

Research article

Open Access

Content validity of manual spinal palpatory exams - A systematic review

Wadie I Najm*¹, Michael A Seffinger², Shiraz I Mishra³,
Vivian M Dickerson⁴, Alan Adams⁵, Sibylle Reinsch⁶, Linda S Murphy⁷ and
Arnold F Goodman⁸

Address: ¹Department of Family Medicine & Geriatrics, University of California, Irvine, Medical Center, 101 City Drive, Orange, CA 92868, USA, ²Department of Osteopathic Manipulative Medicine, Western University of Health Sciences College of Osteopathic Medicine of the Pacific, 309 E. 2nd St., Pomona, CA 91766-1854, USA, ³Office of the Dean, Western University of Health Sciences College of Osteopathic Medicine of the Pacific, 309 E. 2nd St., Pomona, CA 91766-1854, USA, ⁴Department of Obstetrics & Gynecology, University of California, Irvine, Medical Center, 101 City Drive, Orange, CA 92868, USA, ⁵Office for Academic Affairs and Office of the Provost, 212 Westcott Building, Tallahassee, Florida 32306, USA, ⁶Department of Physical Medicine & Rehabilitation, University of California, Irvine, Medical Center, 101 City Drive, Orange, CA 92868, USA, ⁷Science Library Reference Department, University of California, Irvine, P.O. Box 19557, Irvine, CA 926233-9557, USA and ⁸Center for Statistical Consulting, University of California, Irvine, 4900 Berkeley Place, Irvine, CA 92697, USA

Email: Wadie I Najm* - winajm@uci.edu; Michael A Seffinger - mseffinger@westernu.edu; Shiraz I Mishra - smishra@earthlink.net; Vivian M Dickerson - vdickerson@socal.rr.com; Alan Adams - aadamas@mailier.fsu.edu; Sibylle Reinsch - sreinsch@uci.edu; Linda S Murphy - lmurphy@uci.edu; Arnold F Goodman - agoodman@uci.edu

* Corresponding author

Published: 7 May 2003

Received: 28 February 2003

BMC Complementary and Alternative Medicine 2003, **3**:1

Accepted: 7 May 2003

This article is available from: <http://www.biomedcentral.com/1472-6882/3/1>

© 2003 Najm et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

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, published between 1965–2002, was undertaken. Two blinded reviewers abstracted pertinent data from the retrieved papers, using a specially developed quality-scoring instrument. Five papers met the inclusion/exclusion criteria.

Results: Three of the five papers included in the review explored the content validity of motion tests. Two of these papers focused on identifying the level of fixation (decreased mobility) and one focused on range of motion. All three studies used a mechanical model as a reference standard. Two of the five papers included in the review 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 [1,2]. Approximately 10 percent of the adult population has neck pain at any one time [3], and 80% of the population will experience low back pain (LBP) at some time in their lives [4]. Five to 10 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 [5]. Patients with mechanical spinal pain often seek and receive spinal manipulation by chiropractic, osteopathic and allopathic clinicians, physical therapists or other health care professionals [6].

Health care professionals have utilized spinal palpatory diagnostic procedures and manual manipulative treatment for several millennia to treat back injury and pain [7,8]. Along with the history of illness and physical exam, examiners utilize specific spinal palpatory diagnostic tests in order to identify spinal neuro-musculoskeletal dysfunction. Spinal neuro-musculoskeletal dysfunction refers to an alteration of spinal joint position, motion characteristics and/or related palpable paraspinal soft tissue changes. The scientific committee of the International Federation of Manual Medicine has stated: "beneficial 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" [9].

Spinal palpatory procedures have been described in journals [10–12] and textbooks [13–20]. 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 [21–28]. 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 [29,30]. One study was a limited review of chiropractic literature on palpatory diagnostic procedures for the lumbar-pelvic spine [29] and the other concentrates on validity studies at the sacroiliac joint [30].

