12-1-2013

Inter-rater Reliability of Lumbar Segmental Instability Tests and the Subclassification

Faisal Mohammad Alyazedi
Loma Linda University

LOMA LINDA UNIVERSITY
School of Allied Health Professions
in conjunction with the
Faculty of Graduate Studies

The Inter-rater Reliability of Lumbar Segmental Instability Tests and the Subclassification

by

Faisal Mohammad Alyazedi

A Dissertation submitted in partial satisfaction of
the requirements for the degree
Doctor of Science in Physical Therapy

December 2013

Each person whose signature appears below certifies that this dissertation in his/her opinion is adequate, in scope and quality, as a dissertation for the degree Doctor of Science.

____, Chairperson
Everett B. Lohman III, Professor of Physical Therapy, Assistant Dean for Graduate Academic Affairs, and Director of Post- professional Programs

____
Khaled Bahjri, Assistant Professor, Epidemiology, Biostatistics & Population Medicine, Research / Statistical Services Specialist, Research Consultant Group, School of Public Health

____
R. Wesley Swen, Assistant Professor of Physical Therapy

ACKNOWLEDGEMENTS

I extend my sincere gratitude to my committee members who supported me throughout this study, by providing me with prompt feedback and continuous encouragement.

I sincerely thank my mentor, Dr. Lohman, for sparing much of his valuable time to provide me with valuable feedback and support. I highly appreciate his involvement in my graduate research as an expert examiner, skillful editor and presentation trainer.

I also thank Dr. Swen, a skillful examiner and editor, for his valuable input that enabled me to organize my work. His suggestion that I examine five rather than two patients per week enabled me to complete this research before the expiration of my scholarship.

I also express my gratitude to Dr. Khaled, a brilliant statistician, who was the mind behind my research methodology and data analysis.

I also thank my father, my first and most caring teacher. He taught, supported and encouraged me to do well as a physical therapist, and as an educated person who can serve the community. Finally, I thank my mother, wife and family, who sincerely prayed for me and offered me a healthy educational environment, friendship and companionship.

CONTENTS

TABLES

FIGURES

Figures Page

ABBREVIATIONS

CI	Confidence interval
CPR	Clinical prediction rule
EZ	Elastic zone
FABQ	Fear Avoidance beliefs questionnaire
FABQ-PA	Fear Avoidance beliefs questionnaire scale of physical activity
ICC	Intraclass Correlation Coefficient
k	Kappa coefficient
L	Lumbar vertebra
-LR	Negative likelihood
+LR	Positive likelihood
LSI	Lumbar segmental instability
NPRS	Numeric Pain Rating Scale
NZ	Neutral zone
OSW	Oswestry Low Back Pain Disability Questionnaire
PA	Posterior-anterior
PABAK	Prevalence and bias adjusted kappa
PIT	Prone instability test
PLET	Passive lumbar extension test
R/CLBP	Recurrent or chronic lower back pain
ROM	Range of Motion
SLR	Straight Leg Raising

SPSS	Statistical package for social sciences
T	Thoracic vertebra

ABSTRACT OF THE DISSERTATION

The Inter-rater Reliability of Lumbar Segmental Instability Tests and the Subclassification

by

Faisal Mohammad Alyazedi

Doctor of Science Graduate Program in Physical Therapy
Loma Linda University, December 2013
Dr. Everett B. Lohman III, Chairperson

Objectives: This study investigated the inter-rater reliability of three structural end-range lumbar segmental instability tests with the highest positive Likelihood Ratio against flexion-extension radiographs, and three functional mid-range clinical tests that predict the success of lumbar stabilization exercises in patients with recurrent or chronic low back pain (R/CLBP). It also investigated the reliability of lumbar segmental instability subclassification as: Functional, Structural and Combined Instability.

Method: 40 adult with R/CLBP patients (30 men and 10 women), 18 to 80 years of age, underwent repeated measurements of specific clinical tests for structural or functional lumbar segmental instability.

Results: Other than the Lack of Hypomobility with PA Glide test, which was found to be unreliable (percentage agreement = 37.5, and k= - 0.02), all the other tests demonstrated high Kappa coefficients and percentage agreements. The sub-classification categories of lumbar segmental instability (functional, structural, and combined) were found to be significantly reliable (k= 0.722, adjusted k= .7 and k =0.84, respectively).

Discussion: All the investigated tests (except lack of hypomobility with PA glide test); as well as, the categories of lumbar segmental instability sub-classification, were significantly reliable in predicting lumbar stabilization.

Key words: Low back pain, Segmental instability, Reliability, Physical examination, Clinical prediction rule.

CHAPTER ONE

INTRODUCTION

Low back pain (LBP) is a common musculoskeletal condition that affects up to 80 % of the general population during their life time.[1-3] It has high recurrence rates of up to 30% and 45% at the first and the third year after the first episode, respectively.[4] Lumbar segmental instability is believed to be one of the main causes of the high recurrence rates.[2, 5, 6] The prevalence of LBP due to lumbar segmental instability is about 33% among patients treated for mechanical LBP in a physical therapy setting.[7] The prevalence is even higher, at 57%, among patients who have already been referred for flexion-extension radiography due to high suspicion of lumbar segmental instability.[8, 9]

Hides *et al* investigated the long-term effects of stabilizing exercises on the LBP recurrence rate for subjects who are having their first episode of low back pain.[5] They established that subjects who were treated with stabilization exercise had a lower LBP recurrence than the control group. In recent meta-analyses, lumbar stabilization was found to be superior to several other treatment methods with regard to reducing disability in patients with recurrent or chronic low back pain at short, intermediate and long term periods.[10]

Panjabi [11, 12] argued that for the spine to move safely, the passive, active and neural spinal subsystems have to act interdependently to maintain spinal mechanical stability. Other previous studies also supported the argument that spinal stability is paramount during all musculoskeletal tasks and ranges of motion.[1, 6, 11-15] The Passive

Subsystem consists of the vertebrae, facet joints, intervertebral discs, and spinal ligaments. This subsystem passively resists spinal instability at *end-range* of motion. It can withstand a low critical load of about 20 lbs. [11, 12] Conversely, the Active Subsystem consists of the spinal musculotendinous tissues which provide mechanical stability around *the neutral position* of a spinal segment with critical load capacity exceeding 337 lbs.[11, 13] The Neural Control subsystem receives the proprioceptive (position sense) information from various structures in the passive and active subsystems, computes the magnitude and timing of muscular contraction needed for segmental stability, then generates efferent information to the muscular subsystem to provide dynamic stability.[6] The poor quality of proprioceptive afferent information, weak or fatigued spinal muscles and error in motor control subsystem can be the underlying causes of lumbar segmental instability.[6, 11, 12]

In 1944, Knutson [15] recommended the use of flexion-extension radiographs to identify and quantify abnormal anterior-to-posterior translation of the motion segment at the *end-range* of spinal flexion and extension. This imaging modality has become the diagnostic standard of structural lumbar segmental instability or instability due to disruption of passive stabilizers.[8, 15-20] His work in lumbar flexion-extension radiography became the subject of original researches and systematic reviews that were conducted to establish the most accurate test that predicts the flexion-extension radiograph findings. [8, 9, 21-23] On the other hand, the diagnostic standard that can quantify the functional instability *around the neutral position* was lacking.[6, 8, 17] A number of studies attributed this functional instability to the lack of neuromuscular control of the joint during daily living activities.[6, 8, 20] Some physical therapy investigators studied the clinical tests that might predict the success of stabilization exercises that were developed exclusively to improve spinal motor control (stiffness) around the neutral position of spinal.[5, 7, 10] They came up with four predictors that, together, form the clinical prediction rule (CPR) of stabilization exercise.[7] The latest systematic review of the spinal CPRs supports its validity by providing level 4 evidence; the highest level of score in the hierarchy evidence, among all spinal CPRs.[24] However, its reliability in patients with recurrent or chronic LBP (R/CLBP) is yet to be established. Additionally, the reliability of Average SLR >91° test, one of the criteria that make the CPR, is still unknown.

Because it is important to establish the reliability of the most accurate to date tests, this study investigated the Inter-rater reliability of six clinical lumbar instability tests that showed highest positive likelihood ratio (+LR) in predicting either the findings of the flexion-extension radiograph, or the treatment outcome for recurrent or chronic LBP patients. The clinical tests consisted of: 1) the Prone Instability Test (PIT); 2) the Aberrant Motion Test; 3) the Average SLR (>91°) Test; 4) Lumbar Flexion ROM >53°; 5) the Lack of Hypomobility with PA Glide test, and 6) the Passive Lumbar Extension Test.[7, 9, 23]

Recently, a group of researchers investigated the reliability of CPR and the passive lumbar extension test in the general population of LBP subjects.[25] To further explore the reliability of CPR, this study investigates the reliability of CPR criterion, and in addition, the reliability of the 3 tests that showed highest positive likelihood ratio to predict the findings of the radiographic structural instability in recurrent or chronic LBP patients.[9, 23] It also studied the inter-rater reliability of examiners to sub-classify lumbar instability subjects into different lumbar instability categories: structural, functional and combined.

Figure 1 illustrates the main lumbar segmental instability categories. The dashed line indicated the structural instability category because it is not finalized. Future research is needed to finalize the cluster of tests that highly predict the radiographic structural instability gold standard (flexion-extension X-ray).

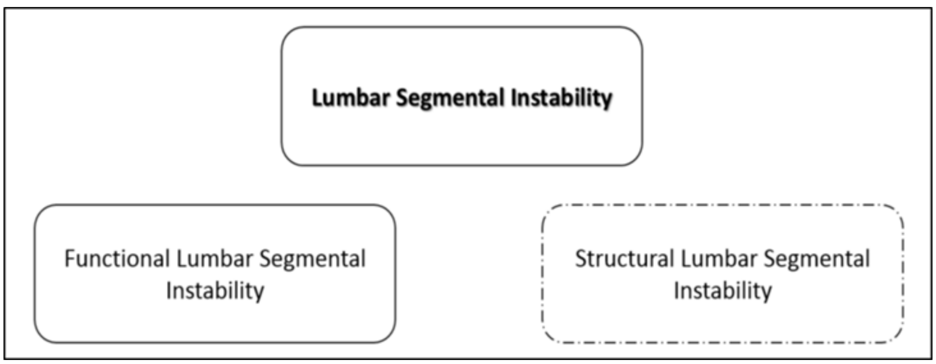

Figure 1: Categories of lumbar segmental instability

Methods

Study Participants

The study participants were 40 subjects (30 men and 10 women) who were between 18 to 80 years of age and had recurrent/chronic low back pain R/CLBP all recruited from San Bernardino community.

Inclusion and Exclusion Criteria

The inclusion criteria for this study consisted of 1) having a new episode of low back pain and, 2) having experienced a similar episode of low back pain before; whereby the first episode of back pain ever experienced was at least three months before the date of recruitment, or 3) currently experiencing persistent low back pain for at least three months duration.[22]

Conversely, the exclusion criteria consisted of 1) having undergone previous spinal fusion surgery, 2) history of traumatic fracture of the spine that resulted in permanent neurological deficit, 3) scoliosis greater than 20°, 4) pregnancy, 5) inability to actively flex and extend the spine adequately to permit an assessment of segmental motion due to pain or muscle spasm, and 6) medical "red flags," such as Cauda equine syndrome, tumor, systemic inflammatory conditions.

