

## **Content Validity of Manual Spinal Palpatory Exams. A Systematic Review.**

### **Wadie I. Najm, MD**

Associate Clinical Professor

University of California, Irvine

101 The City Drive, Bldg. 200, #512, Rt. 81

Orange CA, 92868

714-456-5171

714 456-7984 (fax)

[winajm@uci.edu](mailto:winajm@uci.edu)

### **Michael A. Seffinger, D.O.**

Assistant Professor

Western University of Health Sciences

College of Osteopathic Medicine of the Pacific

309 E. Second Street / College Plaza

Pomona, CA 91766

[mseffinger@westernu.edu](mailto:mseffinger@westernu.edu)

### **Shiraz I. Mishra, MD, PhD**

Director, Research Development and Curriculum

Western University of Health Sciences

College of Osteopathic Medicine of the Pacific

309 E. Second Street

Pomona, CA 91766-1854

[smishra@earthlink.net](mailto:smishra@earthlink.net)

**Vivian Dickerson, MD**

Associate Clinical Professor

University of California, Irvine

101 The City Drive

Orange, CA 92868

[vdickerson@social.rr.com](mailto:vdickerson@social.rr.com)

**Alan Adams, DC, MS**

Faculty Administrator,

Florida State University

110J-Westcott Building

Tallahassee, FL 32306

[aadams@mailier.fsu.edu](mailto:aadams@mailier.fsu.edu)

**Sibylle Reinsch, PhD**

Research Associate, Dept. of PM&R

University of California, Irvine

101 The City Drive

Orange, CA 92868

[sreinsch@uci.edu](mailto:sreinsch@uci.edu)

**Linda L. Suk-Ling MURPHY, MLS**

Health Sciences Librarian

University of California, Irvine

231 Science Library

Irvine, CA 92697

[lmurphy@uci.edu](mailto:lmurphy@uci.edu)

**Arnold F. Goodman, Ph.D.**

Associate Director Center for Statistical Consulting

4900 Berkeley Place

Irvine, CA 92697

[agoodman@uci.edu](mailto:agoodman@uci.edu)

**Address correspondence to:**

Wadie I. Najm, MD

University of California, Irvine

101 The City Drive Bldg. 200 #512, Rt. 81

Orange, CA 92868

## Abstract

**Background:** Many health care professionals use spinal palpatory exams as a primary and well-accepted part of the evaluation of spinal pathology. However, few studies have explored the validity of spinal palpatory exams. To evaluate the status of the current scientific evidence, we conducted a systematic review to assess the content validity of spinal palpatory tests used to identify spinal neuro-musculoskeletal dysfunction. **Methods:** Review of eleven databases, and a hand search of peer-reviewed literature were undertaken from 1965-2001. Five papers met the inclusion/exclusion criteria. Two blinded reviewers abstracted pertinent data from the five papers, using a specially developed quality-scoring instrument. **Results:** Two papers focused on identifying the level of fixation and one focused on range of motion. All three studies used a mechanical model as a reference standard. Two papers explored the validity of pain assessment using the visual analogue scale or the subjects' own report as reference standards. Overall the sensitivity of studies looking at range of motion tests and pain varied greatly. Poor sensitivity was reported for range of motion studies regardless of the examiner's experience. A slightly better sensitivity (82%) was reported in one study that examined cervical pain. **Conclusions:** The lack of acceptable reference standards may have contributed to the weak sensitivity findings. Given the importance of spinal palpatory tests as part of the spinal evaluation and treatment plan, effort is required by all involved disciplines to create well-designed and implemented studies in this area.

## **Background:**

Injury of the spine and back are classified as the most frequent cause of limited activity among people younger than 45 years<sup>1-2</sup>. Approximately 10 percent of the adult population has neck pain at any one time<sup>3</sup>, and 80% of the population will experience low back pain (LBP) at some time in their lives<sup>4</sup>. Five to ten percent of the workforce is off work annually because of LBP. Indeed, LBP is second only to headache among the leading causes of pain. Approximately 80-90% of LBP is mechanical (non-organic musculoskeletal dysfunction) in origin<sup>5</sup>. Patients with mechanical spinal pain often seek and receive spinal manipulation by chiropractic, osteopathic and allopathic clinicians, physical therapists or other health care professionals<sup>6</sup>.

Health care professionals have utilized spinal palpatory diagnostic procedures and manual manipulative treatment for several millennia to treat back injury and pain<sup>7-8</sup>. Along with the history of illness and physical exam, examiners utilize specific spinal palpatory diagnostic tests in order to identify spinal neuromusculoskeletal dysfunction. Spinal neuromusculoskeletal dysfunction refers to an alteration of spinal joint position, motion characteristics and/or related palpable paraspinal soft tissue changes. Outcomes and effectiveness of spinal manipulative procedures rely on appropriate and skilled treatment that is based on an accurate diagnosis, which in turn depends upon the accuracy of the palpatory procedures used.

Spinal palpatory procedures have been described in journals<sup>9-11</sup> and textbooks<sup>12-19</sup>. Static palpation of anatomical landmarks for symmetry, palpation of spinal vertebral

joints before, during and after active and passive motion tests, spinal and paraspinal soft tissue palpatory assessment for abnormalities or altered sensitivity are most common.

Several narrative reviews of the literature on the validity and reliability of spinal palpatory diagnostic procedures have been published<sup>20-27</sup>. However, most reviews are discipline-specific despite the fact that similar spinal palpatory procedures are used across disciplines. Only two systematic reviews of spinal palpatory validity studies have been published<sup>28-29</sup>. One study was a limited review of chiropractic literature on palpatory diagnostic procedures for the lumbar-pelvic spine<sup>28</sup> and the other concentrates on validity studies at the sacroiliac joint<sup>29</sup>. An annotated bibliography<sup>30</sup> and a systematic review of the primary reliability research studies published between 1971 and 2001 are in progress.

Validity and reliability are concepts that are often used interchangeably, but the concepts are quite different. Validity is the accuracy of a measurement of the true state of a phenomenon<sup>31</sup>, while reliability measures the concordance, consistency or repeatability of outcomes<sup>24</sup>. However, even if a measurement is consistent and reliable, it is not necessarily valid (e.g., an arrow may consistently hit the target area, but never hit the bulls-eye).

There are various types of validity studies. The concept of validity differs in qualitative and quantitative research<sup>31</sup>. Though it can be argued that palpatory

diagnostic procedures are subjective and therefore qualitative, investigators in the field believe they can measure a physiological phenomenon that can be detected by objective means. They maintain that studies addressing the validity of spinal palpatory diagnostic tests are quantitative studies. The types of quantitative validity studies can be distinguished as follows: face validity; construct validity, criterion validity and content validity.

Face validity is the extent to which a test *appears* to measure what it is supposed to measure. In other words, whether the proposed test seems to provide a reasonable measure of the concept it is intended to measure. For example, spinal vertebral joint motion palpation tests, which aim to detect the presence of hypomobility, have face validity because they seem to be reasonable measures of the concept they are intended to measure<sup>32</sup>. Face validity studies have been criticized for being subjective, intuitive and unsubstantiated. Troyanovich and Harrison<sup>32</sup> pointed out that in spite of the common perception or belief that motion tests are valid and reliable for assessment of presence or absence of restricted vertebral motion, there was no evidence to support this concept. Thus, palpatory vertebral motion diagnostic tests are prime examples of tests accepted on face validity.

Construct validity is the extent to which a test identifies the concept or trait of that which is being measured. A construct is a hypothetical or conceptual idea that may be used to label or explain observed phenomenon<sup>33</sup>. For example, taking a dysfunctional vertebral joint as the concept, a test demonstrating the ability to identify the presence or

absence of that concept or its related components, is said to have construct validity. Feinstein describes construct validity as an appraisal of the effectiveness with which a measure does its job in describing an existing or established construct; i.e. does the measure behave the way one would predict on the basis of the concept it represents? For example, Jull et al<sup>34</sup> compared cervical spinal static palpation to diagnostic nerve blocks with anesthesia. The construct is that tenderness upon provocative palpation is related to local nerve irritation and nerve conductivity. A local anesthetic nerve block of related spinal segments showed that the identified tender spots no longer elicited a pain response. They thus demonstrated that there is a high degree of correlation between the palpatory test that identified a tender spot and the ability of the anesthesia to reverse the results of the provocative test. Therefore, the pain provocative palpatory tests used were demonstrated to have high construct validity.

