# Literature Review

# A Review of Intraexaminer and Interexaminer Reliability of Static Spinal Palpation: A Literature Synthesis

Michael T. Haneline, DC, MPH,[a] and Morgan Young, DC[b]

## Abstract

**Objective:** The purposes of this study were to locate articles that assessed the reliability of static palpation of the spine and sacroiliac joints, to appraise the quality of these studies, and synthesize their results.

**Methods:** A structured literature search was conducted of chiropractic and medical databases PubMed, Manual Alternative and Natural Therapy System, Index to Chiropractic Literature, and Cumulative Index to Nursing and Allied Health Literature from 1965 through October 2007. Reference sections were inspected for additional citations. Only peer-reviewed articles in English containing information about static palpation of the spine or sacroiliac joints were selected. The resulting studies were appraised for quality by both of the authors using a 6-point scale instrument developed to assess the quality of reproducibility studies.

**Results:** The search generated 343 citations, and another 7 were harvested from the reference lists. After removing articles not meeting the inclusion criteria, 29 were retained. A total of 14 studies focused on the reliability of locating painful or tender points, 10 on the location of landmarks, and 5 on position or alignment of bone structures. A higher proportion of studies that assessed painful or tender points reported acceptable levels of reliability. However, there were no significant differences between methods of palpation when considering the proportions of high-quality studies that reported good reliability. Thus, no form of static palpation could be considered to be superior.

**Conclusion:** Reported indices of agreement were generally low. More of the pain palpation studies reported acceptable $\kappa$ levels, although no one method of palpation could be deemed clearly superior. (J Manipulative Physiol Ther 2009;32:379-386)

**Key Indexing Terms:** *Palpation; Chiropractic; Spine; Sacroiliac Joint*

Static spinal palpation is commonly used by musculoskeletal practitioners as a component of the evaluative process used in making patient management decisions. It may be used to determine the position of bone structures, sites of pain or tenderness, or particular spinal levels. Knowing the actual position of a vertebral segment would be useful to those aiming to reposition the bone to a more normal alignment. Being able to recognize the sites of pain or tenderness is helpful in directing therapies to the correct regions of the spine. The recognition of precise spinal levels is important when one wants to transfer information gained from a patient's spinal examination or x-ray findings to their treatment.

To be of value in the clinical decision-making process, spinal palpation should be reproducible when 1 examiner repeatedly evaluates the same group of subjects (intraexaminer reliability) and when more than 1 examiner evaluates the same group of subjects (interexaminer reliability). A number of studies have evaluated the reliability of static spinal palpation, which have generally found it to be low. Several subsequent reviews have included spinal palpation studies, which highlight these findings, although palpation of tender or painful segments was reported to be more reliable than other forms of static palpation.[1-3]

The purposes of this review were to locate studies that have assessed the reliability of static palpation of the spine and sacroiliac (SI) joints, to appraise the quality of these studies, and synthesize their results. To our knowledge, no previous reviews have focused on this topic. We dealt with the reliability of static palpation in this review, although its validity will be reported elsewhere.

[a] Professor, Palmer College of Chiropractic West, San Jose, Calif.

[b] Instructor, Palmer College of Chiropractic West, San Jose, Calif.

Submit requests for reprints to: Michael T. Haneline, DC, MPH, Professor, 90 E. Tasman Drive, San Jose, CA 95134 (e-mail: *michael_haneline@imu.edu.my*).

Paper submitted November 25, 2008; in revised form January 2, 2009; accepted January 12, 2009.

0161-4754/$36.00



doi:10.1016/j.jmpt.2009.04.010

## Methods

A search was conducted of the MEDLINE-PubMed, Manual Alternative and Natural Therapy System (MANTIS), Index to Chiropractic Literature (ICL), and Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases for the years 1965 through October 2007. The PubMed search involved the following strategy: "(palpation NOT motion) and (reliability or validity or accuracy) and (spine or back or neck or thoracic or lumbar or SI)." MANTIS was searched using the Controlled Supplemental Vocabulary terms "palpation and Spine and (reliability or validity)." The CINAHL and ICL searches were purposefully broad, using "palpation" as the only query term. The reference sections of the literature generated by the searches were inspected for additional relevant citations. Only peer-reviewed articles in the English language that purported to contain information on the subject of static palpation of the spine or SI regions were selected for review.

### Inclusion and Exclusion Criteria

To be included in the literature review, articles had to be original studies that investigated the intra- and/or interexaminer reliability of manual static palpation of the spine or SI regions and published in a refereed journal between 1965 and October 2007 in the English language. Articles that did not meet the inclusion criteria were excluded.

Studies that investigated the validity of manual static palpation of the spine or SI regions were not included because these will be dealt with in another review. Literature reviews, commentaries or editorial articles, letters to the editor, conference proceedings, abstracts, and articles that presented unclear data were excluded.

