Original articles

Interexaminer reliability of thoracic motion palpation using confidence ratings and continuous analysis

Robert Cooperstein MA, DC^{a,*}, Michael Haneline MS, DC^b, Morgan Young DC^c

^aProfessor, Palmer Chiropractic College, San Jose, CA

^bProfessor, International Medical University, Kuala Lumpur, Malaysia

^cInstructor, Palmer Chiropractic College, San Jose, CA

Received 11 December 2009; received in revised form 28 May 2010; accepted 16 June 2010

Key indexing terms:

Palpation;
Reproducibility of
results;
Spine;
Movement;
Chiropractic

Abstract

Objective: Motion palpation is integral to most chiropractic techniques and can be found in curricula of most every chiropractic college. Paradoxically, most studies do not show strong reliability for motion palpation. The purpose of this study was to determine if allowing motion palpators to rate their confidence in their findings, as well using a continuous data analytic method, would influence the level of concordance.

Methods: Subjects were 52 asymptomatic chiropractic student volunteers. Two palpators assessed posterior to anterior glide of T3–T10 in the prone position, alternating in their order and blinded as to each other's results. Each examiner identified the location of maximal restriction in this range and also whether they were "very confident" or "not confident" in their finding.

Results: For all subjects combined, the examiners' calls were "poor": intraclass correlation coefficient [2,1] = .3110 (95% CI, .0458–.5358). In contrast, interexaminer agreement was "good" when both examiners were very confident: intraclass correlation coefficient [2,1] = .8266 (95% CI, 0.6257–0.9253).

Conclusion: When each examiner was "very confident" as to the most fixated thoracic segment, the levels they identified were very close. This corresponds to "good" agreement, an uncommon result in most interexaminer motion palpation studies. Thus, the confidence level of examiners had an effect on the interexaminer reliability of thoracic spine. Our novel continuous measures, statistical methodology, and subtyping the subjects according to the confidence of the palpators seem more capable than level-by-level discrete analysis of detecting interexaminer agreement.

© 2010 National University of Health Sciences.

* Corresponding author. Palmer Chiropractic College, San Jose, 90 East Tasman Drive, San Jose, CA 95134. Tel.: +1 408 944 6009; fax: +1 408 944 6118.

E-mail address: robert.cooperstein@palmer.edu (R. Cooperstein).

Introduction

The concept of vertebral misalignment and, thus, static listings was present in 1895 at the very beginning

of chiropractic. Palmer,¹ describing his first adjustment, said: "An examination showed a vertebra racked from its normal position. I reasoned that if that vertebra was replaced, the man's hearing should be restored." The concept of joint fixation and, thus, dynamic listings is almost as old, having been described as early as 1906 by Smith et al,² "A simple subluxated vertebra differs from a normal vertebra only in its field of motion and the center of its field of motion; because of its being subluxated, its various positions of rest are differently located than when it was a normal vertebra ... its field of motion may be too great in some directions and too small in others."³

Although motion palpation (MP), the examination procedure most targeted at identifying joint fixation, was established early in the profession's history, it traditionally received far less emphasis than models of chiropractic subluxation based on vertebral misalignment. Nonetheless, the European Henri Gillet was a leading proponent of dynamic analysis throughout his career^{4,5} and strongly impacted on the practice and teaching of an influential American exponent of motion palpation, Faye.⁶ Improper motion (too much, too little, or improper coupling patterns) is now seen as a vitally important component of the chiropractic vertebral subluxation complex.⁷

Motion palpation in one form or another is integral to most chiropractic techniques and can be found taught within the core curriculum of virtually every chiropractic college. Paradoxically, the preponderance of information from dozens of reliability studies show MP to be unreliable, in that palpators do not generally show concordance much above chance levels.⁸ Indeed, literature reviews on the subject have reported kappa values suggesting only slight interexaminer reliability and moderate intraexaminer reliability.⁹⁻¹² On this basis, Troyanovich and Harrison¹³ have opined that chiropractic colleges should desist teaching MP and chiropractors should give up the practice. The subsequent review of Hestbaek and Leboeuf-Yde¹⁴ came to the same conclusion after reviewing 15 studies of motion palpation for the lumbar spine and 6 for the sacroiliac joint, "the esteem chiropractors have for motion palpation in particular has not been substantiated by scientific data." Panzer¹¹ came to a similar conclusion as did Russell¹⁵ in their reviews.