Ethical Issues

Participation in this study was voluntary. All participants were briefed about the study, and were issued with a copy of a consent form approved by the Loma Linda University Institute of Review Board. They reviewed and signed it accordingly.

Examiners

Three physical therapy examiners were involved in the study. They all received a one half hour-training regarding all written and clinical procedures of the study. One of them, who had thirteen years of musculoskeletal clinical experience, recorded the baseline data of the participants. The role of remaining two examiners involved performing and interpreting the various clinical tests on all subjects. These examiners were considered to be fit for their assigned role because they each had more than 20 years of musculoskeletal clinical experience.

Data Collection

The data collector therapist collected all baseline data, which consisted of: Informed consent, demographic information, self-reported history and self- reported outcome measures. Then the two clinical examiners, who were blinded to each other's test results, performed the clinical tests and determined the test results for each subject.

Each participant completed three self- reported outcome assessment tools to assess their degree of functional limitation that is attributed to back pain. These included the Numeric Pain Rating Scales (NPRS), the Modified Oswestry Low Back Pain Disability Questionnaire (OSW), and the Fear Avoidance Beliefs Questionnaire (FABQ). [26-28]

The Numerical Pain Rating Scale (NPRS) was used to assess the severity of low back pain by using an 11 point (0-10) scale. On the other hand, the Modified Oswestry Low Back Pain Disability Questionnaire (OSW) has 10 sections; one section for pain severity and the other nine representing various functional activities. [27] This questionnaire indicates the degree of LBP-attributed limitation in the specified activities. Each section contains 6 responses, scored from 0–5. Each section score was summed to obtain the final score, which was then multiplied by 2, and the degree of disability was expressed as a percentage. The Fear Avoidance Beliefs Questionnaire FABQ assesses the level of fear-avoidance beliefs associated with low back pain. [28] It consists of 4-items on physical activity (FABQ-PA), with a magnitude range from 0–24, and 7-items on the scale of work (FABQ-W), potentially ranging from 0–42. Subjects rated their agreement with each statement related to either physical activity or work from 0-6, where 0 is "completely disagree," and 6 is "completely agree".

After filling the assessment tools, each participant underwent a musculoskeletal examination which comprised of the following tests:

1) Assessment of Aberrant Motion; whereby the subject was required to attain a standing position and flex the trunk forward as far as possible. The examiner observed the subject's movement in an effort to identify any of the following abnormalities: painful arc of motion, an instability catch, "thigh climbing" (Gower's sign), or a reversal of lumbopelvic rhythm. If any of these movements was present, then the test was determined to be positive. Previous studies showed that this test is moderately reliable in assessing functional limitation due to low back pain (K=.60 (95% CI, .43–.73).[17]

2) The True Lumbar Flexion Range of Motion (ROM) was measured using a single bubble inclinometer. With the subject in the standing position, the examiner held the inclinometer on the T12-L1 reference point and asked the subject to bend forwards as far as possible toward the toes. The test begins with the subject being in a standing position to allow the examiner to take baseline readings of sacral and lumber flexion. Keeping the subject fully flexed, the end range of T12-L1 (total lumbar flexion) was recorded first; after which, the end range of S2 (sacral flexion) was recorded. The sacral range was then subtracted from the total lumbar ROM to identify the true lumbar flexion. Using an inclinometer to establish lumbar flexion ROM is moderately reliable (ICC = 0.60).[9, 20]

However, using the cut-off value of 53° to measure the test reliability has not been reported in literature.

3) Passive Lumbar Extension (PLE) Test: For this test, the subject was instructed to lie in the prone position. Then the examiner elevated the subject's legs concurrently until the heels were positioned about 30 cm from the bed level, while maintaining the knees in full extension, the legs were gently pulled. The test was considered to be positive if the subject reported a feeling of severe low back pain; a feeling of heaviness on the lower back or a feeling described as the "low back about to come off". Previous literature showed that this test is highly valid in predicting the flexion-extension radiographic results of lumbar structural instability; with a sensitivity of 84.2%, specificity of 90.4%, and positive Likelihood Ratio of 8.84.[23] Even though Kasai et al reported excellent agreement between examiners, no specific reliability coefficients were reported.[23] Recently, another study reported the reliability of the test on general LBP patients to be about 0.76 [25] However, its reliability on R/CLBP has not been investigated.

4) Lack of Hypomobility with PA Glide test: This test was begun with the subject in prone position. Then the examiner located the subject's spinous processes for each lumbar segment, and exerted a posterior-anterior force on the lumbar segment with the hypothenar eminence of his hand. He, then, judged and recorded the accessory movements for each segment as normal, hypermobile, or hypomobile. The test was considered positive if all lumbar spine segments were judged to be not stiff (hypomobile).[9]

The examiners also judged the mobility of the painful segment by using a grading scale developed by Stanley Paris, which correlated scores of ≤ 2, 3, and ≥4 with stiffness, normal mobility and hypermobility, respectively.[17, 29]

5) Prone Instability Test (PIT): The subject bended over the examining table such that his/her torso lay in a prone position on the examining table, while bending the waist so that the feet rested on the floor. While the subject rested in this position, the examiner contacted the subject's lumbar spinous processes with the hypothenar eminence and exerted a posterior-anterior force to each level of the lumbar spine. Any provocation of pain was recorded. The subject was then asked to slightly lift his feet off the floor and the passive intervertebral motion testing was reapplied to any segments that had been identified as painful. The test was considered positive if the pain was provoked during the first part of the test but disappeared when the test was repeated with the legs off the floor. This test is reported to be found to be significantly reliable; k=0.87 and k=0.69. [9, 17]

6) The Average Straight Leg Raising (>91°) test: With the subject in supine position, the inclinometer was positioned on the tibial crest just below the tibial tubercle. Then the leg was passively raised slowly to the maximum tolerated level, and then the maximum SLR degree was recorded. This procedure was repeated for the other leg. The test was considered positive if the average SLR was more than 91°. [9]

After performing all the clinical tests, each examiner classified the subjects into one of the three instability subcategories (structural, functional, or combined instability). The subjects were classified as structurally unstable if the subject had any of these tests positive: Positive passive lumbar extension test[23] positive lumbar flexion ROM (> 53°) test, or lack of hypomobility with PA glide.[9] The study shows that if the second and third tests are positive, then the subject is 12.8 times more likely to have positive radiographic instability.[9, 20]

The subject was considered functionally unstable if three out of four predictors of functional instability (CPR) were present: 1) Age < 40 years, 2) positive prone instability test (PIT), 3) aberrant motion present and 4) average SLR (>91°). If three out of four predictors are present, then the likelihood of success with lumbar stabilization exercises is LR 4.0. [7]

The subject was considered to have combined instability if the subject had both subcategories (structural and functional instability).

Data Analysis

Data was analyzed using the Statistical Package for Social Sciences (SPSS IBM Corporation 1989, 2011.Version 20).

The baseline outcome measures (numeric pain rating scales, modified OSW, and FABQ), as well as self-reported history (age, duration of the symptoms, and number of the LBP episodes) were summarized using means and standard deviations. For the baseline data items that were not normally distributed, the median and range were used as measures of statistical distribution. These variables are presented in Table 1.

The inter-rater reliability for the various lumbar instability tests was evaluated using Kappa correlation coefficients in order to establish the inter-rater reliability that is above chance agreement. The percentages of agreement for the various tests were also reported. Alpha was set at the level of 0.05.

The cut-off values for both Lumbar Flexion ROM and average SLR, >53°and > 91°, respectively, were used to determine the test results as positive if the ROM passed the cut-off values and negative if it did not table 2.

The reliability of the categories was analyzed by Kappa for functional and combined lumbar instability categories. For structural instability adjusted Kappa and the percent of agreements were calculated to adjust the effect of prevalence and bias indices on the Kappa (Table 3).

Table 1: Demographic information

Variables	**Outcomes**
Age (y)	
Median	31
Range	21 - 71
Gender	
Male	30
Female	10
Modified ODI score	
Minimal disability (number, range)	(28) 0% - 20%
Moderate disability (number, range)	(12) 21% - 40%
FABQ score range	
Physical activity (median, range)	3.50 (0.25 – 5.25)
Chronic LBP (number of subjects)	6
Recurrent LPB (number of subjects)	34
Abbreviations: ODI, Oswestry Disability Index; FABA, fear Avoidance Believe Questionnaires	

Results

The percentage agreement and k value for all the clinical prediction rule tests; PIT, Aberrant motion test and Average SLR (>91°) test showed a high percentage agreement (90.5, 97.5 and 95, respectively) and showed substantial Kappa coefficient (0.73, 0.78 and 0.77, respectively).

The lumbar flexion ROM >53° and lumbar extension test showed a fairly high percentage agreement (82.5 and 72.5, respectively) and moderate Kappa coefficient (0.48 and 0.46, respectively).

The Lack of Hypomobility with PA Glide test is found to be poorly reliable with a low percentage agreement and low Kappa coefficient (37.5, 0.02, respectively). Further examination of the data revealed that the 42% of mobility judgment disagreement between the raters was due to inconsistence regarding the grading of normal (grade 3) and slightly restricted mobility (grade 2) across all lumbar segments. However, locating the painful segment was moderately reliable with Kappa = 0.41.

The functional and combined lumbar segmental instability subcategory was found to be substantially reliable (Kappa = 0.72 and 0.84 respectively), while the adjusted Kappa for structure instability was also substantial reliability (PABAK = 0.7).

Table 2: The reliability coefficient for all lumbar segmental instability tests investigated in this study

Variables	**Kappa (95%CI)**	**Percentage agreement**	**Examiner 1 -ve/+ve**	**Examiner 2 -ve/+ve**
1) PLET	0.46	0.725	19/21	26/14
2) Flexion ROM >53°	0.48	0.0.821	7/33	10/30
3) lack of hypomobility	0.020	0.250	29/11	11/29
4) Aberrant motion	0.79	0.975	37/3	38/2
5) PIT	0.71	0.900	9/31	9/31
6) Average SLR > 91°	0.77	0.950	35/5	35/5

Abbreviations: PLET, Passive lumbar extension test; ROM, range of motion; PA glide, posterior-anterior glide; PIT, prone instability test; SLR, straight leg raising.

Discussion

In this inter-rater reliability study, we investigated the consistence of the most valid tests (highest + LR) in the literature to identify lumbar segmental instability subjects who have structural or functional instability.

