ORIGINAL ARTICLE

Lumbar Spine Segmental Mobility Assessment: An Examination of Validity for Determining Intervention Strategies in Patients With Low Back Pain

Julie M. Fritz, PhD, PT, ATC, Julie M. Whitman, DSc, PT, OCS, John D. Childs, PhD, PT, OCS, CSCS

ABSTRACT. Fritz JM, Whitman JM, Childs JD. Lumbar spine segmental mobility assessment: an examination of validity for determining intervention strategies in patients with low back pain. Arch Phys Med Rehabil 2005;86:1745-52.

Objective: To examine the predictive validity of posterior-anterior (PA) mobility testing in a group of patients with low back pain (LBP).

Design: Randomized controlled trial.

Setting: Outpatient physical therapy clinics.

Participants: Patients with LBP (N=131; mean age ± standard deviation, 33.9±10.9y; range, 19–59y), and a median symptom duration of 27 days (range, 1–5941d). Patients completed a baseline examination, including PA mobility testing, and were categorized with respect to both hypomobility and hypermobility (present or absent), and treated for 4 weeks.

Intervention: Seventy patients were randomized to an intervention involving manipulation and 61 to a stabilization exercise intervention.

Main Outcome Measures: Oswestry Disability Questionnaire (ODQ) scores were collected at baseline and after 4 weeks. Three-way repeated measures analyses of variance (ANOVAs) were performed to assess the effect of mobility categorization and intervention group on the change on the ODQ with time. Number-needed-to-treat (NNT) statistics were calculated.

Results: Ninety-three (71.0%) patients were judged to have hypomobility present and 15 (11.5%) were judged with hypermobility present. The ANOVAs resulted in significant interaction effects. Pairwise comparisons showed greater improvements among patients receiving manipulation categorized with hypomobility present versus absent (mean difference, 23.7%; 95% confidence interval [CI], 5.1%–42.4%), and among patients receiving stabilization categorized with hypermobility present versus absent (mean difference, 36.4%; 95% CI, 10.3%–69.3%). For patients with hypomobility, failure rates were 26% with manipulation and 74.4% with stabilization (NNT=2.1; 95% CI, 1.6–3.5). For patients with hypermobility, failure rates were 83.3% and 22.2% for manipulation and stabilization, respectively (NNT=1.6; 95% CI, 1.2–10.2).

Conclusions: Patients with LBP judged to have lumbar hypomobility experienced greater benefit from an intervention including manipulation; those judged to have hypermobility were more likely to benefit from a stabilization exercise program.

Key Words: Diagnosis; Low back pain; Physical examination; Rehabilitation.

PHYSICAL THERAPISTS FREQUENTLY include segmental mobility testing of the lumbar spine using posterior-anterior (PA) forces in the examination and treatment of patients with low back pain (LBP).[1-4] The application of PA forces is performed with the patient prone. The therapist places some aspect of his/her hand on the spinous process of a lumbar vertebra and produces an anteriorly directed force.[4] The 2 most common reasons for using PA forces as an examination procedure are to reproduce the patient's symptoms and determine the mobility of a segment of the lumbar spine.[4-7] Integration of the information obtained from the use of PA forces as an examination procedure into clinical decision making typically focuses on the judgment of mobility.[5,7] For example, many manual therapy approaches recommend mobilization or manipulation interventions if patients are judged to lack mobility with PA testing and no contraindications to mobilization/manipulation are present.[4,8,9] Conversely, some form of stabilization exercise may be suggested when therapists find excess mobility.[10,11]

Studies have generally failed to support the reproducibility of mobility judgments between different examiners,[6,12,13] leading some to suggest that PA mobility testing has little value as an examination procedure.[6,12,14] However, recent studies have suggested that PA mobility testing may improve decision making when combined with other examination information.[13,15-17] In a prospective cohort study of 71 subjects with nonradicular LBP, Flynn et al[17] reported that a finding of hypomobility in the lumbar spine with PA mobility testing, combined with several other historical and physical examination findings, formed a clinical prediction rule that was predictive of a successful reduction in disability with a manipulation intervention. A randomized trial by Childs et al[18] validated this prediction rule and its usefulness in predicting which patients with LBP are most likely to improve with manipulation. Hicks et al[15] studied 57 patients with nonradicular LBP, and found that a judgment of hypermobility was a factor in a multivariate clinical prediction rule that was predictive of a reduction in disability with a stabilization exercise program. Fritz et al[16] studied the diagnostic accuracy of various findings from the history and physical examination for predicting radiographic lumbar segmental instability and reported that a judgment of hypermobility was predictive of radiographic instability, and

---

From the Division of Physical Therapy, University of Utah, and Clinical Outcomes Research Scientist, Intermountain Health Care, Salt Lake City, UT (Fritz); US Army-Baylor University Postprofessional Doctoral Program in Orthopaedic and Manual Physical Therapy, and Affiliate Faculty, Regis University, Denver, CO (Whitman); and Department of Physical Therapy, Wilford Hall Medical Center, San Antonio, TX (Childs).

Presented to the Fifth Interdisciplinary World Congress on Low Back and Pelvic Pain, November 13, 2004, Melbourne, Australia.

Supported by the Foundation for Physical Therapy.

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Air Force or Department of Defense.

Reprint requests to Julie M. Fritz, PhD, PT, ATC, Div of Physical Therapy, University of Utah, 520 Wakara Way, Salt Lake City, UT 84108.

0003-9993/05/8609-9753$30.00/0
doi:10.1016/j.apmr.2005.03.028

that the predictive accuracy of PA mobility judgments were enhanced when combined with other examination findings.

