Preliminary Development of a Clinical Prediction Rule for Determining Which Patients With Low Back Pain Will Respond to a Stabilization Exercise Program

Gregory E. Hicks, PhD, PT, Julie M. Fritz, PhD, PT, ATC, Anthony Delitto, PhD, PT, Stuart M. McGill, PhD


Objective: To develop a clinical prediction rule to predict treatment response to a stabilization exercise program for patients with low back pain (LBP).

Design: A prospective, cohort study of patients with nonradicular LBP referred to physical therapy (PT).

Setting: Outpatient PT clinics.

Participants: Fifty-four patients with nonradicular LBP.

Intervention: A standardized stabilization exercise program.

Main Outcome Measure: Treatment response (success or failure) was categorized based on changes in the Oswestry Disability Questionnaire scores after 8 weeks.

Results: Eighteen subjects were categorized as treatment successes, 15 as treatment failures, and 21 as somewhat improved. After using regression analyses to determine the association between standardized examination variables and treatment response status, preliminary clinical prediction rules were developed for predicting success (positive likelihood ratio [LR], 4.0) and failure (negative LR, .18). The most important variables were age, straight-leg raise, prone instability test, aberrant motions, lumbar hypermobility, and fear-avoidance beliefs.

Conclusions: It appears that the response to a stabilization exercise program in patients with LBP can be predicted from variables collected from the clinical examination. The prediction rules could be used to determine whether patients with LBP are likely to benefit from stabilization exercises.

Key Words: Classification; Exercise; Low back pain; Rehabilitation.

© 2005 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

A PATHOLOGY-BASED DIAGNOSTIC approach to low back pain (LBP) sets forth the idea that LBP is defined by a deviation of the lumbar pelvic complex from its normative physiologic and anatomic state.1 On application of the pathology-based model to LBP, one must first identify a structural fault and then treatment can be directed at correction of the fault. As a result of this treatment and subsequent correction, the signs and symptoms should dissipate.2 A pathology-based approach to diagnosis for LBP has proven difficult because of the inability to identify a structural pathology in the vast majority of patients with LBP.3

Despite the difficulties in identifying structural pathology for most patients, many authors continue to suggest that patients with LBP are not a homogenous group, but rather should be classified into subgroups that share similar clinical characteristics (eg, age, symptom duration, distribution).4-6 This type of classification system could guide diagnosis and treatment and improve overall decision making in the management of LBP patients.7 It may also improve research by denoting homogenous subgroups for treatment outcome studies.8

Patients with lumbar segmental instability (LSI) have been proposed as a unique subgroup of patients with LBP.7-9,11 LSI has been defined as a condition in which there is a loss of stiffness between spinal motion segments, such that normally tolerated external loads result in pain or deformity or place neurologic structures at risk.10 The diagnosis of LSI has traditionally relied on the use of lateral flexion and extension radiographs.12-14 Thresholds for defining LSI have been developed from studies that have examined the magnitude of rotational and translational movement normally available at each spinal motion segment during flexion and extension of the lumbar spine.14,15 However, the use of these criteria for defining LSI has proven unsatisfactory because of high false-positive rates.16,17 For example, Hayes et al17 found that 42% of asymptomatic subjects had at least 1 segment exceeding the instability thresholds. The inadequacies associated with the use of imaging to diagnosis LSI have led to a greater emphasis on clinical characteristics that may be indicative of the condition. Historical variables, presence of an “instability catch,” or excessive intervertebral mobility have been suggested as diagnostic.18-20 However, the validity of these findings is mostly unknown. Because of the current lack of a true reference standard for confirmation of LSI, it is not possible to accurately diagnose LSI clinically; therefore, another approach should be considered.