An annotated bibliography [31] 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 [32], while reliability measures the concordance, consistency or repeatability of outcomes [25]. 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 [32]. 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 [33]. Face validity studies have been criticized for being subjective, intuitive and unsubstantiated. Troyanovich and Harrison [33] 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 [34]. 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 [35] 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. Thus, they 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 [27]. 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 [36], electromyography [37], and coronary angiography [38] 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 phenomenon 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 [39] and Tarr [40] 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 and comprehensively measures what it claims to be measuring. Although Troyanovich and Harrison [41] 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 [42]. 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 of the same phenomenon 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 [43]. Based on their critical review of the literature, Troyanovich and Harrison [44] 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 phenomenon being palpated. Instead they 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.

Inclusion / Exclusion Criteria

The study inclusion/ exclusion criteria were adapted and modified from those published previously by the Cochrane Collaboration [45] and others [46,47]. 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 (content validity); 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 palpation 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 Additional file 2). After review of the retrieved papers, a secondary exclusion criterion, inappropriate statistical tests used, was applied. Appropriate statistical tests included: sensitivity and specificity, predictive value, likelihood ratio, diagnostic odds ratio, and Receiver Operating Characteristic curve (ROC curves) analysis.

Search strategy

A comprehensive strategy was designed to conduct a detailed search of pertinent literature that addressed the study question, "What is the content validity of spinal palpation tests used to identify spinal neuro-musculoskeletal dysfunction?" Specifics on the search strategy are described in another paper [48]. 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 (see Additional file 3). This defined the research question into

four key concepts: validity/validity findings, spine, palpation procedure, and neuro-musculoskeletal 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 September 30, 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.

Review strategy

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 (see Additional file 1). 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).

Review instruments

Two instruments were developed to extract the data and assess the quality of the studies reviewed. The instruments

Table 2: Quality scoring criteria, total weight and total score assigned.

	CRITERIA	TOTAL WEIGHT	TOTAL SCORE
1	<u>STUDY SUBJECTS</u> Study Subjects Adequately Described Presentation Characteristics Spectrum of severity of Symptoms Subject Selection Criteria Number of Subjects in Study detailed Sample Size Determined by Power Analysis Number Subjects Completed Study Recruitment Procedure	15%	15
2	<u>EXAMINERS</u> Selection Criteria for Examiners Described Background of Examiners Described	10%	10
3	<u>REFERENCE STANDARD</u> Was Reference Standard used Reference standard procedure described / referenced Expected Outcome Described Validity of reference Standard Reliability of Reference Standard Positive or Negative Test Result Defined	15%	15
4	<u>PALPATORY TEST</u> Description of palpatory test Procedure Expected Outcome Described Reliability of Test Described Positive or Negative Test Result Defined	15%	15
5	<u>STUDY CONDITIONS</u> Time Interval for Test/ Retest procedure Examiner blinded to Clinical findings Examiner & Subject blinded to previous study findings Examiner & Subject blinded to Std. reference results	15%	15
6	<u>DATA ANALYSIS</u> Statistical Analysis Used	15%	15
7	<u>RESULTS</u> Results Completely Displayed or Described P-Value Displayed or Described Confidence Interval Displayed or Described Study Bias Identified	15%	15

were developed taking into consideration previously published guidelines [49,50], and instruments [51–55]. 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 (see Table 1).

Detailed information on the 7 components identified to denote internal validity and quality of a study were abstracted and scored. 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 se-

lection 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 (title, names of 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] × 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 [55].

A pilot test of the data extraction and quality assessment instruments was conducted 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 journal, 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 topic in contention by consensus.

Results

Study description

A total of five studies, from the 48 articles retrieved and reviewed, met our inclusion criteria for content validity and are discussed in this study (TABLE 2). The remaining 43 studies [56–98] were retrieved, reviewed and excluded from our study because they either 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 (see Additional file 1). Four studies were published in 4 different journals and the fifth study included is a dissertation [99]. Two studies were unfunded (1 dissertation [99] and 1 did not report any funding [100]). Two studies [101,102], were funded by a Research Council and a liability insurance provider, and one study, [103] was funded by the Chiropractic Advancement Association.