Possible explanations for the general poor reliability of MP have involved variation in procedure,¹⁶ poor interexaminer spinal level localization leading to possible misreported discrepancies^{17,18}, and incorrect landmark rules.¹⁹⁻²¹ It may be theoretically difficult, if not impossible, to determine the reliability of motion

palpation.^{10,22} In motion palpating a research subject, the examiner who goes first most likely alters that subject so that we should not be surprised, nor excessively disheartened, to find that the second examiner does not come up with the same result. The first examiner may attenuate spinal restrictions because the palpatory procedure resembles mobilization, which is, after all, a treatment modality, or the first examiner may leave the subject with aggravated restrictions, with the result of stirring up guarding reactions from injured joints.

Mior et al²³ reported that lack of experience among examiners did not seem to be a viable explanation, nor did providing true or false information regarding the location of pain improve reliability.²⁴ Boline et al²⁵ did not find the use of asymptomatic or slightly symptomatic subjects an apparent confounder compared to using symptomatic subjects. Some researchers have attempted to collapse spinal regions during the analysis of MP data, but this statistical practice has been shown to inflate reliability and be methodologically unacceptable.²⁶ Since not all motion palpation procedures are the same, we hypothesized that the choice of method might impact upon the degree of reproducibility. After categorizing 44 studies as having used either an excursion or end-feel method,²⁷ we found that the high-quality studies did not establish that one method outperformed the other.

Against this backdrop of previous studies, we hypothesized that the study designs that had been heretofore used to investigate the reliability of MP may have been wanting. First, many or even most of the subjects in these studies may have lacked a significant fixation. In the absence of a gold standard as to which of the subjects may have been significantly fixated, the study designs could not distinguish between the following 2 possibilities: the examiners were unable to agree on the location of actual fixations, or the subjects simply lacked detectable fixations. Forcing examiners to say "fixated" or "not-fixated," level by level although in some cases they might have preferred the option to say "I am not sure" may not have given them enough choices and, thus, lowered concordance.

Second, although the profession continues to discuss the etiology of putative fixations, it remains possible that some involve the tethering of spinal segments by muscles and ligaments that span several segments. In such cases, one might expect fixation to manifest as a multilevel rather than strictly segmental finding. Then, asking examiners level by level if a segment is fixated, as every study we have seen has done, may be asking

the wrong question. It may be more relevant to define agreement among examiners as having to do with *how near* their calls are to one another, rather than determining their concordance level by level.

When 2 examiners assess a patient with a musculoskeletal complaint, we would clearly like to distinguish cases in which they completely disagree on the location of a putative vertebral level thought to be clinically relevant and cases in which they *almost* agree on the location. Previous studies have stated that an examiner finding T8 movable and T9 fixated, and another finding T8 fixated and T9 movable, were in complete disagreement, whereas it would have been more illuminating to state they were in *close* agreement, although their calls were not identical. This type of assessment would better capture the essence of how MP is actually done: the palpator examines a region of the spine looking for the most fixated place(s).

The objective of this study was to assess the interexaminer reliability of thoracic MP, taking into account (a) the examiners' confidence in their palpation findings and (b) defining agreement as proximity to each other's findings.

Methods

This study was approved by the Palmer College of Chiropractic Institutional Review Board. All subjects were required to provide written informed consent prior to participation. Subjects were 52 asymptomatic chiropractic students who were selected by convenience during the conduct of a technique laboratory class. Subjects were excluded if they had middle back pain greater than 2 on the 11-point numeric pain scale or if they could not tolerate the palpation procedure for any other reason. None of the potential subjects fulfilled the criteria for exclusion from the study.