If LSI could be accurately diagnosed, the conservative treatment of choice would be a lumbar stabilization program to address the tissue damage and resultant loss of spinal stiffness caused by mechanical overload to the spinal stabilizing system. The goals of stabilization exercise are to train muscular motor patterns to increase spinal stability, restrain aberrant micromotion, and reduce associated pain. Although several studies that observed muscle onset and electromyographic patterns have suggested that certain muscles are important stabilizers,21 measurement of stability was not performed. In contrast, those
studies that have quantified stability\textsuperscript{22,23} unanimously agree that all muscles play a role in ensuring spine stability and that the motor patterns of cocontraction between the full complement of muscles are of utmost importance to ensure stability and minimize pain. Although the exercises used in this study are often described in the literature as targeting a specific muscle group, electromyography studies show that the exercises will challenge groups of muscles in a way that enhances overall spinal stability.

Recent research\textsuperscript{24-26} has shown that exercise programs can be designed to challenge the stabilizing muscles of the spine in a way that appears to be effective for at least certain subgroups of patients with LBP. The development of a valid method for identifying this subgroup of patients would help clinicians in determining which patients with LBP are most likely to benefit from this exercise program. The purpose of this study was to explore the validity of various demographic, historical, and clinical examination variables for predicting outcome after participation in a program of spinal stabilization exercises. The reference standard used to determine success was change in disability.

METHODS

We conducted a prospective cohort study involving patients referred to physical therapy (PT) with LBP. Subjects were recruited from new patients referred for treatment of LBP to 1 of 3 outpatient PT clinics located in the Pittsburgh, PA, area or the outpatient PT clinic at Keesler Air Force Base Medical Center (Biloxi, MS). All new patients with complaints of LBP, with or without leg pain, over the age of 18 were eligible for participation. Exclusion criteria were as follows: previous spinal fusion surgery; LBP attributable to current pregnancy; acute fracture, tumor, or infection; and presence of 2 or more of the following signs of nerve root, compressive, diminished lower-extremity strength, sensation, or reflexes. All subjects reviewed and signed a consent form approved by the University of Pittsburgh Health Sciences Institutional Review Board.

Therapists

Four licensed physical therapists participated in the examination and treatment of subjects in this study. All therapists received written instructions and specific training regarding all study procedures. Each therapist had to pass a written test to show an acceptable level of knowledge regarding subject recruitment, clinical examination, and exercise progression.

Examination Procedures

At baseline, subjects were evaluated by the treating physical therapist. The evaluation consisted of the collection of basic demographic information, subject self-report measures, questions of history, and physical examination measures.

Each subject completed a numeric pain rating to assess severity of pain by using an 11-point (0–10) scale.\textsuperscript{27} The Fear Avoidance Beliefs Questionnaire (FABQ)\textsuperscript{28} was used to measure the level of fear-avoidance beliefs. The FABQ has 2 subscales that measure fear-avoidance beliefs about work (7-item scale) and physical activity (4-item scale). The Modified Oswestry Low Back Pain Disability Questionnaire (ODQ)\textsuperscript{29} is a disease-specific measure of disability among patients with LBP. This questionnaire has been used extensively in randomized trials and has shown excellent reliability and good construct validity in comparison with other pain and disability measures.\textsuperscript{29-31} The ODQ was assessed at baseline and after 8 weeks of treatment and served as the reference standard for determining the success of the treatment program.

All subjects underwent a standardized history and physical examination. History taking included questions proposed to indicate the presence of LSI\textsuperscript{7,18,19} including mode of onset, duration of symptoms, number of previous episodes of LBP, and response to various interventions for previous episodes. Subjects were also asked about the distribution of symptoms for the current episode and were asked to rank sitting versus standing and walking and morning versus evening with respect to their symptoms.