We chose three tests that showed the highest positive likelihood ratio against the instability radiographic gold standard; flexion- extension X ray. The selected tests included lumbar extension test (+LR 8.84); the combination of both; lack of hypomobility during lumbar intervertebral motion testing, and lumbar flexion ROM >53° (+LR 12.8).[8, 9, 20, 23] As well as, three tests that predict the treatment outcome results of stabilization exercise category this include; Average SLR (>91°), aberrant motion test and prone instability test +LR 4.0.[7, 24, 30]

The Passive Lumbar extension test was validated by Kasai et al for elderly subjects (age range = 39 to 88 years, mean = 68.9 years) who had known chronic pathologies such as lumbar stenosis, lumbar spondylolisthesis and lumbar degenerative scoliosis.[23] In this study, we included younger subjects ranging from 21 to 71 years of age, median age =31, who have recurrent or chronic LBP. However this test shows acceptable inter-rater reliability even for the younger age group (with Kappa = 0.46). Recently, Rabin et al investigated the reliability of selected lumbar instability tests, one of which was the PLE test on general LBP subjects and found it to be substantially reliable (Kappa=0.76).[25] On the contrary, this study specifically examined subjects with recurrent or chronic LBP. Thus, its findings can be considered to be representative of the patients with recurrent or chronic LBP with mild to moderate modified ODI scores.

Fritz et al studied the common lumbar segmental instability (LSI) clinical tests against the radiographic flexion-extension diagnostic standard. [17] They concluded that the presence of both findings; lumbar flexion ROM >53° and lack of hypomobility with PA glide, would increase the probability of LSI from 50% to 93%, with positive likelihood ratio of about 12.8. However, the reliability of both tests have not been reported in previous studies.

This study established that the lumbar flexion range > 53° test is moderately reliable (Kappa= 0.46), with a high percentage of agreement (82.5). This finding is in agreement with the findings of previous studies, which also showed a high correlation between lumbar flexion range, the flexion –extension radiograph and functional radiograph.[31, 32] This reliability results confirm the appropriateness of this test, especially because this test replicates the first portion (lumbar flexion) of the radiographic procedure in a standing position.

Two previous studies reported low reliability of the segmental mobility test in prone position, with no more than chance agreement (Kappa= -0.02to 0.26 and -0.20 to 0.17, respectively). [17, 33] Hicks et al reported poor reliability of the judgment of hypomobility with PA glide (Kappa = 0.18), and fair reliability for judgment of any hypermobility with PA glide (Kappa= 0.30).[11] On the other hand, Fritz et al reported fair reliability of judgment of hypomobility with PA glide and moderate reliability for hypermobility judgment (Kappa =0.38 and 0.48, respectively).[9] It is noticeable that the judgment of collapsing the segmental mobility tests into a dichotomous rating, as hyper and hypomobility, improves the test's reliability.

The lack of hypomobility with PA glide is determined to be less than chance agreement (Kappa ranging from -0.22 to 0.18). We believe that inclusion of the normal category added confusion to the already poorly reliable test because this category is at the gray zone between the slightly hypomobility judgment (grade 2) and slightly hypermobility judgment (grade 4). Additionally, we considered the lack of hypomobility glide to be an indirect test because the examiner has to identify the hyper and normal mobility segments. In a case where no segments are judged to have even slight hypomobility dysfunction, the test is considered to be positive. Therefore, it is a test of exclusion of the hypomobility judgment. Moreover, we did not rotate the order of examiners and thus, we cannot rule out the potential influence of the first set of examination performed by the first examiners on the second set of examination performed by the second examiner. We, however, tried to eliminate this influence by allowing at least 15-minutes time delay between the two sets of examinations in order to minimize the potential clinical presentation change due to procedure repetition. This trend of low segmental mobility judgment across the reliability studies indicates the importance of standardizing the amount of pressure/ force applied by all examiners by using a pressure mapping system or similar devices.[17, 33] However, the reliability of identification of the pain provocation segment was found to be moderately reliable (Kappa= 0.4). This finding is in line with the findings of previous studies of pain provocation judgment.[9, 17, 33]

The reliability of aberrant motion test was substantial Kappa = 0.79 which is similar to that found by two previous studies.[17, 25]

The reliability of the PIT was substantially reliable (Kappa= 0.71). This finding is consistent with reports from two previous studies.[9, 25] On the contrary, the finding is lower than that reported by Hicks *et al*.[17] Therefore, the reliability of PIT is substantial to almost perfect, with Kappa ranging from 0.67 to 0.87. [17, 25]

The reliability of average SLR in this study was similar to that found by Rabin *et al* (k= 0.77 and 0.73, respectively).[25] However, Rabin *et al* repeated the test twice before recording the test scores at the third repetition. They stated that they performed the test as described by Hicks *et al* who used a description of the test provided by Waddell *et al*. [7, 34] Both descriptions do not mention the repetition; instead, the examiner is required to record the test's result the first time the test is performed. This helps to avoid any chance of the subjects passing the 91° mark due to the stretch effect produced by repeating the test. Therefore, as much as this study produced similar reliability scores, the procedure that we used eliminated stretch effect. Therefore, the finding of this study is more standardized than the other study.

Additionally, Rabin *et al* [25] reported that the high positive prevalence of CPR in their study could have been due to the younger population that they studied. We believe, in addition to that, the stretching effect of the SLR test increased the chances of subjects passing the >91° mark, thus, caused more subjects to pass the 3 out 4 criteria that was required in order to classify the subjects into the stabilization category. Especially since the reported prevalence of positives for the prone instability test was significantly high in all previous studies. [9, 17, 25]

Table 3: Adjusted Kappa and percent of agreement and disagreement

Instability Categories	Unadjusted Kappa	95% CI for Unadjusted Kappa	Percent Agreement	PABAK Adjusted Kappa	Prevalence Index	Bias Index	Percent of Positive Agreement	Percent of Negative Agreement
Structural	0.19	-0.21, 0.59	85%	0.70	0.80	0.1	92%	25%
Functional	0.72	0.36, 1.09	95%	0.90	0.80	0.0	75%	98%
Combined	0.84	0.55, 1.14	98%	0.95	0.83	0.3	86%	99%

We divided the test of lumbar segmental instability into three categories: Functional instability (neuoromotor control dysfunction), structural instability (disruption of passive stabilizers) and combined instability (involvement of both the neuoromotor control and the passive subsystem).

The percentage of agreement and the Kappa coefficient for the functional instability category was substantial (95%, Kappa= 0.72). However, this result was lower than that found by Rabin *et al* (Kappa=0.86).[25]

The combined instability result was almost perfect (Kappa=0.84). We found that most of subjects who had functional instability also had structural instability, a finding that is consistent with the conventional presumption that younger subjects are more flexible, and are likely to pass both cut-off values for the ROM test; SLR > 91° and lumbar flexion range > 53°. Especially since, passing the SLR cut-off value increases the chance of subject allocation to functional instability. While passing the lumbar flexion ROM cutoff value, directly allocate the subjects into the structural category.

Even though, there was high agreement between the raters (85%), the Kappa value for structural instability was poor (Kappa= 0.19). This phenomena is known as Kappa paradox, because the examiners have more agreement on the subjects who have the condition of interest (positive structural instability, percent of positive agreement= 91%) compared to their agreement on subjects who do not have the condition of interest (negative structural instability, negative percent of agreement = 25%). This imbalance in the percent of agreement between positive and negative ratings skewed the magnitude of Kappa.[35] In addition, there was a high prevalence of positive structural instability, as indicated by the high prevalence index and a high percent of positive agreement. This increased the percent of chance agreement, and thus, reduced the Kappa value.[36]

One way to reduce the skewed influence of prevalence and bias indices is by calculating PABAK or adjusted Kappa.[36, 37] Some statisticians recommended the use of adjusted Kappa to eliminate the adverse effect of prevalence and bias on the true value of Kappa derived from the study. [36] Because of the high prevalence of all lumbar instability categories, we calculated PABAK to find out the true value of Kappa after adjusting the prevalence and bias indices. We found that all the categories rounded up to about 0.18 and 0.11 for functional and combined instability categories, respectively. However, the Kappa value of structural instability dramatically increased by about 0.51 to become substantially reliable (Kappa=0.7). This indicated bigger adverse effects of the prevalence and bias indices on the structural instability category in comparison to the other categories. Thus, the established adjusted Kappa value was more representative of the high observed agreement between the raters.

We recommend that further research efforts should be directed towards establishing clinical examinations that can be used as screening tools for ruling out structural instability among low back pain patients. This is particularly important because the physical therapy profession is moving towards direct access, thus, finding out the cluster of structural instability tests that might predict the radiographic structural gold standard result should be pursued. This can be accomplished by comparing all the highly valid tests to the radiographic gold standard in one comprehensive study.[8, 9]

Furthermore, we observed that the lumbar segmental mobility test is conducted with the patient in a baseline prone position, which is very close to the lumbar end-range (closed pack position). Because of the poor reliability of this test, we agree with previous research opinions that there is need for standardization of the PA palpation by using a pressure/force device prior to the reliability study. This will help to determine its added effect on the reliability studies. In addition, we support the importance of exploring the reliability of other kinds of lumbar mobility testing, such as the side lying lumbar mobility test.[17, 33]

Lastly, we would like to mention some of the limitations of this study. First, the 30 minutes' training session for the examiners was rather short, and may have led to inconsistencies in the performance. However, we expect that the examiners' experience might have helped to counter such inconsistencies. Secondly, we chose to include subjects who had recurrent or chronic LPB with or without leg pain. As such, the study results may not be fit for generalization to other groups of LBP subjects. Lastly, the 95% confidence interval for the Kappa coefficient was noticeably wide and might have affected the Kappa precision.

Conclusion

We studied the inter-rater reliability of six clinical tests that might predict the radiographic diagnostic standard, or the outcome of stabilization therapy in 40 subjects who had R/CLBP. The Kappa correlation coefficient values of the functional instability of lumbar spine "CPR tests" confirmed that these tests are substantially reliable. The lumbar flexion ROM and passive lumbar extension tests were also found to be adequately reliable. On the contrary, lack of hypomobility with PA glide was found to be unreliable, and, in many cases, worse than chance. Thus, relying on this test alone to allocate LSI subjects must be restricted, unless other clinical findings confirm or increase the suspicion of lumbar instability. In other words, it should only be used as a component in a cluster of tests or examination procedures in order to increase its value. However, the sub-classification of patients into lumbar stability categories was adequately reliable, as depicted by their high values of Kappa and adjusted Kappa.

References

1. Cairns, Mindy C., Nadine E. Foster, and Chris Wright. "Randomized controlled trial of specific spinal stabilization exercises and conventional physiotherapy for recurrent low back pain." *Spine* 31, no. 19 (2006): E670-E681.

2. Hauggaard, Annika and Ann L. Persson, "Specific spinal stabilisation exercises in patients with low back pain – a systematic review." *Physical Therapy Reviews* 12, no. 3 (2007): 233-248.

3. Waddell, Gordon, "Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain." *Spine* 12, no.7 (1987): 632-644.

4. Andersson, Gunnar B.J., "Epidemiological features of chronic low-back pain." *Lancet* 354, no. 9178 (1999): 354 (9178): 581-585.

5. Hides, Julie, Gwendolen A Jull, and Carolyn A. Richardson, "Long-term effects of specific stabilizing exercises for first-episode low back pain." *Spine* 26, no.11 (2001): 26(11): 243-248.

