

# Spinal palpatory diagnostic procedures utilized by practitioners of spinal manipulation: annotated bibliography of content validity and reliability studies

Michael Seffinger, DO\*  
 Alan Adams, DC\*\*  
 Wadie Najm, MD†  
 Vivian Dickerson, MD††  
 Shiraz I Mishra, MD, PhD§  
 Sibylle Reinsch, PhD§§  
 Linda Murphy, MLS¶¶

*The diagnosis of spinal neuro-musculoskeletal dysfunction is a pre-requisite for application of spinal manual therapy. Different disciplines rely on palpatory procedures to establish this diagnosis and design treatment plans. Over the past 30 years, the osteopathic, chiropractic, physical therapy and allopathic professions have investigated the validity and reliability of spinal palpatory procedures. We explored the literature from all four disciplines looking for scientific papers studying the content validity and reliability of spinal palpatory procedures. Thirteen databases were searched for relevant papers between January 1966 and October 2001. An annotated bibliography of these articles is presented and organized by the type of test used. (JCCA 2003; 47(2):93-109)*

*Le diagnostic d'un dysfonctionnement neuromusculosquelettique spinal est une condition préalable à l'application d'un traitement manuel de la colonne vertébrale. Les différentes disciplines se basent sur des procédures palpatoires pour établir ce diagnostic et des plans de traitement. Au cours des 30 dernières années, les ostéopathes, chiropracteurs, physiothérapeutes et allopathes ont étudié la validité et la fiabilité des procédures palpatoires spinales. Nous avons fouillé les revues spécialisées de ces quatre disciplines pour trouver des articles scientifiques étudiant la validité et la fiabilité des procédures palpatoires spinales. Nous avons exploré treize bases de données à la recherche d'articles pertinents entre janvier 1996 et octobre 2001. Une bibliographie annotée de ces articles est présentée et organisée par type d'examen utilisé. (JACC 2003; 47(2):93-109)*

KEY WORDS : manual therapy, palpation, spinal manipulation.

MOTS CLÉS : traitement manuel, palpation, manipulation spinale.

\* Assistant Professor, Western University of Health Sciences, College of Osteopathic Medicine of the Pacific.

\*\* Professor and Vice President, Southern California University of Health Sciences.

† Associate Clinical Professor, University of California, Irvine.

†† Associate Clinical Professor, University of California, Irvine.

§ Associate Adjunct Professor, University of California, Irvine.

§§ Research Associate, University of California, Irvine.

¶¶ Librarian, University of California, Irvine.

Work was done at the Susan Samuelli Center for Complementary and Alternative Medicine, University of California, Irvine.

Work supported by a grant from the 41st Fund, University of California, Irvine.

Address correspondence to: Wadie Najm, MD, University of California, Irvine,

101 The City Drive Bldg. 200 #512, Rt.81, Orange, CA, USA 92868. Tel: 714-456-5171. Fax: 714-456-7984. winajm@uci.edu

© JCCA 2003.

## Introduction

Professions that employ manual manipulative procedures use their own terminology to describe the diagnostic entity that responds to manipulation therapy.<sup>1</sup> Spinal neuromusculoskeletal dysfunction is the term used in our paper to encompass these various terms employed by the different disciplines. Spinal neuromusculoskeletal dysfunction refers to an alteration of spinal joint position, motion characteristics and related palpable paraspinal soft tissue changes. Spinal palpatory diagnostic procedures typically entail static palpation of anatomical landmarks for symmetry, palpation of spinal vertebral joints before, during and after active and passive motion tests, and spinal and paraspinal soft tissue palpatory assessment for abnormalities or altered sensitivity.

Outcomes and effectiveness of manipulative treatments rely partly on the validity and reliability of the palpatory procedures used to diagnose spinal neuromusculoskeletal dysfunction. Investigation of the validity and reliability of spinal palpatory diagnostic tests has been in progress for the past 70 years. A complete review and analysis of these studies is lacking. A preliminary evaluation reveals an inconsistency in the focus, methodology, palpatory procedures and statistical analysis used. Focused narrative reviews from this literature have been previously published.<sup>2-10</sup>

This paper is an annotated bibliography of primary research studies on the content validity and reliability of spinal diagnostic palpatory procedures from the literature accessed between 1966 and 2001. The annotated bibliography is intended simply to provide the reader with a summary of these studies. Further analysis of data and review of results will be highlighted in a planned systematic review.

To identify these studies, the authors identified specific MeSH and key-word terms used for manual palpatory procedures of the cervical, thoracic and lumbar spine in chiropractic, osteopathic, allopathic and physical therapy journals.<sup>11</sup> Using PubMed, an initial search template was formulated and subsequently applied to 12 additional selected databases from January 1, 1966 through October 1, 2001. These include MANTIS, EMBASE and CINAHL. Manual bibliographic searches and content experts identified further citations (for additional details see Murphy<sup>11</sup>).

Inclusion criteria were: primary research studies published in peer-reviewed journals, or dissertations, that had

data analysis to determine content validity, intra-examiner and/or inter-examiner reliability, of spinal palpatory diagnostic procedures used to identify a manipulatable spinal neuromusculoskeletal dysfunction. Articles were not limited by language. The documents included were limited to cervical, thoracic and lumbar spinal regions. Exclusion criteria were: documents retrieved that were inconsistent with the inclusion criteria, abstracts, proceedings from conferences, anecdotal, speculative or editorial articles and studies of non-manual procedures. These criteria were adapted from criteria previously published<sup>12,13</sup> by the Cochrane Library. Four content validity and 59 reliability articles met the inclusion criteria and authors' annotated summaries were created.

Historically, content validity was not systematically investigated until the 1990s when the chiropractic and then physical therapy professions took on the challenge. Prior to that, spinal palpatory procedures were accepted and used on face validity. The inter-examiner and intra-examiner reliability of spinal palpatory procedures were first investigated by the osteopathic profession as early as 1949<sup>14</sup> but not systematically studied until the 1970s. In successive decades, the chiropractic, physical therapy and allopathic professions followed with their own series of investigations into their spinal palpatory methods. Since the different professions investigated their own procedures, it is not possible to compare the results of the various studies. However, the authors believe that there is merit in combining the available literature from all disciplines and professions and evaluating the quality of the research itself in hopes of globally improving the research literature database.

The content validity studies are grouped according to the reference standard utilized: a) plastic spinal model (SM); b) patient's pain scale (PS). The reliability studies are grouped into categories related to the palpatory procedures used: motion tests (M), pain provocation tests (P), paraspinal soft tissue palpation tests (ST) and assessments of the position of anatomical landmarks (L). Within each subsection the papers are arranged in chronological order to provide the reader with a glimpse of the historical progression of the studies.

The definitions, appropriateness and applicability of the various statistical tests used in the following reliability studies have been addressed in the literature and are beyond the scope of this paper.<sup>15-18</sup> In this article the details

offered in each annotation reflects the level of information presented in the original papers. An in-depth review and analysis of these studies, their research and clinical implications are in progress, and planned for future publication.

### Content Validity Studies

#### Spinal model

**SM1 Harvey D, Byfield D. Preliminary studies with a mechanical model for the evaluation of spinal motion palpation. *Clinical Biomechanics* 1991; 6:79–82.**

The study assessed the ability of experienced chiropractors (at least 5 years clinical practice) and first year chiropractic students to detect the presence or absence of lumbar spine intersegmental motion restriction by palpating a spinal model equipped with artificial fixators. This was the first such study, using a model, to be published. Twenty-seven examiners performed a total of 300 blind palpation observations of fixed and non-fixed segments on the model. Although no training of the test procedure occurred, a standard protocol was followed. Using the spinous process as a contact point, each examiner moved the model through ranges of motion in the sagittal and/or coronal planes assessing presence or absence of relative separation and approximation of the spinous processes within a three minute time period. In determining the exact level of random fixation, student sensitivity/specificity was 53.8%/85.5%. The experienced chiropractor sensitivity/specificity was 47.8%/88%. Results demonstrated that examiners were more accurate at identifying non-fixed segments than fixated segments on a spinal model using this test procedure. The test procedure used was more specific than sensitive.

