priorities. Spine 1998;23:1992–6.
7. Bourdillon JF, Day EA, Bookhout MR. Spinal Manipulation. 5th ed. Oxford: Butterworth-Heinemann, 1992;156–8.
8. Bouter LM, van Tulder MW, Koes BW. Methodologic issues in low back pain research in primary care. Spine 1998;23:2014–20.
9. Broadhurst NA, Bond MJ. Pain provocation for the assessment of sacroiliac joint dysfunction. J Spinal Dis 1998;11:341–5.
10. Cherkin DC, Deyo RA, Battie M, et al. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. N Engl J Med 1998;339:1021–9.
11. Chesworth BM, Padfield BJ, Helewa A, et al. A comparison of hip mobility in patients with nonspecific low back pain. Physiother Can 1994;46:267–74.
12. Cibulka MT, Delitto A, Koldehoff R. Changes in innominate tilt after ma-

nipulation of the sacroiliac joint in patients with low back pain. Phys Ther 1988;68:1359–63.
13. Cibulka MT, Koldehoff R. Clinical usefulness of a cluster of sacroiliac joint tests in patients with and without low back pain. J Orthop Sports Phys Ther 1999;29:83–92.
14. Cibulka MT, Sinacore DR, Cromer GS, et al. Unilateral hip rotation range of motion asymmetry in patients with sacroiliac joint regional pain. Spine 1998; 23:1009–15.
15. 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.
16. Delitto A, Cibulka MT, Erhard RE, et al. Evidence for use of an extension-mobilization category in acute low back syndrome: a prescriptive validation pilot study. Phys Ther 1993;73:216–28.
17. 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.
18. 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.
19. DonTigny R. Dysfunction of the sacroiliac joint and its treatment. J Orthop Sports Phys Ther 1979;1:23–9.
20. Dreyfuss P, Dreyer S, Griffin J, et al. Positive sacroiliac screening tests in asymptomatic adults. Spine 1994;19:1138–43.
21. Dreyfuss P, Michaelsen M, Pauza K, et al. The value of medical history and physical examination in diagnosing sacroiliac joint pain. Spine 1996;21: 2594–602.
22. Erhard RE, Delitto A, Cibulka MT. Relative effectiveness of an extension program and a combined program of manipulation and flexion and extension exercises in patients with acute low back syndrome. Phys Ther 1994; 74:1093–100.
23. Fritz JM, George S. The use of a classification approach to identify subgroups of patients with acute low back pain: inter-rater reliability and short-term treatment outcomes. Spine 2000;25:106–14.
24. 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.
25. 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.
26. Gibson T, Grahame R, Harkness J, et al. Controlled comparison of short-wave diathermy treatment with osteopathic treatment in non-specific low-back pain. Lancet 1985;1:1258–60.
27. Glover JR, Morris JG, Kholsa T. Back pain: a randomized clinical trial of rotational manipulation of the trunk. Br J Ind Med 1974;31:59–64.
28. Godfrey CM, Morgan PP, Schatzker J. A randomized trial of manipulation for low back pain in a medical setting. Spine 1984;9:301–4.
29. Greenman P. Osteopathic manipulation of the lumbar spine and pelvis. In: White A, Anderson A, eds. Conservative Care of Low Back Pain. Baltimore, MD: Williams & Wilkins, 1991:210–5.
30. Hadler NM, Curtis P, Gillings DB, et al. A benefit of spinal manipulation as an adjunctive therapy for acute low-back pain: a stratified controlled study. Spine 1987;12:703–5.
31. Hagen MD. Test characteristics: how good is that test? Med Decis Making 1995;22:213–33.
32. Hogan QH, Abram SE. Neural blockade for diagnosis and prognosis: a review. Anesthesiology 1997;86:216–41.
33. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA 1994;271:389–91.
34. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? Pain 1994;58:387–92.
35. Klenerman L, Slade PD, Stanley IM, et al. The prediction of chronicity in patients with an acute attack of low back pain in a general practice setting. Spine 1995;20:478–84.
36. Koes BW, Bouter LM, van Mameren H, et al. The effectiveness of manual therapy, physiotherapy and treatment by the general practitioner for back and neck complaints: a randomized clinical trial. Spine 1992;17:28–35.
37. Laslett M, Williams M. The reliability of selected provocation tests for sacroiliac joint pathology. Spine 1994;19:1245–9.
38. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. JAMA 1997;277:488–94.
39. Leboeuf-Yde C, Lauritsen JM, Lauritzen T. Why has the search for causes of low back pain largely been nonconclusive? Spine 1997;22:877–81.
40. Levangie PK. Four clinical tests of sacroiliac joint dysfunction: the association of test results with innominate torsion among patients with and without low back pain. Phys Ther 1999;79:1043–57.