Operational definitions for components of the physical examination are included in appendix 1. The physical examination included lumbar spine range of motion (ROM) measured with a single inclinometer. Waddell et al\textsuperscript{32} found these measurements to be highly reliable (intraclass correlation coefficient [ICC] range, .86–.95). Any aberrant motions believed to be associated with LSI occurring during the performance of lumbar ROM were noted, including an instability catch,\textsuperscript{33,34} painful arc of motion,\textsuperscript{35} “thigh climbing” (Gower’s sign), or a reversal of lumbopelvic rhythm. Subjects were categorized as either having or not having aberrant motions present, a determination found to be reliable in a previous study.\textsuperscript{36} Posterior-anterior lumbar mobility testing\textsuperscript{77} was performed at each spinal level. Although the reliability coefficients for these judgments generally have been shown to be only fair,\textsuperscript{38,39} percentage agreement among examiners is high\textsuperscript{30} and mobility judgments have been shown to be useful for determining which patients with LBP may respond to a manipulation intervention.\textsuperscript{40} Generalized ligamentous laxity was assessed on a 9-point scale described by Beighton and Horan.\textsuperscript{41} Higher numbers indicate greater laxity. We previously found the reliability of this scale to be excellent (ICC = .79).\textsuperscript{36} Two special tests proposed to indicate LSI, the prone instability\textsuperscript{7} and posterior shear\textsuperscript{7} tests, were performed. We found the prone instability test (κ = .87) to be more reliable than the posterior shear test (κ = .35),\textsuperscript{36} but the validity of these tests has not been studied previously.

Muscle endurance of the spinal extensors and lateral flexors was assessed by using the modified Biering-Sorensen extensor endurance test and the side support test, respectively, which have been found to be highly reliable (ICC range, .93–.99).\textsuperscript{43} Two strength tests, the active sit-up and active straight-leg raising (SLR) tests, were also used. These tests have been found to be both reliable (κ range, .48–.77) and valid in a previous study.\textsuperscript{32}
null
any potentially useful predictor variables. Any continuous variable found to have a significant univariate association with outcome was further explored by constructing a receiver operator characteristic (ROC) curve. The ROC curve was used to determine a cutoff point defining a positive test. The point nearest the upper left hand corner of an ROC curve represents the cutoff score with the best diagnostic accuracy. This point was therefore selected as the cutoff score defining a positive test. Accuracy statistics (sensitivity, specificity, positive and negative likelihood ratio [LR] values) with the corresponding 95% confidence interval (CI) were calculated based on this cutoff score. For categoric variables with significant univariate relationships, accuracy statistics with 95% CI were calculated.

For the prediction of success, the “target condition” used to calculate accuracy statistics was success with stabilization. The focus for the prediction of success was on maximizing the positive LR because this value represents the change in odds favoring success given a positive diagnostic test result. For the prediction of failure, the “target condition” was not failing with stabilization (ie, experiencing either improvement or success). The focus for the prediction of failure was therefore on minimizing the negative LR value. A test with a small negative LR would serve as a useful screening tool because this indicates that a negative result would substantially decrease the odds of a favorable result (improvement or success) with stabilization treatment.

All variables with a significant univariate relation to treatment outcome were then used to develop multivariate clinical prediction rules. If more than 5 variables were found to have a significant relation to an outcome, then the variables were entered into a forward stepwise logistic regression equation to reduce the number of predictors. A significance of .15 was required to enter, and a significance of .20 was used to remove a variable from the equation. More liberal significance levels were used because this is the first approximation of a clinical prediction rule. Once the number of predictor variables was determined, the clinical prediction rule was developed by examining the accuracy statistics for various combinations of the retained variables.

### RESULTS

A total of 57 subjects gave informed consent for participation between September 2001 and May 2002. Three subjects initially agreed to participate but withdrew after several days because they did not want to commit to the time required. Because no data were obtained from these subjects, they were not included in the analysis. One subject began the stabilization program but dropped out after 2 weeks for an unknown reason. This subject was included in the analysis and considered a stabilization failure. A total of 54 subjects were therefore included in the analysis.