**SM2 Moruzzi S. Accuracy of two different motion palpation procedures for determining fixations in the lumbar spine using an articulated spinal model: *Anglo-European College of Chiropractic, 1992–1993 (Thesis)*.**

The study investigated the accuracy of two different motion palpation procedures for determining fixations in the lumbar spine using an articulated spinal model. The two motion palpation procedures tested were posterior to anterior spinal springing and left and right lateral flexion. The

articulated spinal model was fitted with intersegmental fixators, which were randomly fixed for each examiner. The examiners included 50 senior chiropractic student interns who randomly used one of the two palpation tests on the spinal model. Sensitivity (average 42%) and specificity (average 62%) indicated that there was no significant difference in the accuracy of the two types of motion palpation in detecting the presence or absence of fixators in the lumbar spine. The author concluded that neither of the palpation procedures as performed was a valid diagnostic test.

**SM3 Jensen K, Gemmell H, Thiel H. Motion palpation accuracy using a mechanical spinal model. *European J Chiropractic* 1993; 41:67–73.**

The study determined the accuracy of student interns and experienced chiropractors to detect the presence or absence of fixations in a spinal model equipped with artificial segmental fixators using motion palpation. The examiners comprised of 30 student interns and 15 practicing chiropractors, all of whom performed manual palpation on a spinal model. The study assessed construct and content validity. Results indicated that both, sensitivity (72%) and specificity (83.2%) of the students, was much higher than that of the experienced chiropractors (sensitivity 52.6%; specificity 78.6%). The results demonstrated that since the specificity was higher than the sensitivity, the determination of the presence of non-fixation in a mechanical spinal model was more accurate than the determination of the presence of fixation. The authors concluded that motion palpation was a specific but not sensitive method of spinal analysis.

#### Pain scale

**PS1 Sandmark H, Nisell R. Validity of five common manual neck pain provoking tests. *Scand J Rehabil Med* 1995; 27:131–136.**

This study assessed the validity of five pain provocation tests, one of which was a palpation procedure. A blinded experienced registered physiotherapist assessed 75 patients with acute neck pain (< one week) using standard palpation procedures of the cervical spine, facet joints, and paraspinal soft tissues for provocation of pain. Standard reference was the patient's verbalized acknowledgment of pain. The pain provocation palpation procedure had a sen-

sitivity of 82%, specificity of 79%, positive predictive value of 62% and negative predictive value of 91%.

### Reliability Studies

#### Spinal Motion Assessment Procedures

**M1 Johnston W. Inter-examiner reliability in palpation. J Am Osteopath Assoc 1976; 76:286–287.**

The study evaluated inter-examiner reliability of 2 trained osteopathic medical students using three passive thoracic spinal motion tests on an unspecified number of subjects. The addition of precision level landmarks improved inter-examiner reliability by ensuring that the examiners more accurately identified the spinal level they were evaluating.

**M2 Johnston W, Hill J, Sealey J, Sucher B. Palpatory findings in the cervicothoracic region: variations in normotensive and hypertensive subjects. A preliminary report. J Am Osteopath Assoc 1980; 79:300–308.**

The study evaluated inter-examiner reliability of an experienced osteopathic physician (30 years clinical practice) specializing in manipulation and 2 osteopathic medical students trained (for 6 months) in the utilization of 3 types of motion tests of the cervical and thoracic spine. Data analysis on 132 human subjects revealed examiner agreements at a level significantly beyond chance alone: observed 39.5 +/- 4.81 (+/- 95% confidence level) vs. expected 26.0 +/- 4.32. Marking of spinal levels improved each examiner's ability to define positive results.

**M3 Beal MC, Goodridge JP, Johnston WL, McConnell DG. Inter-examiner agreement on patient improvement after negotiated selection of tests. J Am Osteopath Assoc 1980; 79:432–440.**

The study analyzed inter-examiner reliability in assessing back pain patients' improvement over time using osteopathic palpatory evaluation of the spine. Three osteopathic physicians specializing in manipulation used negotiated and agreed upon (unspecified) spinal palpatory test procedures. These tests included gross and segmental motion testing, paraspinal soft tissue palpation and assessment of anatomical landmarks for levelness, adding individual test procedures as desired. Six symptomatic patients were evaluated. Results were displayed graphically. The obser-

vation that examiners displayed good qualitative agreement on long-term improvement must be tempered by the fact that visit-by-visit agreement among the three examiners was not above the level of chance expectation.

**M4 Johnston WL, Elkiss ML, Marino RV, Blum GA. Passive gross motion testing. II. A study of inter-examiner agreement. J Am Osteopath Assoc 1982; 81:304–308.**

The study assessed the extent to which minimal findings within the study population contributes to lack of inter-examiner reliability. Two osteopathic medical students and 1 faculty osteopathic physician specializing in manipulation were trained in performing six passive gross motion tests of the cervical spine before examining 161 volunteers. Based on calculated observed vs. expected agreement and percent of disagreement, the authors suggested that subjects with inconsistent findings\* have a significant influence in evaluating inter-examiner reliability. \*(Defined as only one examiner or none can agree on the findings after each of three repetitions of a passive gross motion test on the same subject; that is, the findings were unstable and unreliable; not fixed).

**M5 Johnston WL, Beal MC, Blum GA, Hendra JL, Neff DR, Rosen ME. Passive gross motion testing: Part III. Examiner agreement on selected subjects. J Am Osteopath Assoc 1982; 81:309–313.**

The study evaluated inter-examiner agreement in assessing regional asymmetry using gross motion tests of neck rotation. Permutation and level of agreement analyses were performed. The authors found a high level of agreement on 14 of 70 volunteers examined by two trained faculty osteopathic physician examiners specializing in manipulation regarding direction and intensity of asymmetrical restrictive response to the motion tests. Direction of cervical motion asymmetry in the 14 subjects as determined by a blinded second group of examiners (3 osteopathic medical students), trained in performing the same motion tests and evaluating the direction and intensity of vertebral segment response, one out of three showed a high level of agreement with the first group of examiners.



**M6 Beal MC, Goodridge JP, Johnston WL, McConnell DG. Inter-examiner agreement on long-term patient improvement: an exercise in research design. J Am Osteopath Assoc 1982; 81:322–328.**

The study evaluated long-term inter-examiner reliability on diagnosing aggravation or improvement of musculoskeletal conditions of the spine. Three osteopathic physician faculty specializing in manipulation performed unspecified spinal palpation and manual treatment procedures on 3 patients over 17 months. Palpatory evaluation included regional and segmental motion testing and palpation of paraspinal soft tissue. Line graphs using a plus-minus scale demonstrated over-all agreement of findings on 2 out of 3 patients. The authors also stated that an improvement in palpatory test results correlated with patients' reported symptomatic improvement.

**M7 Johnston W, Hill J, Elkiss M, Marino R. Identification of stable somatic findings in hypertensive subjects by trained examiners using palpatory examination. J Am Osteopath Assoc 1982; 81:830–836.**

The study evaluated the inter-examiner reliability of spinal palpation for dysfunction in the cervico-thoracic spine in hypertensive (high blood pressure) vs. normotensive (normal blood pressure) patient populations. Three blinded and trained examiners (one faculty osteopathic physician specializing in manipulation and two osteopathic medical students) examined 307 volunteer adult subjects using three passive motion tests to detect asymmetrical response between C5 and T7. Where there was a high level of inter-examiner agreement on findings, 52.7 % of the cohort were hypertensive. Where there was less agreement, 22.2% of the cohort were hypertensive.

**M8 Gonnella C, Paris SV, Kutner M. Reliability in evaluating passive intervertebral motion. Phys Ther 1982; 62:436–444.**

The study evaluated inter- and intra-examiner reliability in assessing spinal passive mobility of the lumbar spine. Five experienced physical therapists trained in a passive motion grading criteria randomly examined 5 subjects (physical therapy students). Since the grading scale was not continuous, data was largely descriptive (means and standard deviations were calculated). Intra-examiner reliability was found to be greater than inter-examiner reliability.