Construct validity, however, is an artificial framework that is not directly observable<sup>27</sup>. To establish construct validity of a test or measure, the researcher must determine the extent to which the measure correlates with other measures designed to measure the same thing and whether the measure behaves as expected. Construct validity studies do not measure the same phenomena that palpatory procedures are designed to measure (i.e., resistance to digital pressure or motion), but similar phenomena that are believed to be related to the palpable phenomena. Many construct validity studies on diagnostic spinal palpatory tests compare a test's results to another measurement of abnormal physiology in the same region. Studies using thermography<sup>35</sup>, electromyography<sup>36</sup>, and coronary angiography<sup>37</sup> fall into this category.



There are other examples of construct validity studies using instruments to measure skin temperature, electrical skin resistance and/or gross range of motion to discern a dysfunctional vertebral segment. These measurements are then compared to those obtained by another examiner who utilizes one or several palpatory procedures that assess resistance to joint motion or paraspinal soft tissue abnormalities to help to discern a dysfunctional vertebral segment. Or one examiner uses pain provocation and the other palpatory motion restriction sense to assess for a dysfunctional vertebral segment.

Criterion validity measures the extent to which an intervention allows a researcher to predict behavioral or pathological outcomes. Criterion validity studies, therefore, do not measure the phenomena being palpated, but attempt to correlate the findings of a palpatory procedure (e.g.) with another measurable outcome like diagnosed visceral disease. For example, Beal<sup>38</sup> and Tarr<sup>39</sup> studied the ability of physicians using spinal palpatory procedures to identify, or predict, which patients had visceral disease related to the spinal findings of altered structure, motion and/or soft tissue.

Content validity is the extent to which a measure adequately, comprehensively covers what it claims to be measuring. Although Troyanovich and Harrison<sup>40</sup> consider face and content validity as synonymous, there is an important distinction: content validity studies employ a reference standard.

A reference standard (also called “gold standard”) is a measure accepted by consensus of content experts as the best available for determining the presence or absence of a particular phenomenon. When there is no perfect reference standard, as in the case of measurement of a patient’s sense of pain provocation, i.e., pressing on a “tender point” or “trigger point”, then pragmatic criteria can be used as a reference standard<sup>41</sup>. The visual analog pain scale has been used as a pragmatic reference standard for palpatory pain provocation tests.

Ideally, content validity studies attempt to compare a test with a reference standard measure of the same phenomena as that which is being palpated, i.e., palpable abnormalities in structure, motion and soft tissue. The Chiropractic Mercy Center Consensus Conference held in January 1993 identified and rated the value of various measurement instruments related to spinal joint functional assessment that could be used as reference standards<sup>42</sup>. Based on their critical review of the literature, Troyanovich and Harrison<sup>43</sup> suggested postural assessment instruments and radiographic measurement as valid, reliable and clinically useful objective measurement tools to help identify dysfunctional spinal vertebral joints.

Based on this brief review, it appears that construct and criterion validity studies do not measure the phenomena being palpated, but attempt to correlate the findings of a palpatory procedure with another measurable outcome. On the other hand, content validity studies measure the same phenomenon as that which is being palpated. Given how important it is to know whether the diagnostic tests used in palpatory exams are

valid, we conducted a systematic review to assess the content validity of spinal palpatory tests used to identify spinal neuro-musculoskeletal dysfunction.

## **METHODS**

### ***Study Setting***

The study was conducted at the Susan Samueli Center for Complementary and Alternative Medicine (University of California-Irvine [UCI]). A multi-disciplinary team of clinicians, researchers, a statistician, and a health sciences librarian participated in the systematic review. The clinicians represented content area expertise in osteopathic and chiropractic medicine, family medicine, and clinical research. In addition, the researchers had expertise and experience in evidence-based medicine, research design and methodology.

The study inclusion/ exclusion criteria were adapted and modified from those published previously by the Cochrane Collaboration<sup>44</sup> and others<sup>45-46</sup>. Studies included in the review met the following four criteria: 1) the studies pertained to manual spinal (cervical, thoracic, lumbar, and surrounding para-spinal soft tissue but not the sacrum or pelvis) palpation procedures; 2) the studies included measurement of validity or accuracy of spinal palpation, where validity was defined as the capability of the manual spinal palpation procedure to do what it is supposed to do and accuracy was defined as a measure of how well it actually does that; 3) the studies were dissertations or a primary research studies published in a peer-reviewed journal; 4) the document could be written in any language; 5) the primary research must have been published or

accepted for publication; and 6) all studies were made available between January 1, 1966 and September 30, 2002. Studies were excluded from the review based on the following criteria. First, the data pertained to non-manual procedure(s). Second, the studies included a whole regimen of tests or methods; without separate data for each test, and/or the data for spinal palpatory procedure could not be retrieved. Third, although the document retrieved was relevant to the subject matter, it was anecdotal, speculative, or editorial in nature. Fourth, the document retrieved was inconsistent with the inclusion criteria (see Appendix I).

We designed a comprehensive strategy to conduct a detailed search of pertinent literature that addressed the study question, “What is the content validity of spinal palpatory tests used to identify spinal neuro-musculoskeletal dysfunction”. Specifics on the search strategy are described in another paper<sup>47</sup>. In brief, our search strategy included both online and manual searches for appropriate literature. For the online search of literature, we defined a detailed search template, which we applied to appropriate databases. The basic search template included MeSH, Descriptors (from MANTIS, Biosis, etc.), Medical Subject headings from CINAHL, and related key terms generated by the investigators from the review team (Appendix II). This defined the research question into four key concepts: validity/validity findings, spine, palpation procedure, and neuromusculoskeletal dysfunctions.

Limits for the search template included: human studies, publications in all languages, journal articles (research articles and conference proceedings if in press), dissertations, and publications between January 1, 1966 and November 1, 2002. We applied the search template, with minor modifications to optimize and enhance the

search outcome of individual databases, to 11 databases that had a potential coverage for the areas of osteopathic medicine, allopathic medicine, chiropractic, and physical therapy. The databases accessed by the project included: PubMed MEDLINE, MANTIS, CINAHL, Web of Science, Current Contents, BIOSIS, EMBase OCLC FirstSearch, Cochrane, Osteopathic Database, and Index to Chiropractic Literature. The selection of databases was based primarily on the availability of online resources that we could access from our affiliated institution libraries.

In addition to the online literature search strategy, we used manual methods to identify appropriate literature. These manual methods included gleaning references that were cited in studies selected from the online search, consulting experts in the fields of chiropractic and osteopathic medicine, contacting authors of eligible conference abstracts, and manually searching bibliographies of osteopathic text-books and review articles on somatic dysfunction.

We used a three-step selection process to identify articles for the systematic review. First, we reviewed titles identified through the online search, and excluded those which gave no indication that the studies pertained to validity. Second, we reviewed the abstracts of all the remaining studies identified through the application of our search template, and excluded studies that did not meet the inclusion criteria. Third, we reviewed the complete paper and applied the inclusion/exclusion criteria to studies included at step two.

In all, based on the online and manual searches, 48 studies were fully reviewed. Five studies met the inclusion/exclusion criteria for the systematic review. The remaining 43 studies were excluded, because they did not study spinal palpation procedures, did not assess content validity, and/or did not use appropriate statistical tests. Several of the abstracts reviewed at step two of the selection process did not provide clarity towards a study's focus (spinal palpation, type of validity studied).