### Quality Assessment

The studies included in this review were appraised for quality by both of the authors using a 6-point scale developed by Stochkendahl et al,[3] which was specifically developed to assess the quality of reproducibility studies (Appendix A). A maximum of 6 points are possible in this scale for interexaminer reliability studies, whereas 4 points maximum are possible for studies dealing with intraexaminer reliability. The final score may be converted to a percentage, as was done in this review. Disagreements about the ratings of the articles' quality were resolved by consensus. Interexaminer studies that scored 3 or more points (≥50%) were considered to be high quality, and those scoring less than 3 (<50%) low quality. Intraexaminer studies that scored 2 or more points (≥50%) were considered to be high quality, and those scoring less than 2 (<50%) low quality. Data from the included articles were extracted and recorded by both of the authors.

The extent of examiner reliability for the included studies was characterized using the following interpretation of $\kappa$ values: 0, none; 0 to 0.2, slight; 0.2 to 0.4, fair; 0.4 to 0.6, moderate; 0.6 to 0.8, substantial; 0.8 to 1.0, almost perfect.[4] Studies reporting ranges of $\kappa$ values 0.4 or greater were considered to have shown an acceptable level of reliability.[5] Some studies only provided the percentage of agreement between examiners, which was considered to be inconclusive, because of the inherent limitations associated with assessing examiner agreement using this statistic.[6] The intraclass correlation coefficient (ICC) was used in 1 study, wherein we used the following interpretation: greater than 0.75, good reliability; 0.40 to 0.75, fair to good reliability; less than 0.40, poor reliability.[7] The F statistic and the coefficient of variation were also used to denote the degree of examiner agreement. Published interpretations of these statistics with regard to the findings of reproducibility studies are not available, however, so studies using these indices of agreement were considered to be inconclusive.

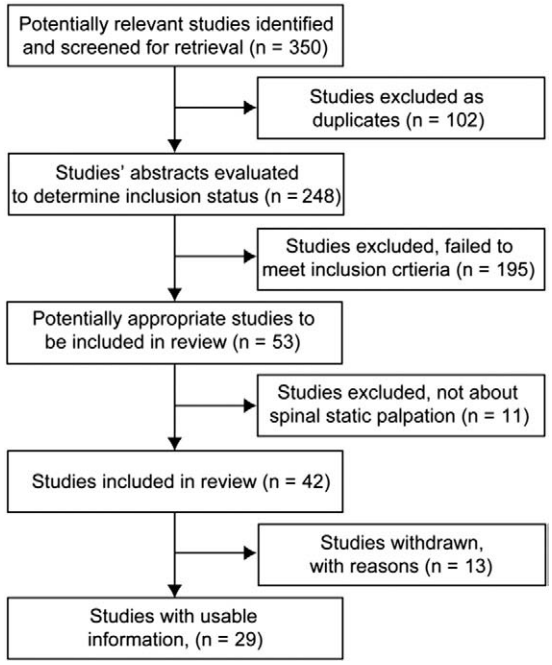


**Fig 1.** *Flowchart of the article selection process.*

### Comparative Analysis

The studies' ranges of $\kappa$ values were depicted in graphical form to permit visual comparison of the outcomes. Studies that did not use the $\kappa$ statistic to represent the degree of agreement were not included in the graphs.

### Levels of Evidence

Levels of evidence in support of the various types of static spinal palpation were categorized according to the following rating system,[3] which takes into consideration the total number of studies, the methodological quality of

**Table 1.** *Studies that focused on the assessment of the reliability of pain provocation in the spine and/or SI region*