**M9 Mior S, King R, McGregor M, Bernard M. Intra and inter-examiner reliability of motion palpation in the cervical spine. J Can Chiropr Assoc 1985; 29:195–199.**

This pilot study assessed the intra- and inter-examiner reliability of motion palpation of the upper cervical spine. Two chiropractic interns examined sixty-two randomly selected students. Examiners were trained in the palpation technique using joint play of the atlas in lateral bending and rotation. The Kappa score for inter-examiner agreement was poor at 0.15. Intra-examiner agreement was fair at 0.37 and 0.52 for the two examiners.

**M10 DeBoer K, Harmon R, Tuttle C, Wallace H. Reliability study of detection of somatic dysfunctions in the cervical spine. J Manipulative Physiol Ther 1985; 8:9–16.**

The study assessed levels of intra- and inter-examiner agreement on cervical spine palpation under normal clinical conditions. Three chiropractic faculty members, using preferred individual methods, examined 40 asymptomatic chiropractic students on two occasions. Weighted Kappa scores were calculated for intra-examiner and inter-examiner agreement for cervical fixation, pain and muscle spasm. Generally, intra-examiner was better than inter-examiner agreement. Agreement on fixation findings was good in the lower cervical spine and poor in the middle cervical area. There was fair level of agreement for fixation findings in the upper cervical spine and also for the other palpation parameters that were used (e.g., pain and muscle spasms) throughout the cervical spine.

**M11 Grant A, Spadon R. An inter- and intra-examiner reliability study using lateral flexion motion palpation of the lumbar spine in the prone position. Anglo-European College of Chiropractic; 1985 (Thesis).**

The study evaluated reliability in locating spinal fixations in the lumbar spine. Four third and fourth year chiropractic students examined 60 chiropractic student subjects (convenience sample) using a prone method of lumbar side bending (Gillet's method of motion palpation). Inter-examiner agreement on motion was 66.7% with highest agreement at L3 (80.8%) and lowest at L5 (52.5%). Inter-examiner agreement of direction was 65.7% for left and 67.3% for right. There was high intra-examiner agreement

for all parameters, ranging from 85 to 90% for all 3 examiners.

**M12 Bergstrom E, Courtis G. An inter- and intra-examiner reliability study of motion palpation of the lumbar spine in lateral flexion in the seated position. Euro J Chiropractic 1986; 134:121–141.**

The study evaluated the inter- and intra-examiner reliability for assessment of spinal segmental mobility in the lumbar spine. Two chiropractic students, pre-trained in the examination procedure examined 100 randomly selected students. Inter-examiner concordance was 81.8% for positive findings based on both level and direction; concordance was 74.8% when only one level of spinal segment was examined. Intra-examiner reliability based on level and direction was 95.4% for both examiners independently.

**M13 Love RM, Bordeur RR. Inter-examiner and intra-examiner reliability of motion palpation for the thoracolumbar spine. J Manipulative Physiol Ther 1987; 10:1–4.**

The study evaluated intra- and inter-examiner reliability using a standardized motion palpation scan on the thoracic and lumbar spine. Eight senior chiropractic students using standardized motion palpation as described by Gillet and Liekens examined 32 asymptomatic chiropractic student volunteer subjects. Intra-examiner reliability using the Pearson Product Moment Correlation Coefficient ( $r$ ) revealed significant levels of agreement for each of the examiners ( $r = 0.300$ ;  $p < 0.05$ ). Inter-examiner reliability using the Index of Association Statistic ( $R$ ) was not statistically significant.

**M14 Boline P, Keating J, Brist J, Denver G. Inter-examiner reliability of palpatory evaluations of the lumbar spine. Am J Chiropractic Med 1988; 1:5–11.**

The study investigated inter-examiner reliability of palpatory procedures for intervertebral motion, paraspinal tonicity and pain in the lumbar spine. Two non-blinded examiners, one a chiropractic intern and the other a recent chiropractic graduate, examined 23 symptomatic low back pain patients and 27 voluntary asymptomatic subjects. The Kappa and percent agreement scores were calculated. There was poor inter-examiner reliability on lumbar palpatory procedures with none of the dimensions signifi-

cant at all segmental levels. Pain was slightly more reliable than motion or tonicity evaluation. Examiners agreed more on abnormal findings of the upper and middle lumbar spine.

**M15 Rhudy T, Sandefur M, Burk J. Inter-examiner intertechnique reliability in spinal subluxation assessment: a multifactorial approach. Am J Chiropractic Med 1988; 1:111–114.**

The study evaluated inter-examiner reliability in detecting spinal dysfunction. Three chiropractors examined 14 symptomatic volunteer patients using full spine postural radiography, vertebral motion palpation and Gonstead chiropractic diagnostic procedures. Reliability was scored by percentage of strength of agreement (kappa score/number of subjects X100). Overall reliability was poor, although the level of agreement differed according to spinal area examined.

**M16 Nansel DD, Peneff AL, Jansen RD, Cooperstein R. Inter-examiner concordance in detecting joint-play asymmetries in the cervical spines of otherwise asymptomatic subjects. J Manipulative Physiol Ther 1989; 12:428–433.**

The study assessed inter-examiner reliability for motion-based palpation of the cervical spine. Two pairs of chiropractors examined 270 healthy, pain-free chiropractic students for joint play asymmetries. Kappa score for inter-examiner reliability from pooled data was 0.013 reflecting little difference from what would be expected by chance. The authors suggest that motion based palpation of the cervical spine may not be a reliable predictor of vertebral dysfunction in healthy patients.

**M17 Mootz RD, Keating JC, Kontz HP, Milus TB, Jacobs GE. Intra observer and inter-observer reliability of passive motion palpation of the lumbar spine. J Manipulative Physiol Ther 1989; 12:440–445.**

The study assessed the intra- and inter-examiner reliability of passive motion palpation in detecting joint fixation (a pair of consecutive vertebrae that have decreased motion) in the lumbar spine. Two experienced chiropractors evaluated 60 symptomatic and asymptomatic chiropractic students. Kappa scores of paired lumbar segments for intra-examiner reliability ranged from  $-0.11$  to  $0.46$ ; for

inter-examiner reliability, the scores were  $-0.19$  to  $0.17$ . There was poor inter-examiner reliability for all segments with fair to moderate intra-examiner reliability at the L1–2 and the L4–5 segments.

**M18 Leboeuf C, Gardner V, Carter A, Scott T. Chiropractic examination procedures: A reliability and consistency study. J Austral Chiropractor Assoc 1989; 19:101–104.**

The study investigated intra- and inter-examiner reliability for certain chiropractic tests: pain on spinous process palpation, interspinous ligament palpation and spinous process percussion, and motion palpation, sign of the rising thumb and resiliency on extension in the lumbar spine. Two chiropractic students examined 39 subjects with chronic low back pain. The ability of examiners to agree on the presence/absence of positive findings in these subjects was generally good. Only motion palpation had significantly better intra-examiner agreement when compared with inter-examiner agreement values at the first visit. Some tests had significantly better results at the fifth visit than at the first visit. There was also a high inter-examiner rate of agreement per segment ( $> 70\%$ ) with the majority of consensus being on negative findings.

**M19 Keating JC, Jr. Bergmann TF, Jacobs GE, Finer BA, Larson K. Inter-examiner reliability of eight evaluative dimensions of lumbar segmental abnormality. J Manipulative Physiol Ther 1990; 13:463–470.**

The study examined the inter-examiner reliability in detecting lumbar spine segmental abnormality. Three experienced chiropractors used uniform tests to examine 25 asymptomatic and 21 symptomatic low back pain patients. Marginal-to-good (Kappa =  $0.19$  to Kappa =  $0.48$ ) and significant concordance for pain on palpation over osseous structures was noted for all examiner-pairs at most segmental levels. Slightly weaker and less frequent, but still marginal-to-good (Kappa =  $0.10$  to K =  $0.59$ ; mean Kappa =  $0.30$ ) and significant agreement between examiners was noted for pain in soft-tissues at most segments. The strongest concordance for pain findings tended to occur at L4–5 and L5–S1. Tests for passive range of motion, misalignment, muscle tension and active/passive motion fixation produced low inter-examiner reliability.