6. Demoulin, C., V. Distrée, M. Tomasella, J. Crielaard and M. Vanderthommen, "Lumbar functional instability: a critical appraisal of the literature." *Ann Readapt Med Phys* 50, no. 8 (2007): 677-684, 669-676.

7. Hicks, Gregory, Julie Fritz, Anthony Delitto, Stuart McGill, "Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program." *Archives of Physical Medicine and Rehabilitation* 86, no. 9 (2005): 1753-1762.

8. Alqarni, Abdullah, Anthony Schneiders and Paul Hendrick, "Clinical tests to diagnose lumbar segmental instability: a systematic review." Journal of Orthopedic and Sports Physical Therapy 41, no. 3 (2011): 130-40.

9. Fritz, Julie, Sara Piva and John Childs, "Accuracy of the clinical examination to predict radiographic instability of the lumbar spine." European Spine Journal 14, no. 8 (2005): 743-750.

10. Bystrom, Gustaf, Eva Rasmussen-Barr and Wilhelmus Grooten, "Motor control exercises reduces pain and disability in chronic and recurrent low back pain: a meta-analysis." Spine 38, no. 6 (2013): E350-E358.

11. Panjabi Manohar M, "The stabilizing system of the spine. Part II. Neutral zone and instability hypothesis." Journal of Spinal Disorders and Techniques 5, no. 4 (1992): 390-396.

12. Panjabi Mahonar, "Clinical spinal instability and low back pain." Journal of Electromyography and Kinesiology 13, no. 4 (2003): 371-9.

13. Akuthota, Venu, Andrea Ferreiro, Tamara Moore and Michael Fredericson, “Core stability exercise principles.” Core stability exercise principles 7, no. 1 (2008): 39-44.

14. O'Sullivan, Peter B., Phyty, Dip Manip Grad, Twomey T. Lance and Allison T. Garry, "Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylolysis or spondylolisthesis." *Spine* 22, no. 24 (1997): 2959-2967.

15. Fritz, Julie, Richard E Erhard and Brian F Hagen, “Segmental instability of the lumbar spine.” Physical Therapy 78, no. 8 (1998): 889-896.

16. Leone, Antonio, Giuseppe Guglielmi, Victor N. Cassar-Pullicino and Lorenzo Bonomo, "Lumbar intervertebral instability: a review." *Radiology* 245, no. 1 (2007): 62-77.

17. Hicks, Gregory, Julie Fritz, Anthony Delitto and John Mishock. “Inter-rater reliability of clinical examination measures for identification of lumbar segmental instability.” *Archives of Physical Medicine and Rehabilitation* 84, no. 12 (2004): 1858-1864.

18. Posner, Ira, Augustus A White, Thomas W Edwards, Wilson C Hayes, "A biomechanical analysis of the clinical stability of the lumbar and lumbosacral spine." *Spine* 7, no. 4 (1982): 374-89.

19. Dupuis, PR. "Radiologic diagnosis of degenerative lumbar spinal instability." *Spine* 10 (1985): 262-276.

20. Beazell, James, Melise Mullins and Terry Grindstaff, “Lumbar instability: an evolving and challenging concept.” *Journal of Manual and Manipulative* 18, no. 1 (2010): 9-14.

21. Maigne, Jean-Yves, Eric Lapeyre, Gerald Morvan and Gilles Chatellier, "Pain immediately upon sitting down and relieved by standing up is often associated with radiologic lumbar instability or marked anterior loss of disc space." Spine 28, no.12 (2003): 1327-1334.

22. Abbott, Haxby J., Brendan McCane, Peter Herbison, Graeme Moginie, Cathy Chapple and Tracy Hogarty, "Lumbar segmental instability: a criterion-related validity study of manual therapy assessment. BMC Musculoskeletal Disorders 6 (2005): 56.

23. Kasai, Yuichi, Koichiro Morishita, Eiji Kawakita, Tetsushi Kondo and Atsumasa Uchida, “A new evaluation method for lumbar spinal instability: passive lumbar extension test.” Physical Therapy 86, no. 12 (2006): 1661-7.

24. May, Stephen and Richard Rosedale, “Prescriptive clinical prediction rules in back pain research: a systematic review.” Journal of Manual and Manipulative Therapy 17, no. 1: 36-45.

25. Rabin, Alon, Anat Shashua, Koby Pizem, Gali Dar, "The interrater reliability of physical examination tests that may predict the outcome or suggest the need for lumbar stabilization exercises." Journal of Orthopaedics and Sports Physical Therapy 43, no. 2 (2013): 83-90.

26. Childs, John D., Sara R. Piva and Julie M. Fritz, "Responsiveness of the numeric pain rating scale in patients with low back pain." Spine 30, no. 11(2005): 1331-1334.

27. Fritz, Julie M., and James J. Irrgang, "A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale." *Physical Therapy* 81, no. 2 (2001): 776-788.

28. Waddell, Gordon, Mary Newton, Iain Henderson, Douglas Somerville, Chris J. Main, "A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability." Pain 52, no. 2 (1993): 157-168.

29. Gonnella, Carmella, Stanley V. Paris and Michael Kutner, "Reliability in evaluating passive intervertebral motion." *Physical Therapy* 62, no. 4 (1986): 436-444.

30. Childs, John D. and Joshua A. Cleland, "Development and application of clinical prediction rules to improve decision making in physical therapist practice." *Physical Therapy* 86, no. 1(2006): 122-13.

31. Mayer, T. G., A. F. Tencer, S. Kristoferson, and V. Mooney. "Use of noninvasive techniques for quantification of spinal range-of-motion in normal subjects and chronic low-back dysfunction patients." *Spine* 9, no. 6 (1984): 588.

32. Saur, Petra, Ensink Franz-Bernhard M., Frese Knut, Seeger Dagmar and Hildebrandt Jan, "Lumbar range of motion: reliability and validity of the inclinometer technique in the clinical measurement of trunk flexibility." *Spine* 21, no. 11 (1996): 1332-1338.

33. Schneider, Michael, Richard Erhard, Jennifer Branch, William Tellin, Franks Imbarlina and Anthony Delitto, "Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability study." *Journal of Manipulative and Physiological Therapeutics* 31, no. 6 (2008): 465-473.

34. Waddell, G., D. Somerville, I. Henderson, and M. Newton. "Objective clinical evaluation of physical impairment in chronic low back pain." *Spine* 17, no. 6 (1992): 617.

35. Cicchetti, Domenic V., and Alvan R. Feinstein," High agreement but low kappa: II. Resolving the paradoxes." *Journal of Clinical Epidemiology* 43, no.6 (1990): 551-558.

36. Sim, Julius, and Chris C Wright, "The kappa statistic in reliability studies: use, interpretation, and sample size requirements." Physical Therapy 85, no. 3 (2005): 257-268.

37. Byrt, Ted, Janet Bishop, and John B. Carlin, "Bias, prevalence and kappa." Journal of Clinical Epidemiology 46, no. 5 (1993): 423-429.

CHAPTER TWO

LITRATURE REVIEW

Low back pain (LBP) is a common condition that affects 80% of the general population at some point in life, with an estimated cost of about $15 to $50 billion per year in the USA. [1] Approximately 80-90% of the affected patients recover spontaneously within 6 weeks.[2, 3] However, the recurrence rate after the acute episode is estimated to range from 60% to 86%, which is very high.[4-6] Lumbar segmental instability is believed to be one of the main causes of the high recurrence rates.[3, 7, 8]

Although low back pain due to mechanical causes is a short-lived self-limited condition, it is complicated by a high recurrence rate, pain, and disability in the long-run.[5, 6, 9] Studies have shown that lumbar stabilization exercises reduce the recurrence rate and disability of LBP. They also comprise the second best non-invasive treatment option after spinal manual therapy all time periods; short-term, intermediate and long-term. [7, 10]

The most widely used definition for clinical instability of the spine is Panjabi's definition. He defined it as "a significant decrease in the spine's stabilizing system's capacity to maintain the intervertebral neutral zones within the physiological limits, so that there is no neurological dysfunction, major deformity, or incapacitating pain." [1, 11, 12] In order to understand the instability definition, two terminologies must be clarified, the "stabilizing system" and its parts and the "neutral zone".

The Spinal Stability System

The spinal stability system can be divided into three subsystems:

1- The Passive Subsystem (spinal column), which consists of: vertebrae, intervertebral discs, facet joints, joints' capsules and spinal ligaments. Below are the main characteristics of this subsystem:

- It provides passive resistance at the end of range.
- In vitro studies have shown that the spinal passive system has load-carrying capacities of less than 90 N (about 20 Ibs). [1, 8, 13]
- It possesses inherent joint stiffness, particularly at the end range of motion. [8]

2- The Active Subsystem (spinal muscles), consists of: the musculotendinous units that are attached to, or influence, the spinal column. [14] They can be divided to:

A- Slow-twitch fibers (local muscle system) which includes transversus abdominis, multifidui, internal oblique, deep transversospinals, and the pelvic floor muscles. These muscles are suited for inter-segmental motion control, and they respond to posture and extrinsic loads. [12, 15]

B- Fast-twitch fibers (global muscle system) which includes erector spinae, external oblique, rectus abdominis and quadrates lumborum. [8, 12, 16] These muscles possess large lever arms, so they can generate large amounts of torque and control trunk movements. [12, 15] Thefollowing are the main characteristics of this system:

- It provides mechanical stability to the spinal column. [1]
- In vivo, those muscles can provide mechanical stability for loads exceeding 1500 N. [1, 8]

3- Neuromuscular Subsystem (the control unit), which consists of sensory receptors in the spinal structures, their central connections, and cortical and subcortical control centers. [14] The main characteristics of this subsystem are:

- It centralizes the proprioceptive afferent neurons from several mechanoreceptors present in passive (ligaments, intervertebral disc, and joint capsules) and active (muscles) structures. It subsequently computes the required body movement, coordinates the incoming efferent signals, and generates efferent signals that activate the stabilizer muscles depending on the required mechanical stability.
- It also allows for a compromise between the respective requirements for, stability and mobility.
- In the absence of external loads the poor neuromuscular control may explain the recurrence of acute LBP." [8]

The Neutral Zone

The Neutral Zone (NZ) is defined by Panjabi as "the part of the physiological intervertebral motion, measured from the neutral position, within which the spinal motion is produced with a minimal internal resistance; it is the zone of high flexibility or laxity." [1] It is different from the Elastic Zone (EZ) that Panjabi defined as "the part of physiological intervertebral motion, measured from the end of the neutral zone up to the physiological limit." [1] Within the EZ, spinal motion is produced against a significant internal resistance; it is the zone of high stiffness" [1, 8]

Studies have been shown that spinal injury can increase both the Neutral Zone and the range of motion (ROM) of the spine. However, increases in the neutral zone were larger than the corresponding increases in ROM. This indicates that the NZ is more sensitive and significant than ROM in reflecting spinal instability. [8, 11] On the other hand, significantly decreasing the NZ increase spinal stability. This cab e achieved by training the spinal stabilizing muscles or using an external fixator. [1, 7, 8, 11, 13]

These three subsystems are functionally interdependent in maintaining spinal stability and intervertebral motion. The whole subsystem may dysfunction in case of compromise to any of the subsystems. For example, an injury or breakdown in the passive subsystem, such as a fracture, disc herniation or degeneration, may decrease the inherent stability of the spine and alter segmental motion patterns. In such a case, enhancement of the neural and active subsystems may compensate for this loss and partially restore stability. [13, 14]

Previous expert reviews and systematic reviews [8, 17, 18] suggested subdivision of the LSI into structural and functional segmental instability, where structural (radiographic) instability refers to disruption of passive stabilizers and decreased structural integrity. This concept was first proposed by Knutson. [19] He proposed the use of a flexion-extension radiograph to identify and quantify abnormal anterior-to-posterior translation of the motion segment at the end range of spinal flexion and extension. Therefore, theoretically, it only would detect the dysfunction of passive stabilizing subsystems, which include the inert structures of the spine such as the vertebral bodies, zygoapophyseal joints, joint capsules and spinal ligaments. [17, 19] Functional Instability, on the other hand, is defined as a lack of neuromuscular control of the joint during daily living activity at the middle range of motion. [8, 17, 18] Using the structural instability gold standard (flexion-extension X-ray), which measures the instability at the end range, was not considered to be a valid measure for diagnosing functional instability that occurs at midrange.