Table 2 provides descriptive information. The mean age ± standard deviation (SD) was 42.4±12.7 years; 57.4% were women and 85.2% were white. Thirty-eight (70.4%) subjects reported a history of LBP that required activity modification. The mean ODQ score at baseline was 29.7±13.7 (table 3). Eighteen (33.3%) subjects experienced a 50% or greater improvement in ODQ score over 8 weeks and were categorized as stabilization successes. Of the remaining 36 subjects, 15 (27.8%) experienced less than a 6-point improvement on the ODQ and were categorized as stabilization failures. The remaining 21 (38.9%) subjects were categorized as improved with stabilization treatment. Initial and follow-up ODQ scores for the 3 outcome groups are shown in figure 1. The mean number of PT visits attended during the 8-week treatment period was 9.7±2.2, and there was no association between the number of visits and change in ODQ score over the 8-week period (P>.05).

Tables 3 and 4 show the results of the history and demographic and physical examination findings. Four variables were significantly related to success and were retained as potential predictors: age (cutoff value defining a positive test <40y), average SLR (cutoff value defining a positive test >91°), the presence of aberrant movement during lumbar ROM, and a positive prone instability test (table 5). Nine variables were retained as potential predictors of failure (table 6), including the FABQ physical activity subscale (cutoff value defining a positive test >8 points), pain rating (cutoff value defining a positive test >2), lumbar flexion ROM (cutoff value defining a positive test >33°), discrepancy in SLR (cutoff value defining a positive test <10°), 3 or more previous episodes of LBP, increasing frequency of episodes of LBP, aberrant movement during lumbar ROM, the presence of hypermobility during lumbar spring testing, and a positive prone instability test.

Among the individual predictor variables for success, age less than 40 years had the greatest positive LR (3.7; 95% CI, 1.9 to 7.2), average SLR (cutoff value defining a positive test >91°), and the presence of aberrant movement during lumbar ROM (cutoff value defining a positive test >33°) were the most powerful predictors of success. Failure was significantly correlated with increasing age (cutoff value defining a negative test ≥40y), a history of LBP requiring activity modification (cutoff value defining a negative test ≥1), a positive diagnostic test for increased lumbar motion (eg, spring testing, lumbar motion testing) (cutoff value defining a negative test ≥1), and the presence of hypermobility during spring testing (cutoff value defining a negative test ≥1).
Abbreviation: Signif, significance. NOTE. Values are accuracy statistics (95% CIs).

If 2 or more of the 4 variables were present, the odds of some improvement with stabilization treatment decreased by a factor of .18. If 2 or more of the 4 variables were present, the odds of some improvement with stabilization increased by 6.3.

**DISCUSSION**

Exercises designed to improve spinal stabilization have gained popularity in the conservative treatment of patients with LBP; however, the evidence for the effectiveness of this approach is sparse and equivocal. A recent systematic review of exercise therapy for LBP concluded that exercise therapy