**M20 Hardy G, Napier J. Inter- and intra-therapist reliability of passive accessory movement technique. New Zealand J Physiother 1991; 19:22–24.**

The study analyzed inter and intra-examiner reliability in performing passive accessory movement examination on a materials testing machine. Five physiotherapists specializing in manual therapy and two physiotherapy students were trained in the Maitland Grades of Passive Accessory Movement using a machine model. Bartlett's test of intra-examiner reliability was statistically significant ( $p < 0.01$ ) only for mobilization grades I and IV; F-tests for inter-examiner reliability revealed a significant variability for all grades ( $p < 0.001$ ). Results demonstrate significant differences for both inter- and intra-examiner variability in performing each grade of mobilization.

**M21 Mastriani P, Woodman K. Reliability of passive lumbar segmental motion: MGH Institute of Health Professions, 1991 (Thesis).**

The study determined inter-examiner reliability of passive intervertebral motion tests of the lumbar spine. Three experienced physical therapists, trained in a uniform test procedure, examined 16 volunteers with low back pain. Percent agreement was  $66\%$  between examiners using a 3-point rating scale, and  $62\%$  using a 7-point scale. Regardless of scale and direction of movement, L3–4 ( $71\%$ ) had the greatest reliability, followed by L4–5 ( $63\%$ ) and L5-S1 ( $58\%$ ).

**M22 Richter T and Lawall J. Reliability of diagnostic findings in manual medicine. Manuelle Medizin, 1993; 31:1–11.**

This pilot study (in German) investigated the inter- and intra-examiner reliability of a three-step diagnostic procedure (testing joint movements, identifying irritation points and conducting provocative testing of irritation points) as well as muscle tests in the lumbar-pelvic region. Five physicians trained in manual medicine examined 61 symptomatic ambulatory patients with low-back pain. The 5 examiners examined each patient within one hour. One of the examiners was the investigator; 4 additional examiners were grouped for comparison as one "examiner x". Kappa statistics were calculated. Inter-examiner reliability of left side bending at L1 and L2 were good ( $0.69$ – $0.72$ ). Other range of motion tests at each lumbar spinal level showed poor to moderate reliability ( $0.08$ – $0.47$ ). Identifying irrita-

tion points and provocative testing of irritation points for pain varied widely from poor to good inter-examiner agreement (0.00–0.65). Authors stated that intra-examiner agreement was very good (0.8).

**M23 Maher C, Adams R. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. *Phys Ther* 1994; 74:801–811.**

The study evaluated inter-examiner agreement in diagnosing joint stiffness and pain in the lumbar spine. Six experienced physical therapists examined 90 patients with low back pain, using the Posteroanterior (PA) Central Pressure test. The percent agreement between examiner's judgments of the subject's pain ranged from 31 to 43%. Intraclass Correlation Coefficients (ICC) ranged from 0.67 to 0.72. Inter-examiner agreement for joint stiffness was 21 to 29% (ICC ranged from 0.03 to 0.37). There was, therefore, good reliability between examiners for assessment of a subject's pain, but not for joint stiffness.

**M24 Binkley J, Stratford PW, Gill C. Inter-rater reliability of lumbar accessory motion mobility testing. *Phys Ther* 1995; 75:786–792.**

The study evaluated inter-rater reliability of accessory motion mobility testing of the lumbar spine in subjects with low back pain. Six orthopedic physical therapists examined 18 subjects with chronic mechanical low back pain. Kappa scores and Intraclass Correlation Coefficient (ICC) statistical analyses were used. The ICC for determination of the marked level was  $R_{(2,1)} = 0.69$  (95% confidence interval = 0.53–0.82). The ICC for mobility findings at the marked level was  $R_{(2,1)} = 0.25$  (95% confidence interval = 0–0.44). A secondary Kappa analysis, to determine agreement on treatment decision-making, demonstrated similarly low levels of agreement. There is poor inter-rater reliability of posterior-anterior (P-A) accessory motion mobility testing in the absence of corroborating clinical data. There was poor inter-examiner agreement determination of the segmental level of a marked spinous process.

**M25 Insoe E, Witt P, Gross M, Mitchell R. Reliability in evaluating passive intervertebral motion of the lumbar spine. *J Man & Manipulative Ther* 1995; 3:135–143.**

The study assessed intra- and inter-examiner reliability in the use of passive intervertebral motion (PIVM) measure-

ments of the lumbar spine during forward bending. Two blinded, experienced physical therapists examined 6 subjects with lower back pain. Using percent agreement and Scott's pi ratio, there was 66.67% and 75% intra-examiner agreement respectively, with agreement greater than chance of 41.89% and 61.29% respectively. Inter-examiner agreement was 48.61% and agreement greater than chance was 18.35%. The authors conclude that the study does not offer strong support for inter-examiner reliability using PIVM in the lumbar spine.

**M26 Haas M, Raphael R, Panzer D, Peterson D. Reliability of manual end-play palpation of the thoracic spine. *Chiropractic Tech* 1995; 7:120–124.**

The study assessed inter-examiner reliability in performing end-play palpation on the thoracic spine. Two experienced chiropractors examined 73 asymptomatic and symptomatic volunteers using a validated and referenced test procedure. A blind randomized repeated procedures design found poor inter-examiner reliability with Kappa scores ranging from 0.14 to 0.19. Intra-examiner reliability was found to be moderate with Kappa scores averaging 0.55 and 0.43 for the respective examiners. The authors report that these results were consistent with the findings of previous studies in other regions of the spine.

**M27 Lindsay DM, Meeuwisse WH, Mooney ME, Summersides J. Inter-rater reliability of manual therapy assessment techniques. *Physiotherapy Canada, Summer* 1995; 4(3):173–180.**

The study, in part, examined inter-examiner reliability of 20 standard active and passive lumbar motion tests. Two experienced physical therapists specializing in manual therapy examined 8 cross-country ski team volunteers. Percent agreement and Kappa statistics were calculated. Seventy-percent agreement level was adopted as the minimum criteria for acceptable inter-examiner test reliability. This level was achieved on 8 of the 20 tests, with 3 tests (compression and right and left torsion) achieving 100% agreement (Kappa not calculated). Kappa scores for 16 of the remaining 17 tests showed poor reliability (–0.5 to 0.3). The active flexion test had moderate to good reliability (0.6). Agreement was most likely when test outcomes were negative. Disagreement on positive findings occurred 3 times more often than agreement.



**M28 Tuchin P, Hart C, Johnson C, et al. Inter-examiner reliability of chiropractic evaluation for cervical spine problems – a pilot study. Part 1: Graduates from one institution. *Australian Chiropractic & Osteopathy* 1996; 5:23–29.**

The study evaluated the reliability of examiners in palpation of the cervical spine. Eight experienced chiropractors, using individual palpatory methods including static and/or motion palpation, and vertebral springing, examined 53 asymptomatic volunteer student subjects for cervical spine dysfunction. There was poor inter-examiner agreement with C6 being the segment of the highest disagreement.

**M29 Phillips DR, Twomey LT. A comparison of manual diagnosis with a diagnosis established by a uni-level lumbar spinal block procedure. This study was presented in part at the 8th Biennial Conference of the MPAA in 1993. *Manual Therapy* 1996; 1:82–87.**

This study investigated the inter-examiner reliability and validity of lumbar spine manual palpation in the diagnosis of patients with low back pain using a randomized cross-over design with a prospective and retrospective part. Two manipulative physiotherapists evaluated 63 symptomatic and 9 asymptomatic volunteer subjects for abnormal quantity and quality of passive intervertebral motion and vertebral response to digital pressure. The authors used percent agreement, Kappa and weighted Kappa analysis to determine inter-examiner reliability. There was poor inter-therapist reliability for motion ratings (weighted Kappa ranged from  $-0.15$  to  $0.32$ ) and vertebral response to pressure (Kappa ranged from  $-0.16$  to  $0.28$ ).

**M30 McPartland JM, Goodridge JP. Counterstrain and traditional osteopathic examination of the cervical spine compared. *J Bodywork Movt Ther* 1997; 1:173–178.**

The study evaluated inter-examiner reliability of strain-counterstrain (S-CS) diagnosis (specific tender points) and other osteopathic palpatory tests (segmental motion and soft tissue texture changes) in the cervical spine. Two experienced osteopathic physicians specializing in manipulation examined 7 symptomatic and 11 asymptomatic patients. Inter-examiner reliability for S-CS tender points was 72.7% agreement (Kappa = 0.45) for symptomatic patients and 59.4% agreement (Kappa = 0.19) for asymptomatic patients.

