
UNIVERSITY OF NORTH TEXAS
HEALTH SCIENCE CENTER AT FORT WORTH
1970

Russo, David P., Interexaminer Reliability of Somatic Palpatory Findings Associated with Chronic Low Back Pain. Master of Science, May 2002, 15 pp., 1 table, 1 figure, bibliography, 26 titles

Osteopathic physicians rely on specific clinical palpatory tests to diagnose somatic dysfunction of the neuromusculoskeletal system. The purpose of this study is to compare the interexaminer reliability of six common osteopathic clinical tests to severity ratings of somatic dysfunction in six body regions. Ten trained and clinically supervised predoctoral osteopathic manipulative medicine fellows collected palpatory data using the Standardized Outpatient Osteopathic Soap Note Form (SNF) and recorded findings for six pre-selected osteopathic clinical diagnostic tests as part of a randomized controlled trial of osteopathic manipulative treatment for chronic low back pain. Kappa coefficients were used to assess overall examiner agreement for the osteopathic clinical tests. Intraclass correlational coefficients (ICC) and Chronbach's alpha were used to assess examiner agreement for the severity ratings. Kappa values for the six clinical tests ranged from 0 to 0.32. The single item ICC was 0.32, average item ICC was .74, and the coefficient alpha for internal consistency of the six body region scores was 0.80. These results indicate that diagnostic impressions of somatic dysfunction severity may be more reliable than outcomes from isolated osteopathic clinical tests.

Major Professor: Scott Stoll, D.O., Ph.D.

INTEREXAMINER RELIABILITY OF SOMATIC PALPATORY FINDINGS

ASSOCIATED WITH CHRONIC LOW BACK PAIN

David P. Russo, BA

APPROVED:

Major Professor

Committee Member

Committee Member

University Member

Graduate Advisor

Dean, Graduate School of Biomedical Sciences

INTEREXAMINER RELIABILITY OF SOMATIC PALPATORY FINDINGS ASSOCIATED WITH CHRONIC LOW BACK PAIN

THESIS

Presented to the Graduate School of Biomedical Sciences

University of North Texas Health Science Center at Fort Worth

in Partial Fulfillment of the Requirements

for the degree of

Master of Science

By

David P. Russo, BA

Fort Worth, TX

May 2002

ACKNOWLEDGEMENTS

I am indebted to Michele Langan, Ph.D. and Kimberly Fulda, M.P.H. for their assistance in this project. I would also like to thank Scott Stoll, D.O., Ph.D. for his years of support and mentorship in osteopathic manipulative research methods.

Interexaminer Reliability of Somatic Palpatory Findings Associated with Chronic Low Back Pain

David P. Russo, BA *

Scott T. Stoll, DO, PhD *

John C. Licciardone, DO, MS, MBA †

Russell Gamber, DO *

* Department of Osteopathic Manipulative Medicine, Texas College of Osteopathic Medicine, University of North Texas Health Science Center

† Department of Family Medicine, Texas College of Osteopathic Medicine, University of North Texas Health Science Center

Corresponding author: Scott T. Stoll, DO, PhD; University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Fort Worth, TX 76107; 817-735-2461; 817-735-2270 (fax); sstoll@hsc.unt.edu (email)

Word count: 3,473; May 1, 2002

Running heading: Palpatory Findings

ABSTRACT

Background: Osteopathic physicians rely on specific clinical palpatory tests to diagnose somatic dysfunction of the neuromusculoskeletal system. The purpose of this study is to compare the interexaminer reliability of six common osteopathic clinical tests to severity ratings of somatic dysfunction in six body regions.

Methods: Ten trained and clinically supervised predoctoral osteopathic manipulative medicine fellows collected palpatory data using the Standardized Outpatient Osteopathic Soap Note Form (SNF) and recorded findings for six pre-selected osteopathic clinical diagnostic tests as part of a randomized controlled trial of osteopathic manipulative treatment for chronic low back pain.