### Table 4: Physical Examination Variables Assessed at Baseline

<table>
<thead>
<tr>
<th>Physical Examination Variables</th>
<th>All Subjects (N=54)</th>
<th>Succeeded With Treatment (n=18)</th>
<th>Improved With Treatment (n=21)</th>
<th>Failed With Treatment (n=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total flexion (deg)</td>
<td>88.5±22.9</td>
<td>94.2±25.8</td>
<td>84.2±25.9</td>
<td>87.5±12.3</td>
</tr>
<tr>
<td>Pelvic flexion (deg)</td>
<td>47.3±18.4</td>
<td>49.9±23.1</td>
<td>42.2±17.9</td>
<td>51.2±10.6</td>
</tr>
<tr>
<td>Lumbar flexion (deg)</td>
<td>41.5±13.6</td>
<td>44.3±12.4</td>
<td>42.3±15.7</td>
<td>36.9±11.1</td>
</tr>
<tr>
<td>Percentage of total flexion from lumbar spine (deg)</td>
<td>47.8±14.0</td>
<td>48.2±14.3</td>
<td>50.8±15.2</td>
<td>42.0±10.6</td>
</tr>
<tr>
<td>Total extension</td>
<td>25.9±9.9</td>
<td>26.3±7.5</td>
<td>23.8±10.4</td>
<td>28.3±11.6</td>
</tr>
<tr>
<td>Left side–bending (deg)</td>
<td>27.6±8.4</td>
<td>28.3±6.7</td>
<td>28.0±9.3</td>
<td>26.2±9.2</td>
</tr>
<tr>
<td>Right side–bending (deg)</td>
<td>27.9±9.4</td>
<td>27.2±9.8</td>
<td>28.9±10.0</td>
<td>27.3±8.5</td>
</tr>
<tr>
<td>Average side-bending (deg)</td>
<td>28.0±8.1</td>
<td>28.5±6.7</td>
<td>28.4±9.2</td>
<td>26.7±8.3</td>
</tr>
<tr>
<td>Side-bending discrepancy* (deg)</td>
<td>4.7±5.2</td>
<td>5.4±7.3</td>
<td>3.9±3.8</td>
<td>4.8±3.8</td>
</tr>
<tr>
<td>Left SLR (deg)</td>
<td>77.8±13.7</td>
<td>82.4±11.0</td>
<td>75.4±15.6</td>
<td>74.7±13.0</td>
</tr>
<tr>
<td>Right SLR (deg)</td>
<td>77.9±14.6</td>
<td>82.6±13.9</td>
<td>76.4±14.4</td>
<td>74.2±15.2</td>
</tr>
<tr>
<td>Average SLR (deg)</td>
<td>77.7±13.7</td>
<td>82.5±12.1</td>
<td>75.9±14.7</td>
<td>74.5±13.5</td>
</tr>
<tr>
<td>SLR discrepancy* (deg)</td>
<td>5.6±4.3</td>
<td>5.2±4.3</td>
<td>4.9±4.4</td>
<td>7.2±3.9</td>
</tr>
<tr>
<td>Aberrant motion during lumbar ROM (% yes)</td>
<td>59.3</td>
<td>77.8</td>
<td>66.7</td>
<td>26.7</td>
</tr>
<tr>
<td>Beighton scale score</td>
<td>0.9±1.7</td>
<td>1.1±1.9</td>
<td>0.9±1.4</td>
<td>0.6±2.1</td>
</tr>
<tr>
<td>Posterior shear test (% positive)</td>
<td>42.6</td>
<td>33.3</td>
<td>57.1</td>
<td>33.3</td>
</tr>
<tr>
<td>Prone instability test (% positive)</td>
<td>51.9</td>
<td>72.2</td>
<td>61.9</td>
<td>13.3</td>
</tr>
<tr>
<td>Active sit-up test (% positive)</td>
<td>13.0</td>
<td>11.1</td>
<td>14.3</td>
<td>11.1</td>
</tr>
<tr>
<td>Active SLR test (% positive)</td>
<td>20.3</td>
<td>16.7</td>
<td>26.8</td>
<td>13.3</td>
</tr>
<tr>
<td>Any hypermobility present</td>
<td>20.3</td>
<td>22.2</td>
<td>33.3</td>
<td>0</td>
</tr>
<tr>
<td>during spring testing (% yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Side support left (s)</td>
<td>27.7±24.9</td>
<td>32.1±23.2</td>
<td>27.1±26.8</td>
<td>23.3±25.1</td>
</tr>
<tr>
<td>Side support right (s)</td>
<td>26.3±24.6</td>
<td>29.7±21.9</td>
<td>25.9±27.0</td>
<td>22.7±25.4</td>
</tr>
<tr>
<td>Average side support (s)</td>
<td>27.0±24.5</td>
<td>30.9±22.2</td>
<td>26.5±26.5</td>
<td>23.0±25.0</td>
</tr>
<tr>
<td>Side support discrepancy* (s)</td>
<td>6.2±5.5</td>
<td>5.6±6.1</td>
<td>7.1±5.8</td>
<td>5.6±4.3</td>
</tr>
<tr>
<td>Extensor endurance (s)</td>
<td>49.0±45.8</td>
<td>51.7±38.5</td>
<td>48.3±43.7</td>
<td>46.7±58.4</td>
</tr>
<tr>
<td>Ratio average side support/</td>
<td>94.1±109.7</td>
<td>74.7±47.6</td>
<td>102.2±133.4</td>
<td>108.8±135.3</td>
</tr>
<tr>
<td>extensor endurance (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE. Values are mean ± SD unless otherwise indicated.