Subjects

The three motion palpation studies were done in the United Kingdom. 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 [101], recruited only pregnant female subjects (n = 200, representing a 90% response rate: 200/222), while the other study [102] 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 [101], 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 [101] blinding was not described.

Measurement

Among the three studies using mechanical models, 2 [100,103] looked at intersegmental motion restriction, and one [99] looked at the ability to determine fixation levels. The mechanical model was the reference standard used.

Table 3: Included studies: details Examiner / Subject / Design / & Examiner 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 (<= 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).

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 [103] 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) (see Table 3). 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.

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 [99] looked only at one group of examiners using two different motion test procedures.

Two studies [100,103] 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 (0.720).

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 50.0% in both studies, for both groups (42.3–46.2%) and for each subgroup. While the NPV was

Table 5: Statistical analysis for Motion Palpation Studies using students and experienced practitioners

Examiners	Test	Harvey (1991)	Jensen (1993)
Both Groups	PPV	0.431	0.459
	NPV	0.9	0.902
	+LR	3.893	3.49
	-LR	0.564	0.403
Student	PPV	0.437	0.367
	NPV	0.898	0.936
	+LR	3.71	4.23
	-LR	0.54	0.337
Practitioner	PPV	0.423	0.454
	NPV	0.903	0.830
	+LR	4.05	2.45
	-LR	0.592	0.603

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

greater than 80% (83–93.7%) supporting the above statement of better capability of these tests at detecting unrestricted than restricted motion (Table 4).

The third motion palpation study [99] 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 (< 31.0%) 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 patient report), provocation test used and population studied.

The cervical study [102] 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 [101] 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 provoca-

tion 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 84.4%. 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 78.4% (see Table 5). 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.

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 phenomenon 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 phenomenon 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 de-

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 single and multiple intersegmental motion restrictions	Sensitivity Interns: 72%; Practitioners: 52.6%; Specificity Interns: 83.2%; Practitioners 78.6% (PPV Interns 46.2%; Pract. 45.5%; NPV Interns 93.7%; Pract 83%)	Motion palpation is an accurate method for determining non-fixed segments but not accurate for determining fixed 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

vice 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 [104]. 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 neuro-musculoskeletal 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 [105–107].

Given that face, construct and criterion validity studies do not measure the phenomenon 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 phenomenon 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 content validity (as defined in this paper) of these widely used tests. Among the few validity studies identified, motion palpation tests were evaluated only by chiropractors and pain studies by physical therapists.

In this review 5 studies focused on three types of tests: fixation (#2), range of motion (#1) and pain (#2). 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 [108] 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 [9]. 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. Also, argument could be advanced that the use of a mechanical model may not have external validity when applied to human subjects. 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.

To translate these results into the clinical setting, additional studies exploring the content validity of spinal palpatory exams, using accepted reference standards are needed.

Identifying a perfect (error free) reference standard for each palpatory test is challenging. Even widely accepted reference standards are imperfect (e.g. histopathology)[109]. Therefore identifying a perfect reference standard is not as important as identifying an acceptable reference standard. Content experts in this field should come to an agreement on acceptable reference standards for spinal palpatory tests.

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 [110] and CONSORT [111,112] 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 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 another systematic review [27], one notes that our scores are consistently lower.

Conclusion

Despite the use of manual spinal palpation by many health care disciplines, very few studies investigated their ability to measure what they intend to measure (content validity). Given the high frequency of spinal pathology and the use of these diagnostic methods to investigate them, 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.

Additional material

Additional File 2

Inclusion and Exclusion Criteria for developed and used for this study review.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6882-3-1-S2.doc>]

Additional File 3

Key terms and search strategy used to identify articles in the databases used in this review.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6882-3-1-S3.doc>]

Additional File 1

Studies reviewed and excluded from this study organized by type of validity they were classified to.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6882-3-1-S1.doc>]

Acknowledgment

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

The authors would like to thank D.V. Gokhale, Ph.D. for his statistical input, Joseph Vorro, Ph.D. and William L. Johnston, D.O. for their review and comments.