### **INSERT TABLE 1 ABOUT HERE**

We developed two instruments to extract the data and assess the quality of the studies reviewed. The instruments were developed taking into consideration previously published guidelines<sup>48-49</sup>, and instruments<sup>50-54</sup>. To maximize objectivity in the evaluation of paper quality, a checklist of quality factors was developed and transformed into a quality assessment instrument. The factors were grouped into 7 major components of quality: study subjects, examiner characteristics, the reference standard used, palpatory test, study conditions, data analysis and presentation of results.

### **INSERT TABLE 2 ABOUT HERE**

We abstracted and scored detailed information on the 7 components identified to denote internal validity and quality of a study. In terms of the subject characteristics, we considered criteria such as their socio-demographic description, presentation characteristics and severity of symptoms, selection criteria and sample size determination procedures, sample size and recruitment procedures. Information regarding the examiners pertained to their selection criteria, sample size, and background. The reference standard (if used) and palpatory procedure information

pertinent to the quality scoring included a description of the tests, their reliability and expected outcomes, and definition of positive or negative test results. The study conditions were documented with regards to consensus on and description of the palpatory procedure, the training of examiners in the procedure, and blinding of examiners and subjects. For information on the data analysis and results, we abstracted information on the type of statistical procedure(s) used to assess validity and how the results were displayed and described.

The quality assessment instrument focused mainly on the internal validity, taking into consideration biases reported previously namely: selection, performance, measurement, and attrition bias. A weight was assigned to each criterion based on a group consensus. A maximum score of 100 points was set. In designing this instrument we differentiated between quality of an article (i.e. conduct of the trial and reproducibility) and validity, which relates to the ability of the study to answer the research question.

The data extraction and quality assessment instruments were structured to mirror each other and facilitate the review and scoring.

Using the quality assessment instrument, each article was reviewed and scored on the seven major components, discussed above, by two-blinded reviewers (Author(s) and journal were removed). The quality scores included an “absolute” score (i.e., total points received on all seven components of the quality assessment form) and a

“relative” score (i.e., [absolute score/ total score that could be obtained] X 100). The relative score was especially important for studies wherein certain aspects of the quality scoring components were inapplicable (i.e., the subjects’ criteria was inapplicable for studies which used mechanical models or measures). An article’s score (absolute or relative) indicated its quality in terms of its internal validity criteria (whether conclusions drawn from study are likely to be unbiased) and the authors’ explicit description of the study. Although important, the quality score does not imply a study’s significance or impact (in terms of findings, relevance to the discipline). Based on prior recommendations, the overall quality of studies was assessed through the summary scores and the relevant methodological issues pertinent toward internal validity of a study were assessed individually and their influence explored<sup>55</sup>.

We conducted a pilot test of the data extraction and quality assessment instruments on four articles randomly selected from the 48 studies evaluated during step three of the study selection process. After completion of the pilot test, we made changes to further clarify and simplify the instruments. For the final review, the articles were blinded to title and author, and randomly assigned to a pair of reviewers. In all, six reviewers (three pairs) conducted the final review, abstracted pertinent data and scored each article based on the quality assessment instrument.

We used descriptive statistics on the quality assessment data to determine agreement/disagreement among a pair of reviewers, and to present the data. The descriptive statistics included standard deviation (S.D.) / Mean ratio, histogram and



variability. To achieve a consensus between the pair of reviewers on the scoring of each article, we calculated the standard deviation (S.D.) to mean score percentage.

Agreement on quality scores was defined as less than 10% variance (S.D./Mean ratio), in the paired reviewers' scores on each article. When the S.D./Mean ratio variance between the paired reviewers' score was equal to, or exceeded 10%, the pair of reviewers attempted to reach a consensus on each of the criteria where disagreement existed. When reviewers failed to arrive at a consensus on the quality score, two content experts reviewed and scored the article.

## **Results**

### **Study description:**

A total of five studies, from the 48 articles retrieved and reviewed, met our inclusion criteria for content validity and were included in this study (TABLE 3). The remaining 43 studies were retrieved, reviewed and excluded from our study because either they did not address manual palpation procedure(s), did not pertain to content validity but focused on either construct, predictive, or criterion validity, or used inappropriate statistics (Table 1). Four studies were published in 4 different journals and the fifth study (Moruzzi) included is a dissertation. Two studies (Moruzzi & Harvey) were unfunded (1 dissertation and 1 did not report any funding). Two studies (Kristiansson & Sandmark), were funded by the Research Council and a liability insurance provider, and one study, (Jensen) was funded by the Chiropractic Advancement Association.

**INSERT TABLE 3 ABOUT HERE**

### **Subjects:**

The three motion palpation studies were done in the UK. All three studies utilized mechanical models as the study subjects as well as the reference standard. The two pain studies were done in Sweden. One study (Kristiansson), recruited only pregnant female subjects (n=200, representing a 90% response rate: 200/222), while the other study recruited an entirely male population (n=75, they failed to report the response rate) with acute (< 1 week) neck pain,

### **Examiners**

Senior chiropractic students and/or experienced (>3 yrs) practitioners were the examiners in the three motion palpation studies. One physical therapist was the examiner in the cervical spine pain provocation study. The lumbar spine pain provocation study (Kristiansson), did not specify the background of the examiner(s).

### **Design**

All the studies used a prospective study design. In 4 studies the examiners were blinded to fixation levels or clinical presentation. In one pain study (Kristiansson) blinding was not described.

### **Measurement**

Among the three studies using mechanical models, 2 (Harvey & Jensen) looked at intersegmental motion restriction, and one (Moruzzi) looked at the ability to determine fixation levels. The mechanical model was the reference standard used.

The two pain studies used digital pressure and percussion to elicit pain. Visual Analog Scale (VAS) and pain reported by subjects were used as reference standards.

Reliability of the palpation procedure was not reported in any papers with the exception of 1 (Jensen) looking at motion palpation in a mechanical model.

### **Quality Scoring Findings:**

In general the quality score would indicate the rigor with which the science was presented in the paper. Quality scores of included studies ranged from 45.5 to 82 out of a possible 100. The overall quality of the included studies was good for those focusing on motion palpation (69.5 - 82), and fair for those looking at pain (45.5 - 55.5).

Discussion of examiners and study conditions were the two major areas where weakness was noted in the two pain studies, but not in the motion palpation studies.

Statistical tests used were adequate for all studies (this was one of the inclusion criteria). All studies were done in the 1990's; hence the time factor was not felt to be contributive.

### **INSERT TABLE 4 ABOUT HERE**

#### **Study findings:**

***Motion Palpation Tests:*** The three studies examining motion palpation were similar in using a mechanical model as the reference standard and focusing on the lumbar spine only. While two studies used similar examiner groups and motion test, the third study (Moruzzi) looked only at one group of examiners using two different motion test procedures.

Two studies (Harvey & Jensen) looked at intersegmental motion restriction, using sagittal and coronal motion as determined by two groups of chiropractic examiners with

different experience levels (senior students and practitioners). Both studies presented data on sensitivity (ability of a test to detect correctly restricted motion segments) and specificity (ability of a test to detect correctly unrestricted motion segments). The sensitivity for both groups in each study varied between 0.510 and 0.636, and the specificity from 0.868 to 0.902, indicating less ability to detect restricted motion segments than unrestricted motion segments. The sensitivity for practitioners in both studies was poor (0.478 and 0.526). For students, the sensitivity was lower in the Harvey study (0.538) than the Jensen study (0.72).

Based on the data provided in each of the studies we calculated the positive and negative predictive (PPV; NPV) values and the likelihood ratio (LR) for each group. The PPV was less than 50% in both studies, for both groups (42.3-46.2%) and for each subgroup. While the NPV was greater than 80% (83-93.7%) supporting the above statement of better capability of these tests at detecting unrestricted than restricted motion.