The inter-examiner reliability for the motion and soft tissue feel exam was 67.5% agreement (Kappa = 0.45) for symptomatic patients and 73.7% agreement (Kappa = 0.34) for asymptomatic patients. Tender points were a more reliable indicator in symptomatic patients than in asymptomatic patients.

**M31 Strender LE, Sjoblom A, Sundell K, Ludwig R, Taube A. Inter-examiner reliability in physical examination of patients with low back pain. *Spine* 1997; 22:814–820.**

The study evaluated inter-examiner reliability of 29 agreed upon clinical test procedures commonly used in the evaluation of patients with low back pain, including 7 palpatory vertebral motion and tenderness tests of the lumbar spine. Two experienced physical therapists certified in manual medicine that worked together for many years and 2 physicians, unfamiliar with each other, but both experienced in caring for the selected patient population, examined 71 patients with low back pain. The two pairs of examiners were evaluated for inter-examiner reliability separately. The authors calculated percent agreement, Kappa coefficient and the 95% confidence interval for the difference in prevalence of pathological findings, as an evaluation of inter-examiner bias. All but two palpatory test procedures had poor reliability. Both groups had moderate agreement (Kappa coefficient  $> 0.40$ ) only on intersegmental tenderness evaluation. Using weighted Kappa at the 95% confidence interval, only the physical therapy pair had moderate agreement on intersegment mobility of the lumbosacral joint (Kappa = 0.75; 0.60–0.90) and the segment above the lumbosacral joint (Kappa = 0.66; 0.45–0.86).

**M32 Olson KA, Paris SV, Spohr C, Gorniak G. Radiographic assessment and reliability study of the craniovertebral side-bending test. *J Manual & Manipulative Ther* 1998; 6:87–96.**

The study analyzed the effects of patient positioning on the inter-examiner reliability of passive motion testing and end-feel assessment of the cervical spine. Six physical therapists assessed 10 asymptomatic subjects for passive cranial vertebral side bending in five positions. Kappa scores on mobility grade assessment ranged from  $-0.031$  to  $0.182$  and  $-0.022$  to  $0.137$  for inter-examiner and intra-examiner reliability respectively. For end-feel assessment, Kappa scores ranged from  $-0.043$  to  $0.194$  and  $0.01$  to

0.308 for inter-examiner and intra-examiner reliability respectively. The authors concluded that there was poor inter and intra-examiner reliability for the cranio-vertebral side bending passive motion test.

**M33 Maher CG, Latimer J, Adams R. An investigation of the reliability and validity of posteroanterior spinal stiffness judgments made using a reference-based protocol. *Phys Ther* 1998; 78:829–837.**

The reliability portion of this study assessed the agreement in stiffness estimates in the lumbar spine. Two blinded experienced physical therapists, examined 40 asymptomatic volunteers at L3, using their preferred palpation method. Interclass correlation coefficient (ICC) with 95% confidence intervals was 0.5–0.62 with a low standard error of measurement.

**M34 Hawk C, Phongphua C, Bleecker J, Swank L, Lopez D, Rubley T. Preliminary study of the reliability of assessment procedures for indications for chiropractic adjustments of the lumbar spine. *J Manipulative Physiol Ther* 1999; 22:382–389.**

The study evaluated inter- and intra-examiner reliability in assessing indications for chiropractic adjustment of the lumbar spine. Four licensed chiropractic examiners (2 with 20+ and 2 with 3 or less years of experience), examined 18 (2 symptomatic and 16 asymptomatic) volunteer subjects. Examiners were trained in a standardized flexion-distraction technique. They also used manual assessment procedures used in everyday clinical practice including 1) hypo or hyper mobility of each segment, 2) changes in tissue texture or tension of the skin and underlying tissue, 3) palpable temperature changes, and 4) tenderness elicited on palpation. Intra-examiner reliability Kappa scores ranged from –0.17 to 0.85. For intra-examiner reliability there was considerable variation by segment and among the four examiners. Inter-examiner reliability scores ranged from –0.42 to 0.44. The authors noted that intra-examiner reliability appeared to be greater than inter-examiner reliability and that training of examiners in a standardized assessment procedure did not enhance agreement on specific segments of the spine to be adjusted.

**M35 Fjellner A, Bexander C, Faleij R, Strender L-E. Inter-examiner reliability in physical examination of the cervical spine. *J Manipulative Physiol Ther* 1999; 22:511–516.**

The study evaluated the reliability of examination procedures used commonly by physiotherapists in patients with neck problems. Two blinded physiotherapists examined 47 symptomatic and asymptomatic subjects using cervical standard regional and intersegmental passive range of motion and end-feel assessment procedures. Weighted Kappa statistics and percent agreement were used. Acceptable inter-examiner reliability (weighted Kappa > 0.40) was noted in six of eight regional range of motion tests (extension, right and left lateral flexion, right and left rotation, right rotation at maximum flexion). For regional motion tests of end-feel, only 3 (left and right rotation and right lateral flexion) of 8 showed acceptable reliability. Generally, there was poor reliability for 58 passive intersegmental procedures (only 5 tests obtained acceptable weighted Kappa values). More clinical tests were reliable among symptomatic vs. asymptomatic patients.

**M36 Lundberg G, Gerdle B. The relationships between spinal sagittal configuration, joint mobility, general low back mobility and segmental mobility in female homecare personnel. *Scan J Rehab Med* 1999; 31:197–206.**

The study reported on the reliability of manual palpation of mobility and pain provocation in the thoracic and lumbar spine. Two blinded experienced physical therapists, examined 156 subjects using a pre-determined protocol. The weighted Kappa scores for inter-rater reliability varied from 0.42 to 0.75 for segmental mobility; 0.71 at L4–L5 and 0.67 at L5–S1 for pain provocation. The authors concluded that manual testing of the lumbar spine, using a highly structured protocol, was reliable but might not generalize to clinical practice.

**M37 French S, Green SAF. Reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low back pain. *J Manipulative Physiol Ther* 2000; 23:231–238.**

The study analyzed intra- and inter-examiner reliability of commonly used chiropractic procedures. Five chiropractors examined 19 patients with chronic mechanical low back pain using neurological, motion palpation and x-ray

procedures. Overall, intra-examiner agreement on decision to manipulate ranged from 73 to 92 percent, with Kappa values of 0.13 to 0.73. Inter-examiner reliability was low at lower thoracic and lumbar spinal levels, with the mean percent agreement ranging from 48 to 83% and Kappa values ranging from -0.16 to 0.27.

**M38 Schöps P, Siebert U, Schmitz U, Friedle AM, Beyer A. Reliabilität nichtinvasiver diagnostischer Untersuchungsmethoden zur Erfassung schmerzhafter Halswirbelsäulensyndrome. Manuelle Medizin 2000; 38(1):17–32. (German)**

This German study evaluated inter-rater reliability of procedures for painful musculo-skeletal conditions and dysfunctions of the cervical spine. Five physicians examined 20 symptomatic and 20 asymptomatic patients, using a standardized and blinded protocol. Inter-examiner reliability, as measured by Kappa scores, ranged from poor to very good agreement. Inter-examiner reliability on inspection was moderate (0.29 to 0.52), on palpation to elicit pain was poor to moderate (0.2 to 0.4) except at T1 where it was good (0.6–0.75). Reliability of palpation of segmental dysfunction was poor to moderate (0.2 to 0.4). Mobility assessment showed good to very good (0.6–0.8) inter-examiner reliability.

**M39 Smedmark V, Wallin M, Arvidsson I. Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. Manual Therapy, 2000; 5(2):97–101.**

The study examined inter-examiner reliability of 3 common passive intervertebral motion procedures in the cervical spine region. Two experienced physical therapists that worked together for 17 years using the same procedures, examined 61 symptomatic patients with mechanical neck problems at their private clinic. Percent agreement and Kappa statistics showed moderate inter-examiner reliability for all three tests. C1–C2 right rotation had 87% agreement and Kappa = 0.28; C2–C3 lateral flexion had 70% agreement and Kappa = 0.43; and C7 flexion/extension had 70% agreement and Kappa = 0.36.