References

1. Andersson GB **Epidemiologic aspects on low-back pain in industry** *Spine* 1981, **6**:53-60
2. Loeser JD and Volinn E **Epidemiology of low back pain** *Neurosurg Clin N Am* 1991, **2**:713-718
3. Hadler NM **Illness in the workplace: the challenge of musculoskeletal symptoms** *J Hand Surg [Am]* 1985, **10**:451-456
4. Andersson GB **Epidemiological features of chronic low-back pain** *Lancet* 1999, **354**:581-585
5. Deyo RA **Rethinking strategies for acute low back pain** *Emergency Medicine* 1995, 38-56
6. Hart LG, Deyo RA and Cherkin DC **Physician office visits for low back pain. Frequency, clinical evaluation, and treatment patterns from a U.S. national survey** *Spine* 1995, **20**:111-19
7. Schiötz EH and Cyriax JH *Manipulation past and present: with an extensive bibliography* London: Heinemann Medical 1975,
8. Lomax E **A Historical Perspective from Ancient Times to the Modern Era** In: *The Research Status of Spinal Manipulative Therapy* (Edited by: M G) Bethesda: Maryland: US Dept. of Health, Education and Welfare 1975, **(NIH) # 76-998**:11-17
9. Patijn J and Editor **FIMM (International Federation for Manual/Musculoskeletal Medicine) Scientific Committee: Reproducibility and Validity Studies of Diagnostic Procedures in Manual/Musculoskeletal Medicine for Low Back Pain Patients** Protocol formats [<http://www.dsम्म.org/Videnskab/Validity-protocol.doc>]
10. Dinnar U, Beal MC, Goodridge JP, Johnston WL, Karni Z, Mitchell FL Jr, Upledger JE and McConnell DG **Description of fifty diagnostic tests used with osteopathic manipulation** *J Am Osteopath Assoc* 1982, **81**:314-321
11. 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* 1997, **20**:583-589
12. Walker BF **The Reliability of Chiropractic methods used for the Detection of Spinal Subluxation. An Overview of the Literature** *Aust Chiro Osteo* 1996, **5**:12-22
 13. Ward RC *Foundations for osteopathic medicine Philadelphia: Lippincott Williams & Wilkins* 2003,
 14. Chaitow L *Palpation skills : assessment and diagnosis through touch* New York: Churchill Livingstone 1997,
 15. Bergman T, Peterson D and Lawrence D *Chiropractic techniques. Principles and procedures* New York: Churchill-Livingstone 1993,
 16. Kaltenborn F *The spine: basic evaluation and mobilization techniques* Minneapolis, MN: OPTP 1993,
 17. Maitland G and ed *Maitland's vertebral manipulation* Oxford : Butterworth-Heinemann 2000,
 18. Lewit K *Manipulations therapy in rehabilitation of the locomotor system* Boston: Butterworths 1985,
 19. Maigne R and ed *Doleurs d'origine vertébrale et traitements par manipulations. English. Orthopedic medicine : a new approach to vertebral manipulations* Springfield, Ill.: C.C. Thomas 1972,
 20. Wadsworth C *Manual examination and treatment of the spine and extremities* Baltimore: Williams and Wilkins 1988,
 21. Johnston WL **Interexaminer reliability studies: spanning a gap in medical research – Louisa Burns Memorial Lecture** *J Am Osteopath Assoc* 1982, **81**:819-829
 22. Gonnella C, Paris SV and Kutner M **Reliability in evaluating passive intervertebral motion** *Phys Ther* 1982, **62**:436-444
 23. Panzer DM **The reliability of lumbar motion palpation** *J Manipulative Physiol Ther* 1992, **15**:518-524