The 2 examiners used in this study were licensed doctors of chiropractic (MH and RC), each with more than 20 years of clinical experience. Each had used the palpation procedure that was used in this study (described below) in their practices. The patients were not allowed to speak to the examiners during the examination and were unaware of the results of palpation. The order of examiners was randomized by the toss of a coin to prevent order effects. Each examiner was blinded as to the results of the other's examination of the subjects.

Subjects were placed in the prone position on a pelvic bench, one that did not feature drop pieces or an

unlocking abdominal piece. Before palpation, one of the examiners placed horizontal marks at the T3, T10, and S1 spinous process levels with a skin marking pencil. The MP examination was carried out between the T3 and T10 levels. The MP procedure involved the examiners applying downward thumb pressure over the articular processes of each of the targeted thoracic vertebrae to assess the quality of motion of the segments (ie, joint-play) at end-range. Thus, we sought restrictions in posterior to anterior glide of the posterior thoracic joints. After locating the spinal level of greatest fixation, the first examiner placed a small adhesive backed marker on the subject's skin at the indicated fixation point. The examiner also indicated whether he was "very confident" or "not very confident" in the fixation call. This was a forced call methodology, meaning the examiners simply did not have the option of saying the subject was "not fixated"—they were required to find the most fixated level, even if minimally fixated.

The finding "not very confident" could come about in 2 very different ways: the examiner could not locate a significantly fixated level in the range examined, or there was more than 1 level that the examiner judged to be significantly fixated. A research assistant then measured the distance from the S1 spinous process to the marker with a measuring tape and also recorded the examiner's confidence rating (Fig 1). The sticker was then removed and then the second examiner palpated the subject for the most fixated level and expressed his level of confidence. Some generalized erythema was often induced by the first examiner's palpatory procedure, but it was judged to be fairly uniform and no one location thus became identifiable by the second examiner.

Fig 1. Measuring from S1 to marker identifying site of fixation.

Data analysis

The location of the examiners' marks was measured continuously, so their calls could be compared using correlation statistics, the intraclass correlation coefficient (ICC). The following interpretation scale was used for the reported ICC values: above 0.75 = good reliability, 0.40 to 0.75 = fair to good reliability, below 0.40 = poor reliability.²⁸ The data were entered into an SPSS for Windows (Version 9.01; SPSS, Inc, Chicago, IL) spreadsheet for analysis. The data were exported to a Microsoft Excel (Version 2003, Microsoft Corp, Redmond, Wash) spreadsheet to create the graphs.

Results

Demographics of the 52 study subjects are provided in Table 1. The study outcomes data are provided in Table 2, consisting of 6 groups based on the examiners' confidence ratings, and these groups are portrayed by means of a Venn diagram in Fig 2:

- group 1 (n = 52), all subjects
- group 2 (n = 21) both examiners confident
- group 3 (n = 15), examiner 1 but not examiner 2 confident
- group 4 (n = 6), examiner 2 but not examiner 1 confident
- group 5 (n = 31), only 1 examiner confident
- group 6 (n = 10), neither examiner confident.

Examiner 1 was confident 36% of the time, whereas examiner 2 was confident only 27% of the time. For all subjects combined, group 1, the examiners' calls correlated weakly: ICC [2,1] = .3110 (95% CI, .0458-.5358), and the corresponding index of agreement was judged to be "poor." In contrast, in group 2, in which both examiners were confident, the examiner ratings were highly correlated: ICC [2,1] = .8266 (95% CI, .6257-.9253), and the corresponding index of agreement was judged to be "good." The only other relatively high ICC occurred in group 4, in which the second examiner

Table 2 Study results stratified by group

Group and sample size	ICC [2,1]/95% CI
All subjects	ICC [2,1] = .3110 (95% CI, .0458-.5358)
Both confident	ICC [2,1] = .8266 (95% CI, .6257-.9253)
Ex 1 confident,	ICC [2,1] = -.0431
Ex 2 not confident	(95% CI, -.5183 to .4603)
Group 3, n = 15	
Ex 2 confident,	ICC [2,1] = .6881
Ex 1 not confident	(95% CI, -.0505 to .9484)
Group 4, n = 6	
Ex 1 nor Ex 2 not	ICC [2,1] = -.0748
confident	(95% CI, -.4099 to .2801)
Group 5, n = 31	
Neither Ex 1 nor Ex 2	ICC [2,1] = -.3873
confident	(95% CI, -.7907 to .2729)
Group 6, n = 10	

but not the first was confident: ICC [2,1] = .6881 (95% CI = -.0505 to .9484). The lowest ICC occurred in group 6, in which neither examiner was confident: ICC [2,1] = -.3873 (95% CI, -.7907 to .2729).