These results provide preliminary evidence for the validity of PA mobility testing as a useful examination procedure for patients with LBP; however, the direct relation between mobility judgments and treatment outcomes have not been previously reported. The purpose of this study was to further examine the predictive validity of PA mobility in a group of patients with LBP who were participants in a randomized controlled trial (RCT) comparing a manipulation plus exercise intervention with a stabilization exercise intervention. We hypothesized that patients judged to have hypomobility would respond better to an intervention that included manipulation, whereas patients judged to have hypermobility would preferentially respond to a stabilization exercise program.

METHODS

Participants

This study involved subjects who were participants in an RCT.[18] The purpose of the RCT was to validate a clinical prediction rule developed to identify patients with LBP likely to benefit from spinal manipulation. Subjects were randomized to receive manipulation plus a stabilization exercise intervention, or a stabilization exercise intervention alone. The purpose of this secondary analysis was to examine the relation between a preintervention judgment of PA mobility and the outcomes of the 2 interventions. Subjects were required to be between the ages of 18 and 60 years with a primary complaint of LBP and Oswestry Disability Questionnaire (ODQ) scores of at least 30%. Exclusion criteria were current pregnancy, any red flags for a serious spinal condition[19] (eg, tumor, compression fracture, infection), or prior surgery to the lumbar spine or buttocks. Subjects could have symptoms extending into the lower extremity, but were excluded if any signs consistent with nerve-root compression were present (ie, positive straight-leg raise <45°, or diminished reflexes, sensation, or lower-extremity strength).

A total of 131 subjects provided informed consent approved by the local institutional review board at the site; 70 subjects were randomized to receive the manipulation plus stabilization exercise intervention and 61 subjects to the stabilization exercise intervention. The mean age ± standard deviation (SD) of the subjects was 33.9±10.9 years (range, 19–59y), median symptom duration was 27 days (range, 1–5941d), and 55 (42%) were women. Further descriptive information is provided in table 1. There were no baseline differences between intervention groups for any variable listed in table 1 ($P>.05$).

Table 1: Subject Characteristics at Initial Examination

Characteristics	Values
Age (y)	33.9±10.9 (range, 19–59)
Sex (% female)	42
Prior history of LBP (% yes)	67.9
Duration of current symptoms (median no. of days)	27 (range, 1–5941)
Symptoms distal to the knee (% yes)	23.7
FABQ physical activity scale	17.0±4.3 (range, 3–24)
FABQ work scale	17.0±10.3 (range, 0–42)
ODQ score	41.2±10.4 (range, 30–76)
NRS for pain	5.8±1.6 (range, 0–9)

NOTE. Values are mean ± SD unless otherwise indicated. Abbreviations: FABQ, Fear-Avoidance Beliefs Questionnaire; NRS, numeric rating scale.

Baseline and Outcome Measurements

Before randomization, all subjects completed a set of self-report measures and then underwent a history and physical examination performed by a physical therapist who was masked to the subject's intervention group assignment. We used an 11-point numeric rating scale (NRS) for pain to assess current pain intensity. The NRS used a scale ranging from 0 (no pain) to 10 (worst imaginable pain) and asked subjects to rate their current pain intensity, and the best and worst levels of pain during the last 24 hours. The average of the 3 ratings was used.[20] We used the Fear-Avoidance Beliefs Questionnaire to quantify the patient's fear of pain and beliefs about avoiding activity.[21] The primary outcome for this analysis was the modified ODQ, a region-specific measure of disability for patients with LBP.[22] The modified ODQ contains 10 items (each scored from 0 to 5), with the final score expressed as a percentage. This modified ODQ has been found to have high levels of test-retest reliability (intraclass correlation coefficient [ICC]=.90, for 23 subjects with LBP whose condition remained stable over 4wk), construct validity (Pearson correlations with global patient ratings and other region-specific disability measures >.80), and responsiveness (effect size=1.8 in patients receiving physical therapy interventions for LBP).[22]

Although a variety of physical examination procedures were performed, the focus of this secondary analysis was the assessment of PA mobility. With the patient prone, the examiner contacted the spinous process of the segment to be tested with his or her hypothenar eminence. With the elbow and wrist extended, the examiner was instructed to apply a gentle but firm, anteriorly directed pressure on the spinous process. For each spinal segment, mobility was graded as normal, hypomobile, or hypermobile. A separate judgment was made for each spinal segment of the lumbar spine (L5-1). Examiners were instructed to base their judgment of mobility on the anticipation of what normal mobility would feel like at the tested spinal level for the particular patient being examined, and in comparison with the mobility detected in the spinal segment above and below. Each subject was then categorized as to the presence of hypomobility and hypermobility. If hypomobility was judged to be present at any level of a subject's lumbar spine, the subject was categorized as having hypomobility. If hypomobility was judged not to be present at any level (ie, all levels either normal or hypermobile), the subject was categorized as not having hypomobility. A second categorization was made for each subject with respect to hypermobility. Subjects were categorized as having hypermobility present (ie, a judgment of hypermobility was made for at least 1 lumbar level) or not having hypermobility present (ie, all segments judged to be either normal or hypomobile). We have examined the reliability of these categorizations in previous studies of patients with LBP. Kappa values for judging the presence of hypomobility have ranged between .13 and .30, with percentage agreements between 59% and 78%.[16,23,24] For judgments of hypermobility, κ values have ranged between .18 and .48, and percentage agreements between 76% and 77%.[16,23]