 24. Keating JC Jr and Jacobs GE **Inter- and intraexaminer reliability of palpation for sacroiliac joint dysfunction** *J Manipulative Physiol Ther* 1989, **12**:155-158
 25. Haas M **The reliability of reliability** *J Manipulative Physiol Ther* 1991, **14**:199-208
 26. Russell R **Diagnostic palpation of the spine: a review of procedures and assessment of their reliability** *J Manipulative Physiol Ther* 1983, **6**:181-183
 27. Alley J **The clinical value of motion palpation as a diagnostic tool** *Journal of the Canadian Chiropractic Association* 1983, **27**:97-100
 28. Huijbregts P **Spinal motion palpation: a review of reliability studies** *The Journal of Manual & Manipulative Therapy* 2002, **10**:24-39
 29. Hestbaek L and 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**:258-275
 30. van der Wurff P, Meyne W and Hagmeijer RH **Clinical tests of the sacroiliac joint** *Man Ther* 2000, **5**:89-96
 31. Seffinger M, Adams A, Najm W, Dickerson V, Mishra S, Reinsch S and Murphy L **Spinal palpatory diagnostic procedures utilized by practitioners of spinal manipulation: Annotated bibliography of reliability studies** *Journal of the Canadian Chiropractic Association* 2003, **47**:89-105
 32. Winter G **A comparative discussion of the notion of 'validity' in qualitative and quantitative research** *The Qualitative Report* 2000, **4**:[\[http://www.nova.edu/ssss/OR/OR4-3/winter.html\]](http://www.nova.edu/ssss/OR/OR4-3/winter.html)
 33. Troyanovich SJ, Harrison DD and Harrison DE **It's time to accept the evidence** *J Manipulative Physiol Ther* 1998, **21**:568-571
 34. Feinstein A *Clinometrics* New Haven: Yale University Press 1987,
 35. Jull G, Bogduk N and Marsland A **The accuracy of manual diagnosis for cervical zygapophysial joint pain syndromes** *Med J Aust* 1988, **148**:233-236
 36. Kelso AF, Grant RG and Johnston WL **Use of thermograms to support assessment of somatic dysfunction or effects of osteopathic manipulative treatment: preliminary report** *J Am Osteopath Assoc* 1982, **82**:182-188
 37. Vorro J and Johnston WL **Clinical biomechanic correlates of cervical dysfunction: Part 4. Altered regional motor behavior** *J Am Osteopath Assoc* 1998, **98**:317-323
 38. Kelso AF, Larson NJ and Kappler RE **A clinical investigation of the osteopathic examination** *J Am Osteopath Assoc* 1980, **79**:460-467
 39. Beal MC **Palpatory testing for somatic dysfunction in patients with cardiovascular disease** *J Am Osteopath Assoc* 1983, **82**:822-831
 40. Tarr RS, Feely RA, Richardson DL, Mulloy AL, Nelson KE, Perrin WE, Allin EF, Efrusy ME, Greenstein SI and Vatt RD **A controlled study of palpatory diagnostic procedures: assessment of sensitivity and specificity** *J Am Osteopath Assoc* 1987, **87**:296-301
 41. Troyanovich SJ **The reliability and validity of chiropractic assessment procedures** *Chiropractic Technique* 1996, **8**:10-13
 42. Knottnerus J and ed *The evidence base of clinical diagnosis* London : BMJ Books 2002,
 43. Haldeman S, Chapman-Smith D, Peterson DJ and eds *Guidelines for chiropractic quality assurance and practice parameters* Burlingame, CA: Aspen Publishers 1993,
 44. Troyanovich SJ and Harrison DD **In reply to letter to the editor** *J Manipulative Physiol Ther* 1999, **22**:182-183
 45. Clarke M, Oxman A and eds *The Reviewers' Handbook 4.1.6* Oxford: The Cochrane Collaboration 2003,