At the 2 extremes, the mean of the absolute value of examiner differences was 2.00 cm (about 1 vertebral level) in the n = 21 group 2 where both doctors were confident, and 7.10 cm in the n = 10 group 6 in which neither doctor was confident. Although we recorded the side of fixation in 33 of the subjects, visual inspection of the data showed poor interexaminer agreement, and we decided to not perform statistical analysis. Fig 3 depicts these data using scatterplots. The upper left plot shows a somewhat dispersed cloud reflecting poor agreement for all subjects, whereas the tighter cloud in the upper right plot confirms good agreement for

Table 1 Subject demographics

Males	41 (79%)
Females	11 (21%)
Mean	
Age	25.8 y
Height	172.2 cm
Weight	76.1 kg
Pain level	0.7/10

Fig 2. Venn diagram representation of examiners' ratings.

Fig 3. Scatterplots (mm) depicting interexaminer agreement in 4 representative groups. Axes denominate distance of palpatory finding from sacral marker. Clockwise, starting with upper left: all subjects, both examiners confident, only one doctor confident, neither doctor confident.

subjects where both doctors were confident. The lower 2 plots, the left for situations where only 1 doctor was confident and the right for those in which neither doctor was confident, show essentially no agreement above chance levels.

Discussion

To accomplish *motion palpation*, the examiner introduces motion into joints to assess the range, pattern, and quality of movement. Perhaps the broadest distinction that might be made is between the motion palpation of intersegmental range of motion (ie, excursion), as compared with unisegmental motion (ie, end-feel²⁹). In palpating for excursion, the examiner usually contacts elements of 2 or 3 vertebrae, using the fingers of the palpating hand to assess intersegmental movements, whereas the other hand imparts motion into the articulation(s). This is quantitative analysis, whereby the examiner estimates the amount of movement and generally categorizes the motion segments as hypermobile, normal, or hypomobile. In palpating end-feel, the examiner contacts a single segment, using the fingers of the palpating hand to apply overpressure into rotation, flexion-extension, and lateral flexion, whereas the other hand either stabilizes or assists in imparting motion. This is a more qualitative analysis whereby the results are

interpreted in terms of the unisegmental character of movement; it may lack “springiness” or have a “hard end-feel.” In a previous systematic review of the literature,²⁷ we noted a trend for the end-feel method to be more reliable than the excursion method, although there was no statistically significant advantage when quality ratings of the relevant literature were taken into account. The trend for the end-feel method to outperform the excursion method determined our choice of palpatory methods in this study.

Visual inspection of the scatter plot in Fig 3 is clear: when each examiner was “very confident” as to the most fixated thoracic segment (upper right scatterplot), the levels they identified were very close (within 1 vertebral level). This corresponds to “good” agreement, a result not seen, to our knowledge, in other interexaminer MP studies. The data for all subjects (upper left scatterplot) show much less concordance, and the lower plots in which at least 1 doctor lacked confidence show essentially no concordance. In addition to demonstrating that examiners can, under certain circumstances, agree in their fixation findings, our data suggest that minimally symptomatic and asymptomatic subjects do indeed manifest spinal findings that a palpator may experience as fixation, even in the absence of significant presenting patient complaints. Not surprisingly, among subjects found either barely fixated or significantly fixated at multiple levels, the examiners did not agree above chance levels.