On the other hand, the gold standard for functional instability around the neutral position was lacking. [8, 17, 20] Therefore, Hicks et al. investigated clinical tests that might predict the successful outcomes of the stabilization exercises that were developed to improve spinal motor control (stiffness) around the neutral position of the spine.[7, 10, 21] They devised four predictors, which include age <40 years; positive prone instability test; presence of aberrant movement and an average straight leg raise > 91° If three out of four of these predictors of stabilization are present, clinical prediction rules (CPR) that the positive likelihood of success increase four times (+LR = 4).[21] the latest systematic review of all spinal CPRs confirm its accuracy.

This research identified four predictors that together form the clinical prediction rule (CPR) of stabilization exercise.[8] Although the latest systematic review of the spinal CPRs supports its accuracy by providing level 4 evidence, its reliability in patients with R/CLBP is yet to be established. [22]

The reported prevalence of low back pain due to potential functional instability is about 33% for patients with potential functional instability,[7] compared with 57% for patients with evidence of structural instability indicated by positive flexion-extension X-rays.[8, 9]

The purpose of this study was to investigate the inter-rater reliability of the most valid clinical test. Such reliability will be indicated by the highest positive likelihood ratio against either the radiographic reference standard of structural instability or against successful stabilization outcomes (stabilization CPR) for functional instability. This study also investigates the inter-rater reliability of the sub-classification of lumbar segmental instability in to structural, functional, and combined instability.

The literature search conducted during this study yielded three tests that matched the selected criteria for structural instability, which had the highest +LR against the radiographic gold standard. They include: 1) passive lumbar extension test (PLET) with +LR = 8.84, [23] 2) lumbar flexion ROM >53°; and 3) the lack of hypomobility with PA glide test. The positive likelihood ratio of the combination of the second and third tests is 12.8.[24]. Conversely, three functional clinical tests that can predict the successful stabilization outcomes were identified with a combined +LR of 4. These functional instability tests include: 1) the prone instability test (PIT), 2) the aberrant motion test, 3) the average straight leg raising (SLR) test (>91°). [21]

References

1. Panjabi Mahonar, "Clinical spinal instability and low back pain." *Journal of Electromyography and Kinesiology* 13, no. 4 (2003): 371-9.

2. Cairns, Melinda, Nadine Foster and Chris Wright, "Randomized controlled trial of specific spinal stabilization exercises and conventional physiotherapy for recurrent low back pain." *Spine* 31, no. 19 (2006): 670-681.

3. Hauggaard, Annika and Ann Persson, "Specific spinal stabilisation exercises in patients with low back pain – a systematic review." *Physical Therapy Reviews* 12, no.3 (2007): 233-248.

4. Suni, Jaana, Rinne Marjo, Natri Antero, Statistisian Matti Pasanen, Parkkari Jari and Alaranta Hannu, "Control of the lumbar neutral zone decreases low back pain and improves self-evaluated work ability: a 12-month randomized controlled study." *Spine* 31, no. 18 (2006): 611-620.

5. Hestbaek, Lisse, Charlotte Leboeuf-Yde, and Claus Manniche, "Low back pain: what is the long-term course? A review of studies of general patient populations." *European Spine Journal* 12, no 2 (2003): 149-165.

6. Pengel, LisEt, Robert D Herbet, Chris G Maher and Kathryn M Refshauge, "Acute low back pain: systematic review of its prognosis." *British Medical Journal* 327, no. 7410 (2003): 323.

7. Hides, Julie, Gwendolen A Jull, and Carolyn A. Richardson, "Long-term effects of specific stabilizing exercises for first-episode low back pain." *Spine* 26, no.11 (2001): 26(11): 243-248.

8. Demoulin, Christophe, V. Distrée, M. Tomasella, J. Crielaard and M. Vanderthommen, "Lumbar functional instability: a critical appraisal of the literature." Ann Readapt Med Phys 50, no. 8 (2007): 677-684, 669-676.

9. Waddell, Gordon, "Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain." *Spine* 12, no.7 (1987): 632-644.

10. Bystrom, Gustaf, Eva Rasmussen-Barr and Wilhelmus Grooten, "Motor control exercises reduces pain and disability in chronic and recurrent low back pain: a meta-analysis." *Spine* 38, no. 6 (2013): 350-358.

11. Panjabi Manohar, "The stabilizing system of the spine. Part II. Neutral zone and instability hypothesis." *Journal of Spinal Disorders and Techniques* 5, no. 4 (1992): 390-39.

12. Akuthota, Venu, Andrea Ferreiro, Tamara Moore and Michael Fredericson, "Core stability exercise principles." *Core stability exercise principles* 7, no. 1 (2008): 39-44.

13. McGill M Stuart, "Low back stability: from formal description to issues for performance and rehabilitation" *Exercise and Sport Sciences Reviews* 29, no. 1 (2001): 26-31.

14. Standaert, Christopher, Stuart Weinstein, and John Rumpeltes, "Evidence-informed management of chronic low back pain with lumbar stabilization exercises." *Spine Journal* 8, no.1 (2008):114-120.

15. O'Sullivan Peter, "Lumbar segmental 'instability': clinical presentation and specific stabilizing exercise management." *Manual Therapy* 5, no. 1 (2002): 2-12.

16. Ebenbichler, Gerold, Oddsson Lars, Kollmitzer, Josef and Erim Zeynep, "Sensory-motor control of the lower back: implications for rehabilitation." *Medicine and Science in Sports and Exercise* 33, no.11 (2001): 1889-1898.

17. Alqarni, Abdullah, Anthony Schneiders and Paul Hendrick, "Clinical tests to diagnose lumbar segmental instability: a systematic review." *Journal of Orthopedic and Sports Physical Therapy* 41, no. 3 (2011): 130-40.

18. Beazell, James, Melise Mullins and Terry Grindstaff, "Lumbar instability: an evolving and challenging concept." *Journal of Manual and Manipulative* 18, no. 1 (2010): 9-14.

19. Fritz, Julie, Richard E Erhard and Brian F Hagen, "Segmental instability of the lumbar spine." *Physical Therapy* 78, no. 8 (1998): 889-896.

20. Hicks, Gregory, Julie Fritz, Anthony Delitto and John Mishock. "Inter-rater reliability of clinical examination measures for identification of lumbar segmental instability." *Archives of Physical Medicine and Rehabilitation* 84, no. 12 (2004): 1858-1864.

21. Hicks, Gregory, Julie Fritz, Anthony Delitto, Stuart McGill, "Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program." *Archives of Physical Medicine and Rehabilitation* 86, no. 9 (2005): 1753-1762.

22. May, Stephen and Richard Rosedale, "Prescriptive clinical prediction rules in back pain research: a systematic review." *Journal of Manual and Manipulative Therapy* 17, no. 1: 36-45.

23. Kasai, Yuichi, Koichiro Morishita, Eiji Kawakita, Tetsushi Kondo and Atsumasa Uchida, "A new evaluation method for lumbar spinal instability: passive lumbar extension test." *Physical Therapy* 86, no. 12 (2006): 1661-7.

24. Fritz, Julie, Sara Piva and John Childs, "Accuracy of the clinical examination to predict radiographic instability of the lumbar spine." *European Spine Journal* 14, no. 8 (2005): 743-50.

CHAPTER THREE

JORNAL MANUSCRIPT CHAPTER

The Interrater Reliability of Six Clinical Tests that Best Predict the Structural Instability or Successful Stabilization Exercise Outcomes Due to Functional Instability

Faisal M. Alyazedi, DS.c, PT, Everett B. Lohman, DS.c, PT, R. Wesley Swen, DPT, PT, Khaled Bahjri, MD, MPH , J Man Manip Ther ; submitted for publication.

Abstract

Objectives: This study investigated the interrater reliability of three structural end-range lumbar segmental instability tests with the highest positive likelihood ratio against flexion-extension radiographs, and three functional mid-range clinical tests that predict the success of lumbar stabilization exercises in patients with recurrent or chronic low back pain (R/CLBP). It also investigated the reliability of lumbar segmental instability subclassification as: Functional, Structural, and Combined Instability.

Method: 40 adult with R/CLBP patients (30 men and 10 women), 18 to 80 years of age, underwent repeated measurements of specific clinical tests for structural or functional lumbar segmental instability.

Results: Other than the lack of hypomobility with PA glide test, which was found to be unreliable (percentage agreement = 37.5, and k= -0.02), all the other tests demonstrated high Kappa coefficients and percentage agreements. The subclassification categories of lumbar segmental instability (functional, structural, and combined) were found to be significantly reliable (k = 0.722, adjusted k = 0.7, and k = 0.84, respectively).

Discussion: All the investigated tests (except lack of hypomobility with PA glide test), as well as the categories of lumbar segmental instability subclassification, are significantly reliable in predicting lumbar stabilization.