Intervention

After the initial examination, subjects were randomized to either the manipulation plus exercise group or the stabilization exercise intervention group. Both groups attended a total of 5 therapy sessions, with subjects in both groups completing a home exercise program on the days they did not attend a

Table 2: Comparison of Intervention Groups for Mobility Judgments at Initial Examination and ODQ Scores

Variable	Manipulation Plus Exercise Group (n=70)	Stabilization Exercise Group (n=61)
No. (%) with hypomobility present	50 (71.4)	43 (70.5)
No. (%) with hypermobility present	6 (8.6)	9 (14.8)
Initial ODQ score	41.4±10.1	40.9±10.8
Four-week ODQ score*	17.7±16.6	26.0±17.6
Percentage change in ODQ scores*	57.3±36.6	34.0±42.9
No. (%) of subjects with successful outcome*	44 (62.9)	22 (36.1)

NOTE. Values are mean ± SD unless otherwise indicated.
*Difference between intervention groups ($P<.05$).

therapy session. All subjects received an instruction booklet outlining the proper performance of each exercise and were advised to maintain their usual activity within the limits of pain. Two therapy sessions took place in the first week after randomization, with weekly sessions occurring over the next 3 weeks. The NRS and ODQ were reassessed after 1 week (after 2 therapy sessions) and again after 4 weeks (completion of therapy) (table 2).

Subjects in the manipulation plus stabilization exercise intervention group received a spinal manipulation procedure and instruction in a range of motion (ROM) exercise at each of the first 2 sessions. The manipulation technique has been described in the literature[17,25] and was performed with the subject supine. The subject's trunk was side-bent toward the side of the pelvis to be manipulated, and then rotated in the opposite direction. The therapist delivered a quick thrust to the pelvis in a posterior and inferior direction. Subjects were then instructed in a ROM exercise performed by tilting the pelvis anterior and posterior in a supine position. Beginning at the third therapy session, subjects in the manipulation intervention group began the same exercise program outlined below for the stabilization exercise intervention group.

Subjects in the stabilization exercise intervention group received a stabilization exercise intervention with no manipulation performed during any of the 5 treatment sessions. Stabilization exercises included abdominal "drawing in" as described by Richardson et al,[26,27] performed with the subject supine, standing, and while performing bilateral hip extension from a hook-lying position (ie, "bridging"). Subjects also performed extension of the upper and lower extremities while in quadruped and the horizontal isometric side bridge described by McGill.[28,29] Subjects also performed a ROM exercise by rocking forward and backward while in a quadruped position. The stabilization exercise program also included low-stress aerobic activity such as treadmill walking or stationary cycling. The aerobic exercise component was performed at a subject-selected pace with an initial goal of 10 minutes daily.

Data Analysis

Each subject's outcome was determined based on change in the ODQ after the 4-week intervention period. Data were analyzed using intention-to-treat principles with the last score carried forward for subjects with missing 4-week ODQ scores. To assess the effect of intervention group and mobility judgment on the intervention outcome, a 3-way repeated measures analysis of variance (ANOVA) was performed. The dependent variable was the ODQ. The within-subjects factor was time with 2 levels (initial, 4wk), and the between-subject factors were hypomobility categorization (present, absent) and intervention group (manipulation plus exercise, stabilization exercise). The first- and second-order interactions as well as main effects were examined. Post hoc pairwise comparisons were performed if a significant interaction effect was found. A separate 3-way repeated-measures ANOVA was performed substituting the hypermobility categorization for hypomobility.

To further examine the clinical meaningfulness of the results, we dichotomized the intervention outcome and calculated number-needed-to-treat (NNT) statistics.[30] Subjects with at least 50% improvement in their ODQ score were judged to have a successful outcome. Otherwise, they were judged to have an unsuccessful outcome. Percentage improvement was calculated by subtracting the final score from the initial score, then dividing by the initial score. This criterion for defining success has been used in previous studies performed to determine clinical variables predictive of success with manipulation and stabilization interventions,[15,17] and is based on earlier studies that found percentage improvements on the ODQ for patients with LBP receiving interventions that were not believed to be matched to the patients' clinical presentation, ranging from 20% to 38% over the 1- to 4-week intervention periods.[25,31,32] The relations between the baseline hypomobility and hypermobility categorizations and intervention success were examined for the entire sample and within each intervention group using chi-square tests of association. We further explored the relation between mobility categorization and intervention outcome using NNT statistics with the associated 95% confidence interval (CI).[33] For subjects categorized as having hypomobility present, failure rates were calculated within each intervention group. This procedure was repeated for subjects categorized with hypermobility present. We used a significance level of .05 for all comparisons.

RESULTS

Of the 131 subjects participating in the trial, 93 (71.0%) and 15 (11.5%) were judged to have hypomobility and hypermobility present, respectively. The distribution of subjects with hypomobility or hypermobility did not differ between the intervention groups. Three subjects (2.3%, 1 in the manipulation plus stabilization exercise group, 2 in the stabilization exercise group) were judged to have both hypomobility and hypermobility. These subjects were included in each analysis. The

Table 3: ODQ Scores Related to Mobility Judgments in Manipulation Plus Exercise Intervention Group

	Hypomobility		Hypermobility	
Outcome	Present (n=50)	Absent (n=20)	Present (n=6)	Absent (n=64)
Initial ODQ score	41.2±9.8	41.9±11.1	47.3±15.3	40.9±9.5
Four-week ODQ score	14.4±15.4	26.0±16.9	32.0±23.0	16.4±15.4
Percentage change in ODQ scores	64.1±36.9	40.4±30.6	37.5±32.8	59.2±36.6
No. (%) of subjects with successful outcome	37 (74.0)	7 (35.0)	1 (16.7)	43 (67.2)

NOTE. Values are mean ± SD unless otherwise indicated.