#### **Insert Table 5 About Here**

The third motion palpation study (Moruzzi) looked at intersegmental motion restriction as determined by lateral flexion and posterior-anterior springing (PAS). Examiners were 50 senior chiropractic students. Sensitivity for lateral flexion was 41.2% and for PAS 42.8%, while specificity for lateral flexion was 61.5% and PAS was 62.2% indicating that the motion palpation procedures utilized were neither sensitive nor specific for detecting spinal segmental motion restriction. The calculated PPV (< 31%) and NPV (73.7% for both tests) supported this conclusion.

### **Pain Provocation:**

The two studies differed in procedure location (cervical vs. thoracic & lumbar), reference standard (VAS vs. subjective pt report), provocation test used and population studied.

The cervical study (Sandmark) assessed presence or absence of pain as reported by the subjects upon palpation of their facet joints. Sensitivity (ability of the test to identify presence of pain in subjects reporting pain symptoms) was 82% and specificity (ability of the test to identify the absence of pain in asymptomatic subjects) was 79%, the PPV was 62% and NPV was 91%. The results indicate that the test procedure, as performed, is moderately good at identifying subjects with neck pain and very good at identifying asymptomatic subjects.

The thoracic and lumbar spine study (Kristianson) used the VAS as the reference standard and assessed the relationship between the clinical back status and reported pain locations during and after pregnancy. Two types of pain provocation tests were used: digital pressure (within 5 cm of the midline) and lumbar percussion. In the thoracic region, digital pressure (DP) sensitivity was 17.8%, specificity was 98.5%, calculated PPV was 72.2% and NPV was 84.44%. In the lumbar region: DP sensitivity was 21.2%, specificity was 96.19%, calculated PPV was 61.76% and NPV was 80.83%; lumbar percussion sensitivity was 5.1%, specificity was 100%, calculated PPV was 100% and NPV was 78.44%. These results suggest that the thoracic DP test was better at

identifying asymptomatic than symptomatic subjects. Both tests performed in the lumbar region were unable to discriminate adequately between subjects.

### **INSERT TABLE 6 ABOUT HERE**

#### **Discussion:**

To the best of our knowledge, this is the first comprehensive systematic review of literature on the content validity of spinal palpatory procedures. To reiterate, it is imperative to focus on studies assessing content validity of procedures since, by definition, they attempt to measure the same phenomena as that which is being palpated. Studies with a focus on other forms of validity (i.e., face, construct and criterion), although important, provide information which does not directly answer the question, “Does the procedure (i.e., palpation) measure (or assess) the phenomena it is supposed to assess?” but attempt to correlate the findings of a palpatory procedure with another measurable outcome.

The systematic review revealed several methodological, reporting and research issues which severely constrained integrative, qualitative and quantitative evaluations such as systematic reviews and meta-analysis. The evaluation of the validity of spinal palpatory procedures has a number of methodological challenges. In particular, there is no agreed upon reference or “gold” standard measuring device for spinal palpatory procedures. A reference standard is the best available independently established test/procedure used to determine the presence or absence of a phenomenon. In the absence of well-established reference standards, one would use other research designs, such as pragmatic criteria (e.g., pain scales), independent expert panels,

clinical follow-up (delayed type cross sectional study), standardized protocols or prognostic criteria<sup>56</sup>. One may also use the most reproducible and reliable test or the most experienced examiner as a reference standard. Some designs utilize invasive procedures, e.g., surgery, histopathology or angiography or a combination of tests to serve as a reference standard.

It is important to identify a reference standard to which a palpatory diagnostic test is compared to ensure that it actually measures what it purports to measure (i.e., that a test for resistance to motion actually measures resistance to motion). Spinal palpatory diagnostic procedures, like vertebral joint motion restriction assessment, are difficult to objectively measure in humans. The concept of a neuromusculoskeletal spinal dysfunction that is corrected by non-invasive manual spinal manipulation has no agreed upon reference standard. Typically, a conglomerate of findings of altered position, motion characteristics and paraspinal soft tissue feel is necessary to make the diagnosis. X-rays can be validated by altered position. . Altered motion has been difficult to validate due to the difficulty of finding a suitable reference standard. However, in order to assess an examiner's ability to discern resistance to vertebral joint motion, the plastic spinal model with an artificially fixed vertebral segment has been employed as a reference standard. Altered tissue feel can be validated in part by measuring skin moisture, temperature, friction, and resistance to pressure. A reference standard used for palpatory pain provocation tests has been the visual analog or numeric pain scale<sup>57-</sup>

<sup>59</sup>.

Given that face, construct and criterion validity studies do not measure the phenomena being palpated, but attempt to correlate the findings of a palpatory procedure with another measurable outcome, only content validity studies, which attempt to measure the same phenomena as that which is being palpated were included in this systematic review.

Physicians (orthopedists, physiatrists, neurologists, emergency medicine, family medicine, sports medicine, etc.), chiropractors, massage therapists, osteopaths, and physical therapists use manual palpatory exams regularly in their practice. However very few studies (#5) have attempted to assess the validity of these widely used tests. Among the few validity studies identified, motion palpations were carried out only by chiropractic and pain studies by physical therapy.

In this review three types of studies: 2 on fixation, one on range of motion (ROM) and two on pain were identified. The quality scores of motion palpation studies were good, however all the tests had poor sensitivity. This indicates that the motion palpatory tests (intersegmental, lateral flexion and posterior-anterior springing) are not able to identify areas of fixation or motion restriction. A poor positive predictive Value (PPV) supported this finding. The pain provocation studies reported good validity for evaluation of pain in the cervical region but not in the lumbar area. This result confirms the results of a previously published<sup>60</sup> study indicating a higher sensitivity for identifying pain in the cervical region compared to the lumbar spine.



Unfortunately, most of the research study results reported are not comparable due to variability in the palpatory tests, terminology, research design, methodology and statistical analysis utilized. These inconsistencies make it difficult to rate the relative value of their results. There is a worldwide concerted effort underway to rectify this problem. The International Federation of Manual Medicine (FIMM), an international organization of physicians and surgeons who practice manual medicine held their General Assembly in Chicago in July 2001. At that meeting, their Scientific Committee reported that their top priority is to promote validity, reliability, sensitivity and specificity studies of spinal palpatory diagnostic procedures. They recently developed guidelines (“Protocol Formats”) on how to perform high quality validity and reliability studies of spinal palpatory procedures, which are available on their web site <http://www.fimm-online.org/>. They recommend the use of valid palpatory tests so that homogeneous populations with spinal musculoskeletal dysfunction can be selected and treated as part of a controlled clinical trial. The results of these trials can subsequently be combined using meta-analysis and would help formulate guidelines for the practice of spinal manipulation.

It is difficult to translate these results into the clinical setting due to the limited number of studies, focused anatomical sites and populations studied. All three-motion palpation studies used a mechanical model as the subjects and reference standard, and focused on the lumbar spine. Findings indicate poor validity of the motion palpation tests. The 2 pain studies are of fair to poor quality. One focused on examining pain in the lumbar

spine of pregnant women, and the other on pain in the cervical spine among men with acute injuries.

Additional studies exploring the validity of spinal palpatory exams are needed. To do so, identifying an acceptable reference standard is a paramount condition.

This review is unique a) by the cooperative work among a multidisciplinary team of researchers and content-experts; b) the review was not limited to any specific discipline or language; c) the focus on content validity is practical and clinically relevant to practitioners and researchers; d) a great effort and detail went into the development of the search strategy, inclusion/exclusion criteria and quality scoring instrument.

The search strategy included 11 databases and was done three times using general and specific keywords and strategies to verify results. The quality-scoring instrument was developed taking into consideration strengths and weakness of published instruments, recommendations by the QUOROM<sup>61</sup> and CONSORT<sup>62-63</sup> statements as well as the Cochrane criteria. In addition this study makes a contribution to the field of manipulation and medicine, in general, by highlighting the limited research and reference standards in this field. It also provides future researchers with a guideline to follow to design a successful content validity study.