| Author | Region | Examiners, experience | Subjects | Study type | Quality score[a] (%) | Findings | Degree of reliability |
|---|---|---|---|---|---|---|---|
| Boline et al (1988)[32] | T12-S1 | 1 DC, <1 y, 1 St | 50 (23 Sx, 27 Asx) | Inter | 83 | $\kappa$ = −0.03 to 0.49<br>% = 60 to 90 | None to moderate |
| Boline et al (1993)[33] | L1-S1 | 3 DC, Exp | 28 Sx | Inter | 50 | $\kappa$ = 0.48 to 0.90<br>% = 79 to 96 | Moderate to almost perfect |
| Christensen et al[34] | T1-T8 | 2 DC, Exp | 107 (51 Sx angina, 56 Asx) | Inter | 100 | $\kappa$ = 0.38 to 0.70 | Fair to substantial |
| | T1-T8 | 2 DC, Exp | 107 (51 Sx angina, 56 Asx) | Intra | 100 | $\kappa$ = 0.34 to 0.77 | Fair to substantial |
| Deboer et al[35] | C1-C7 | 3 DC, Exp | 40 Asx | Inter | 50 | $\kappa$ = −0.04 to 0.48 | None to moderate |
| | C1-C7 | 3 DC, Exp | 40 Asx | Intra | 25 | $\kappa$ = 0.20 to 0.56 | Fair to moderate |
| Hubka and Phelan[36] | C2-C7 | 2 DC, 1-5 y | 30 Sx | Inter | 50 | $\kappa$ = 0.68 | Substantial |
| Keating et al[37] | T12-S1 | 3 DC, >2.5 y | 46 (21 Sx, 25 Asx) | Inter | 67 | $\kappa$ = 0.19 to 0.48 | Slight to moderate |
| Lundberg and Gerdle[38] | T10-S1 | 2 PT, Exp | 150 | Inter | 50 | $\kappa$ = 0.67 to 0.71 | Substantial |
| McCombe et al[39] | L1-L5, SI | 3 MD, 1 PT, Exp | 83 Sx | Inter | 17 | $\kappa$ = 0.28 to 0.47 | Fair to moderate |
| Paydar et al[40] | SI | 2 St | 32 Asx | Inter | 50 | $\kappa$ = 0.73<br>% = 90.6 | Substantial |
| | SI | 2 St | 32 Asx | Intra | 25 | $\kappa$ = 0.91<br>% = 96.8 | Almost perfect |
| Strender et al [41] | C0-C3 | 2 PT, ≥21 y | 50 (25 Sx, 25 Asx) | Inter | 67 | $\kappa$ = 0.31 to 0.52<br>% = 58 to 68 | Fair to moderate |
| Strender et al[42] | L5-S1 | 2 MD, 2 PT, Exp | 71 Sx | Inter | 67 | $\kappa$ = 0.06 to 0.71<br>% = 73 to 88 | Slight to substantial |
| Van Suijlekom[43] | CO - C7 | 2 neuro, Exp | 24 Sx | Inter | 17 | $\kappa$ = 0.14 to 0.31 | Slight to fair |
| Viikari-Juntura[44] | C1-C7 | 1 MD, 1 PT, Exp | 52 Sx | Inter | 17 | $\kappa$ = 0.47 to 0.56 | Moderate |
| Waddell et al[45] | L1-S1 | 4 MD, Exp | 475 Sx, 335 Asx | Inter | 33 | $\kappa$ = 1.0<br>% = 100 | Almost perfect |

C indicates cervical; T, thoracic; L, lumbar; S, sacral; Sx, symptomatic; Asx, asymptomatic; Inter, interexaminer reliability; Intra, intraexaminer reliability; %, percent agreement; DC, doctor of chiropractic; MD, doctor of medicine; PT, physical therapist; St, student; Exp, experienced.

[a] Percentage rounded off the nearest whole number.

the studies (ie, quality scores), and the consistency of the study outcomes:

- *Strong evidence*: provided by generally consistent findings in multiple (≥2) high-quality studies
- *Moderate evidence*: provided by generally consistent findings in 1 high-quality study and 1 or more low-quality studies or in multiple (≥2) low-quality studies
- *Preliminary evidence*: only 1 study available
- *Conflicting evidence*: inconsistent findings in multiple (≥2) studies
- *No evidence*: no studies were identified.

Consistency was determined by comparing study outcomes presented in the tables and figures.

### Data Analysis

Differences between proportions were tested for statistical significance using Yates corrected $\chi^2$ test. Data analyses were carried out using SPSS for Windows (Version 15.0.1; SPSS, Inc, Chicago, Ill).

## Results

The search was carried out on October 26, 2007, generating 343 citations (143 from PubMed, 89 from MANTIS, 58 ICL, and 53 from CINAHL); another 7 were harvested from the reference lists. One hundred two duplicates were found, and 195 articles did not meet the inclusion criteria, resulting in the elimination of 297

**Table 2.** *Studies that focused on the assessment of the reliability of locating landmarks in the spine and/or SI region*