Previous studies that used a forced call paradigm, in which the rater had to find the subjects fixated or not at each spinal level, were very demanding for the examiners, in that they were required to identify all fixations, under the implicit assumption that they were of the same severity. It is unlikely that examiners will agree with one another when the signal to noise ration is very low. These studies analyzed their data using the κ statistic. However, the value of the kappa statistic is changeable when the prevalence of the attribute being tested varies and/or when bias (the degree of disagreement between raters on the proportion of positive or negative cases) is present.³⁰

When 2 examiners assess a patient with a musculoskeletal complaint, we would clearly distinguish cases in which they completely disagree on the location of a putative vertebral level that is relevant and cases in which they *almost* agree on the location. Previous studies would have stated that an examiner finding T8 movable and T9 fixated, and another finding T8 fixated and T9 movable, disagreed, whereas in our study, we would find them to be close agreement, although their calls are not identical.

Among the some 4 dozen MP studies, we discussed in an annotated review of motion palpation⁸ Potter et al³¹ were the only investigators to have used a most fixated segment paradigm similar to ours. Like ourselves, these investigators used ICC for the purposes of analysis. Since theirs was an intraexaminer study, unlike ours, and furthermore used findings in addition to MP to assess agreement, we can not otherwise compare the results of their study with our own.

In our study, the palpators did not have any verbal interaction with the subjects, so that findings of fixation could be considered central to their identification of dysfunctional spinal segments. We wanted to avoid confounding our objective findings with subjective information about pain or tenderness. Although the motion palpation procedures described and tested by Jull et al^{32,33} are sometimes regarded to be valid, their interpretation is questionable because in their work the examiner's finding of restriction is commingled with other findings, such as patient-reported tenderness and soft-tissue textural changes. Thus, it is not clear that the finding of fixation per se is central to their identification of dysfunctional spinal segments.

We should note that at least 1 other MP reliability study acknowledged the excessive stringency of assessing interexaminer agreement at individual motion segments. In a study by Christensen et al,³⁴ examiners were considered to be in agreement when their calls were within ± 1 spinal segment of each other.

Intraexaminer reliability was reported to be good ($\kappa = 0.59$ to 0.77), whereas interexaminer reliability was low ($\kappa = 0.24$ and 0.22). In addition, Humpreys et al³⁵ used a "most fixated level" method in a validity study, which assessed the accuracy of blinded palpators in detecting fixation in 3 subjects with congenital block vertebrae as a reference standard.

Limitations

The sample size was relatively small after it was stratified by degree of confidence. Nonetheless, there was a robust contrast between reported indices of agreement when examiners were very confident compared with when they were not confident. The use of convenience samples (mostly asymptomatic subjects) in MP studies has been previously criticized^{14,18} although there is some evidence it makes no difference.²⁴ It appeared our study featured a mix of subjects with varying degrees of fixation, partially satisfying the acknowledged condition that diagnostic studies include a spectrum of subjects with the target disorder.³⁶

We did not explore the question as to whether an examiner's lack of confidence was related to the absence of palpable fixations or the existence of multiple fixations (wherein no maximally fixated segment could be identified) because the subgroups were too small. It is not obvious why examiner agreement was limited to level and did not include side-specificity. The examiners did not call out many such multiple fixations, and when they did, poststudy discussion suggested the data were inconsistently recorded, precluding analysis. The results of this study of thoracic MP may not be relevant to studies of cervical and lumbar MP using similar methodology. With data collection in progress, we will report on similar studies of these regions in other publications.

The most glaring limitation of this study, at least as we see it, is not related to its methodology or findings so much the uncertain clinical value of MP in determining the optimal locations to target adjustive and other therapeutic procedures. In other words, to our knowledge it has not been demonstrated that the information provided by MP improves the outcome of clinical care. Indeed, at least 1 study by Haas et al³⁷ suggests that end-play assessment does not contribute to same-day clinical improvement in the cervical spine, although the investigators do not rule out possible contribution over a longer term. Moreover, the study design did not allow distinguishing between MP being intrinsically not useful, the

adjustor being nonspecific, or the motion palpator being inaccurate.