**M40 Van Suijlekom HA, De Vet HC, Van Den Berg SG, Weber WE. Interobserver reliability in physical examination of the cervical spine in patients with headache. Headache 2000; 40:581–586.**

The study evaluated inter-examiner reliability for palpation of the cervical spine in symptomatic patients. Two experienced neurologists examined 24 patients diagnosed with cervicogenic headache. There was moderate inter-examiner reliability (Kappa scores ranged from 0.28 to 0.46) for range of motion, and moderate to good reliability in provocation of pain during movement (0.53 to 0.67). Slight to fair agreement was obtained for zygapophyseal joint pressure pain (0.14 to 0.37). Reliability was variable for elicitation of tender points by digital pressure (0.00 to 1.00). Good reliability (Kappa > 0.6) was obtained for pressure pain on the mastoid process and 3–4 cm posterior to the sternocleidomastoid muscle border.

**M41 Comeaux Z, Eland D, Chila A. Measurement challenges in physical diagnosis: refining inter-rater palpation, perception and communication. J Bodywork & Movt Ther 2001; 5:245–253.**

The study assessed inter-examiner reliability of palpatory findings and diagnosis using a well-defined protocol. Comparisons of the tissue texture cues at rest and during gentle provocative regionally induced passive motions were used to assess dysfunction. Three experienced (10+ years of clinical experience) osteopathic physicians specializing in manipulation examined 54 asymptomatic volunteers according to a consensus procedure for lower cervical and upper thoracic spine palpation. Inter-examiner reliability was poor to moderate (Kappa coefficients ranged from 0.12 to 0.56) between various tests and examiners.

**M42 Marcotte J, Normand MC. Standardisation de la palpation dynamique en chiropractique: une etude de fiabilité pour la région cervicale. J Can Chiropr Assoc 2001; 45(2):106–112. (French)**

The study examined the inter- and intra-examiner reliability of 6 passive vertebral segmental motion tests in the cervical spine. Three blinded experienced chiropractors, two trained and one not trained in the procedures used, evaluated 12 patients with chronic mechanical neck problems. Percent agreement and Kappa calculations showed moderate to very good inter-examiner reliability. Inter-

examiner reliability between the two trained examiners was greater (93% agreement; Kappa = 0.85) than for either of these two with the (third) untrained examiner, who used his own test procedures (82% and 84% agreements; Kappa = 0.57 and 0.61 respectively). Intra-examiner reliability for one of the trained examiners was good (91% agreement; Kappa = 0.78).

#### **Pain or sensitivity provocation procedures**

**P1 Waddell G, Main CJ, Morris EW, et al. Normality and reliability in the clinical assessment of backache. *BMJ (Clin Res Ed)* 1982; 284:1519–1523.**

The study evaluated the inter-examiner reliability of history and physical examination procedures and clinical assessment in patients with back pain. Five orthopedic surgeons examined 810 patients with backache. This included elicitation of lumbar tenderness by spinal palpation. Most study groups compared two examiners. The un-weighted Kappa scores for reliability of spinal palpation for tenderness on 8 patients was 1.0 ( $p < 0.001$ ). The un-weighted kappa scores for reliability on physical exam ranged from 0.41 to 1.0 and on psychological and behavioral assessment from 0.27 to 0.94 ( $p < 0.05$ ).

**P2 DeBoer K, Harmon R, Tuttle C, Wallace H. Reliability study of detection of somatic dysfunctions in the cervical spine. *J Manipulative Physiol Ther* 1985; 8:9–16.**

See M10

**P3 Viikari-Juntura E. Inter-examiner reliability of observations in physical examinations of the neck *Phys Ther* 1987; 1526–1532.**

The study assessed inter-examiner reliability of palpatory procedures of the cervical spine. A physical medicine and rehabilitation specialist (physiatrist) and a physical therapist examined 69 consecutive symptomatic patients using a conventional neurological evaluation, palpation, and evocative tests for pain, numbness and paresthesias. Agreement on palpation for tenderness was reported for 51 subjects using empirical value of Kappa scores and proportion of significant agreement ( $p_s$ ). The Kappa score was fair for upper spinous processes at 0.47 ( $p_s$  0.56), as well as for lower spinous processes at 0.52 ( $p_s$  0.67); however, was poor for right cervical paraspinal soft tissues at 0.24

( $p_s$  0.33). For left cervical paraspinal soft tissues the prevalence was less than 10% so Kappa was not used ( $p_s$  0.00).

**P4 Boline P, Keating J, Brist J, Denver G. Inter-examiner reliability of palpatory evaluations of the lumbar spine. *Am J Chiropractic Med* 1988; 1:5–11. See M14**

**P5 Leboeuf C, Gardner V, Carter A, Scott T. Chiropractic examination procedures: A reliability and consistency study. *J Austral Chiropractor Assoc* 1989; 19:101–104. See M18**

**P6M Keating JC, Jr. Bergmann TF, Jacobs GE, Finer BA, Larson K. Inter-examiner reliability of eight evaluative dimensions of lumbar segmental abnormality. *J Manipulative Physiol Ther* 1990; 413:463–70. See M19**

**P7M Nice DA, Riddle DL, Lamb RL, Mayhew TP, Rucker K. Intertester reliability of judgments of the presence of trigger points in patients with low back pain. *Arch Phys Med Rehabil* 1992; 73:893–898.**  
The study evaluated inter-examiner reliability in assessing the presence of trigger points in the lumbar spine region. Twelve experienced physical therapists evaluated 50 patients with low back pain using the Travell and Simon assessment examination using pain as the endpoint. The Kappa score for inter-examiner reliability ranged from 0.29 to 0.38; percent agreement ranged from 76% to 79%; the observed proportion of positive agreement ranged from 0.43 to 0.52. The authors concluded that there was poor inter-examiner reliability in the assessment of the presence of trigger points in patients with low back pain.

**P8 Boline PD, Haas M, Meyer JJ, Kassak K, Nelson C, Keating JC Jr. Inter-examiner reliability of eight evaluative dimensions of lumbar segmental abnormality: Part II. *J Manipulative Physiol Ther* 1993; 16:363–374.**

The study evaluated inter-examiner reliability using several measurements including lumbar spinal palpatory procedures. Three experienced chiropractors examined 28 symptomatic patients with chronic low back pain. Palpa-



tion for osseous pain produced percent agreement that ranged from 79% to 96% with Kappa coefficients ranging from 0.48 to 0.98. Palpation for soft tissue pain produced percent agreement ranging from 75% to 93% with Kappa coefficients ranging from 0.40 to 0.79. Good reliability was demonstrated for inter-examiner provocative palpatory procedures for elicitation of both osseous and paraspinal soft tissue pain in the study population.

**P9 Richter T and Lawall J. Reliability of diagnostic findings in manual medicine. *Manuelle Medizin*, 1993; 31:1–11.**  
See M22

**P10 Hubka MJ, Phelan SP. Inter-examiner reliability of palpation for cervical spine tenderness. *J Manipulative Physiol Ther* 1994; 17:591–595.**  
The study evaluated the inter-examiner reliability of palpation for cervical spine tenderness using a within subjects (repeated measures) design. Two experienced chiropractors examined 30 patients with mechanical neck pain. Inter-examiner reliability as assessed by percent agreement was 76.6% with a Kappa score of 0.68. The authors found that manual palpation of the cervical spine for tenderness is a reliable examination tool.

**P11 Maher C, Adams R. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. *Phys Ther* 1994; 74:801–811.**  
See M23

**P12 McPartland JM, Goodridge JP. Counterstrain and traditional osteopathic examination of the cervical spine compared. *J Bodywork Movt Ther* 1997; 1:173–178.**  
See M30

**P13 Strender LE, Lundin M, Nell K. Inter-examiner reliability in physical examination of the neck. *J Manipulative Physiol Ther* 1997; 20:516–520.**  
The study evaluated inter-examiner reliability of cervical spinal palpatory tests. Two experienced physiotherapists examined 50 volunteer symptomatic and asymptomatic subjects using 10 tests agreed upon by the examiners to determine if there was a difference between left and right sides. A Kappa coefficient value > 0.40 was considered

as acceptable (“moderate”) for agreement greater than chance. Only two tests met this criterion; one of three tests used in assessment of pain on palpation and the foramen compression test.