Key words: Low back pain; Segmental instability; Reliability; Physical examination; Clinical prediction rule

Introduction

Low back pain (LBP) is a common musculoskeletal condition that affects up to 80% of the general population during their lifetime.[1-3] The reported recurrence rate of LBP is high; about 73% in 12 months.[38, 39] Lumbar segmental instability is believed to be one of the main causes of the high recurrence rates.[2, 5, 6] The estimated prevalence of LBP due to lumbar segmental instability is about 33% for patients with potential functional instability,[7] compared with 57% for patients with evidence of structural instability indicated by positive flexion-extension X-rays.[8, 9]

Panjabi [11, 12] hypothesized that in order for the spine to move safely, the passive osseoligamentous subsystem, the active musculotendinous subsystem, and the neural subsystems have to act interdependently to maintain spinal mechanical stability. The passive subsystem resists spinal instability at *end-range* of motion.[11, 12] Conversely, the active subsystem provides dynamic stability around *the neutral position* of a spinal segment.[11, 13] The neural control subsystem activates the muscular subsystem to provide dynamic stability.[6] Panjabi suggested that the loss of passive subsystem integrity might lead to segmental instability unless the neuromuscular subsystem compensates for that loss.[11, 12, 15, 20, 40]

In 1944, Knutson [15] recommended the use of flexion-extension radiographs to identify and quantify abnormal anterior-to-posterior translation of the motion segment at the *end-range* of spinal flexion and extension. This imaging modality has become the diagnostic standard of structural lumbar segmental instability or instability due to disruption of passive stabilizers.[8, 15-20] Conversely, the diagnostic standard that can quantify the functional instability *around the neutral position* was lacking.[6, 8, 17] A number of studies attributed this functional instability to the lack of neuromuscular control of the joint during daily living activities.[6, 8, 11, 20, 40] Some physical-therapy researchers studied the clinical tests that might predict the success of the stabilization exercises that were developed to improve spinal motor control (stiffness) around the neutral position of the spine.[5, 7, 10] These researchers came up with four predictors that together form the clinical prediction rule (CPR) of stabilization exercise.[7] Even though the latest systematic review of the spinal CPRs supports its validity by providing level 4 evidence,[24] its reliability in patients with R/CLBP has yet to be established.

Because it is important to establish the reliability of the most accurate tests, this study investigated the inter-rater reliability of six clinical lumbar instability tests that showed the highest positive likelihood ratio (+LR) in predicting either the findings of the flexion-extension radiograph or the treatment outcome for recurrent or chronic LBP patients. The clinical tests consisted of 1) the prone instability test (PIT), 2) the aberrant motion test, 3) the average straight leg raising (SLR) test (>91°), 4) lumbar flexion ROM >53°; 5) the lack of hypomobility with PA glide test, and 6) the passive lumbar extension test (PLET).[7, 9, 23]

Recently, a group of researchers investigated the reliability of CPR and the passive lumbar extension test in the general population of LBP subjects.[25] To further explore the reliability of CPRs, this study investigates the reliability of CPR criterion, as well as the reliability of the three tests that showed the highest positive likelihood ration to predict the findings of the radiographic structural instability in recurrent or chronic LBP patients.[9, 23] This study also explored the inter-rater reliability of examiners to subclassify lumbar instability subjects into different lumbar instability categories: structural, functional, and combined. To the best of our knowledge, this is the first clinical study that subclassifies lumbar segmental instability subjects into different categories. Figure 1.

Methods

Study Participants

Forty subjects from the San Bernardino community who had low back pain with or without leg pain, age range 31-71, participated in the study

Inclusion and Exclusion Criteria

The inclusion criteria for this study consisted of 1) having a new episode of LBP and 2) having experienced a similar episode of LBP before; whereby the first episode of back pain ever experienced was at least three months before the date of recruitment, or 3) currently experiencing persistent LBP for at least a three-month duration.[22]

Conversely, the exclusion criteria consisted of 1) having undergone previous spinal fusion surgery, 2) history of traumatic fracture of the spine that resulted in permanent neurological deficit, 3) scoliosis greater than 20°, 4) pregnancy, 5) inability to actively flex and extend the spine adequately to permit an assessment of segmental motion due to pain or muscle spasm, and 6) medical "red flags" such as cauda equine syndrome, tumor, and systemic inflammatory conditions.

Ethical Issues

Participation in this study was voluntary. All participants were briefed about the study and were provided with a copy of a consent form approved by the Loma Linda

University Institute of Review Board. Participants reviewed and signed the form accordingly.

Examiners

Three physical-therapy examiners were involved in the study. They all received 30-minute trainings regarding all written and clinical procedures of the study. One of them, who had 13 years of musculoskeletal clinical experience, recorded the baseline data of the participants. The role of the remaining two examiners involved performing and interpreting the various clinical tests on all subjects. These examiners had more than 20 years of musculoskeletal clinical experience each.

Data Collection

The physical therapist who collected all baseline data, which consisted of: informed consent, demographic information, self-reported history, and self-reported outcome measures. Then the two clinical examiners, who were blinded to each other's test results, performed the clinical tests and determined the test results for each subject.

Each participant completed three self-reported outcome questionnaires. These included the Numeric Pain Rating Scales (NPRS), the Modified Oswestry Low Back Pain Disability Questionnaire (OSW), and the Fear Avoidance Beliefs Questionnaire (FABQ).[26-28]

The NPRS was used to assess the severity of LBP by using an 11-point (0–10) scale. The OSW has 10 sections: one section for pain severity and the other nine representing various functional activities.[27] This questionnaire indicates the degree of

LBP-attributed limitation in the specified activities. The FABQ assesses the level of fear-avoidance beliefs associated with LBP.[28] It consists of four items on physical activity (FABQ-PA) and seven items on the scale of work (FABQ-W).

After filling out the assessment tools, each participant underwent a specific musculoskeletal examination as listed in Table 4.

After performing all of the clinical tests, each examiner classified the subjects into one of the three instability subcategories (structural, functional, or combined instability). The subjects were classified as structurally unstable if the subject tested positive for any of the following: passive lumbar extension test,[23] lumbar flexion ROM (>53°) test, or lack of hypomobility with PA glide.[9] The study shows that if the first test is positive, then the subject is about 9 times more likely to have positive radiographic instability, compared with about 13 times if second and third tests are positive.[8, 9, 20, 23]

The subject was considered functionally unstable if three out of four predictors of functional instability (CPR) were present: 1) age < 40 years, 2) positive prone instability test (PIT), 3) aberrant motion present, and 4) average SLR (>91°). If three out of four predictors are present, then the likelihood of success with lumbar stabilization exercises is LR 4.0.[7]

The subject was considered to have combined instability if the subject had both subcategories (structural and functional instability).

Data Analysis

Data were analyzed using the Statistical Package for Social Sciences (SPSS IBM Corporation 1989, 2011; Version 20).

The demographic information was illustrated in Table 1.

The interrater reliability for the various lumbar instability tests was evaluated using Kappa correlation coefficients in order to establish the interrater reliability that is above chance agreement. The percentages of agreement for the various tests were also reported. Alpha was set at the level of 0.05. This is shown in Table 2.

The reliability of the categories was analyzed by Kappa for functional and combined lumbar instability categories. For structural instability, adjusted Kappa and the percent of agreements were calculated to adjust the effect of prevalence and bias indices on the Kappa. This is shown in Table 3.

Results

The percentage agreement and k value for all the clinical prediction rule tests; PIT, aberrant motion test, and average SLR (>91°) test showed a high percentage agreement (90.5, 97.5, and 95, respectively) and showed substantial Kappa coefficients (0.73, 0.78 and 0.77, respectively).

The lumbar flexion ROM >53° and lumbar extension test showed a fairly high percentage agreement (82.5 and 72.5, respectively) and moderate Kappa coefficients (0.48 and 0.46, respectively).

The lack of hypomobility with PA glide test is found to be poorly reliable with a low percentage agreement and a low Kappa coefficient (38%, 0.02, respectively). Further examination of the data revealed that the 42% of mobility judgment disagreement between the raters was due to inconsistence regarding the grading of normal (grade 3) and slightly restricted mobility (grade 2) across all lumbar segments. However, locating the painful segment was moderately reliable with Kappa = 0.41.

The functional and combined lumbar segmental instability subcategory was found to be substantially reliable (Kappa = 0.72 and 0.84 respectively), while the adjusted Kappa for structure instability was also substantial reliability (PABAK = 0.7).

Discussion

We investigated the consistence of the most valid tests (highest + LR) in the literature to identify lumbar segmental instability subjects who have structural or functional instability.

We chose three tests that showed the highest positive likelihood ratio against the instability radiographic gold standard: flexion-extension X-ray. The selected tests included the lumbar extension test (+LR 8.84), the combination of both tests; the lack of hypomobility with PA glide test, and lumbar flexion ROM >53° (+LR 12.8).[8, 9, 20, 23] Also, three tests that predict the treatment outcome results of the stabilization exercise include: average SLR (>91°), aberrant motion test, and prone instability test +LR 4.0.[7, 24, 30]

PLET was validated by Kasai et al. for elderly subjects (age range = 39 to 88 years, mean = 68.9 years) who had experienced chronic pathologies such as lumbar stenosis, lumbar spondylolisthesis, and lumbar degenerative scoliosis.[23] In this study, we included younger subjects ranging from 21 to 71 years of age, median age = 31, who have recurrent or chronic LBP. However, this test shows acceptable inter-rater reliability even for the younger age group (with Kappa = 0.46). Recently, Rabin et al. investigated the reliability of the PLE test on general LBP subjects and found it to be substantially reliable (Kappa = 0.76).[25]

The lumbar flexion range > 53° test was found to be moderately reliable (Kappa = 0.46), with a high percentage of agreement (82.5). This finding is in agreement with the findings of previous studies, which also showed a high correlation between lumbar flexion range, the flexion-extension radiograph, and the functional radiograph.[31, 32] It is noticeable that this test replicates the first portion (lumbar flexion) of the radiographic procedure in a standing position.

Two previous studies reported low reliability of the segmental mobility test in prone position, with no more than chance agreement (Kappa = -0.02 to 0.26 and -0.20 to 0.17, respectively).[17, 33] Hicks et al. reported poor reliability of the judgment of hypomobility with PA glide (Kappa = 0.18), and fair reliability for judgment of any hypermobility with PA glide (Kappa = 0.30). [11] Conversely, Fritz et al. reported fair reliability of judgment of hypomobility with PA glide and moderate reliability for hypermobility judgment (Kappa = 0.38 and 0.48, respectively).[9]

The lack of hypomobility with PA glide is determined to be less than chance agreement (Kappa ranging from - 0.22 to 0.18). We believe that inclusion of the normal category added confusion to the already poorly reliable test because this category is at the gray zone between the slightly hypomobility judgment (grade 2) and slightly hypermobility judgment (grade 4). Additionally, we considered the lack of hypomobility glide to be an indirect test because the examiner has to identify the hyper- and normal-mobility segments. In a case where no segments are judged to have even slight hypomobility dysfunction, the test is considered to be positive. Therefore, it is a test of exclusion of the hypomobility judgment. Moreover, we did not rotate the order of examiners; however, we allowed at least 15 minutes' time delay between the two sets of examinations in order to minimize the potential clinical presentation change due to procedure repetition.