Results: Kappa coefficients were used to assess overall examiner agreement for the osteopathic clinical tests. Intraclass correlational coefficients (ICC) and Chronbach's alpha were used to assess examiner agreement for the severity ratings. Kappa values for the six clinical tests ranged from 0 to 0.32. The single item ICC was 0.32, average item ICC was .74, and the coefficient alpha for internal consistency of the six body region scores was 0.80.

Discussion: These results indicate that diagnostic impressions of somatic dysfunction severity may be more reliable than outcomes from isolated osteopathic clinical tests. The authors discuss the implications of these results for the design and analysis of future palpatory reliability studies

Keywords: Interexaminer reliability, palpatory findings, osteopathic clinical research

BACKGROUND

Palpation is one of the fundamental methods used in the physical examination of patients. While sophisticated medical technology has lessened the degree to which physicians depend exclusively upon physical examination data in formulating a clinical diagnosis for most diseases, these technologies remain limited in determining the etiology of most musculoskeletal pain syndromes such as chronic low back pain [1, 2]. Somatic dysfunction is broadly defined as "impaired or altered function of related components of the somatic system: Skeletal, arthrodial, myofascial structure, and related vascular, lymphatic, and neural elements [1]." Osteopathic principles assert that somatic dysfunction plays a key role in musculoskeletal pain syndromes as well as systemic diseases.

Osteopathic physicians have developed a number of specialized palpatory methods to assess somatic dysfunction and to diagnose neuromusculoskeletal disorders. In general, these palpatory tests involve the application of variable manual pressure to the surface of the body in order to determine shape, size, temperature, consistency, position, and mobility of underlying somatic structures. By using these palpation techniques to evaluate flexibility, muscular tone, and ligamentous stability, osteopathic physicians gather information about the structural and physiological characteristics of the neuromusculoskeletal system unavailable through other means. Osteopathic physicians combine the results of these palpatory tests with information from a standard medical

history and physical examination to make an assessment of the total burden of somatic dysfunction in a patient and plan manipulative treatment.

It is widely acknowledged that a key element in the acceptance and use of any diagnostic modality is the level of agreement that can be achieved among multiple examiners [3]. In the osteopathic literature alone, over fifty diagnostic tests have been identified in widespread clinical use [4]. A review of common osteopathic musculoskeletal diagnostic tests has shown that osteopathic physicians do not use all 50 tests uniformly in their practice [5]. While a general consensus within the osteopathic profession exists about the theoretical basis for many tests and palpatory techniques, widespread idiosyncrasies in their actual use and interpretation has made it difficult to interpret the significance of many research reports [6].

There are significant costs and unintended consequences associated with using unreliable diagnostic systems. Unreliable diagnostic systems can yield misleading information and can misdirect research into unproductive endeavors. For example, in the area of osteopathic manipulative research, information about the detectability, stability, reliability, and validity of palpatory findings is essential for classifying patients with somatic dysfunction and determining their appropriateness for manipulative intervention. Yet, basic information about palpatory tests commonly used by osteopathic physicians for detecting somatic dysfunction is lacking. This information is needed to guide the development of future clinical research projects in manipulative medicine.

Moreover, it is not known to what extent the kind of palpatory procedures and tests used by osteopathic physicians contribute to manipulative treatment outcome. As

such, there is a need to develop standard methods to establish the reliability of neuromusculoskeletal tests commonly cited in osteopathic manipulative medicine literature. Assuming that a measurable phenomenon does not change between observations, then reliability can be thought of as an interaction of three separate components: The true observation (the target), the clinical test's sensitivity, and test's associated measurement error [7]. It follows that in order to maximize reliability for a given clinical test, the target must be detectable and stable, the clinical test must be sensitive to the target, and the test's measurement error must be low. A derangement in the interaction of any of these three components can negatively impact a test's reliability.