As with a majority of reviews, this is a retrospective review, which makes it susceptible to potential sources of bias (publication quality). The focused definition used for content validity limits the studies that are included in this review. However, this

strategy allowed more clarity since only content validity studies, which attempt to measure the same phenomena as that which is being palpated, were included in this systematic review. Despite the number of safeguards used to be inclusive (multiple databases, hand search, review by experts, and multiple searches) in our search, a few studies published but not included in these databases could have been missed.

The quality assessment tool, used for this review, was developed by this team of researchers based on their evaluation of the literature, feedback from methodologists and statisticians. Although we feel that the instrument is well balanced and unbiased, it might have over or underestimated the quality of certain papers. When comparing the quality scores assigned to studies included in this paper to scores assigned to the same papers in other systematic reviews<sup>27</sup>, one notes that our scores are consistently lower.

### **Conclusion:**

Despite the use of manual palpatory spinal palpation by many health care disciplines, very few studies investigated its ability to measure what it intends to measure (content validity). Given the high frequency of spinal pathology and the use of these diagnostic methods, well-designed studies are needed. For the practice of evidence-based medicine, it is important to assess the efficacy and effectiveness of procedures usually and customarily used in clinical practice. To this end, established benchmarks for the validity and reliability of procedures are essential.

This comprehensive systematic review has highlighted serious gaps in our knowledge about the accuracy of spinal palpatory procedures. The findings have

implications for research, clinical practice, and policy. From the research perspective, researchers across discipline need to incorporate more rigor towards the definition of the study questions, methods and measures, implementation procedures, and reporting. The absence of well identified reference standards and possible technical difficulties conducting these studies might have contributed to this scarcity.

From the clinical perspective, the findings suggest poor sensitivity of the range of motion and pain diagnostic tests in the evaluation of spinal dysfunction. From a policy perspective, given that manual procedures are a cornerstone towards diagnostic and therapeutic interventions across disciplines, professional societies and associations need to enact continuing medical education and research guidelines to address the efficacy of spinal palpatory procedures.

**Competing interests:**

None declared

**Authors' contributions:**

All listed authors worked collectively on the design of the study and development of the review instruments. WIN, MAS, and SIM carried out article reviews, data extraction, analysis of the data, and drafted the manuscript. VD carried out paper reviews, data extraction and analysis. SR carried out article reviews and worked on the paper search methodology. LSM oversaw the database searches, and carried out all the article tabulation. AA participated in the design of the study, instrument development, and methodology. AFG participated in development of the review instruments and advised on statistical conduct of the study.

**Acknowledgment:** This work is supported by a grant from the 41<sup>st</sup> Fund, and done at the Susan Samueli Center for Complementary and Alternative Medicine, University of California, Irvine.

## References

1. Andersson GBJ: Epidemiologic aspects of low back pain in industry. *Spine* 1981;6:53.
2. Loeser JD, Volinn E: Epidemiology of low back pain. *Neurosurg Clin North Am* 1991;2:713-8.
3. Hadler NM. Illness in the workplace: the challenge of musculoskeletal symptoms. *J Hand Surg* 1985 Jul;10(4):451-6.
4. Andersson GBJ. Epidemiological features of chronic low-back pain. *Lancet* 1999;354:581-85.
5. Deyo RA. Rethinking strategies for acute low back pain. *Emergency Medicine*. Nov 1995; pp.38-56.
6. Hart LG, Deyo RA, Cherkin DC. Physician office visits for low back pain. Frequency, clinical evaluation, and treatment patterns from a U.S. national survey. *Spine* 1995; 20(1):11-19.
7. Schiötz, EH and Cyriax, J. Manipulation. Past and Present. London: William Heinemann Medical Books LTD, 1975.
8. Lomax, E. Manipulative Therapy: A Historical Perspective from Ancient Times to the Modern Era. In Goldstein M, ed., The Research Status of Spinal Manipulative Therapy, Bethesda: US Dept. of Health, Education and Welfare (NIH) # 76-998, 1975; pp. 11-17.
9. Dinnar, U et al: Description of fifty diagnostic tests used with osteopathic manipulation. *JAOA* 81: 314-321 Jan 1982.

10. Walker BF and Buchbinder R. Most commonly used methods of detecting spinal subluxation and the preferred term for its description: A survey of chiropractors in Victoria, Australia. *J Manipulative Physiol Ther*, Nov/Dec, 1997; 20(9):583-589.
11. Walker BF. The reliability of Chiropractic methods used for the detection of spinal subluxation. An overview of the literature. *ACO* March 1996; 5(1):12-22.
12. Ward RC Ex Ed. *Foundations for Osteopathic Medicine*. Philadelphia: Lippincott, Williams and Wilkins. 2003.
13. Chaitow L. *Palpation skills: Assessment and diagnosis through touch*. New York: Churchill-Livingstone. 1997.
14. Bergman TF, Peterson DH and Lawrence DJ. *Chiropractic techniques. Principles and procedures*. New York: Churchill-Livingstone. 1993.
15. Kaltenborn FM. *The spine: Basic evaluation and mobilization techniques*. Minneapolis, MN: OPTP. 1993.
16. Maitland GO ed. *Maitland's vertebral manipulation*. Oxford: Butterworth-Heinemann. 2000.
17. Lewit K. *Manipulations therapy in rehabilitation of the locomotor system*. Boston: Butterworths. 1985.
18. Maigne R. *Orthopedic medicine. A new approach to vertebral manipulation*. WT Liberson, trans. Springfield, IL: Charles Thomas. 1972.
19. Wadsworth CT. *Manual examination and treatment of the spine and extremities*. Baltimore: Williams and Wilkins. 1988.

20. Johnston, W.L. Interexaminer reliability studies: spanning a gap in medical research. Louisa Burns Memorial Lecture. J Am Osteopath Assoc 1982, 81:819-829.
21. Gonnella C, Paris SV, and Kutner M: Reliability in evaluating passive intervertebral motion. Phys. Ther. 1982, 62:436-444.
22. Panzer, DM: The Reliability of Lumbar Motion Palpation. J Manipulative Physiol Ther. October 1992, 15(8): 518-524.
23. Keating JC, Jacobs GE. Inter- and intraexaminer reliability of palpation for sacroiliac joint dysfunction. JMPT 1989;12(2):155-8.
24. Haas M. The reliability of reliability. JMPT 1991;14(3):199-208.
25. Russell R. Diagnostic palpation of the spine: A review of procedures and assessment of their reliability. J Manipulative Physiol Ther 1983; 6:181-183.
26. Alley JR. The clinical value of motion palpation as a diagnostic tool: A review. The Journal of the CCA, Sept. 1983;27(3):97-100.
27. Huijbregts PA. Spinal motion palpation: A review of reliability studies. The Journal of Manual and Manipulative Therapy, 2002. Vol. 10(1): 24-39.
28. Hestboek L, Leboeuf YC: Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A Systematic Critical Literature Review. JMPT, May 1, 2000, 23(4): 258-75.
29. Van der Wurff P, Meyne W, Hagemeyer RHM. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 2. Validity. Manual Therapy 2000, 5(2): 89-96.



30. Seffinger MA, Adams AA, Najm WI, Dickerson V, Reinsch S, Goodman A, Gokhale DV, Murphy L, Mishra S. Spinal Palpatory Diagnostic Procedures Utilized by Practitioners of Spinal Manipulation: Annotated Bibliography of Reliability Studies. (Accepted for publication, JCCA 2003.)
31. Winter G. A comparative discussion of the notion of 'validity' in qualitative and quantitative research. The Qualitative Report. Vol. 4, Numbers 3 & 4, March 2000 (<http://www.nova.edu/ssss/QR/QR4-3/winter.html>)
32. Troyanovich SJ and Harrison DD. Motion palpation: Its time to accept the evidence. Commentary. J Manipulative Physiol Ther, Oct 1998; 21(8):568-571.
33. Feinstein AR. Clinometrics. New Haven: Yale University Press; 1987.
34. Jull G, Bogduk N, Marsland A. The accuracy of manual diagnosis for cervical zygapophysial joint pain syndromes. Med J Aust. 1988 Mar 7;148(5):233-6.
35. Kelso AF, Grant RG, Johnston WL. Use of thermograms to support assessment of somatic dysfunction or effects of osteopathic manipulative treatment: preliminary report. J Am Osteopath Assoc. 1982 Nov;82(3):182-8.
36. Vorro J, Johnston WL. Clinical biomechanic correlates of cervical dysfunction: Part 4. Altered regional motor behavior. J Am Osteopath Assoc. 1998 Jun;98(6):317-23.
37. Kelso AF, Larson NJ, Kappler RE. A clinical investigation of the osteopathic examination. J Am Osteopath Assoc. 1980 Mar;79(7):460-7.
38. Beal MC. Palpatory testing for somatic dysfunction in patients with cardiovascular disease. J Am Osteopath Assoc. 1983 Jul;82(11):822-31.