*Discrepancy values calculated as the absolute value of the right side minus the left side.

### Table 5: Significant Univariate Predictors of Stabilization Success

<table>
<thead>
<tr>
<th>Variable Associated With Success</th>
<th>Signif</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive LR</th>
<th>Negative LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (&lt;40y)</td>
<td>.082</td>
<td>.61 (.39–.80)</td>
<td>.83 (.68–.92)</td>
<td>3.7 (1.6–8.3)</td>
<td>.47 (0.26–0.85)</td>
</tr>
<tr>
<td>Average SLR (&gt;91°)</td>
<td>.069</td>
<td>.28 (.13–.51)</td>
<td>.92 (.78–.97)</td>
<td>3.3 (1.90–12.4)</td>
<td>.79 (0.58–1.1)</td>
</tr>
<tr>
<td>Aberrant movement present</td>
<td>.050</td>
<td>.78 (.55–.91)</td>
<td>.50 (.35–.66)</td>
<td>1.6 (1.0–2.3)</td>
<td>.44 (0.18–1.1)</td>
</tr>
<tr>
<td>Positive prone instability test</td>
<td>.034</td>
<td>.72 (.49–.88)</td>
<td>.58 (.42–.73)</td>
<td>1.7 (1.1–2.8)</td>
<td>.48 (0.22–1.1)</td>
</tr>
</tbody>
</table>

NOTE. Values are accuracy statistics (95% CIs).

Abbreviation: Signif, significance.
overall was not effective for patients with acute LBP but may be helpful for those with chronic LBP. The review also noted that the evidence could not be used to examine the effectiveness of specific types of exercise (eg, flexion, extension, strengthening, stabilization) because of heterogeneity in study populations and interventions. The few studies that have examined specific stabilization exercise programs such as that used in this study in more homogenous populations have shown promising results. O’Sullivan et al randomized patients with chronic, recurrent LBP who had a radiologic diagnosis of spondylolysis or spondylolisthesis to receive either stabilization exercises or usual care directed by a general practitioner and found statistically significant reductions in pain and disability at a 30-month follow-up. Hides et al studied patients with acute, first-time episode of unilateral LBP, comparing stabilization exercises particularly targeting the multifidus muscle with usual medical management. No significant differences in disability or pain were found after 4 weeks, but the stabilization group experienced significantly fewer recurrences at 2- to 3-year follow-up. These studies indicate that a specific stabilization exercise approach may be effective for certain subgroups of patients. This study is an initial step in identifying the subgroup of patients for whom this approach may be particularly effective.

The preliminary prediction rule for success with stabilization treatment contains 4 variables: positive prone instability test, aberrant movements present, average SLR greater than 91°, and age greater than 40 years old. These variables seem reasonable based on the available theoretical literature regarding LSI and stabilization. The prone instability test works on the premise that if pain is present on passive provocation testing of the vertebral levels but disappears when the spinal extensors are active, then the muscle activity may be able to stabilize the segment and reduce pain, and the individual may benefit from stabilization exercises. Several authors have suggested that the observation of aberrant movement patterns during active trunk motion is valuable in the diagnosis of LSI. The presence of aberrant movements may represent an inability to adequately control lumbar motion and indicate a need for stabilization exercises. Decreased SLR ROM is related to the presence of radiculopathy and a generally worse prognosis.