Patients with chronic, nonspecific low back pain are believed to have a high burden of somatic dysfunction, especially of the lumbar spine and pelvis. Thus, this patient population should possess a variety of detectable palpatory targets suitable for a reliability study. Given the chronic nature of somatic dysfunction and disability in these patients, it is expected that these palpatory targets would remain stable in the absence of manipulation or other clinical intervention. Examiner training sessions have been shown to minimize measurement error for a variety of clinical tests [8]. Thus, when these sources of variability are constrained, any remaining discrepancies in reliability can be attributed to the clinical test itself. This purpose of this study was to compare the reliability of six common osteopathic clinical tests and severity ratings of somatic dysfunction in six body regions.

METHODS

This study was conducted at the University of North Texas Health Science Center at Fort Worth from January 2000 through February 2001 as part of a randomized controlled trial of osteopathic manipulative treatment (OMT) in patients with chronic non-specific low back pain [9]. The Institutional Review Board at the University of North Texas Health Science Center at Fort Worth approved the research protocol.

Patient selection

Various low back pain descriptors were used in patient recruiting materials including terms such as "back-ache," "lumbago," and "sciatica." Participants with low back pain symptoms were targeted for recruitment through advertisement in local newspapers, through referrals from institution-affiliated clinics, and contacts with other local physicians.

Respondents were immediately excluded from further participation if they were younger than 21 or older than 69 years old, had a history of spinal surgery within the previous three months, were pregnant, were current or past employees of the study clinic, were involved in a Workman's compensation claim or had any other insurance claim related to their back pain under litigation. Other automatic exclusions included a diagnosis of low back pain attributable to serious medical causes such as: Cancer, excluding non-malignant skin cancer; spinal osteomyelitis; spinal fracture; herniated disc;

ankylosing spondylitis; cauda equina; Participants with "red flags" for any of the six underlying causes of low back pain were given appropriate medical referral and excluded from the study.

Non-excluded respondents were explained the research protocol and administered verbal and written informed consent. These respondents then underwent a more thorough a standardized clinical assessment conducted by a clinically supervised predoctoral manipulative medicine research fellow. Predoctoral manipulative medicine fellows complete an additional year of undergraduate medical training devoted entirely to osteopathic manipulative medicine. The clinical assessment was adapted from the Clinical Practice Guideline on Acute Low Back Pain Problems in Adults [10]. Other baseline assessment included a focused medical history and physical examination, neurologic screening evaluation, and osteopathic structural examination. The osteopathic structural examination was repeated on the first treatment visit prior to any manipulative intervention. The palpatory findings from the baseline visit and re-evaluation at the first treatment served as the available data for reliability analysis.

Data collection, test selection, and examiner training

Palpatory data and structural findings were collected at the baseline visit and at the first treatment visit one week later. Palpatory data were collected using a modified version of the Standardized Outpatient Osteopathic SOAP Note Form (SNF) [11]. This form is an easy to complete, valid method of recording osteopathic structural and palpatory findings believed to be essential information for osteopathic diagnosis and treatment. A validation

study demonstrated that the SNF accurately reflected the same information recorded in unstandardized physician progress notes and that data extracted from the SNF were suitable for osteopathic research [11].

The SNF requires examiners to make a qualitative assessment of the severity of somatic dysfunction on a four point scale (0 = no or background levels of somatic dysfunction, 1 = greater than background levels or minor tissue texture changes, 2 = obvious tissue texture changes, 3 = key lesions with stand-out tissue texture changes). These data were extracted for reliability analysis. Additionally, six common osteopathic clinical tests believed to be responsive to somatic dysfunction associated with chronic low back pain were also selected for evaluation. Descriptions of these tests and their procedures are summarized in Table 1. The results of each test were dichotomized into positive/negative findings. Tests involving a bilateral comparison were coded such that a positive finding on one side did not preclude a positive finding occurring on the other side.