Table 4: ODQ Scores Related to Mobility Judgments in Stabilization Exercise Intervention Group

	Hypomobility		Hypermobility	
Outcome	Present (n=43)	Absent (n=18)	Present (n=9)	Absent (n=52)
Initial ODQ score	39.7±10.3	43.8±11.8	47.3±16.3	39.8±9.3
Four-week ODQ score	29.4±16.7	18.0±17.3	15.6±18.8	27.8±16.9
Percentage change in ODQ scores	23.2±41.1	59.6±36.5	67.9±34.7	28.1±41.7
No. (%) of subjects with successful outcome	11 (25.6)	11 (61.1)	7 (77.8)	15 (28.8)

NOTE. Values are mean ± SD unless otherwise indicated.

initial, final, and percentage change in ODQ scores for each intervention group based on mobility categorizations are listed in tables 3 and 4. Six (4.5%) subjects dropped out of the study before the 4-week reassessment (2 in the manipulation plus stabilization exercise group, 4 in the stabilization exercise group). All patients reported non–study-related reasons for dropping out (eg, time constraints, family issues). These subjects were included in the analysis by carrying forward the last available ODQ score.

The 3-way repeated-measures ANOVA performed using hypomobility judgments resulted in significant 3-way interaction effects between time, intervention, and hypomobility category ($P<.001$). The 2-way interactions between time and intervention ($P=.31$), and between time and hypomobility category ($P=.49$) were not significant. The 3-way interaction is graphed in figure 1. Post hoc pairwise comparisons found differences between subjects in the manipulation plus stabilization exercise group with hypomobility either present or absent (mean difference, 10.9 points favoring subjects with hypomobility present, $P=.013$; 95% CI, 2.4–19.4). For subjects in the stabilization exercise group, a difference was found between subjects with hypomobility present and subjects with hypomobility absent (mean difference, 15.5 points favoring subjects without hypomobility, $P=.003$; 95% CI, 5.4–25.6). These results indicate that subjects receiving the manipulation plus stabilization exercise intervention experienced greater improvement in ODQ when hypomobility was present, whereas subjects in the stabilization exercise group experienced greater improvement when hypomobility was absent.

The 3-way interaction between hypermobility categorization, intervention, and time was also significant ($P=.003$) (fig 2). The 2-way interactions between hypermobility category and time (.27) and between intervention and time ($P=.69$) were not significant. Post hoc pairwise comparisons failed to detect a difference between subjects in the manipulation plus exercise group with hypermobility present and subjects in the group with hypermobility absent ($P=.20$). For subjects in the stabilization exercise intervention group, a difference was found between subjects with hypermobility present and subjects with hypermobility absent (mean difference, 19.8 points favoring subjects with hypermobility, $P=.003$; 95% CI, 6.8–35.0).

Of the 131 subjects in the study, 66 (50.4%) experienced at least a 50% improvement on the ODQ over 4 weeks and were judged to have a successful outcome. A greater percentage of subjects in the manipulation plus stabilization exercise intervention group experienced a successful outcome than subjects in the stabilization exercise intervention (62.9% vs 36.1%, $P=.002$). Of the 93 subjects with hypomobility present, 48 (51.6%) had a successful outcome, which did not differ from the percentage of subjects with successful outcomes among subjects without hypomobility (18/38 [47.4%], $P=.66$). Likewise, no difference was found between the percentage of subjects with successful outcomes among subjects with hypermobility present or absent (53.3% vs 50%, respectively, $P=.81$).

Differences in success rates based on mobility judgments were found when the intervention group was taken into consideration. Among subjects in the manipulation plus stabilization exercise group (table 3), a greater percentage of subjects with hypomobility had successful outcomes than subjects without hypomobility (74.0% vs 35.0%, $P=.002$), whereas subjects with hypermobility were less likely to experience a successful outcome than subjects without hypermobility (16.7% vs 67.2%, $P=.014$). Considering subjects in the stabilization exercise group, the opposite pattern was observed (table 4). The percentage of

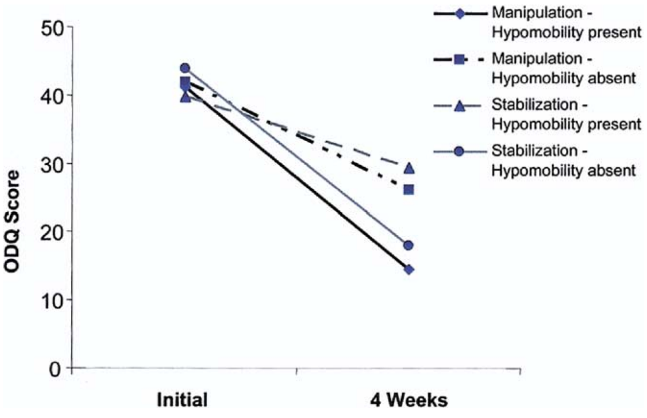

Fig 1. ODQ scores over the 4-week treatment period based on the intervention group and presence or absence of hypomobility. A significant 3-way interaction (*P*<.001) was found between time, intervention, and hypomobility.