The reliability of identification of the pain provocation segment was found to be moderately reliable (Kappa = 0.4). This finding is in line with the findings of previous studies of pain provocation judgment.[9, 17, 33]

The reliability of the aberrant motion test was substantial (Kappa = 0.79), which is similar to that found by two previous studies.[17, 25]

The reliability of the PIT was substantially reliable (Kappa = 0.71). This finding is consistent with reports from two previous studies.[9, 25]

The reliability of average SLR in this study was similar to that found by Rabin et al. (Kappa = 0.77 and 0.73, respectively).[25] However, Rabin et al. repeated the test twice before recording the test scores at the third iteration. They stated that they performed the test as described by Hicks et al., who used a description of the test provided by Waddell et al. [7, 34] Neither description mentions the repetition; instead, the examiner is required to record the test's result the first time the test is performed. This helps to avoid any chance of the subjects passing the 91° mark due to the stretch effect produced by repeating the test. Therefore, we believe, in addition to the younger subjects they had, the stretching effect of the SLR test increased the chances of subjects passing the >91° mark, thereby passing the three-out-of-four criteria, which was required in order to classify the subjects into the stabilization category. Especially since the reported prevalence of positives for the prone instability test was significantly high in all previous studies.[9, 17, 25]

We divided the test of lumbar segmental instability into three categories: functional instability (neuoromotor control dysfunction), structural instability (disruption of passive stabilizers), and combined instability (involvement of both the neuoromotor control and the passive subsystem).

The percentage of agreement and the Kappa coefficient for the functional instability category was substantial (95%, Kappa = 0.72). However, this result was lower than that found by Rabin et al. (Kappa = 0.86).

The combined instability result was almost perfect (Kappa = 0.84). We found that most subjects who had functional instability also had structural instability, a finding that is consistent with the conventional presumption that young and flexible subjects are likely to pass both cut-off values for the ROM test: SLR > 91° and lumbar flexion range > 53°. Especially since, passing the SLR cut-off value increases the chance of subject allocation to functional instability. While passing the lumbar flexion ROM cutoff value directly allocates the subjects into the structural category.

Even though there was high agreement between the raters (85%), the Kappa value for structural instability was poor (Kappa = 0.19). This phenomenon is known as the Kappa paradox, because the examiners have more agreement on the subjects who have the condition of interest (positive structural instability, percent of positive agreement = 91%) than on subjects who do not have the condition of interest (negative structural instability, negative percent of agreement = 25%). This imbalance in the percent of agreement between positive and negative ratings skewed the magnitude of Kappa.[35] In addition, there was a high prevalence of positive structural instability, as indicated by the high prevalence index and a high percentage of positive agreement. This increased the percentage of chance agreement, and thus, reduced the Kappa value.[36]

One way to reduce the skewed influence of prevalence and bias indices is by calculating PABAK or adjusted Kappa.[36, 37] Some statisticians recommended the use of adjusted Kappa to eliminate the adverse effect of prevalence and bias on the true value of Kappa derived from the study.[36] Because of the high prevalence of all lumbar instability categories, we calculated PABAK to find out the true value of Kappa after adjusting the prevalence and bias indices. We found that all of the categories rounded up to about 0.18 and 0.11 for functional and combined instability categories, respectively. However, the Kappa value of structural instability dramatically increased by about 0.51 to become substantially reliable (Kappa = 0.7). This indicated larger adverse effects of the prevalence and bias indices on the structural instability category than in the other categories. Thus, the established adjusted Kappa value was more representative of the high observed agreement between the raters.

We recommend that further research efforts should be directed towards establishing the cluster of structural insanity tests that can be used as screening tools for ruling out structural instability among low back pain patients. This can be accomplished by comparing all the highly valid tests [9, 10] with the radiographic gold standard in one comprehensive study.

Furthermore, because of the poor reliability of lumbar mobility test at the prone position, we agree with previous research findings that recommend exploring the added effect of using a pressure/force device prior to the reliability study. In addition, we support exploring the reliability of other kinds of lumbar mobility testing, such as the side-lying lumbar mobility test.[17, 33]

Lastly, we would like to mention some of the limitations of this study. First, the 30-minute training session for the examiners was rather short and may have led to inconsistencies in the performance. Moreover, the 95% confidence interval for the Kappa coefficient was noticeably wide and might have affected the Kappa precision.

Conclusion

We studied the inter-rater reliability of six clinical tests that might predict the radiographic diagnostic standard, or the outcome of stabilization therapy in 40 subjects who had R/CLBP. The Kappa correlation coefficient values of the functional instability of lumbar spine confirmed that these tests are substantially reliable. The lumbar flexion ROM and passive lumbar extension tests were also found to be adequately reliable. Conversely, lack of hypomobility with PA glide was found to be unreliable, and, in many cases, worse than chance. Finally, the subclassification of patients into lumbar stability categories was adequately reliable, as depicted by their high values of Kappa and adjusted Kappa.

References

1. Cairns, M.C., N.E. Foster, and C. Wright, *Randomized controlled trial of specific spinal stabilization exercises and conventional physiotherapy for recurrent low back pain.* Spine (Phila Pa 1976), 2006. **31**(19): p. E670-81.

2. Hauggaard, A. and A.L. Persson, *Specific spinal stabilisation exercises in patients with low back pain – a systematic review.* Physical Therapy Reviews, 2007. **12**(3): p. 233-248.

3. Waddell, G., *1987 Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain.* Spine (Phila Pa 1976), 1987. **12**(7): p. 632-44.

4. Hestbaek, L., C. Leboeuf-Yde, and C. Manniche, *Low back pain: what is the long-term course? A review of studies of general patient populations.* Eur Spine J, 2003. **12**(2): p. 149-65.

5. Pengel, L.H., et al., *Acute low back pain: systematic review of its prognosis.* Bmj, 2003. **327**(7410): p. 323.

6. Hides, J.A., G.A. Jull, and C.A. Richardson, *Long-term effects of specific stabilizing exercises for first-episode low back pain.* Spine (Phila Pa 1976), 2001. **26**(11): p. E243-8.

7. Demoulin, C., et al., *Lumbar functional instability: a critical appraisal of the literature.* Ann Readapt Med Phys, 2007. **50**(8): p. 677-84, 669-76.

8. Hicks, G.E., et al., *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**(9): p. 1753-62.

9. Alqarni, A.M., A.G. Schneiders, and P.A. Hendrick, *Clinical tests to diagnose lumbar segmental instability: a systematic review.* J Orthop Sports Phys Ther, 2011. **41**(3): p. 130-40.

10. Fritz, J.M., S.R. Piva, and J.D. Childs, *Accuracy of the clinical examination to predict radiographic instability of the lumbar spine.* Eur Spine J, 2005. **14**(8): p. 743-50.

11. Panjabi, M.M., *The stabilizing system of the spine. Part II. Neutral zone and instability hypothesis.* J Spinal Disord, 1992. **5**(4): p. 390-6; discussion 397.

12. Panjabi, M.M., *The stabilizing system of the spine. Part I. Function, dysfunction, adaptation, and enhancement.* J Spinal Disord, 1992. **5**(4): p. 383-9; discussion 397.

13. Akuthota, V., et al., *Core stability exercise principles.* Curr Sports Med Rep, 2008. **7**(1): p. 39-44.

14. Beazell, J.R., M. Mullins, and T.L. Grindstaff, *Lumbar instability: an evolving and challenging concept.* J Man Manip Ther, 2010. **18**(1): p. 9-14.

15. Panjabi, M.M., *The stabilizing system of the spine. Part I. Function, dysfunction, adaptation, and enhancement.* J Spinal Disord, 1992. **5**(4): p. 383-9; discussion 397.

16. Fritz, J.M., R.E. Erhard, and B.F. Hagen, *Segmental instability of the lumbar spine.* Phys Ther, 1998. **78**(8): p. 889-96.

17. Leone, A., et al., *Lumbar intervertebral instability: a review.* Radiology, 2007. **245**(1): p. 62-77.

18. Hicks, G.E., et al., *Interrater reliability of clinical examination measures for identification of lumbar segmental instability.* Arch Phys Med Rehabil, 2003. **84**(12): p. 1858-64.

19. Posner, I., et al., *A biomechanical analysis of the clinical stability of the lumbar and lumbosacral spine.* Spine (Phila Pa 1976), 1982. **7**(4): p. 374-89.

20. Dupuis, P.R., et al., *Radiologic diagnosis of degenerative lumbar spinal instability.* Spine (Phila Pa 1976), 1985. **10**(3): p. 262-76.

21. Bystrom, M.G., E. Rasmussen-Barr, and W.J. Grooten, *Motor control exercises reduces pain and disability in chronic and recurrent low back pain: a meta-analysis.* Spine (Phila Pa 1976), 2013. **38**(6): p. E350-8.

22. May, S. and R. Rosedale, *Prescriptive clinical prediction rules in back pain research: a systematic review.* J Man Manip Ther, 2009. **17**(1): p. 36-45.

23. Kasai, Y., et al., *A new evaluation method for lumbar spinal instability: passive lumbar extension test.* Phys Ther, 2006. **86**(12): p. 1661-7.

24. Rabin, A., et al., *The interrater reliability of physical examination tests that may predict the outcome or suggest the need for lumbar stabilization exercises.* J Orthop Sports Phys Ther, 2013. **43**(2): p. 83-90.

25. Abbott, J.H., et al., *Lumbar segmental instability: a criterion-related validity study of manual therapy assessment.* BMC Musculoskelet Disord, 2005. **6**: p. 56.

26. Childs, J.D., S.R. Piva, and J.M. Fritz, *Responsiveness of the numeric pain rating scale in patients with low back pain.* Spine (Phila Pa 1976), 2005. **30**(11): p. 1331-4.

27. Fritz, J.M. and J.J. Irrgang, *A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale.* Phys Ther, 2001. **81**(2): p. 776-88.

28. Waddell, G., et al., *A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability.* Pain, 1993. **52**(2): p. 157-68.

29. Childs, J.D. and J.A. Cleland, *Development and application of clinical prediction rules to improve decision making in physical therapist practice.* Phys Ther, 2006. **86**(1): p. 122-31.

30. Mayer, T.G., et al., *Use of noninvasive techniques for quantification of spinal range-of-motion in normal subjects and chronic low-back dysfunction patients.* Spine (Phila Pa 1976), 1984. **9**(6): p. 588-95.

31. Saur, P.M., et al., *Lumbar range of motion: reliability and validity of the inclinometer technique in the clinical measurement of trunk flexibility.* Spine (Phila Pa 1976), 1996. **21**(11): p. 1332-8.

32. Schneider, M., et al., *Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability study.* J Manipulative Physiol Ther, 2008. **31**(6): p. 465-73.

33. Waddell, G., et al., *Objective clinical evaluation of physical impairment in chronic low back pain.* Spine (Phila Pa 1976), 1992. **17**(6): p. 617-28.

34. Cicchetti, D.V. and A.R. Feinstein, *High agreement but low kappa: II. Resolving the paradoxes.* J Clin Epidemiol, 1990. **43**(6): p. 551-8.

35. Sim, J. and C.C. Wright, *The kappa statistic in reliability studies: use, interpretation, and sample size requirements.* Phys Ther, 2005. **85**(3): p. 257-68.

36. Byrt, T., J. Bishop, and J.B. Carlin, *Bias, prevalence and kappa.* J Clin Epidemiol, 1993. **46**(5): p. 423-9.

CHAPTER FOUR

CONCLUSION

This study evaluated the inter-rater reliability of six clinical tests that best predict structural lumbar segmental instability, or the functional lumbar segmental instability, in 40 subjects who had recurrent and or chronic Low Back Pain (R/CLBP). It also explored the reliability of the sub-classification of lumbar segmental instability into the functional, structural and combined categories.