An osteopathic manipulative medicine specialist provided a series of eight study-specific training sessions to ensure that the screening and diagnostic protocol was implemented consistently. Specifically, these training sessions focused on reviewing how to perform and record the result of each clinical test in a uniform manner. A total of 10 predoctoral fellows were trained in how to use and interpret tests. Predoctoral fellows rotate three-month OMM-specific educational blocks with clinical training blocks. Thus, each fellow attended 2 training sessions per three-month period for the duration of the study.

Data management and analysis

Data management was performed using the SPSS software package version 10.0 (SPSS Inc, Chicago, IL). For all clinical tests, a binary decision was recorded. For these tests, the raw percent agreement, the prevalence of pathological findings, the kappa coefficient and its 95% confidence interval were computed. The kappa coefficient provides a measure of agreement between examiners corrected for chance [12]. The kappa coefficient has a maximum of 1.00 when agreement is perfect, a value of zero when no agreement exists beyond that expected by chance alone, and a negative value when agreement is worse than chance. The kappa coefficient's value is strongly influenced by the prevalence of findings such that it is attenuated severely toward low values when the prevalence of a finding is either extremely high or low [13]. Kappa values can be optimized when the target finding has an overall prevalence between 36% and 65% [14, 15].

Ten examiners were randomly assigned to either "examiner group A" status or "examiner group B" status. Thus, the kappa values reported in this study represent the interexaminer reliability between two groups of similarly trained examiners. Alternatively, with a study involving ten examiners, it would have been possible to estimate the interexaminer reliability for each of five pairs of examiners. However, the resulting kappa values would be based on calculations from sub-samples of palpatory findings reflecting only one-fifth of the total study-wide prevalence. The computation of such multiple pair-wise comparisons can result in chance factors attenuating kappa

estimates toward lower values due to artificially low prevalence of findings among selected pairs. In the literature, kappa values greater than .20 are considered "fair agreement", values greater than .40 are considered "moderate agreement", and values greater than .60 are considered "good agreement" [16]. Most authors consider values greater than .40 acceptable for clinical or research use.

The intraclass correlation coefficient (ICC) is the appropriate statistic to use when measuring the interexaminer reliability of two sets continuous, ordinal, or interval-like data that share a common metric [17]. As such, the ICC can estimate the interexaminer reliability of data from an ordinal symptom severity rating scale. As described above, the somatic dysfunction severity ratings from SNF range from 0 to 4 for each body region, thus possessing ordinal characteristics. Examiners were asked to judge the severity of somatic dysfunction in each of six body regions: Thoracic spine T1-T4, thoracic spine T5-T9, thoracic spine T10-T12, lumbar spine L1-L5, sacrum/pelvis region, and pelvis/ innominate region. The severity ratings were summed to create a total burden of somatic dysfunction index of the thoracolumbar spine and pelvis.

The ICC measures how closely each rater agrees for every observation. It is the proportion of variance that is attributable to the object of measurement (the target) expressed as a ratio of the inter-target variance (as a percent of the total variance) to the inter-rater error (measurement error). The variances and the ICC are computed using analysis of variance (ANOVA) techniques. ICC analysis includes reliability of raters on a single item, the average reliability of raters for a combination of items, as well the value for Chronbach's alpha for the total scale scores. The appropriate ICC measure to use

depends upon whether one plans to rely on a single rating or a combination of ratings. In general, combining multiple ratings generally produces more reliable measurements. Cronbach's alpha is a measure of a set of items' internal consistency that reflects how well the items measure a single underlying construct. The ICC can range from 0 to 1 representing no agreement to perfect agreement, respectively. In general, ICC values less than 0.40 are considered poor; 0.40 to 0.59 considered fair; 0.60 to 0.74 considered good; and values greater than 0.75 are considered excellent [18]. Values between 0.70-0.80 are considered adequate for applied clinical tests.