It is difficult to discern from the literature which type of examination findings would be most clinically informative on deciding where and how to adjust chiropractic patients. The PARTS acronym (pain, asymmetry, range of motion, tone/texture/temperature, and special tests)³⁸ is widely respected in chiropractic, as is the very similar TART acronym (tissue texture changes, asymmetry, restriction of motion, tenderness) in osteopathy,³⁹ but assessment of their clinical utility awaits outcome studies. In the meantime, the absolute and relative importance of fixation, pain provocation, tenderness, temperature asymmetry, misalignment, functional leg length inequality, or other types of examination findings are unclear.

Conclusions

The confidence level of examiners has an effect on the interexaminer reliability of thoracic spine MP, such that agreement is “good” when examiners are “very confident” in their calls and not above chance levels when at least one of them is not. Looking at the data set as a whole, unstratified by degree of examiner confidence, our results resemble those of other investigators, in that the index of agreement is low. Thus, we believe using continuous measures methodology, and defining subgroups according to the confidence of the palpators, is more capable than level-by-level discrete analysis of detecting interexaminer agreement. We also believe our analytic method better reflects what motion palpators, who presumably look for maximally fixated levels within a spinal region logically related to a patient complaint, actually do.

We would suggest that future studies deploying a similar methodology, using confidence ratings and continuous analysis, use a more representative mix of study subjects, some with and some without clinically significant pain. Moreover, we would avoid using the “not confident” rating to refer to 2 very different clinical situations: the finding of multiple fixations and that of not finding any significant fixations at all. This complicates the analysis greatly because it confounds subjects that seem so fixated that multiple levels would be chosen, and others who seem not fixated at all.

Ultimately, it is desirable that chiropractic education mirror the clinical situations that graduates are likely to encounter as closely as possible. Since we would expect clinicians to detect, make record of, and treat the most fixated level(s) within a range including the area of chief complaint, we would think it reasonable to teach

chiropractic interns to examine patient just that way. This would be more relevant than asking them to agree or disagree, level by level, on the segmental motion or lack thereof, with the instructors or with each other.

Funding sources and potential conflicts of interest

No funding sources or conflicts of interest were reported for this study.

References

1. Palmer DD. The chiropractor's adjuster, the science, art and philosophy of chiropractic. Portland, Oregon: Portland Printing House; 1910.
2. Smith OG, Langworthy SM, Paxson MC. Modernized chiropractic. Cedar Rapids: Lawrence Press; 1906.
3. Leach RA. Stephenson's Principles revisited in 1997. *Dynamic Chiropractic* 1997;15(11):42, 45.
4. Gillet H, Liekens M. Belgian chiropractic research notes. 10th ed.; 1973.
5. Gillet H, Liekens M. The different types of fixation. The Belgian chiropractic research notes. Huntington Beach, CA: Motion Palpation Institute; 1981. p. 13-6.
6. Faye LJ. The subluxation complex. *J Chiropr Humanit* 1999;9(1):1-4.
7. Lantz C. The vertebral subluxation complex—Part I: an introduction to the model and the kinesiological component. *Chiropr Res J* 1989;1(3):23-36.
8. Haneline M, Cooperstein R, Young M, Birkeland K. An annotated bibliography of spinal motion palpation reliability studies. *J Can Chiropr Assoc* 2009;53(1):40-58.
9. Dishman RW. Static and dynamic components of the chiropractic subluxation complex: a literature review. *J Manipulative Physiol Ther* 1988;11(2):98-107 [see comment in *J Manipulative Physiol Ther* 1989;12(2):152].
10. Breen A. The reliability of palpation and other diagnostic methods. *J Manipulative Physiol Ther* 1992;15(1):54-6.
11. Panzer DM. The reliability of lumbar motion palpation. [Erratum in *J Manipulative Physiol Ther* 1992 Nov-Dec;15(9):following table of contents] *J Manipulative Physiol Ther* 1992;15(8):518-24.
12. Haas M, Panzer D, Raphael R. Reliability of manual end-play palpation of the thoracic spine. *Chiropr Tech* 1995;7(4):120-4.
13. Troyanovich SJ, Harrison DD. Motion palpation: it's time to accept the evidence. *J Manipulative Physiol Ther* 1998;21(8):568-71.
14. Hestbaek L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. *J Manipulative Physiol Ther* 2000;23(4):258-75.
15. Russell R. Diagnostic palpation of the spine: a review of procedures and assessment of their reliability. *J Manipulative Physiol Ther* 1983;6(4):181-3.