**P14 Strender LE, Sjoblom A, Sundell K, Ludwig R, Taube A. Inter-examiner reliability in physical examination of patients with low back pain. *Spine* 1997; 22:814–820.**  
See M31

**P15 Hawk C, Phongphua C, Bleecker J, Swank L, Lopez D, Rubley T. Preliminary study of the reliability of assessment procedures for indications for chiropractic adjustments of the lumbar spine. *J Manipulative Physiol Ther* 1999; 22:382–389.**  
See M34

**P16 Lundberg G, Gerdle B. The relationships between spinal sagittal configuration, joint mobility, general low back mobility and segmental mobility in female homecare personnel. *Scan J Rehab Med* 1999; 31:197–206.**  
See M36

**P17 French S, Green SAF. Reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low back pain. *J Manipulative Physiol Ther* 2000; 23:231–238.**  
See M 37

**P18 Hsieh C-YJ, Hong C-Z, Adams AH, et al. Inter-examiner reliability of the palpation of trigger points in the trunk and lower limb muscles. *Arch Phys Med Rehab* 2000; 81:258–264.**  
The study assessed inter-examiner reliability of palpation of lumbar paraspinal trigger points. A physiatrist experienced in trigger point evaluation, 4 chiropractic residents, and 4 physiatry residents examined 52 symptomatic and asymptomatic patients. Inter-examiner reliability for tenderness was best among the trained examiners (Kappa = 0.44). Inter-examiner reliability was poor for other parameters (taut band Kappa = 0.13 and local twitch Kappa = 0.106).

**P19 Schöps P, Siebert U, Schmitz U, Friedle AM, Beyer A. Reliabilität nichtinvasiver diagnostischer Untersuchungsmethoden zur Erfassung schmerzhafter Halswirbelsäulensyndrome. Manuelle Medizin 2000; 38(1):17–32. (German) See M38**

**P20 Van Suijlekom HA, De Vet HC, Van Den Berg SG, Weber WE. Interobserver reliability in physical examination of the cervical spine in patients with headache. Headache 2000; 40:581–586. See M40**

#### Paraspinal soft tissue palpation procedures

**ST1 Beal MC, Goodridge JP, Johnston WL, McConnell DG. Inter-examiner agreement on patient improvement after negotiated selection of tests. J Am Osteopath Assoc 1980; 79:432–440. See M3**

**ST2 Beal MC, Goodridge JP, Johnston WL, McConnell DG. Inter-examiner agreement on long-term patient improvement: an exercise in research design. J Am Osteopath Assoc 1982; 81:322–328. See M6**

**ST3 Johnston WL, Allan BR, Hendra JL, et al. Inter-examiner study of palpation in detecting location of spinal segmental dysfunction. J Am Osteopath Assoc 1983; 82:839–845.**  
The study assessed the reliability of trained examiners in distinguishing presence of deep tissue tension in the thoracic spine and in grading its intensity. An osteopathic faculty physician specializing in manipulation and 5 osteopathic medical students examined 31 volunteer patients. The percent agreement for dullness, decreased rebound, and degree of tightness was 79% to 86%. The authors noted that the standing position proved more stable and suitable for providing standardized landmarks. In this position, a significant gradient was established separating the degree of more major from the less major findings.

**ST4 DeBoer K, Harmon R, Tuttle C, Wallace H. Reliability study of detection of somatic dysfunctions in the cervical spine. J Manipulative Physiol Ther 1985; 8:9–16. See M10**

**ST5 Viikari-Juntura E. Inter-examiner reliability of observations in physical examinations of the neck. Phys Ther 1987; 1526–1532. See P3**

**ST6 Boline P, Keating J, Brist J, Denver G. Inter-examiner reliability of palpatory evaluations of the lumbar spine. Am J Chiropractic Med 1988; 1:5–11. See M14**

**ST7 Leboeuf C, Gardner V, Carter A, Scott T. Chiropractic examination procedures: A reliability and consistency study. J Austral Chiropractor Assoc 1989; 19:101–104. See M18**

**ST8 Keating JC, Jr. Bergmann TF, Jacobs GE, Finer BA, Larson K. Inter-examiner reliability of eight evaluative dimensions of lumbar segmental abnormality. J Manipulative Physiol Ther 1990; 13:463–470. See M19**

**ST9 McPartland JM, Goodridge JP. Counterstrain and traditional osteopathic examination of the cervical spine compared. J Bodywork Movt Ther 1997; 1:173–178. See M30**

**ST10 Rouwmaat PHM, Everaert D, Stappaerts KH, Aufdemkampe G. Reliability of manual skinfold tests in a healthy male population. J Manipulative Physiol Ther. 1998; 21(5):327–332.**

This study examined the inter- and intra-examiner agreement of the paraspinal skinfold thickness/compliance test in the evaluation of thoracic spinal dysfunction. Twelve experienced manipulative physical therapists that practiced the test procedure prior to the study, examined 12 volunteer healthy students at marked thoracic spinal vertebral levels T3, 6, 9 and 12. Intraclass correlation coeffi-

cient (ICC) at the 95% confidence interval was calculated. Intra-examiner reliability was poor to moderate for both skinfold thickness (0.25) and compliance (0.28). Inter-examiner reliability was poor for both skinfold thickness (0.08) and compliance (0.12). Correlation between amount of time an examiner spent in practicing the test procedure and agreement level was not significant.

**ST11 Hawk C, Phongphua C, Bleecker J, Swank L, Lopez D, Rubley T. Preliminary study of the reliability of assessment procedures for indications for chiropractic adjustments of the lumbar spine. J Manipulative Physiol Ther 1999; 22:382–389. See M34**

**ST12 Eriksson EM, Mokhtari M, Pourmotamed L, Holmdahl L, Eriksson H. Inter-rater reliability in a resource-oriented physiotherapeutic examination. Physiotherapy Theory and Practice, 2000; 16:95–103.** This study, in part, evaluated the inter-examiner reliability of palpation of tissue tension of paraspinal muscles in the thoracic and lumbar regions. Two physical therapists were trained in, and practiced the exam procedures prior to the study. They each evaluated 19 healthy volunteers. Kappa and percent agreement were calculated. Inter-examiner reliability was very good (Kappa = 0.82; 94.7%) for lumbar and poor (Kappa = 0.16; 73.6%) for thoracic paraspinal muscle palpation.

**ST13 French S, Green SAF. Reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low back pain. J Manipulative Physiol Ther 2000; 23:231–238. See M37**

**ST14 Hsieh C-YJ, Hong C-Z, Adams AH, et al. Inter-examiner reliability of the palpation of trigger points in the trunk and lower limb muscles. Arch Phys Med Rehab 2000; 81:258–264. See P18**

**ST15 Schöps P, Siebert U, Schmitz U, Friedle AM, Beyer A. Reliabilität nichtinvasiver diagnostischer Untersuchungsmethoden zur Erfassung schmerzhafter Halswirbelsäulensyndrome. Manuelle Medizin 2000; 38(1):17–32. (German) See M38**

**ST16 Comeaux Z, Eland D, Chila A. Measurement challenges in physical diagnosis: refining inter-rater palpation, perception and communication. J Bodywork & Movt Ther 2001; 5:245–253. See M41**

**ST17 Ghoukassian M, Nicholls B, McGlaughlin P. Inter-examiner reliability of the Johnson and Friedman percussion scan of the thoracic spine. J Osteo Med, 2001; 4(1):15–20.**

The study examined inter-examiner reliability of a paraspinal percussion palpatory exam of the thoracic spine used to identify segmental tissue tension. Ten senior post-graduate Australian osteopathic students, experienced and trained in the standardized protocol, examined 19 asymptomatic male volunteers. Kappa scores showed poor (0.07) inter-examiner reliability.