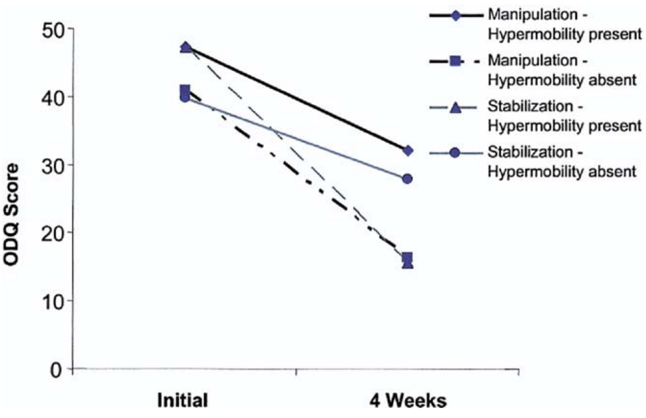

Fig 2. ODQ scores over the 4-week treatment period based on the intervention group and presence or absence of hypermobility. A significant 3-way interaction (*P*=.003) was found between time, intervention, and hypermobility.

Table 5: Rates of Unsuccessful Outcomes With Each Intervention Based on Mobility Judgments Made at Initial Examination

Outcome Rate	Subjects With Hypomobility Present (n=93)	Subjects With Hypermobility Present (n=15)
Rate of unsuccessful outcomes in the manipulation group	13/50 (26.0)	5/6 (83.3)
Rate of unsuccessful outcomes in the stabilization group	32/43 (74.4)	2/9 (22.2)
	ARR: .48 (.29–.63) NNT with manipulation (vs stabilization) to prevent an additional unsuccessful outcome: 2.1 (1.6–3.5)	ARR: .61 (.10–.82) NNT with stabilization (vs manipulation) to prevent an additional unsuccessful outcome: 1.6 (1.2–10.2)

NOTE. Values are n (%) or n (95% CI). The absolute risk reduction (ARR) was calculated as $|\text{Rate}_{\text{manipulation}} - \text{Rate}_{\text{stabilization}}|$. The NNT was calculated as 1/(ARR).

patients having a successful outcome was greater in subjects without hypomobility than those with hypomobility (61.1% vs 25.6%, P=.008), and in subjects with hypermobility versus those without hypermobility (77.8% vs 28.8%, P=.005).

The NNT statistics were calculated from the rates of unsuccessful outcome based on hypomobility or hypermobility judgments[34] (table 5). For subjects with hypomobility, failure rates were 26.0% for subjects receiving manipulation plus stabilization exercise and 74.4% for subjects in the stabilization exercise group, resulting in an NNT of 2.1 (95% CI, 1.6–3.5). This indicates that for a person with LBP judged to have hypomobility at the initial examination, approximately 2 people would need to be treated with the manipulation intervention, rather than the stabilization intervention, for 4 weeks to prevent 1 additional person from experiencing an unsuccessful outcome. For subjects with hypermobility, failure rates were 83.3% and 22.2% for the manipulation and stabilization interventions, respectively, resulting in an NNT of 1.6 (95% CI, 1.2–10.2). Therefore, for those judged to have hypermobility present at the initial examination, approximately 2 people would need to be treated with the stabilization exercise intervention, rather than manipulation, for 4 weeks to prevent 1 additional person from experiencing an unsuccessful outcome.

DISCUSSION

When examining the validity of examination procedures such as PA mobility testing, it is important to consider the therapeutic decisions being made based on the test results.[35] Previous studies examining the validity of the PA mobility testing have focused on comparing judgments made from therapists' assessment of PA mobility in humans with mechanical devices designed to quantify stiffness.[36-40] The results of these studies have shown variable correlations between therapist and mechanical stiffness assessments, but have not addressed the validity of the therapeutic decisions that result from assessments of PA mobility. Maitland[4] popularized the use of PA mobility assessment as a way to detect mobility of the lumbar spinal segments. Maitland and others[8,9,41,42] have proposed that the identification of hypomobility is an important examination finding indicating a need to apply mobilization/manipulation interventions; however, the validity of this decision-making rationale has been largely untested. Similarly, evidence[43,44] supporting the benefits of stabilization exercises has increased interest in identifying the subgroup of patients with LBP most likely to respond to this intervention. The presumption has been that patients judged to have hypermobility would be most likely to benefit from a stabilization intervention.[10,11,45] This decision-making theory has also been untested until recently.

The results of this study support the theorized relations between mobility judgments and preferential response to a specific intervention. We found that subjects with LBP judged to have hypomobility at some level of the lumbar spine experienced greater benefit from an intervention program that included manipulation plus stabilization, and those judged to have hypermobility were more likely to benefit from an intervention focused on stabilization exercises. These findings corroborate the results of earlier research[15,17] and support the predictive validity of PA mobility testing for patients with LBP.

The NNT statistics illuminate the impact that mobility judgments can have on clinical decision making. The NNT was approximately 2 for both the use of the manipulation intervention for patients with lumbar hypomobility, and the use of the stabilization intervention for patients with lumbar hypermobility. These results indicate that for every 2 patients treated over 4 weeks with the intervention matched to the mobility judgment (ie, manipulation for hypomobility or stabilization for hypermobility), an unsuccessful outcome will be avoided in 1 of those 2 patients that would have occurred if the unmatched intervention had been applied instead. Considering the disproportionate amount of health care and workers' compensation costs accounted for by people with LBP who experience persistent disability beyond 4 weeks,[46] the early application of the intervention most likely to result in a rapid, pronounced reduction in disability could result in a substantial cost savings. Further research is needed to study the benefits of using mobility judgments in conjunction with other examination findings for making intervention decisions.