RESULTS

Palpatory data was collected from 91 participants who met study eligibility criteria. For analysis of the binary clinical tests, 72-90 paired observations were available for analysis. Figure 1 displays these kappa values and their respective 95% confidence intervals. Kappa values for the six clinical tests ranged from 0 to 0.32. Tests for ASIS asymmetry could produce two diagnostic impressions that are equivalent depending upon how a given examiner chose to localize their respective findings. Thus, data for ASIS asymmetry were coded such that a finding of an inferior ASIS on the right was coded the same way as a superior ASIS on the left. For consistency, all pelvic findings in this report are coded with localization to the left hemi-pelvis.

For severity of somatic dysfunction ratings, ICC and their 95% confidence intervals were computed for 80 complete observations. The single item ICC, reflecting

the examiner agreement per body region, was 0.32 (CI = 0.20-0.44); the average-item ICC, reflecting agreement on total somatic dysfunction burden across all body regions was 0.74 (CI = 0.59-0.83); and the coefficient alpha for internal consistency of the six body region scores was 0.80.

DISCUSSION

The results of this study suggest that the interexaminer reliability of six isolated osteopathic clinical diagnostic tests is generally poor as evidenced by no kappa value greater than 0.32, while the average interexaminer reliability of cumulated somatic dysfunction severity ratings is good as evidenced by ICC values greater than 0.70. This implies that under conditions similar to those in this study, more general diagnostic ratings about patients' somatic dysfunction performed by multiple raters may be more reliable than isolated palpatory tests for somatic dysfunction performed by pairs of examiners.

These findings need to be interpreted in the context of several methodological limitations. First, this investigation did not occur as a dedicated study of interexaminer reliability, rather it occurred as part of a clinical trial of OMT for chronic low back pain. This could have biased the findings in a number of ways. In a dedicated interexaminer reliability study, examiners would make ratings in immediate succession and not one-week apart. Although we chose palpatory targets believed to be relatively stable in a chronic patient population, it is conceivable that the one week interval between the two

times of measurement could have changed the palpatory findings. Moreover, in a dedicated reliability study of palpatory findings, examiners would be asked to make multiple ratings, allowing for analysis of both inter- and intra-examiner reliability. The logistics of this project did not allow for more than one examiner to make more than one rating per patient visit. Second, the training sessions used in this study did not aim to calibrate examiners to predetermined palpatory competency thresholds. Instead, the training sessions emphasized a general consensus approach to interpreting osteopathic clinical diagnostic tests. That is, even though every examiner in this study shared a consensus about which tests were to be used, how to perform each test, and how to document the outcome, we did not train examiners to use standardized forces to motion test particular body regions nor refine other subtle psychomotor skills associated with each test.

This study raises important questions about how best to design and analyze future osteopathic palpatory studies. In osteopathic clinical research, the goal of measuring interexaminer reliability is to estimate the validity (accuracy) of findings in the absence of a "gold standard." This is a reasonable use of agreement data because if two raters disagree, then at least one of them must be incorrect. However, the analysis of palpatory findings with kappa coefficients affords only a few simplistic ways to describe this amount of agreement. One can measure the proportion of times two or more raters of the same target agree, the proportions of times two or more raters use different rating levels, and so forth. Studies based solely upon kappa coefficients for isolated palpatory tests answer only narrowly defined questions about interexaminer reliability. Using kappa-

based methodology it is difficult to understand why examiners disagreed about a given test or what factors contributed to their disagreement.

There are other ways of quantifying examiner agreement that extend beyond simple comparisons of proportions and may yield more valuable information about the reliability of palpatory findings. For example, it is widely accepted that when osteopathic physicians plan a manipulative intervention, they do so based upon an assessment of a variety of palpatory findings. It appears less meaningful to evaluate the reliability of isolated palpatory findings when the final diagnostic impression and treatment plan depends upon a constellation of clinical findings. What is needed is a better understanding of how osteopathic physicians integrate a variety of sources of palpatory data, recognize patterns, and form general diagnostic impressions in order to arrive at a reliable diagnosis of somatic dysfunction.