#### **Landmark position assessment procedures**

**L1 Beal MC, Goodridge JP, Johnston WL, McConnell DG. Inter-examiner agreement on patient improvement after negotiated selection of tests. J Am Osteopath Assoc 1980; 79:432–440. See M3**

**L2 Keating JC, Jr. Bergmann TF, Jacobs GE, Finer BA, Larson K. Inter-examiner reliability of eight evaluative dimensions of lumbar segmental abnormality. J Manipulative Physiol Ther 1990; 13:463–470. See M19**

**L3 Byfield D, Humphreys K. Intra- and inter-examiner reliability of bony landmark identification in the lumbar spine. Eur J Chiropractic 1992; 72:13–17.**

The study assessed intra- and inter-examiner reliability in locating spinous processes of L1 and L4 by static palpa-

tion. Two experienced chiropractors examined 42 asymptomatic chiropractic students in the sitting and prone positions. The examiners used a skin-marking pen to identify the spinous processes. Intra-examiner percent agreement at L1 with the subject sitting vs. prone was 55% for one examiner and 39% for the other. At L4, intra-examiner percent agreement for both examiners was 62%. Inter-examiner percent agreement was better at L4 (sitting – 79%; prone – 81%) than at L1 (sitting – 55%; prone – 69%) with the subjects in either the sitting or prone positions.

**L4 Binkley J, Stratford PW, Gill C. Inter-rater reliability of lumbar accessory motion mobility testing. Phys Ther 1995; 75:786–792.**

See M24

**L5 McKenzie AM, Taylor NF. Can Physiotherapists locate lumbar spinal levels by palpation? Physiother 1997; 83:235–239.**

The study evaluated intra-examiner and inter-examiner reliability in locating lumbar spinal levels by palpation. Three physiotherapists (intra-examiner) and 14 physiotherapists (inter-examiner) examined 10 volunteer subjects, using their preferred method of palpation. Kappa scores for intra-examiner reliability were 0.61 to 0.90. Kappa scores for inter-examiner reliability was 0.28. There was good to excellent intra-examiner reliability, but poor inter-examiner reliability when palpating for lumbar spine levels.

**L6 Downey BJ, Taylor NF, Niere KR. Manipulative physiotherapists can reliably palpate nominated lumbar spinal levels. Man Ther 1999; 4:151–156.**

The study assessed inter-examiner reliability in palpating lumbar spine levels. Three pairs of experienced physical therapists palpated 60 patients with low back pain, marking the mid-point of a randomly nominated spinous process. Almost perfect overall agreement was achieved among all three pairs in locating the nominated level (weighted Kappa = 0.92).

## Unspecified types of tests

**U1 McConnell DG, Beal MC, Dinnar U, et al. Low agreement of findings in neuromusculoskeletal examinations by a group of osteopathic physicians using their own procedures. J Am Osteopath Assoc 1980; 79:441–450.**

The study examined the inter-examiner reliability in neuromuscular examination procedures, including spinal palpation. Six osteopathic physicians specializing in manipulation using their (unspecified) customary palpation procedures examined 21 symptomatic volunteers. Results revealed low inter-examiner reliability on segmental location and intensity of findings. The authors inferred that inter-examiner agreement would likely improve, if the examiners first agreed upon the following: a) the areas to be examined; b) the test procedures to be used; c) the method of quantifying the intensity of the findings; and d) the method of recording.

**U2 Beal M, Dvorak J. Palpatory examination of the spine: a comparison of the results of two methods and their relationship to visceral disease. Man Med 1984; 1:25–32.**

The study evaluated inter-examiner agreement using two methods of spinal palpation: the conventional American osteopathic and the Manual Medicine Society of Switzerland methods. Two physicians specializing in spinal manipulation, one an American osteopathic physician and the other a Swiss neurologist examined 50 patients in recovery from an acute visceral disease or with a chronic visceral disease. Each examiner recorded findings by which he had identified segmental dysfunction, i.e. changes in tissue texture, asymmetry of bony prominences and restriction in segmental spinal motion. No specific statistical analyses were offered. Results were presented in tabular and graphic form. Agreement level between the examiners improved if based on the absence as well as the presence of somatic dysfunction at each spinal level, and if the examination sites were grouped together.

**U3 Jull G, Zlot G, Trott P, Potter H, Shirley D, Richardson C. Inter-examiner reliability to detect painful upper cervical joint dysfunction. The Austral J of Physiother 1997; 43:125–129.**

The study analyzed inter-examiner reliability in the detec-



tion of painful upper cervical joint dysfunction. The findings of the chief investigator of the trial were compared with those of each of the other 6 experienced independent physiotherapists. The therapists examined 40 symptomatic (headache and neck pain) and asymptomatic volunteer subjects using their own personal test procedures. Additionally, some of the independent examiners were tested against each other. There was complete agreement ( $Kappa = 1.0$ ) in six pairs of examiners and excellent agreement ( $Kappa = 0.78$  and  $Kappa = 0.8$ ) between two pairs. Percent agreement was 70% for inter-examiner reliability on the most dysfunctional joint in symptomatic patients.

## References

- 1 Goldstein M.(ed): The Research Status of Spinal Manipulative Therapy. NINCDS Monograph No.15, DHEW Publication No. (NIH) 76-998, U.S. Department of Health, Education and Welfare, Bethesda, 1975.
- 2 Johnston WL. Interexaminer reliability studies: spanning a gap in medical research. Louisa Burns Memorial Lecture. J Am Osteopath Assoc 1982; 81:819–829.
- 3 Gonnella C, Paris SV, Kutner M. Reliability in evaluating passive intervertebral motion. Phys Ther 1982; 62:436–444.
- 4 Alley J. The clinical value of motion palpation as a diagnostic tool: a review. J Can Chiropr Assoc 1983; 27(3):97–100.
- 5 Russell R. Diagnostic palpation of the spine- a review of procedures and assessment of their reliability. J Manipulative Physiol Ther 1983; 6:181–183.
- 6 Keating J. Inter-examiner reliability of motion palpation of the lumbar spine: a review of quantitative literature. Am J Chiropractic Medicine 1989; 2(3):107–110.
- 7 Haas M. The reliability of reliability. J Manipulative Physiol Ther 1991; 14:199–208.
- 8 Panzer, DM. The reliability of lumbar motion palpation. J Manipulative Physiol Ther 1992; 15(8):518–524.
- 9 Hestbaek L, Leboeuf-Yde C: Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. J Manipulative Physiol Ther 2000; 23(4):258–275.
- 10 Huijbregts PA. Spinal motion palpation: a review of reliability studies. J Manual and Manipulative Therapy 2002; 10(1):24–39.
- 11 Murphy LS, Reinsch S, Najm WI, Dickerson VM, Seffinger MA, Adams A, Mishra SI. Spinal Palpation: The Challenges of Information Retrieval Using Available Databases. J Manipulative Physiol Ther [In press].
- 12 Jadad AR, Carrol D, Moore A, McQuay H, et al. Assessing the quality of reports of randomized controlled trials: is blinding necessary? Control Clin Trials 1996; 17:1–12.
- 13 Koes BW, Assendelft WJJ, van der Heijden GJMG, Bouter LM. Spinal manipulation for low back pain. An updated systematic review of randomized clinical trials. Spine 1996; 21(24):2860–2871.
- 14 Denslow JS. The supporting tissues in normal and abnormal joints. A preliminary report. J Am Osteopath Assoc Feb 1950; 49(6).
- 15 Koran LM. The reliability of clinical methods, data and judgments. N England J Med 1975; 293(14):695–701.
- 16 Haas M. Statistical methodology for reliability studies. J Manipulative Physiol Ther 1991; 14(2):119–132.
- 17 Haas M. How to evaluate intra-examiner reliability using an inter-examiner reliability study design. J Manipulative Physiol Ther 1995; 18(1):10–15.
- 18 Lantz CA. Application and evaluation of the kappa statistic in the design and interpretation of chiropractic clinical research. J Manipulative Physiol Ther 1997; 20(8):521–528.

## Movement, Stability and Low Back Pain – A Holistic Approach

Dr. Andry Vleeming PhD

Chairman: Interdisciplinary World Congress  
on Low Back Pain and Pelvic Pain

Founder and Director: Spine and Joint Centre,  
Rotterdam, The Netherlands

July 12 – 13, 2003

St. John's, Newfoundland

For information contact Dr. Mike Carstensen  
or Dr. Mark Pitcher at 709-738-7800 or email  
docmike@thezone.net