16. Marcotte J, Normand MC, Black P. Measurement of the pressure applied during motion palpation and reliability for cervical spine rotation. *J Manipulative Physiol Ther* 2005;28(8):591-6.
17. Billis EV, Foster NE, Wright CC. Reproducibility and repeatability: errors of three groups of physiotherapists in locating spinal levels by palpation. *Man Ther* 2003;8(4):223-32.
18. Huijbregts PA. Spinal motion palpation: a review of reliability studies. *J Manipulative Physiol Ther* 2002;10(1):24-39.
19. Haneline M, Cooperstein R, Young M, Ross J. Determining spinal level using the inferior angle of the scapula as a reference landmark: a retrospective analysis of 50 radiographs. *J Can Chiropr Assoc* 2008;52(1):24-9.
20. Cooperstein R, Haneline M. Where is the inferior angle of the scapula? *Dynamic Chiropr* 2008;26(8).
21. Cooperstein R, Haneline M. Spinous process palpation using the scapular tip as a landmark vs a radiographic criterion standard. *J Chiropr Med* 2007;6(3):87-93.
22. Cooperstein R. The chiropractic technique-research interface: when neither loyal ignorance nor enlightened despair will do. *The Bartlett* 2001;9(2):7-8, 23.
23. Mior SA, McGregor M, Schut B. The role of experience in clinical accuracy. *J Manipulative Physiol Ther* 1990;13(2):68-71.
24. DeCina P, Mior S. Interexaminer reliability of motion palpation: the effect of knowledge of the location of pain. In: FCER, editor. *Proceedings of the 1992 International Conference on Spinal Manipulation*; 1992. Chicago: FCER; 1992. p. 106.
25. Boline P, Keating J, Brist J, Denver G. Interexaminer reliability of palpatory evaluations of the lumbar spine. *Am J Chiropr Med* 1988;1(1):5-11.
26. Haas M. Statistical methodology for reliability studies. *J Manipulative Physiol Ther* 1991;14(2):119-32.
27. Haneline MT, Cooperstein R, Young M, Birkeland K. Spinal motion palpation: a comparison of studies that assessed intersegmental end feel vs excursion. *J Manipulative Physiol Ther* 2008;31(8):616-26.
28. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2000.
29. Brown J, Cooperstein R. Why motion palpation is so confounding. *J Am Chiropr Assoc* 2001;38(10):34-6.
30. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;304(6840):1491-4.
31. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *Phys Ther* 1985;65(11):1671-5.
32. Jull G, Bogduk N, Marsland A. The accuracy of manual diagnosis for cervical zygapophysial joint pain syndromes. *Med J Aust* 1988;148(5):233-6.
33. Jull G, Bullock M. A motion profile of the lumbar spine in an aging population assessed by manual examination. *Physiother Theor Pract* 1987;3:70-81.
34. Christensen HW, Vach W, Vach K, et al. Palpation of the upper thoracic spine: an observer reliability study. *J Manipulative Physiol Ther* 2002;25(5):285-92.
35. Humphreys BK, Delahaye M, Peterson CK. An investigation into the validity of cervical spine motion palpation using subjects with congenital block vertebrae as a 'gold standard'. *BMC Musculoskelet Disord* 2004;15(5).
36. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994;271(5):389-91.
37. Haas M, Group E, Panzer D, Partna L, Lumsden S, Aickin M. Efficacy of cervical endplay assessment as an indicator for spinal manipulation. *Spine* 2003;28(11):1091-6 [discussion 1096].
38. Bergmann T. P.A.R.T.S. joint assessment procedure. *Chiropr Tech* 1993;5(3):135-6.
39. Cooperstein R. ABS annual meeting: San Francisco, Nov. 16-19, 2005 (Part 1 of 2). *Dynamic Chiropractic* 2006;24(8).