To that end, it might be more useful to develop standardized palpatory protocols with continuous data properties that yield indices of somatic dysfunction per body region or composite indices of global burden of somatic dysfunction. Each element of such a diagnostic palpatory protocol could be assigned a value so as to produce scale scores. Scales have intrinsic measurement properties that would allow osteopathic clinical researchers to construct more robust models to study how palpatory ratings are made and how raters agree or disagree about different kinds of findings. Ultimately, this is the kind of information that will help refine and advance osteopathic manipulative clinical research.

Future palpatory reliability studies should also seek to incorporate rigorous examiner training paradigms that allow examiners to achieve not only consensus on how to perform clinical tests and documents findings, but also allow examiners to calibrate their palpatory senses. These training sessions could include standardizing the amount of force used to illicit a particular finding with pressure sensitive instruments, incorporate static and dynamic anatomical models of structural asymmetries, refine psychomotor skills, and incorporate practice sessions with symptomatic patients similar to those to be recruited in the actual study.

Reviews of palpatory findings in the manual medicine literature have found that: 1) Intra-examiner reliability usually surpasses inter-examiner reliability for most palpatory clinical tests [19]; 2) Experienced and novice clinicians differ in their application of palpatory tests [20]; 3) Consensus and standardization procedures improve the inter-examiner reliability of palpatory tests [21]; and 4) Using clusters of palpatory tests to formulate a diagnostic impression is more reliable than depending upon single tests [22]. Osteopathic clinical researchers need to build upon these findings to start constucting explicit, testable models of examiner agreement that can be evaluated using more than simple kappa coefficients.

Other investigations should extend beyond describing the limits of examiner agreement for specific tests and seek to test competing hypotheses of reliability in order to uncover the sources of examiner disagreement. A variety of statistical techniques other than kappa coefficients could be used to test the reliability of these models, including ICC [23], limits of agreement [24, 25], Rasch analysis [26], and item response theory [7].

Once refined, these palpatory protocols could then be used in manipulative clinical trials as uniquely osteopathic outcome instruments to measure patients' global or regional burden of somatic dysfunction. After it is shown that changes in palpatory findings can be reliably assessed and documented, osteopathic clinical researchers could begin to compare palpatory outcomes to more traditional treatment outcome measures.

In summary, information is needed about the reliability of palpatory diagnosis that more closely reflects how osteopathic physicians synthesize a variety of clinical tests, general diagnostic impressions, and patient report to arrive at a reliable diagnosis of somatic dysfunction. Evaluating the reliability of isolated palpatory findings using kappa values is a reasonable starting place for the construction of more robust models of examiner agreement, but clearly more work is needed to develop reliable palpatory protocols that can be effectively used in outcomes-oriented osteopathic manipulative clinical research. To do so will require a better understanding of the underlying detectability of somatic dysfunction, the psychophysiology of palpatory thresholds, the cognitive science of clinical decision-making, and the optimization of palpatory training procedures. These areas of research should prove to be fertile ground for osteopathic basic science and clinical researchers alike.

REFERENCES

1. Ward, R.C., J.A. Jerome, and American Osteopathic Association, Foundations for osteopathic medicine. 1997, Baltimore: Williams & Wilkins. xxviii, 1162.
2. Korr, I.M., The spinal cord as organizer of disease processes: II. The peripheral autonomic nervous system. J Am Osteopath Assoc, 1979. 79(2): p. 82-90.
3. Russell, R., Diagnostic palpation of the spine: a review of procedures and assessment of their reliability. J Manipulative Physiol Ther, 1983. 6(4): p. 181-3.
4. Dinnar, U., et al., Classification of diagnostic tests used with osteopathic manipulation. J Am Osteopath Assoc, 1980. 79(7): p. 451-5.
5. Dinnar, U., et al., Description of fifty diagnostic tests used with osteopathic manipulation. J Am Osteopath Assoc, 1982. 81(5): p. 314-21.
6. Bronfort, G., Spinal manipulation: current state of research and its indications. Neurol Clin, 1999. 17(1): p. 91-111.
7. Hays, R.D., L.S. Morales, and S.P. Reise, Item response theory and health outcomes measurement in the 21st century. Med Care, 2000. 38(9 Suppl): p. II28-42.
8. Van Dillen, L.R., et al., Reliability of physical examination items used for classification of patients with low back pain. Phys Ther, 1998. 78(9): p. 979-88.
9. Licciardone, J., et al., A Randomized Controlled Trial of Osteopathic Manipulative Treatment in Patients with Chronic Low Back Pain. In press, 2001.