Several investigators[6,12,13,47] have questioned the clinical usefulness of PA mobility judgments based on the results of studies showing questionable reliability. We believe that the usefulness of PA mobility judgments shown in this and previous studies[15-18] may be attributed to 2 primary factors; first, the grading scale used, and second, the examination of predictive validity along with reliability. Maher and Adams[6] examined PA mobility in 90 patients with LBP using an 11-level grading scale ranging from −5 (markedly reduced stiffness) to +5 (markedly increased stiffness). The interrater reliability for judgments at each lumbar level ranged from .03 to .37.[6] Binkley et al[12] used a similar 9-point grading scale to judge mobility of each lumbar level on 18 patients with chronic LBP and reported an overall ICC value of .25. We reduced the clinician's decision making to 3 options (hypomobile, normal, hypermobile). We believe that a 3-level judgment is reflective of the clinical judgments made from the assessment of PA mobility. A principal reason offered for assessing mobility is to improve decision making when selecting the intervention most likely to benefit the patient.[4,5,7] This implies that as a therapist assesses the mobility of a lumbar level there is some threshold level of reduced mobility that, once exceeded, indicates to the therapist that the patient may benefit from an intervention to increase mobility, such as manipulation or mobilization. A similar threshold would exist to indicate that a sufficient amount of hypermobility is present to warrant an intervention designed to improve motor control and spinal stability. Considered in this manner, 3 primary mobility judgments would be relevant—hypomobility sufficient to require a mobilizing intervention, hypermobility sufficient to require a lumbar stabilization intervention, or normal mobility. The clinical utility of judgments of incremental levels of hypomobility or hypermobility has not been shown. The reduction of mobility judgments to 3 levels was supported in a study by Maher et al.[7] They studied 29 therapists from 2 countries using cluster analysis and found that 31 descriptors of PA mobility could be ultimately reduced to 2 "superclusters" that the authors labeled as limited mobility and increased mobility.[7]

We further deviated from prior assessments of PA mobility by considering the entire lumbar spine instead of each individual lumbar level. The therapists in the Maher study judged each lumbar level using the 3-level grading scale, but we further collapsed the mobility judgments into 2 categories. Each subject was categorized as having hypomobility either present at some level of the lumbar spine or absent from all levels. A similar categorization was made for the presence of hypermobility. The rationale for these categorizations was 2-fold; first, previous research has shown that part of the error associated with determining segmental mobility is attributable to disagreements in identifying the spinal level being examined.[12,48] Second, the necessity of specifying the hypomobile or hypermobile lumbar level(s) is questionable. Stabilization exercise interventions are generally not applied to individual levels of the lumbar spine, and there is no indication in the literature that the motor control of an individual lumbar motion segment can be trained in isolation, thereby rendering the specification of a lumbar level associated with a judgment of hypermobility unnecessary. The ability to isolate manipulation techniques to a specific lumbar segment is also a somewhat dubious claim, despite the implicit assumption taught by advocates of manual therapy that localization to a specific segment is a precondition of a successful technique.[42,49] Recent evidence challenges the veracity of the presumption that segment specificity is necessary, or even possible. Studies have found that manual forces applied to 1 lumbar level create considerable movements at all levels of the lumbar spine.[50,51] The mechanisms underlying the benefit of spinal manipulation are poorly understood,[52] and some studies question prevailing biomechanic theories that have attributed the effects of spinal manipulation to the separation of joint surfaces at a specific level of the spine.[53-55] A recent randomized trial by Haas et al[56] compared the immediate effect of manipulation on pain and mobility on patients with neck pain. In 1 group, the level of the manipulation was determined by the clinician's manual assessment of restricted mobility, and in the other group a random assigned level was manipulated. Both groups showed improvements in pain and mobility with no differences between the groups, leading the authors to speculate that "specificity of the site and vector of manipulation is not as important as generally thought."[56(p1095)]

Because of the difficulty in identifying a particular level of the lumbar spine, and the questionable need to do so, we chose to categorize subjects without regard to level. This did result in a small percentage of subjects (2%, n=3) categorized as having both hypermobility and hypomobility. We cannot draw any conclusions on the best intervention approach for patients with LBP judged to have both hypomobility and hypermobility simultaneously.

Despite removing the need to name the particular level of the lumbar spine judged to be hypomobile or hypermobile, our previous work has not found reliability coefficients substantially greater than those reported by other investigators using different mobility scales. The results of our previous studies[16,23,24] that have examined the interrater reliability of judgments of hypomobility or hypermobility made in the same manner as in this study have shown fair to moderate κ values according to Landis and Koch[57] (range for hypomobility, .13–.38; range for hypermobility, .30–.48), with generally good percentage agreement (59%–77% for hypomobility; 77%–78% for hypermobility). The paradox of high agreement but low κ values is an anticipated phenomenon when the prevalence of positive ratings by each examiner is substantially greater or less than 0.5.[58] In this sample of subjects, the prevalence of hypomobility judgments was high (.71), whereas the prevalence of hypermobility judgments was low (.12). The previous studies examining the reliability of these judgments reported similarly extreme prevalence values for positive findings, ranging from as high as .86 for judgments of hypomobility, to a low of .08 for judgments of hypermobility.[23,24] It is likely that these prevalence values have deflated the κ values reported in this study and in previous studies.