The inter-rater reliability coefficient for all the tests that determine the functional instability category were determined to be substantially reliable. Notably, the inter-rater reliability coefficient of two structural clinical tests that replicate the flexion- extension radiograph at end range of flexion-extension; the passive lumbar extension test and the lumbar flexion > 53° were found to be moderately reliable. However, the lack of hypomobility with PA glide test was found to be poorly reliable, and less than chance in some cases. On the other hand, the reliability of the sub-classification scheme for lumbar segmental instability (functional, structural and combined) was reliable as determined by Kappa and adjusted Kappa coefficients.

Thus, the findings of this study supported the importance of establishing a cluster of clinical tests that would successfully rule out structural instability in patients.[8, 9] We propose that future studies should combine all the structural instability tests in one comprehensive reliability and validity study. This would serve three purposes. First, it will reduce the X-ray cost by allowing the spinal physicians to base their referral for the x-ray imaging on a cluster of reliable and valid tests instead of mere suspicion of instability. Secondly, utilization of multiple tests will serve as a tool for identifying patients who may need further x-ray-based radiographic imaging of the spine. It will also minimize the use of X-ray imaging, by spinal physicians, as a guide for sending patients for physical therapy. .

With the current shortage of literature on lumbar segmental instability, we recommend that future studies on the reliability of lumbar segmental mobility tests should explore the added benefit of pressure/force devices, such as the pressure mapping system. [33] We also recommend further exploration of the reliability of other lumbar segmental mobility tests, such as the use of the side-lying position test instead of the prone position test. [17, 33]

References

1. Cairns, M.C., N.E. Foster, and C. Wright, *Randomized controlled trial of specific spinal stabilization exercises and conventional physiotherapy for recurrent low back pain.* Spine (Phila Pa 1976), 2006. **31**(19): p. E670-81.

2. Hauggaard, A. and A.L. Persson, *Specific spinal stabilisation exercises in patients with low back pain – a systematic review.* Physical Therapy Reviews, 2007. **12**(3): p. 233-248.

3. Waddell, G., *1987 Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain.* Spine (Phila Pa 1976), 1987. **12**(7): p. 632-44.

4. Andersson, G.B., *Epidemiological features of chronic low-back pain.* Lancet, 1999. **354**(9178): p. 581-5.

5. Hides, J.A., G.A. Jull, and C.A. Richardson, *Long-term effects of specific stabilizing exercises for first-episode low back pain.* Spine (Phila Pa 1976), 2001. **26**(11): p. E243-8.

6. Demoulin, C., et al., *Lumbar functional instability: a critical appraisal of the literature.* Ann Readapt Med Phys, 2007. **50**(8): p. 677-84, 669-76.

7. Hicks, G.E., et al., *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**(9): p. 1753-62.

8. Alqarni, A.M., A.G. Schneiders, and P.A. Hendrick, *Clinical tests to diagnose lumbar segmental instability: a systematic review.* J Orthop Sports Phys Ther, 2011. **41**(3): p. 130-40.

9. Fritz, J.M., S.R. Piva, and J.D. Childs, *Accuracy of the clinical examination to predict radiographic instability of the lumbar spine.* Eur Spine J, 2005. **14**(8): p. 743-50.

10. Bystrom, M.G., E. Rasmussen-Barr, and W.J. Grooten, *Motor control exercises reduces pain and disability in chronic and recurrent low back pain: a meta-analysis.* Spine (Phila Pa 1976), 2013. **38**(6): p. E350-8.

11. Panjabi, M.M., *The stabilizing system of the spine. Part II. Neutral zone and instability hypothesis.* J Spinal Disord, 1992. **5**(4): p. 390-6; discussion 397.
12. Panjabi, M.M., *Clinical spinal instability and low back pain.* J Electromyogr Kinesiol, 2003. **13**(4): p. 371-9.

13. Akuthota, V., et al., *Core stability exercise principles.* Curr Sports Med Rep, 2008. **7**(1): p. 39-44.

14. O'Sullivan, P.B., et al., *Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylolysis or spondylolisthesis.* Spine (Phila Pa 1976), 1997. **22**(24): p. 2959-67.

15. Fritz, J.M., R.E. Erhard, and B.F. Hagen, *Segmental instability of the lumbar spine.* Phys Ther, 1998. **78**(8): p. 889-96.

16. Leone, A., et al., *Lumbar intervertebral instability: a review.* Radiology, 2007. **245**(1): p. 62-77.

17. Hicks, G.E., et al., *Interrater reliability of clinical examination measures for identification of lumbar segmental instability.* Arch Phys Med Rehabil, 2003. **84**(12): p. 1858-64.

18. Posner, I., et al., *A biomechanical analysis of the clinical stability of the lumbar and lumbosacral spine.* Spine (Phila Pa 1976), 1982. **7**(4): p. 374-89.

19. Dupuis, P.R., et al., *Radiologic diagnosis of degenerative lumbar spinal instability.* Spine (Phila Pa 1976), 1985. **10**(3): p. 262-76.

20. Beazell, J.R., M. Mullins, and T.L. Grindstaff, *Lumbar instability: an evolving and challenging concept.* J Man Manip Ther, 2010. **18**(1): p. 9-14.

21. Maigne, J.Y., et al., *Pain immediately upon sitting down and relieved by standing up is often associated with radiologic lumbar instability or marked anterior loss of disc space.* Spine (Phila Pa 1976), 2003. **28**(12): p. 1327-34.

22. Abbott, J.H., et al., *Lumbar segmental instability: a criterion-related validity study of manual therapy assessment.* BMC Musculoskelet Disord, 2005. **6**: p. 56.

23. Kasai, Y., et al., *A new evaluation method for lumbar spinal instability: passive lumbar extension test.* Phys Ther, 2006. **86**(12): p. 1661-7.

24. May, S. and R. Rosedale, *Prescriptive clinical prediction rules in back pain research: a systematic review.* J Man Manip Ther, 2009. **17**(1): p. 36-45.

25. Rabin, A., et al., *The interrater reliability of physical examination tests that may predict the outcome or suggest the need for lumbar stabilization exercises.* J Orthop Sports Phys Ther, 2013. **43**(2): p. 83-90.

26. Childs, J.D., S.R. Piva, and J.M. Fritz, *Responsiveness of the numeric pain rating scale in patients with low back pain.* Spine (Phila Pa 1976), 2005. **30**(11): p. 1331-4.

27. Fritz, J.M. and J.J. Irrgang, *A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale.* Phys Ther, 2001. **81**(2): p. 776-88.

28. Waddell, G., et al., *A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability.* Pain, 1993. **52**(2): p. 157-68.

29. Gonnella, C., S.V. Paris, and M. Kutner, *Reliability in evaluating passive intervertebral motion.* Phys Ther, 1982. **62**(4): p. 436-44.

30. Childs, J.D. and J.A. Cleland, *Development and application of clinical prediction rules to improve decision making in physical therapist practice.* Phys Ther, 2006. **86**(1): p. 122-31.

31. Mayer, T.G., et al., *Use of noninvasive techniques for quantification of spinal range-of-motion in normal subjects and chronic low-back dysfunction patients.* Spine (Phila Pa 1976), 1984. **9**(6): p. 588-95.

32. Saur, P.M., et al., *Lumbar range of motion: reliability and validity of the inclinometer technique in the clinical measurement of trunk flexibility.* Spine (Phila Pa 1976), 1996. **21**(11): p. 1332-8.

33. Schneider, M., et al., *Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability study.* J Manipulative Physiol Ther, 2008. **31**(6): p. 465-73.

34. Waddell, G., et al., *Objective clinical evaluation of physical impairment in chronic low back pain.* Spine (Phila Pa 1976), 1992. **17**(6): p. 617-28.

35. Cicchetti, D.V. and A.R. Feinstein, *High agreement but low kappa: II. Resolving the paradoxes.* J Clin Epidemiol, 1990. **43**(6): p. 551-8.

36. Sim, J. and C.C. Wright, *The kappa statistic in reliability studies: use, interpretation, and sample size requirements.* Phys Ther, 2005. **85**(3): p. 257-68.

37. Byrt, T., J. Bishop, and J.B. Carlin, *Bias, prevalence and kappa.* J Clin Epidemiol, 1993. **46**(5): p. 423-9.

38. Hestbaek, L., C. Leboeuf-Yde, and C. Manniche, *Low back pain: what is the long-term course? A review of studies of general patient populations.* Eur Spine J, 2003. **12**(2): p. 149-65.

39. Pengel, L.H., et al., *Acute low back pain: systematic review of its prognosis.* Bmj, 2003. **327**(7410): p. 323.

40. Panjabi, M.M., *The stabilizing system of the spine. Part I. Function, dysfunction, adaptation, and enhancement.* J Spinal Disord, 1992. **5**(4): p. 383-9; discussion 397.

APPENDIX A

LUMBAR SEGMENTAL INSTABILITY TESTS DESCRIPTION

TABLE 1	LUMBAR SEGMENTAL INSTABLITY TESTS
1) Aberrant Motion[9, 17]	If any of these movements are observed during forward bending—such as painful arc of motion, an instability catch, thigh climbing, or a reversal of lumbopelvic rhythm—the test is considered to be positive.
2) PIT Test[9, 17]	The subject lay in prone position on the edge of the examining table with the feet on the floor. Examiner performs PA mobility testing on each lumbar segment; if painful segment is identified, the subject is asked to slightly lift the legs off the floor. Then the examiner applies the same amount of pressure to the painful segment. If pain provoked at initial position and subsided at the second one, the test is considered to be positive.
3) Average SLR >91° Test[7, 17]	From supine position, the bubble inclinometer is positioned at tibial crest. The leg is then passively raised to the maximum tolerated level; then the ROM degree is recorded, and the examiner repeats the same process on the second leg. If the average reading of both legs is >91°, the tests are considered positive.
4) Lumbar Flexion ROM >53°[3, 9]	From standing the position, the bubble inclinometer is used to record the baseline reading of S2 and T12-L1 reference point. Then, after the subject bends forward, the end range of T12-L1 is recorded; then the S2 reading is recorded. The true lumbar range is a result of the subtraction of sacral ROM from thoracolumbar ROM. If the result is > 53°, the test is considered to be positive.
5) PLE Test[8, 23]	With subject in prone position, both legs are passively raised about 30cm from bed level and then pulled gently. If the subject experiences severe LBP, or there is a feeling of heaviness on the lower back or a feeling as though the lower back were about to “come off,” the test is considered positive
6) Lack of Hypomobility with PA Glide Test[9, 20]	Subject in prone position. Examiner performs PA glide on the lumbar spinous processes. If all lumbar segments are judged to *not* have stiffness (hypomobility), the test is considered to be positive

Abbreviations: PIT - prone instability test; SLR - average straight leg raising, ROM - range of motion, PLE - passive lumbar extension.