 46. Khan K, ter Riet G, Glanville J, Sowden A, Kleijnen J and eds *undertaking Systematic Reviews of Research on Effectiveness – CRD's Guidance for those Carrying Out or Commissioning Reviews – CRD Report Number 4* York Publishing Services, Ltd 2001,
 47. Deville WL, Buntinx F, Bouter LM, Montori VM, De Vet HC, Van Der Windt DA and Bezemer P **Conducting systematic reviews of diagnostic studies: didactic guidelines** *BMC Med Res Methodol* 2002, **2**:9
 48. Murphy L, Reinsch S, Najm W, Dickerson V, Seffinger M, Adams A and Mishra S **Searching biomedical databases on Complementary and Alternative Medicine: the use of controlled vocabulary among authors, indexers, and investigators** (In review)
 49. Juni P, Altman DG and Egger M **Systematic reviews in health care: Assessing the quality of controlled clinical trials** *Bmj* 2001, **323**:42-46
 50. Deeks JJ **Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests** *Bmj* 2001, **323**:157-162
 51. Shekelle PG, Adams AH, Chassin MR, Hurwitz EL and Brook RH **Spinal manipulation for low-back pain** *Ann Intern Med* 1992, **117**:590-598
 52. Cook DJ, Sackett DL and Spitzer WO **Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis** *J Clin Epidemiol* 1995, **48**:167-171
 53. Koes B, Assendelft WJ, van der Heijden GJ, Bouter LM and Knipschild PG. **Spinal manipulation and mobilization for back and neck pain: a blinded review** *BMJ* 1991, **303**:1298-1303
 54. Schulz KF, Chalmers I, Hayes RJ and Altman DG **Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials** *JAMA* 1995, **273**:408-412
 55. Juni P, Witschi A, Bloch R and Egger M **The hazards of scoring the quality of clinical trials for meta-analysis** *JAMA* 1999, **282**:1054-1060
 56. Beal MC and Kleiber GE **Somatic dysfunction as a predictor of coronary artery disease** *J Am Osteopath Assoc* 1985, **85**:302-307
 57. Beal MC, Vorro J and Johnston WL **Chronic cervical dysfunction: correlation of myoelectric findings with clinical progress** *J Am Osteopath Assoc* 1989, **89**:891-900
 58. Braun B and Schiffman EL **The validity and predictive value of four assessment instruments for evaluation of the cervical and stomatognathic systems** *J Craniomandib Disord* 1991, **5**:239-244
 59. Breen A and Muggleton **Lumbar Spine Motion Palpation Compared with Objective Intervertebral Motion Analysis Preliminary Findings** *In Process*
 60. Brunarski DJ **Chiropractic biomechanical evaluations: validity in myofascial low back pain** *J Manipulative Physiol Ther* 1982, **5**:155-161
 61. Bush KW, Collins N, Portman L and Tillett N **Validity and intertester reliability of cervical range of motion using inclinometer measurements** *The Journal of Manual & Manipulative Therapy* 2000, **8**:52-61
 62. Byfield D and Mathiasen J **A Preliminary Study Investigating the Accuracy of Bony Landmark Identification in the Lumbar Spine** *European Journal Of Chiropractic* 1991, **39**:105-109
 63. Cox JM, Gorbis S, Dick LM, Rogers JC and Rogers FJ **Palpable musculoskeletal findings in coronary artery disease: results of a double-blind study** *J Am Osteopath Assoc* 1983, **82**:832-836
 64. Dott GA, Hart CL and McKay C **Predictability of sacral base levelness based on iliac crest measurements** *J Am Osteopath Assoc* 1994, **94**:383-390