39. Tarr RS, Feely RA, Richardson DL, Mulloy AL, Nelson KE, Perrin WE, Allin EF, Efrusy ME, Greenstein SI, Vatt RD. A controlled study of palpatory diagnostic procedures: assessment of sensitivity and specificity. *J Am Osteopath Assoc.* 1987 Apr;87(4):296-301.
40. Troyanovich SJ and Harrison DD. The reliability and validity of chiropractic assessment procedures. *Chiropractic Technique*, Feb. 1996; 8(1):10-13.
41. Knottnerus JA, ed. The evidence base of clinical diagnosis. London: BMJ books, 2002.
42. Haldeman S, Chapman-Smith D, Peterson DM Jr. eds. Guidelines for chiropractic quality assurance and practice parameters. Proceedings of the Mercy Center Consensus Conference; 1993 January 25-30. Burlingame, CA: Aspen Publishers; 1993.
43. Troyanovich SJ and Harrison DD. In reply to letter to the editor. *J Manipulative Physiol Ther*, March/April 1999; 22(3):182-183.
44. <http://www.cochrane.org/cochrane/hbook.htm>
45. CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD Report Number 4 (2<sup>nd</sup> Edition), March 2001. The University of York.  
<http://www.york.ac.uk/inst/crd/report4.htm>
46. Devillé WL, Buntinx F, Bouter LM, Montori VM, de Vet H, Danielle AWM van der Windt D, P Dick Bezemer PD. Conducting systematic reviews of diagnostic

studies: didactic guidelines. *BMC Medical Research Methodology* 2002 2:9

<http://www.biomedcentral.com/1471-2288/2/9>

47. Murphy LS, Reinsch S, Najm WI, Dickerson VM, Seffinger MA, & Mishra SI. Researching the Validity of Spinal Palpation: the Use of Controlled Vocabulary among Authors, Indexers and Researchers. (In review)
48. Juni P, Altman DG, Egger M. Systematic reviews in health care. Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-46.
49. Deeks JJ. Systematic reviews in health care. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323: 157-162.
50. Shekelle PG, Adams AH, Chassin MR, Hurwitz EL, Brook RH. Spinal manipulation for low-back pain. *Ann Intern Med.* 1992 Oct 1;117(7):590-8.
51. Cook DJ, Sackett DL, Spitzer WO. Methodologic Guidelines for Systematic Reviews of Randomized Control Trials in Health Care from the Postdam Consultation on Meta-analysis. *J Clin Epidemiol* 1995;48(1):167-171.
52. Koes BW, Assendelft WJJ, van der Heijden GJMG, et al. Spinal manipulation and mobilization for back and neck pain: a blinded review. *BMJ* 1991;303:1298-1303.
53. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-412.
54. Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guide for those Carrying Out or Commissioning Reviews. CRD Report Number 4 (2<sup>nd</sup> Edition) March 2001. York Publishing Services Ltd.

55. Juni P, Wistchi A, Bloch R, Egger M. The Hazards of Scoring the Quality of Clinical trials for Meta-analysis. JAMA 1999;282(1):1054-1060.
56. Knottnerus, Andre. The Evidence Base of Clinical Diagnosis. BMJ Books, January 2002.
57. Olson SL, O'connor DP, Birmingham G, Broman P, Herrera L. Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. J Orthop Sports Phys Ther. Jan 2000; 30(1): 13-20
58. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analog scales as ratio scale measures for chronic and experimental pain. Pain. 1983; 17:45-56.
59. Huskisson EC. Measurement of pain. J Rheumatology; 1982; 9:768-769
60. Keating L, Lubke C, Powell V, Young T, Souvlis T, Jull G. Mid-thoracic tenderness: a comparison of pressure pain threshold between spinal regions, in asymptomatic subjects. Man Ther. 2001 Feb;6(1):34-9.
61. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999 Nov 27;354(9193):1896-900.
62. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996 Aug 28;276(8):637 9.

63. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-91.

**Studies included in this review:**

1. Harvey D, Byfield D. Preliminary Studies with a Mechanical Model for the Evaluation of Spinal Motion Palpation. *Clinical Biomechanics* 1991; 6:79-82
2. Jensen K, Gemmell H, Thiel H. Motion palpation accuracy using a mechanical spine model. *European J Chiropractic* 1993;41:67-73.
3. Kristiansson P, Svardsudd K, von Schoultz B. Back Pain During Pregnancy. A Prospective Study. *Spine* 1996;21:702-8.
4. Moruzzi S, Accuracy of Two Different Motion Palpation Procedures For determining Fixations in the Lumbar Spine Using An Articulated Spinal Model. Dissertation. Anglo-European College of Chiropractic, 1992-1993.
5. Sandmark H, Nisell R. Validity of Five Common Manual Neck Pain Provoking Tests. *Scand J Rehab med* 1995;27:131-36.

**TABLE 1**

Studies reviewed and excluded from this study of content validity

Author	Validity of Spinal Palpatory Procedures	Validity	Statistics used	Discipline	Palpatory method used	Reference test used
Beal 1989	No	Construct	Descriptive	DO	Passive ROM; tissue texture; passive mobility at specific spinal levels	Electromyography
Braun 1991	No	Construct	Pearson Correlation coefficient	PT	Cervical Range of Motion instrument	Questionnaire and Muscle Index neck subscale
Bush 2000	No	Construct	Correlation Coefficient	PT	Single, double and stabilization methods measuring Cervical ROM	X-Ray
Gregory 1998	Yes	Construct	Chi square	DC	Motion palpation	X-Ray
Haas 1994	No	Construct	fisher 2 tailed exact t-test/ Prevalence	DC	Vertebral challenge	Response to manual high velocity low amplitude adjustment
Haas 1995	Yes	Construct	Descriptive	DC	Motion palpation using a spinous process contact (Bergmann)	Response to manual high velocity low amplitude rotatory manipulation
Johnston 1985	Yes	Construct	MANOVA	DO	Range of motion; passive and active ROM	Kinematic procedure
Jull 1988	Yes	Construct	Sensitivity & Specificity	PT	Passive accessory intervertebral movements	Cervical nerve block
Kawchuk	No	Construct	Intra-class correlation	DC	Tissue compliance meter	Measurement on a controlled