In our study, greater SLR mobility was associated with stabilization success. The final predictor of success was younger age (<40y). Increased age is associated with decreased lumbar extensor muscle mass. Despite the evidence suggesting detrimental effects to muscle with aging, there is encouraging evidence that older men and women can still increase muscle strength even up to age 90 with appropriate training. Adults over the age of 39 may require more than 8 weeks of training to gain the same benefits as their younger peers with LBP. Additionally, because of the decreased flexibility and ROM associated with aging, it is also possible that older adults may require a combination of mobility-related interventions followed by stabilization exercise to achieve the same gains as their younger counterparts.

We also sought to develop a preliminary clinical prediction rule for patients likely to fail with stabilization treatment. The prediction rule contained 4 variables: negative prone instability test, aberrant movements absent, FABQ physical activity subscale score greater than 9, and no hypermobility with lumbar spring testing. A negative prone instability test may indicate that muscular stabilization is not needed, and the absence of an aberrant movement pattern during active trunk motion may indicate an ability to control movement in a pain-free fashion. The FABQ measures fear-avoidance beliefs in patients with LBP. In this study, lower levels of fear-avoidance beliefs about physical activity were associated with stabilization failure. This is counter to other findings in the literature that suggest higher levels of fear-avoidance beliefs are associated with poorer outcomes. Our findings may suggest that this stabilization program preferentially helps patients with moderate to high levels of fear-avoidance beliefs to overcome their...
The stabilization program used in our study was standardized and each physical therapist involved followed the same protocol, but we acknowledge that clinician-based modifications to this program may produce superior results in certain situations. Although numerous types of stabilization exercises have been proposed, we believe the exercises used in this program are fairly representative and supported by the literature. We also attempted to control for any other types of biases that could temper the results of our study. First, the reference standard was applied to all subjects equally, and the examiner who judged the reference standard (patient response to stabilization) was blinded to the results of all diagnostic tests as well as to the clinical presentation of the subjects to avoid review bias. Additionally, the study design made it impossible for the clinician who judged the diagnostic test to be influenced by the result of the reference standard because it was not known until the 8-week time point. Finally, the inclusion of consecutive subjects with LBP from clinical practice should decrease the impact of spectrum bias and make the results generalizable to patients with LBP seeking outpatient PT treatment.

**CONCLUSIONS**

These results represent a preliminary step in the development of a clinical prediction rule for the use of stabilization exercises. We were able to identify several variables that appear to be important in determining which patients are likely or unlikely to benefit from a stabilization treatment approach. From the history and patient self-report, the variables of age, fear of physical activity, pain intensity, and frequency of previous episodes had some relation to treatment outcome. From the physical examination, the key variables appear to be SLR motion, lumbar mobility testing, aberrant movements during lumbar ROM, and the prone instability test. From these variables, we developed preliminary multivariate prediction rules. The results of this study need to be replicated in a separate sample for confirmation and ultimately examined in a randomized controlled trial before they can be recommended for widespread use.