The results of this study and previous research[15-18] indicate that, although the reliability coefficients associated with judgments of hypomobility and hypermobility do not exceed arbitrary thresholds of "acceptable" reliability,[57] the judgments can provide information that is useful and valid for clinical decision making, particularly when used in combination with other examination findings. Studies by Flynn[17] and Hicks[15] and colleagues found that judgments of hypomobility and hypermobility were significantly related to success with a manipulation or stabilization exercise intervention, respectively, and both studies included mobility judgments in multifactorial clinical prediction rules for predicting success with their respective interventions. In a previous study we also found that a judgment of hypermobility was 81% specific with a positive-likelihood ratio of 2.4 for predicting radiographic lumbar instability.[16] Each of these studies included numerous examination procedures for their potential use in predicting the intervention or radiographic outcome and, although many procedures had greater reliability than mobility judgments, very few showed significant relations with the outcome of interest.[15-17]

Although counterintuitive to the traditional admonition that reliability is a precursor to validity, other studies have reported on diagnostic tests with low reliability coefficients but clinically useful accuracy for predicting a certain disorder or outcome. For example, Etchells et al[59] studied the accuracy of various clinical examination findings for predicting moderate or severe aortic stenosis using a reference standard of Doppler echocardiography. The clinical judgment of delayed carotid upstroke had a low κ coefficient for interrater reliability (.26) but an excellent positive-likelihood ratio (9.2), and was 1 factor in a multifactorial clinical prediction rule with a positive-likelihood ratio of 40 for predicting moderate or severe stenosis.[59] Edelman et al[60] studied the accuracy of clinical examination findings performed by primary care physicians for predicting increased risk of foot ulcers in patients with diabetes. The interrater reliability of judgments of reduced dorsalis pedis and posterior tibialis pulses were low (.36 and .38, respectively) yet the findings were associated with high positive-likelihood ratios (9.5 and 7.4, respectively) for predicting an increased risk of foot ulcers. The reasons for these paradoxical findings could relate to problems with prevalence, or point to the arbitrary nature of determining "acceptable" reliability based on traditional rankings. Improving the reliability of these examination procedures would likely enhance their diagnostic accuracy, yet these examples and our results offer a caution against dismissing examination procedures solely on the basis of reliability with consideration of validity.

CONCLUSIONS

To our knowledge, this is the first study to examine the validity of the PA segmental mobility assessment based on comparing judgments of segmental mobility with the interventions used and patient outcomes. In our study, subjects with LBP who were judged to have lumbar spine hypomobility experienced greater benefit from the manipulation plus stabilization intervention. Similarly, subjects who were judged to have lumbar hypermobility were more likely to benefit from a stabilization exercise program. Further research is needed to determine the benefits of using judgments made from PA mobility assessment in combination with other examination findings for intervention selection and, ultimately, on patient outcomes.

Acknowledgments: We thank the following physical therapists for their assistance with data collection: Kevin Johnson, Guy Majkowski, Maria West, Evan Kelley, David Browder, Mike Blowers, Sherri Morrow, Brian Langford, Jeff McGuire, Cory Middel, and Trevor Petrou.

References

1. Jones MA, Jones HM. Principles of the physical examination. In: Boyling JD, Palastanga N, editors. Grieve's modern manual therapy. Edinburgh: Churchill Livingstone; 1996. p 491-510.
2. Hertling D, Kessler RM. Management of common musculoskeletal disorders. Philadelphia: Lippincott; 1996. p 640-58.
3. Magee DJ. Orthopaedic physical assessment. 3rd ed. Philadelphia: Saunders; 1997. p 407.
4. Maitland GD. Vertebral manipulation. 5th ed. Oxford: Butterworth Heinemann; 1986. p 74-6.
5. Latimer J, Lee M, Adams R, Moran CM. An investigation of the relationship between low back pain and lumbar posteroanterior stiffness. J Manipulative Physiol Ther 1996;19:587-91.
6. Maher C, Adams R. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. Phys Ther 1994;74: 801-11.
7. Maher CG, Simmonds M, Adams R. Therapists' conceptualization and characterization of the clinical concept of spinal stiffness. Phys Ther 1998;78:289-300.
8. Jull GA. Examination of the articular system. In: Boyling JD, Palastanga N, editors. Grieve's modern manual therapy. Edinburgh: Churchill Livingstone; 1996. p 511-27.
9. Grieve GP. Common vertebral joint problems. Edinburgh: Churchill Livingstone; 1989. p 350-68.
10. Grieve GP. Lumbar instability. Physiotherapy 1982;68:2-9.
11. Paris SV. Physical signs of instability. Spine 1985;10:277-9.
12. Binkley J, Stratford P, Gill C. Interrater reliability of lumbar accessory motion mobility testing. Phys Ther 1995;75:786-95.
13. Phillips DR, Twomey LT. A comparison of manual diagnosis with a diagnosis established by a uni-level lumbar spinal block procedure. Man Ther 1996;2:82-7.
14. Matyas TA, Bach TM. Reliability of selected techniques in clinical arthrometrics. Aust J Physiother 1985;31:175-99.
15. Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. Arch Phys Med Rehabil 2005;86:1753-62.
16. Fritz JM, Piva SR, Childs JD. Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. Eur Spine J. In press.
17. Flynn T, Fritz J, Whitman J, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short term improvement with spinal manipulation. Spine 2002;27:2835-43.
18. Childs JD, Fritz JM, Flynn TW, et al. A clinical prediction rule to identify patients with low back pain who will benefit from spinal manipulation: a validation study. Ann Intern Med 2004;141: 920-8.
19. Bigos S, Bowyer O, Braen G, et al. Acute low back problems in adults. Rockville: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services; 1994. AHCPR Publication No. 95-0642.
20. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? Pain 1994;58:387-92.
21. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52:157-68.
22. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Phys Ther 2001;81:776-88.
23. Hicks GE, Fritz JM, Delitto A, Mischock J. Interrater reliability of clinical examination measures for identification of lumbar segmental instability. Arch Phys Med Rehabil 2003;84:1858-64.
24. Fritz JM, Whitman JM, Flynn TW, Wainner RS, Childs JD. Factors related to the inability of individuals with low back pain to improve with a spinal manipulation. Phys Ther 2004;84:173-90.
25. Delitto A, Cibulka MT, Erhard RE, Bowling RW, Tenhula JA. Evidence for use of an extension-mobilization category in acute low back syndrome: a prescriptive validation pilot study. Phys Ther 1993;73:216-28.
26. Richardson CA, Jull GA. Muscle control-pain control: what exercises would you prescribe? Man Ther 1995;1:2-10.
27. Richardson CA, Snijders CJ, Hides JA, Damen L, Pas MS, Storm J. The relation between the transversus abdominis muscles, sacroiliac joint mechanics, and low back pain. Spine 2002;27:399-405.
28. McGill SM. Low back exercises: evidence for improving exercise regimens. Phys Ther 1998;78:754-64.
29. McGill SM. Low back stability: from formal description to issues for performance and rehabilitation. Exerc Sports Sci Rev 2001; 29:26-31.
30. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. Ann Intern Med 1997;126:712-20.
31. 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.
32. 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.
33. Altman DG. Clinical trials and meta-analysis. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. Statistics with confidence. Bristol: BMJ Books; 2000. p 105-19.
34. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. New York: Churchill Livingstone; 2000. p 105-55.
35. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. Phys Ther 2001;81:1546-64.
36. Chiradejnant A, Maher CG, Latimer J. Objective manual assessment of lumbar posteroanterior stiffness is now possible. J Manipulative Physiol Ther 2003;26:34-9.
37. Simmonds MJ, Kumar S, Lechelt E. Use of a spinal model to quantify the forces and motion that occur during therapists' tests of spinal motion. Phys Ther 1995;75:212-22.