1995			coefficient			surface
Mayer 1997	No	Construct	Accuracy	PT	Lumbar sagittal ROM	Instrumented Inclinator
McPartland 1997	Yes	Construct	Percent agreement &Kappa	DO	Tender point evaluation, std osteopathic exam.	Traditional Osteopathic Examination
Nansel 1988	No	Construct	Descriptive	DC	Motion and Static palpation exam	Galvanic Skin Response
Phillips 1996	Yes	Construct	Sensitivity/Specificity	PT	Passive physiologic & Passive accessory intervertebral movements	Spinal Anesthetic Block
Swerdlow 1992	No	Construct	t-test	Technologist	Thermography	Palpation of active trigger points
Vorro 1987	Yes	Construct	MANOVA	DO	Primary rotation movements	Electromyography
Vorro 1998	Yes	Construct	Chi square	DO	Primary rotation movements	Electromyography of isolated muscles
Breen (in Press)	Yes	Face	2 tailed unpaired t- test	DC	Motion Palpation	X-Ray Fluoroscopy sequence
Byefield 1991	Yes	Face	Descriptive	DC	Palpate inferior tip of L3	X-Ray
Dott 1994	No	Face	Descriptive	DO	Leg length discrepancy	X-Ray of Iliac crest discrepancy
Jende 1997	Yes	Face	Kappa	DC	Motion Palpation of the upper cervical complex	X-Ray
Lucchetti 1992	Yes	Face	Descriptive	DC	Flexion - extension motion palpation	X-Ray
Beal 1985	No	Predictive/ Criterion	Sensitivity and specificity; Predictive	DO	Palpatory compression test	Angiogram

			value			
Bjorksten 1997	No	Predictive/ Criterion	Sensitivity and Specificity	PT	Questionnaire / VAS and pain drawings	Clinical examination
Brunarski 1982	Yes	Predictive/ Criterion	Sensitivity & Specificity; Predictive values and accuracy	DC	Motion Palpation & plumbline with lateral bending	Lateral bending dynamic roentgenograms
Cox 1983	Yes	Predictive/ Criterion	Chi square and logistic regression	DO	ROM; Soft tissue texture; Pain;	Angiogram
Ebraheim 1999	No	Predictive/ Criterion	Percent agreement	MD	Observation and palpation of transverse process during surgery	intra-operative X-ray
Gracovetsky 1998	Not specified?	Predictive/ Criterion	Percent agreement	MD	Motion of the skin over the spine	Clinical examination by two expert physicians
Jull 1994	Yes	Predictive/ Criterion	Descriptive	PT	Manual assessment of segmental tissue stiffness	Subjects' independent reports of pain provocation.
Lebouef-Yde 2000	Yes	Predictive/ Criterion	Sensitivity & Specificity	DC	Lumbo-pelvic dysfunctional tests (pain on movement & pain provoking)	Low Back Pain status
Lebouef-Yde 2002	Yes	Predictive/ Criterion	Sensitivity and Specificity	DC	Motion palpation	LBP status and pain reports
Lundberg 1999	Yes	Predictive/ Criterion	Linear regression analysis	MD	Segmental mobility	Sagittal Lumbar mobility measured by Kyphometry
Maher 1998	Yes	Predictive/ Criterion	Pearson Correlation coefficient	PT	PA spinal stiffness	Stiffness Assessment Machine
Mennell 1990	No	Predictive/ Criterion	Descriptive	MD	Manipulative therapy	X-Ray



Olson 2000	Yes	Predictive/ Criterion	Pearson Correlation coefficient	MD	Pressure Algometry for tender point sensitivity; ROM	Visual Analog Scale
Tarr 1987	Yes	Predictive/ Criterion	Sensitivity and Specificity	DO	Diagnostic Osteopathic palpatory technique	Documented Visceral disease
Viikari- Juntura 1998	No	Predictive/ Criterion	Kappa	MD/PT	Standardized Instrument (History and physical: ROM & Pain)	Future Sick leaves
Visscher 2000	Yes	Predictive/ Criterion	Pearson Correlation coefficient	PT	Active/Passive ROM; Dynamic/Static tests; Muscle palpation	Patient report of pain
Vorro 1991	Yes	Predictive/ Criterion	ANOVA; ANCOVA	DO	Primary rotation movements	Kinematic measures
Keating 2001	No	Not validity		DC	Pressure pain threshold using pressure algometry	Patient pain report
Leclaire 1996	No	Not Validity	Receiver Operating Characteristics	MD	Thermography; dynamometry; spinoscopy	Expert Clinicians
Simmonds 1995	No	Not Validity	Descriptive	PT	Forces applied in spinal mobilization	Mechanical model
Osterbauer 1996	No	Correlation	Sensitivity/Specificity/ likelihood ration	DC	3D kinematic parameters of Head motion	VAS/Neck disability index questionnaire
Vehagen 1996	No	Correlation	Mean Whitney U test	PT	Extension and coordination test	Subjects diagnosed with and without Whiplash

**TABLE 2**

QUALITY SCORING CRITERIA, TOTAL WIEIGHT AND TOTAL SCORE ASSIGNED.

	<b>CRITERIA</b>	<b>TOTAL WEIGHT</b>	<b>TOTAL SCORE</b>
1	<b><u>STUDY SUBJECTS</u></b> <ul style="list-style-type: none"> <li>• Study Subjects Adequately Described</li> <li>• Presentation Characteristics</li> <li>• Spectrum of severity of Symptoms</li> <li>• Subject Selection Criteria</li> <li>• Number of Subjects in Study detailed</li> <li>• Sample Size Determined by Power Analysis</li> <li>• Number Subjects Completed Study</li> <li>• Recruitment Procedure</li> </ul>	15%	15
2	<b><u>EXAMINERS</u></b> <ul style="list-style-type: none"> <li>• Selection Criteria for Examiners Described</li> <li>• Background of Examiners Described</li> </ul>	10%	10
3	<b><u>REFERENCE STANDARD</u></b> <ul style="list-style-type: none"> <li>• Was Reference Standard used</li> <li>• Reference standard procedure described / referenced</li> <li>• Expected Outcome Described</li> <li>• Validity of reference Standard</li> <li>• Reliability of Reference Standard</li> <li>• Positive or Negative Test Result Defined</li> </ul>	15%	15
4	<b><u>PALPATORY TEST</u></b> <ul style="list-style-type: none"> <li>• Description of palpatory test Procedure</li> <li>• Expected Outcome Described</li> <li>• Reliability of Test Described</li> <li>• Positive or Negative Test Result Defined</li> </ul>	15%	15
5	<b><u>STUDY CONDITIONS</u></b>	15%	15

	<ul style="list-style-type: none"> <li>• Time Interval for Test/ Retest procedure</li> <li>• Examiner blinded to Clinical findings</li> <li>• Examiner &amp; Subject blinded to previous study findings</li> <li>• Examiner &amp; Subject blinded to Std. reference results</li> </ul>		
6	<p><b><u>DATA ANALYSIS</u></b></p> <ul style="list-style-type: none"> <li>• Statistical Analysis Used</li> </ul>	15%	15
7	<p><b><u>RESULTS</u></b></p> <ul style="list-style-type: none"> <li>• Results Completely Displayed or Described</li> <li>• P-Value Displayed or Described</li> <li>• Confidence Interval Displayed or Described</li> <li>• Study Bias Identified</li> </ul>	15%	15

**Table 3**

Included studies: Examiner / Subject / design / blinding

Author (year)	Examiner (number)	Study Subject	Study Design	Examiner Blinding
Harvey D (1991)	D.C. (n=27)	Mechanical Model	Cross-sectional	Blinded to fixation level and each other's findings
Moruzzi S (1993)	D.C. (n=50)	Mechanical Model	Cross-sectional	Blinded to fixation level and each other's findings
Jensen K (1993)	D.C. (n=45)	Mechanical Model	Cross-sectional	Blinded to fixation level and each other's findings
Sandmark H (1995)	P.T. (n= 1)	75 randomly selected males with acute neck pain ( $\leq$ 1 wk)	Cross-sectional	Blinded to clinical presentation
Kristiansson, P (1996)	Not described	200 pregnant women with back pain	Cohort	Not described

Table 4

**Average Quality Scores given in each of the 7 major criteria and the total and relative scores for each included article.**

Author/Date	Study Subjects (Total 15)	Examiners (Total 10)	Reference Standard (Total 15)	Palpatory Test (Total 15)	Study Conditions (Total 15)	Data Analysis (Total 15)	Results (Total 15)	Total Mean Score	Relative Mean Score
Harvey, D / 1991	0	10	9	9	12	15	14.5	69.5	81.7
Moruzzi, S / 1993	0	10	12	15	15	15	15	82	96.5
Jensen, K / 1993	0	10	15	7	15	15	10	72	84.5
Sandmark, H / 1995	9.5	0	3.5	8	2	15	7.5	45.5	45.5
Kristiansson, P / 1996	8.5	0	9	11	2	15	10	55.5	55.5

Total Mean Score = Average of total absolute score obtained by each study

Relative Score = Total Mean score adjusted to 100% (to reflect “0” score given for subjects when mechanical models were used).