65. Ebraheim NA, Inzerillo C and Xu R **Are anatomic landmarks reliable in determination of fusion level in posterolateral lumbar fusion?** *Spine* 1999, **24**:973-974
66. Gracovetsky SA, Newman NM, Richards MP, Asselin S, Lanzo VF and Marriott A **Evaluation of clinician and machine performance in the assessment of low back pain** *Spine* 1998, **23**:568-575
67. Gregory P, Hayek R and Mann-Hayek A **Correlating motion palpation with functional X-ray findings in patients with low back pain** *Australasian Chiropractic & Osteopathy* 1998, **7**:15-19
68. Haas M, Peterson D, Hoyer D and Ross G **Muscle testing response to provocative vertebral challenge and spinal manipulation: a randomized controlled trial of construct validity** *J Manipulative Physiol Ther* 1994, **17**:141-148
69. Haas M, Panzer D, Peterson D and Raphael R **Short-Term Responsiveness of Manual Thoracic End-Play Assessment to Spinal Manipulation: A Randomized Controlled Trial of Construct Validity** *Journal Of Manipulative And Physiological Therapeutics* 1995, **18**:582-589
70. Jende A and Peterson CK **Validity of static palpation as an indicator of atlas transverse process asymmetry** *European Journal of Chiropractic* 1997, **45**:35-42
71. Johnston WL, Vorro J and Hubbard RP **Clinical/biomechanic correlates for cervical function: Part I. A kinematic study** *J Am Osteopath Assoc* 1985, **85**:429-437
72. Jull G, Bogduk N and Marsland A **The accuracy of manual diagnosis for cervical zygapophysial joint pain syndromes** *Med J Aust* 1988, **148**:233-236
73. Jull G, Treleaven J and Versace G **Manual examination: is pain provocation a major cue for spinal dysfunction** *Australian Physiotherapy* 1994, **40**:159-165
74. Kawchuk G and Herzog W **The reliability and accuracy of a standard method of tissue compliance assessment** *J Manipulative Physiol Ther* 1995, **18**:298-301
75. Keating L, Lubke C, Powell V, Young T, Souvlis T and Jull G **Mid-thoracic tenderness: a comparison of pressure pain threshold between spinal regions, in asymptomatic subjects** *Man Ther* 2001, **6**:34-39
76. Leboeuf-Yde C and Kyvik KO **Is it possible to differentiate people with or without low-back pain on the basis of tests of lumbopelvic dysfunction?** *J Manipulative Physiol Ther* 2000, **23**:160-167
77. Leboeuf-Yde C, van Dijk J, Franz C, Hustad SA, Olsen D, Pihl T, Robech R, Skov Vendrup S, Bendix T and Kyvik KO **Motion palpation findings and self-reported low back pain in a population-based study sample** *J Manipulative Physiol Ther* 2002, **25**:80-87
78. Leclaire R, Esdaile JM, Jequier JC, Hanley JA, Rossignol M and Bourdouxhe M **Diagnostic accuracy of technologies used in low back pain assessment. Thermography, triaxial dynamometry, spinoscopy, and clinical examination** *Spine discussion* 1331. 1996, **21**:1325-1330
79. Lucchetti CA **Palpation of the C7 Vertebral Spinous Process. An Inter- And Intra-examiner Reliability and Accuracy Study.** *Anglo-European College of Chiropractic* 1991-1992
80. Lundberg G and Gerdl B **The relationships between spinal sagittal configuration, joint mobility, general low back mobility and segmental mobility in female homecare personnel** *Scand J Rehabil Med* 1999, **31**:197-206
81. Maher CG, Latimer J and Adams R **An investigation of the reliability and validity of posteroanterior spinal stiffness judgments made using a reference-based protocol** *Physical Therapy* 1998, **78**:829-837
82. Mayer TG, Kondraske G, Beals SB and Gatchel RJ **Spinal range of motion. Accuracy and sources of error with inclinometric measurement** *Spine* 1997, **22**:1976-1984
83. McPartland JM and Goodridge JP **Counterstrain and traditional osteopathic examination of the cervical spine compared** *Journal of Bodywork & Movement Therapies* 1997, **1**:173-178
84. Mennell J **The validation of the diagnosis "joint dysfunction" in the synovial joints of the cervical spine** *Journal Of Manipulative And Physiological Therapeutics* 1990, **13**:7-12
85. Nansel DD and Jansen RD **Concordance between galvanic skin response and spinal palpation findings in pain-free males** *J Manipulative Physiol Ther* 1988, **11**:267-272
86. Olson KA, Paris SV, Spohr C and Gorniak G **Radiographic assessment and reliability study of the craniocervical sidebending test** *Journal of Manual & Manipulative Therapy* 1998, **6**:87-96