10. Bigos, S.J. and United States. Agency for Health Care Policy and Research., Acute low back problems in adults. Clinical practice guideline ; no. 14. 1994, Rockville, Md.: U.S. Dept. of Health and Human Services Public Health Service Agency for Health Care Policy and Research. viii, 160.

11. Sleszynski, S.L., T. Glonek, and W.A. Kuchera, Standardized medical record: a new outpatient osteopathic SOAP note form: validation of a standardized office form against physician's progress notes. J Am Osteopath Assoc, 1999. 99(10): p. 516-29.

12. Cohen, J., A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960. 20: p. 37-46.

13. Steinijans, V.W., et al., Interobserver agreement: Cohen's kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. Int J Clin Pharmacol Ther, 1997. 35(3): p. 93-5.

14. Karras, D.J., Statistical methodology: II. Reliability and variability assessment in study design, Part A. Acad Emerg Med, 1997. 4(1): p. 64-71.

15. Karras, D.J., Statistical methodology: II. Reliability and validity assessment in study design, Part B. Acad Emerg Med, 1997. 4(2): p. 144-7.

16. Welkowitz, J., R.B. Ewen, and J. Cohen, Introductory statistics for the behavioral sciences. 5th ed ed. 2000, Fort Worth: Harcourt Brace College Publishers. xvii, 359.

17. Prieto, L., et al., The evaluation of agreement on continuous variables by the intraclass correlation coefficient. J Epidemiol Community Health, 1997. 51(5): p. 579-81.

18. Cicchetti, D.V. and S.A. Sparrow, Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic, 1981. 86(2): p. 127-37.

19. Mootz, R.D., et al., Intra- and interobserver reliability of passive motion palpation of the lumbar spine. J Manipulative Physiol Ther, 1989. 12(6): p. 440-5.

20. Mior, S.A., M. McGregor, and B. Schut, The role of experience in clinical accuracy. J Manipulative Physiol Ther, 1990. 13(2): p. 68-71.

21. Beal, M.C., et al., Interexaminer agreement on patient improvement after negotiated selection of tests. J Am Osteopath Assoc, 1980. 79(7): p. 432-40.

22. Cibulka, M.T. and R. Koldehoff, Clinical usefulness of a cluster of sacroiliac joint tests in patients with and without low back pain. J Orthop Sports Phys Ther, 1999. 29(2): p. 83-9; discussion 90-2.

23. Bartko, J.J., The intraclass correlation coefficient as a measure of reliability. Psychol Rep, 1966. 19(1): p. 3-11.

24. Bland, J.M. and D.G. Altman, A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med, 1990. 20(5): p. 337-40.

25. Bland, J.M. and D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1986. 1(8476): p. 307-10.

26. Schumacker, R.E., Many-facet Rasch analysis with crossed, nested, and mixed designs. J Outcome Meas, 1999. 3(4): p. 323-38.

Table 1: Palpatory Clinical Tests

Clinical Test	Description	Interpretation
Gross Side-bending Test	The patient is asked to stand with hands at sides. The patient side-bends as far as possible to the left and right by sliding hand down lateral aspect of the thigh toward the knee. The degree of excursion is visually inspected by observing how far down fingertips reach toward floor.	Test is positive for gross somatic dysfunction of the lumbar spine if unequal side-bending motion is observed.
Seated Flexion Test	The patient is seated upright with feet flat on floor on ladder-back stool. Examiner's thumbs are place inferior to the patients' PSIS. Patient is asked to bend forward.	Test is positive for somatic dysfunction of the pelvis if thumbs are not level at end of motion.