**APPENDIX 1: OPERATIONAL DEFINITIONS FOR PHYSICAL MEASURES**

<table>
<thead>
<tr>
<th>Physical Measures</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Range of motion</strong></td>
<td></td>
</tr>
<tr>
<td>Flexion ROM$^{32}$</td>
<td>The patient stands and the inclinometer is held at T12-L1. The patient is asked to reach down as far as possible toward the toes while keeping the knees straight.</td>
</tr>
<tr>
<td>Extension ROM$^{32}$</td>
<td>The patient stands and the inclinometer is held at T12-L1. The patient is asked to arch backward as far as possible.</td>
</tr>
<tr>
<td>Right and left side-bending</td>
<td>The patient stands with the inclinometer aligned vertically in line with the spinous processes of T9 and T12. The patient is asked to lean over to 1 side as far as possible with the fingertips reaching down the side of the thigh.</td>
</tr>
</tbody>
</table>
### Physical Measures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right and left SLR ROM&lt;sup&gt;32&lt;/sup&gt;</td>
<td>The patient is supine. The inclinometer is positioned on the tibial crest just below the tibial tubercle. The leg is raised passively by the examiner, whose other hand maintains the knee in extension. The leg is raised slowly to the maximum tolerated straight leg raise (not the onset of pain).</td>
</tr>
<tr>
<td>Muscle performance tests</td>
<td></td>
</tr>
<tr>
<td>Side support test&lt;sup&gt;44&lt;/sup&gt;</td>
<td>The patient is side lying with legs extended and the top foot in front of the lower foot. While resting on the lower elbow for support, the patient lifts the hips off the table with only the elbow and feet remaining in contact with the table. The patient is instructed to hold this position as long as possible. The test is done for both sides, and the performance time is recorded in seconds.</td>
</tr>
<tr>
<td>Extensor endurance test&lt;sup&gt;44&lt;/sup&gt;</td>
<td>The patient is asked to lie prone while holding the sternum off the floor for as long as possible. A small pillow is placed under the lower abdomen to decrease the lumbar lordosis. The patient also needs to maintain maximum flexion of cervical spine and pelvic stabilization through gluteal contraction. The patient is asked to hold this position as long as possible not to exceed 5 minutes. The performance time is recorded in seconds.</td>
</tr>
<tr>
<td>Active sit-up test&lt;sup&gt;32&lt;/sup&gt;</td>
<td>The patient is supine and is asked to flex the knees to 90° and place the soles of the feet flat on the surface. The examiner holds both feet down with 1 hand. The patient is instructed to reach up with the fingertips of both hands to touch (not hold) both knees and hold the position for 5 seconds. If the patient cannot maintain this position for 5 seconds, the test is positive.</td>
</tr>
<tr>
<td>Active bilateral SLR test&lt;sup&gt;32&lt;/sup&gt;</td>
<td>The patient is supine and is asked to lift both legs together 6 inches (15.24 cm) off the floor, away from the examining table. Both heels and calves should be clearly visible from the examining surface. If the patient cannot maintain this position for 5 seconds, the test is positive.</td>
</tr>
<tr>
<td>Special tests</td>
<td></td>
</tr>
<tr>
<td>Posterior shear test&lt;sup&gt;7&lt;/sup&gt;</td>
<td>The patient is standing with arms across the lower abdomen. The examiner stands at 1 side of the patient and places 1 arm around the patient's abdomen, over the patient's crossed hands. The heel of the opposite hand is placed on the patient's pelvis for stabilization. The examiner produces a posterior force through the patient's abdomen and an anteriorly directed stabilizing force with the opposite hand. The test is repeated at all lumbar levels. A positive test is determined by the provocation of symptoms.</td>
</tr>
<tr>
<td>Prone instability test&lt;sup&gt;43&lt;/sup&gt;</td>
<td>The patient lies prone with the body on the examining table and legs over the edge and feet resting on the floor. While the patient rests in this position, the examiner applies posterior to anterior pressure to the lumbar spine. Any provocation of pain is reported. Then the patient lifts the legs off the floor (the patient may hold table to maintain position) and posterior compression is applied again to the lumbar spine. If pain is present in the resting position but subsides in the second position, the test is positive.</td>
</tr>
<tr>
<td>Lumbar segmental testing for mobility&lt;sup&gt;f&lt;/sup&gt;</td>
<td>The patient is prone. The L1 spinous process is contacted with the examiner’s thenar eminence, and an anteriorly directed force is applied. The procedure is repeated at each lumbar level. Mobility is judged as hypermobile or not hypermobile.</td>
</tr>
</tbody>
</table>

### References

SUCCESS WITH STABILIZATION, Hicks


58. Sackett DL. A primer on the precision and accuracy of the clinical examination. JAMA 1992;267:2638-44.