38. 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.
39. Latimer J, Lee M, Adams R. The effects of training with feedback on physiotherapy students' ability to judge lumbar stiffness. Man Ther 1996;1:266-70.
40. Maher C, Adams R. Is the clinical concept of spinal stiffness multidimensional? Phys Ther 1995;75:854-64.
41. Kaltenborn FM. Manual mobilization of the extremity joints. Minneapolis: Orthopedic Physical Therapy Products; 1989. p 5-48.
42. Greenman P. Osteopathic manipulation of the lumbar spine and pelvis. In: White A, Anderson A, editors. Conservative care of low back pain. Baltimore: Williams & Wilkins; 1991. p 210-5.
43. O'Sullivan PB, Phyty GD, Twomey LT, Allison GT. Evaluation of specific stabilizing exercises in the treatment of chronic low back pain with radiologic diagnosis of spondylosis or spondylolisthesis. Spine 1997;22:2959-67.
44. Hides JA, Jull GA, Richardson CA. Long-term effects of specific stabilizing exercises for first-episode low back pain. Spine 2001; 26:E243-8.
45. Delitto A, Erhard RE, Bowling RW. A treatment based classification approach to low back syndrome: identifying and staging patients for conservative treatment. Phys Ther 1995;75:470-89.
46. 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;40:1110-9.
47. Riddle DL. Measurement of accessory motion: critical issues and related concepts. Phys Ther 1992;72:865-74.
48. Downey B, Taylor N, Niere K. Can manipulative physiotherapists agree on which lumbar level to treat based on palpation? Physiotherapy 2003;89:74-81.
49. Triano J. The mechanics of spinal manipulation. In: Herzog W, editor. Clinical biomechanics of spinal manipulation. New York: Churchill Livingstone; 2000. p 92-190.
50. Powers CM, Kulig K, Harrison J, Bergman G. Segmental mobility of the lumbar spine during posterior to anterior mobilization: assessment using dynamic MRI. Clin Biomech (Bristol, Avon) 2003;18:80-3.
51. Lee R, Evans J. An in vivo study of the intervertebral movements produced by posteroanterior mobilization. Clin Biomech (Bristol, Avon) 1997;12:400-8.
52. Evans DW. Mechanisms and effects of spinal high-velocity, low-amplitude thrust manipulation: previous theories. J Manipulative Physiol Ther 2002;25:251-62.
53. Flynn TW, Fritz JM, Wainner RS, Whitman JM. The audible pop is not necessary for successful spinal high-velocity thrust manipulation in individuals with low back pain. Arch Phys Med Rehabil 2003;84:1057-60.
54. Suter E, Herzog W, Conway PJ, Zhang YT. Reflex response associated with manipulative treatment of the thoracic spine. J Neuromusculoskeletal Syst 1994;2:72-6.
55. Herzog W, Scheele D, Conway PJ. Electromyographic responses of back and limb muscles associated with spinal manipulative therapy. Spine 1999;24:146-53.
56. Haas M, Groupp E, Panzer D, Partna L, Lumsden S, Aickin M. Efficacy of cervical endplay assessment as an indicator for spinal manipulation. Spine 2003;28:1091-6.
57. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
58. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-9.
59. Etchells E, Glenns V, Shadowitz S, Bell C, Siu S. A bedside clinical prediction rule for detecting moderate or severe aortic stenosis. J Gen Intern Med 1998;13:699-704.
60. Edelman D, Sanders LJ, Pogach L. Reproducibility and accuracy among primary care providers of a screening examination for foot ulcer risk among diabetic patients. Prev Med 1998;27:274-8.