41. Levangie PK. The association between static pelvic asymmetry and low back pain. Spine 1999;24:1234–42.
42. MacDonald RS, Bell CMJ. An open controlled assessment of osteopathic manipulation in nonspecific low-back pain. Spine 1990;15:364–70.
43. Maher CG, Latimer J, Adams R. An investigation of the reliability and validity of posteroanterior spinal stiffness judgments made using a reference-based protocol. Phys Ther 1998;78:829–37.
44. Maigne J-Y, Aivaliklis A, Pfefer F. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. Spine 1996;21:1889–92.
45. Maigne J-Y, Boulahdour H, Chatellier G. Value of quantitative radionuclide bone scanning in the diagnosis of sacroiliac joint syndrome in 32 patients with low back pain. Eur Spine J 1998;7:328–31.
46. Mann NH, Brown MD, Hertz DB, et al. Initial-impression diagnosis using low-back pain patient pain drawings. Spine 1993;18:41–53.
47. Mathews W, Morkel M, Mathews J. Manipulation and traction for sciatica and lumbago: physiotherapeutic techniques in two controlled trials. Physiother Pract 1987;4:201–6.
48. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guide to the medical literature: XXII. How to use articles about clinical decision rules. JAMA 2000; 284:79–84.
49. McKenzie RA. The Lumbar Spine: Mechanical Diagnosis and Therapy. Waikanae, New Zealand: Spinal Publications, 1989.
50. Meijne W, van Neerbos K, Aufdemkampe G, et al. Intraexaminer and interexaminer reliability of the Gillet test. J Manipulative Physiol Ther 1999;22:4–9.
51. Mellin G. Correlation of hip mobility with degree of back pain and lumbar spinal mobility in chronic low back pain patients. Spine 1988;13:668–70.
52. Postacchini F, Facchini M, Palieri P. Efficacy of various forms of conservative treatments in low back pain: a comparative study. Neuro Orthop 1988;6: 28–35.
53. Potter NA, Rothstein JM. Intertester reliability of selected clinical tests of the sacroiliac joint. Phys Ther 1985;65:1671–5.
54. Rasmussen GG. Manipulation in treatment of low-back pain: a randomized clinical trial. Manual Med 1979;1:8–10.
55. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research: getting better but still not good. JAMA 1995;274: 645–51.
56. Rose SJ. Physical therapy diagnosis: role and function. Phys Ther 1989;69: 535–7.
57. Sackett DL. A primer on the precision and accuracy of the clinical examination. JAMA 1992;267:2638–44.
58. Sackett DL, Haynes RB, Guyatt GH, et al. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd ed. Boston, MA: Little Brown, 1991:119–39.
59. Schwarzer AC, Aprill CN, Derby R, et al. The false positive rate of uncontrolled diagnostic blocks of the lumbar zygapophyseal joints. Pain 1994;58: 195–200.
60. Sikorski JM. A rationalized approach to physiotherapy for low back pain. Spine 1985;10:571–9.
61. Slipman CW, Sterenfeld EB, Chou LH, et al. The predictive value of provocative sacroiliac joint stress maneuvers in the diagnosis of sacroiliac joint syndrome. Arch Phys Med Rehabil 1998;79:288–92.
62. Sutton SE. Postural imbalances: examination and treatment utilizing flexion tests. J Am Osteopath Assoc 1978;77:456–65.
63. Triano JJ, McGregor M, Hondras MA, et al. Manipulative therapy versus education programs in chronic low back pain. Spine 1995;20:948–55.
64. Van Dillen LR, Sahrmann SA, Norton BJ, et al. Reliability of physical examination items used for classification of patients with low back pain. Phys Ther 1998;78:979–88.
65. van Tulder MW, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain: a systematic review of randomized controlled trials of the most common interventions. Spine 1997;22:2128–56.
66. van Tulder MW, Malmivaara A, Esmail R, et al. Exercise therapy for low back pain: a systematic review within the framework of the Cochrane Collaboration Back Review Group. Spine 2000;25:2784–96.
67. Waddell G. Low back pain: a twentieth century health care enigma. Spine 1996;21:2820–5.
68. Waddell G, McCulloch JA, Kummel E, et al. Nonorganic signs in low-back pain. Spine 1980;5:117–25.
69. Waddell G, Newton M, Henderson I, et al. A Fear-Avoidance Beliefs Questionnaire (FABQ), the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52:157–68.
70. Werneke M, Hart DL, Cook D. A descriptive study of the centralization phenomenon: a prospective analysis. Spine 1999;24:676–83.
71. Wreje U, Nordgren B, Aberg H. Treatment of pelvic joint dysfunction in primary care: a controlled study. Scand J Prim Care 1992;10:310–5.