**Table 5**

**Statistical analysis for Motion Palpation Studies using students and experienced practitioners**

<b>Examiners</b>	<b>Test</b>	<b>Harvey</b>	<b>Jensen</b>
<b>Both Groups</b>	<b>PPV</b>	0.431	0.459
	<b>NPV</b>	0.9	0.902
	<b>+LR</b>	3.893	3.49
	<b>-LR</b>	0.564	0.403
<b>Student</b>	<b>PPV</b>	0.437	0.367
	<b>NPV</b>	0.898	0.936
	<b>+LR</b>	3.71	4.23
	<b>-LR</b>	0.54	0.337
<b>Practitioner</b>	<b>PPV</b>	0.423	0.454
	<b>NPV</b>	0.903	0.830
	<b>+LR</b>	4.05	2.45
	<b>-LR</b>	0.592	0.603

PPV = Positive Predictive Value  
NPV = Negative Predictive Value  
+LR = Positive Likelihood Ratio  
-LR = Negative Likelihood Ratio

**Table 6**

Spinal focus of the study, Reference standard used, Primary outcome, statistics, and author's conclusion.

Author (Year)	Spinal Focus	Reference Standard	Primary Outcome	Statistics	Author's Conclusion
Harvey D (1991)	Lumbar spine	Mechanical Model	Detect presence or absence of lumbar spine intersegmental motion restriction	Sensitivity Intern: 53.8%; Practitioner: 47.8%; Specificity Intern: 85.5%; Practitioner: 88% (PPV Pract. 42.3%, Interns 43.7%; NPV Pract. 90.3%, Interns 89.8%; +LR Pract. 4.05, Interns 3.7; -LR Pract. 0.592; Interns 0.54)	Intersegmental motion restriction palpation is more specific than sensitive
Moruzzi S (1993)	Lumbar spine	Mechanical Model	Detect accuracy of two types of spinal motion palpation procedures in correctly determining fixation	Sensitivity Lateral Flexion: 41.2%; PA springing: 42.8%; Specificity LF: 61.5%; PAS: 62.2% (PPV Post-Ant 28.6%; Lat. Flex. 30.6%; NPV Post-Ant 73.7%, Lat. Flex. 73.7%)	The palpation procedures as performed were not valid tests.
Jensen K (1993)	Lumbar spine	Mechanical Model	Detect presence or absence of	Sensitivity Interns: 72%; Practitioners: 52.6%;	Motion palpation is an accurate

			single and multiple intersegmental motion restrictions	Specificity Interns: 83.2%; Practitioners 78.6% (PPV Interns 46.2%; Pract. 45.5%; NPV Interns 93.7%; Pract 83%)	method for determining non-fixated segments but not accurate for determining fixated segments.
Sandmark H (1995)	Cervical spine	Pain reported by subjects	Assess presence or absence of pain upon palpation of facet joint	Sensitivity 82%; Specificity 79%; Positive Predictive Value=62%; NPV=91%	Palpation over the facet joint had better sensitivity and specificity than motion tests in study.
Kristiansson P (1996)	Lumbar spine	Visual Analog Scale	Assess the relationship between clinical back status and reported pain locations during and after pregnancy.	Thoracic DP Tenderness: Sensitivity 17.8%, Specificity 98.5%, Positive Predictive Value 72.2%, Negative Predictive Value 84.44%; Lumbar DP Tenderness: Sens. 21.2%, Spec. 96.19%, PPV 61.76%, NPV 80.83%; Lumbar Percussion: Sens 5.1%, Spec. 100%, PPV 100%, NPV 78.44%.	Pain provocation tests were better at discriminating LBP than tests of configuration or mobility

DP = Digital Pressure

+LR = positive Likelihood ratio

Pract. = Practitioners

-LR = negative Likelihood ratio



Sens. = Sensitivity

PPV = positive predictive value

NPV = negative predictive value

### Appendix I

Inclusion and Exclusion Criteria for Validity Studies on Spinal Palpation as Applied to Title and Abstract Review

INCLUSION CRITERIA	EXCLUSION CRITERIA
The document must pertain to manual spinal (Cervical; Thoracic; Lumbar & surrounding para-spinal soft tissue, but not Sacrum or Pelvis) palpation procedures.	The data pertains to non-manual procedure(s).
The document includes measurement of validity or accuracy of spinal palpation. <i>(Validity is the capability of the manual spinal palpation procedure to do what it is supposed to do;</i> <i>Accuracy is a measure of how well it actually does that!)</i>	The document included a whole regimen of tests or methods; without separate data for each test, and/or the data for spinal palpatory procedure could not be retrieved.
The document must be a primary research study published in a peer reviewed journal, or dissertation. The document can be written in any language.	Although the document retrieved was relevant to the subject matter, it is anecdotal, speculative, or editorial in nature.
The primary research or monograph must have been published or accepted for publication. All	The document retrieved was inconsistent with the inclusion criteria.

documents must have been made available between January 1, 1966 and September 30, 2002.	
---	--

## Appendix II

### MeSH and Other Key Terms Used as Part of the Search Strategy

<b>Validity Terms</b>	<b>Spinal Terms</b>	<b>Procedure Terms</b>	<b>Neuromuscular/Musculoskeletal Dysfunctions</b>
Validation	Spine [mh][sh][de]	Palpation [mh][sh][de]	Fixation / Tissue fixation [mh][de]
Validation Studies [pt][sh]	Spinal	Palpatory	Subluxation [sh][de]
Validity [sh]	Cervical / Cervical vertebrae [mh] [sh] [de]	Palpate*	Somatic Dysfunction [de]
Sensitivity and Specificity [mh][sh][de]	Thoracic / Thoracic vertebrae [mh][sh][de]	Manual exam*	Blockage
Sensitivity	Lumbar / Lumbar vertebral [mh][sh][de]	Manual diagnosis	Asymmetry
Specificity	Vertebral	Manipulated	Back pain [mh][sh][de]
Valid	vertebrae	Manipulation, chiropractic [mh][sh][de]	Hypomobility

Accuracy, accurate	Para-spinal	Manipulation, orthopedic [mh][sh][de]	Joint Instability [mh][sh][de]
Prediction, predict	Neck [mh][sh][de]	Manipulation, osteopathic [mh][sh][de]	Manipulable lesion
Predictive Value		Manipulation, spinal [mh][de]	Range of Motion, articular [mh][de] / Range of motion [sh]
Predictive value of tests [mh][sh][de]			Quality of Motion
Likelihood ratio			Tissue texture
Likelihood functions [mh][de]			Muscle tension / Muscle contraction [mh][sh][de]
ROC			Mobility
Discriminant validity [de]			Stiffness
Discriminant analysis [mh][de]			Myofascial pain syndromes [mh][sh][de]
Gold Standard			Apophyseal
Reference Standards [mh][de] [mh]			Zygapophyseal joint [mh] / Zygapophysial
Instrumentation [sh]			Motion unit
Analysis of variance [mh]			Tender point*
Convergent			Trigger Points [de]
Differential			Neck pain [mh][sh][de]

Predictability			
Construct			
Criterion / Criterion related validity [sh]			#5: Validity Findings
Judgment [sh]			Measurement
Representation			Skin resistance / galvanic skin response [mh][fr]
Stability			Physiologic parameters
Generaliza*			Thermography [mh][sh][de]
Content / Content validity [sh]			Temperature
			Pain measurement [mh][sh][de]
			Radiography [mh] [de] [sh] X-rays [mh] [sh], x-ray [de]

[mh] MeSH -- Medicine Medical Subject Headings and is used for MEDLINE search

[de] descriptor - used for MANTIS and Biosis descriptor search

[sh] Subject headings - used for CINAHL Subject Headings search