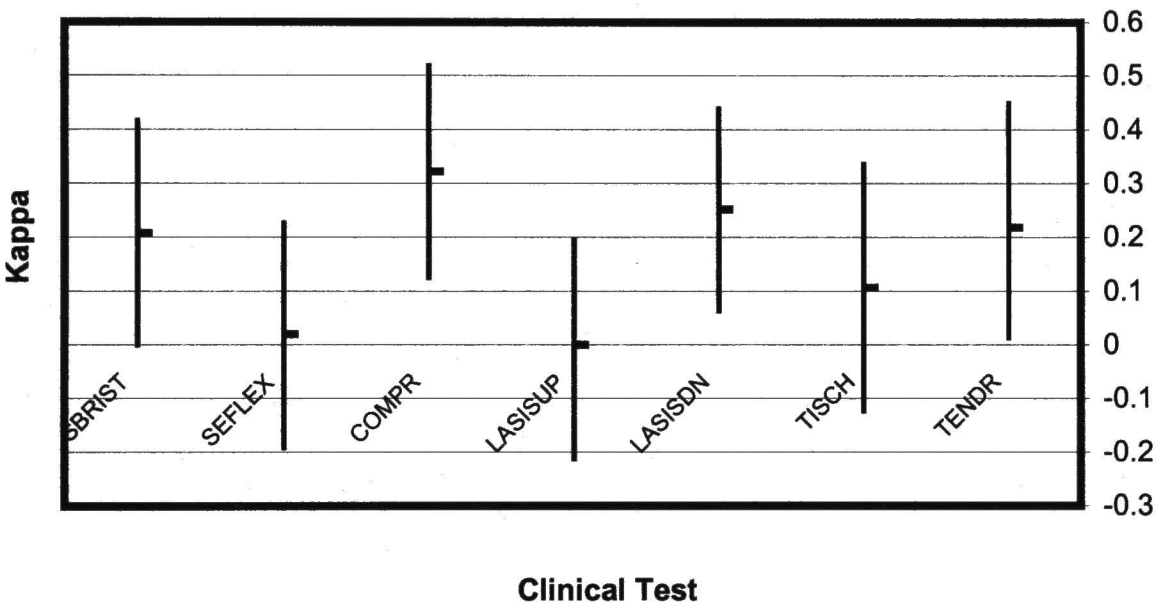
	Levelness of PSIS is assessed at beginning and end of active motion.	
ASIS Compression Test	The patient is supine. Examiner places palms over each ASIS and stabilizes the opposite ASIS while introducing a springing motion in a postero-medial direction.	Test is positive for somatic dysfunction of the sacroiliac joint and/or pelvis if a lack of resiliency is encountered.
ASIS Asymmetry Test	The patient is supine. Examiner places thumbs under inferior aspect of ASIS and inspects levelness.	Test is positive for somatic dysfunction of the pelvis if ASIS are observed to be unlevel.
Lumbar Tenderness	The patient is prone. The lumbar paraspinal muscles are palpated.	Test is positive for somatic dysfunction of lumbar spine if lumbar paraspinal muscle patient reports subjective

		tenderness to palpation.
Lumbar Tissue Texture Changes	The patient is prone. The lumbar paraspinal muscles are palpated.	Test is positive for somatic dysfunction of lumbar spine if lumbar paraspinal muscle tissues are tight, boggy, ropey, doughy compared to other paraspinal tissues.

ASIS – Anterior superior iliac spines

PSIS – Posterior superior iliac spines

Figure 1: Kappa values and 95% confidence intervals for clinical palpatory tests



SBRIST = Sidebending restriction
SEFLEX = Seated flexion test
COMPR = ASIS compression tests
LASISUP = Left ASIS elevated
LASISDN = Left ASIS depressed
TISCH= Lumbar tissue texture changes present
TNDR = Lumbar tenderness present