87. Olson SL, O'Connor DP, Birmingham G, Broman P and Herrera L **Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain** *J Orthop Sports Phys Ther* 2000, **30**:13-20
88. Osterbauer PJ, Long K, Ribaudo TA, Petermann EA, Fuhr AW, Bigos SJ and Yamaguchi GT **Three-dimensional head kinematics and cervical range of motion in the diagnosis of patients with neck trauma** *J Manipulative Physiol Ther* 1996, **19**:231-237
89. Phillips DR and 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 MPA, in 1993** *Manual Therapy* 1996, **1**:82-87
90. Simmonds MJ, Kumar S and Lechelt E **Use of a spinal model to quantify the forces and motion that occur during therapists' tests of spinal motion** *Phys Ther* 1995, **75**:212-222
91. Swerdlow B and Dieter JN **An evaluation of the sensitivity and specificity of medical thermography for the documentation of myofascial trigger points** *Pain* 1992, **48**:205-213
92. Tarr R, Feely R and Richardson D **A controlled study of palpatory diagnostic procedures: assessment of sensitivity and specificity** *Journal Of American Osteopathic Association* 1987, **87**:296-302
93. Verhagen AP, Lanser K, de Bie RA and de Vet HC **Whiplash: assessing the validity of diagnostic tests in a cervical sensory disturbance** *J Manipulative Physiol Ther* 1996, **19**:508-512
94. Viikari-Juntura E, Takala EP, Riihimaki H, Malmivaara A, Martikainen R and Jappinen P **Standardized physical examination protocol for low back disorders: feasibility of use and validity of symptoms and signs** *J Clin Epidemiol* 1998, **51**:245-255
95. Visscher CM, Lobbezoo F, de Boer W, van der Zaag J, Verheij JG and Naeije M **Clinical tests in distinguishing between persons with or without craniomandibular or cervical spinal pain complaints** *Eur J Oral Sci* 2000, **108**:475-483
96. Vorro J and Johnston WL **Clinical biomechanic correlates for cervical function: Part II. A myoelectric study** *J Am Osteopath Assoc* 1987, **87**:353-367
97. Vorro J, Johnston WL and Hubbard RP **Clinical biomechanic correlates for cervical function: Part III. Intermittent secondary movements** *J Am Osteopath Assoc* 1991, **91**:145-146
98. Vorro J and Johnston WL **Clinical biomechanic correlates of cervical dysfunction: Part 4. Altered regional motor behavior** *J Am Osteopath Assoc* 1998, **98**:317-323
99. 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
100. Harvey D and Byfield D **Preliminary Studies with a Mechanical Model for the Evaluation of Spinal Motion Palpation** *Clinical Biomechanics* 1991, **6**:79-82
101. Kristiansson P and Svardsudd K **Discriminatory power of tests applied in back pain during pregnancy** *Spine discussion* 1996, **21**:2337-2343
102. Sandmark H and Nisell R **Validity of five common manual neck pain provoking tests** *Scand J Rehabil Med* 1995, **27**:131-136
103. Jensen K, Gemmel H and Thiel H **Motion Palpation Accuracy Using a Mechanical Spinal Model** *European Journal Of Chiropractic* 1993, **41**:67-73
104. Knottnerus A *The evidence base of clinical diagnosis* London: BMJ Books 2002.
105. Olson SL, O'Connor DP, Birmingham G, Broman P and Herrera L **Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain** *J Orthop Sports Phys Ther* 2000, **30**:13-20
106. Price DD, McGrath PA, Rafii A and Buckingham B **The validation of visual analogue scales as ratio scale measures for chronic and experimental pain** *Pain* 1983, **17**:45-56
107. Huskisson EC **Measurement of pain** *J Rheumatol* 1982, **9**:768-769
108. Keating L, Lubke C, Powell V, Young T, Souvlis T and Jull G **Mid-thoracic tenderness: a comparison of pressure pain threshold between spinal regions, in asymptomatic subjects** *Man Ther* 2001, **6**:34-39
109. Walter S, Les Irwig and Glasziou PP **Meta-Analysis of Diagnostic Tests With Imperfect Reference Standards** *J Clin Epidemiol* 1999, **52**:943-951

110. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D and Stroup DF **Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses** *Lancet* 1999, **354**:1896-1900
111. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D and Stroup DF **Improving the quality of reporting of randomized controlled trials. The CONSORT statement** *JAMA* 1996, **276**:637-639
112. Moher D, Schulz KF and Altman D **The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials** *JAMA* 2001, **285**:1987-1991

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1472-6882/3/1/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