Appendix

Special Tests for SI Dysfunction

	Procedure	Criteria for Positive
Tests for symmetry		
PSIS symmetry standing[17]	Palpation of right and left PSIS with the patient standing	One PSIS judged to be higher than the other
ASIS symmetry standing[17]	Palpation of right and left ASIS with the patient standing	One ASIS judged to be higher than the other
Iliac crest symmetry standing[17]	Palpation of right and left iliac crest with the patient standing	One iliac crest judged to be higher than the other
Landmark symmetry standing[17]	Comparison of PSIS, ASIS, and iliac crest findings in standing	All landmarks are not level and all landmarks are not high on the same side
PSIS symmetry sitting[13]	Palpation of right and left PSIS with the patient sitting	One PSIS judged to be higher than the other
Pubic tubercle symmetry supine[29]	Palpation of the right and left tubercle with the patient supine	One pubic tubercle judged to be higher than the other
Ischial tuberosity symmetry in prone[29]	Palpation of the right and left ischial tuberosity with the patient prone	One ischial tuberosity judged to be higher than the other
Motion tests		
Standing flexion test[13,62]	The patient is standing and the relative heights of the PSIS are assessed. The patient is asked to flex forward as far as possible, with the examiner continuing to palpate the PSIS.	A change in the relative relationship of the PSIS is found in the fully flexed position.
Seated flexion test[17,62]	The patient is seated and the relative heights of the PSIS are judged. The patient is asked to bend forward as far as possible, with the examiner continuing to palpate the PSIS.	A change in the relative relationship of the PSIS is found in the fully flexed position.
Long-sitting test[4,13]	The patient is supine with hips and knees extended. The examiner grasps around each ankle with the thumbs below the medial malleoli. A visual estimation of leg length is made. The patient is assisted to a long-sitting position, and the examiner reexamines the relative leg lengths.	A change in the relative position of medial malleoli occurs.
Prone knee bend test[12, 17]	The patient is prone. The relative leg lengths are assessed by looking at the heels. The examiner passively flexes the patient's knees to approximately 90°. The relative leg lengths are assessed a gain in this position	A change in relative lengths occurs between the two positions.
Gillet test[50]	The patient is standing. The examiner places one thumb under the PSIS on the side being tested, with the other thumb over the S2 spinous process. The patient is instructed to stand on one leg and flex the other hip and knee, bringing the leg toward the chest.	The PSIS fails to move posterior and inferior with respect to S2.
Patrick test—range of motion[9,50]	The patient is placed in the test position by flexing, abducting, and externally rotating the hip of the tested leg, placing the lateral malleolus on the knee of the opposite leg. Overpressure is applied to the medial aspect of the knee. The amount of motion available in the tested extremity is compared with the opposite side.	A difference in the range of motion exists between the two sides.
Provocation tests		
Gaenslen test[26,45,50]	The patient is supine with both legs extended. The leg being tested is passively brought into full hip and knee flexion, while the opposite hip is maintained in an extended position. Overpressure is applied to the flexed extremity.	Pain is reproduced in either SI joint region with performance of the test.
Posterior shear test[9,26,50]	The patient is supine. The hip is flexed to 90° and adducted. The examiner applies an axial force through the femur at different angles of hip adduction–abduction.	Buttock pain is produced.
Compression/distraction test[26,50]	The patient is supine. Pressure is applied first in a posterior and lateral direction (compression) on the ASIS simultaneously. Pressure is then applied in an anterior and medial direction on the ASIS (distraction).	Pain is reproduced in the SI joint region with either maneuver.
Patrick test—Buttock pain[26]	The patient's hip is flexed, abducted, and externally rotated by placing the lateral malleolus on the knee of the opposite leg. Overpressure is applied to the medial aspect of the knee while the pelvis is stabilized.	Buttock or low back pain is produced.
Patrick test—Groin pain[26]	Same as above.	Groin pain is produced.
Resisted hip abduction[9]	The patient is supine with the hip in about 30° of abduction. The examiner pushes the leg medially to cause an isometric contraction of the hip abductors.	Buttock pain is produced.
Sacral sulcus test[21]	The patient is prone. The examiner palpates with firm pressure in the region directly medial to the PSIS.	Pain is reproduced in the SI region.
Sacral thrust test[26]	The patient is prone. The examiner delivers an anteriorly directed thrust directly over the sacrum.	Pain is reproduced in the SI region.

ASIS = anterior superior iliac spine; PSIS = posterior superior iliac spine; SI = sacroiliac.