| Author | Region | Examiners, experience | Subjects | Study type | Quality Score[a] (%) | Findings | Degree of reliability |
|---|---|---|---|---|---|---|---|
| Keating et al[37] | T12-S1 | 3 DC, >2.5 y | 46 (21 Sx, 25 Asx) | Inter | 67 | $\kappa$ = −0.08 to 0.03 | None to slight |
| Billis et al[46] | C5, T6, L5 | 17 PT, ≥2 y, 13 PT St | 9 Asx | Inter | 67 | F = 18.43<br>*P* = .001 | Inconclusive |
| | C5, T6, L5 | 17 PT, ≥2 y, 13 PT St | 9 Asx | | 50 | F = 2.09<br>*P* = 0.161 | Inconclusive |
| Binkley et al[47] | L1-S1 | 6 PT, at least 6 y | 18 Sx | Inter | 50 | $\kappa_w$ = 0.30<br>ICC = 0.69<br>(CI, 0.53-0.82) | Fair<br>Fair to good |
| Broadbent et al[48] | T12-S1 | 2 MD, NI | 100 Sx | Inter | 50 | $\kappa_w$ = 0.43-0.63 | Moderate to substantial |
| Byfield and Humphreys[49] | L1, L4 | 2 DC, Exp | 42 Asx | Inter | 17 | % = 55-81 | Inconclusive |
| | L1, L4 | 2 DC, Exp | 42 Asx | Intra | 0 | % = 39-62 | Inconclusive |
| Downey et al[50] | L1-L5 | 6 PT, >7 y | 20 Sx | Inter | 33 | $\kappa_w$ = 0.44-0.98 | Moderate to almost perfect |
| Holmgren and Waling[51] | L5 and SI | 3 PT, ~15 y | 25 Sx | Inter | 67 | $\kappa$ = 0.11-0.17 | Slight |
| McKenzie and Taylor[52] | L1-L5 | 14 PT, Inexp | 5 Asx | Inter | 17 | $\kappa$ = 0.28<br>% = 56 | Fair |
| | L1-L5 | 3 PT, >5 y | 5 Asx | Intra | 25 | $\kappa$ = 0.61-0.9<br>% = 84-96 | Substantial to almost perfect |
| O'Haire and Gibbons[53] | SI | 10 DO, fifth-year St | 10 Asx | Inter | 50 | $\kappa$ = 0.04-0.08 | Slight |
| | SI | 10 DO, fifth-year St | 10 Asx | Intra | 25 | $\kappa$ = −0.05 to 0.58 | None to moderate |
| Simmonds and Kumar[54] | L4, SI | 20 PT, St | 20 Asx | Inter | 33 | CoefVar = 0.48-0.65 | Inconclusive |
| | L4, SI | 20 PT, St | 20 Asx | Intra | 25 | CoefVar = 0.28-0.78 | Inconclusive |

C indicates cervical; T, thoracic; L, lumbar; S, sacral; Sx, symptomatic; Asx, asymptomatic; Inter, interexaminer reliability; Intra, intraexaminer reliability; %, percent agreement; CI, 95% confidence interval; CoefVar, coefficient of variation; DC, doctor of chiropractic; MD, doctor of medicine; DO, doctor of osteopathic medicine; PT, physical therapist; St, student; Exp, experienced; NI, no information presented.

[a] Percentage rounded off the nearest whole number.

references. Thus, 52 articles were considered to be potentially relevant, although, after reading the articles' full-text, another 24 were eliminated for the following reasons: did not involve spine or SI region,[8-11] validity study,[12] review or commentary article,[13,14] abstract only,[15] motion palpation,[16-20] combined tests,[21-24] not palpation,[25,26] not reliability,[27] involved cadavers,[28,29] and unclear data.[30,31] Twenty-nine articles were ultimately included in this review and were appraised for quality (Fig 1).

A total of 14 of the included studies focused on the assessment of the reliability of locating painful or tender points[32-45] (Table 1), 10 focused on the location of landmarks[37,46-54] (Table 2), and 5 focused on position or alignment of the bone structure[55-59] (Table 3). One of the studies[37] considered pain provocation as well as landmark identification and was therefore included in both Tables 1 and 2.

Of the 14 studies that assessed the location of painful or tender points, 6 reported acceptable levels of interexaminer reliability,[33,36,38,40,44,45] although 2 of them[44,45] were considered to be low quality. One of the pain location studies reported acceptable intraexaminer reliability,[40] but it was deemed to be low quality.

Of the 10 studies that assessed locating landmarks, 2 reported acceptable levels of interexaminer reliability,[48,50] and 1 reported acceptable intraexaminer reliability.[52] Nevertheless, only 1 of these 3 studies was above the cutoff for acceptable quality,[48] scoring 50%. Billis et al[46] also reported acceptable intraexaminer reliability, but their use of the F statistic was inconclusive in determining examiner agreement.

None of the 6 studies that assessed the position or alignment of spinal structures reported acceptable levels of interexaminer reliability, and only 1 reported acceptable intraexaminer reliability.[56] However, the quality of the single study that reported acceptable intraexaminer reliability was below the preestablished standard.

There were no statistically significant differences in any of the pairwise comparisons of the proportions of

**Table 3.** *Studies that focused on the assessment of the reliability of spine and/or SI position or alignment*

| Author | Region | Examiners, experience | Subjects | Study type | Quality score[a] (%) | Findings | Degree of reliability |
|---|---|---|---|---|---|---|---|
| Collaer et al[55] | Lumbar | 3 PT, >11 y | 30 Sx | Inter | 67 | $\kappa$ = 0.18-0.39<br>% = 63-76 | Slight to fair |
| Fryer and O'Keefe[56] | SI | 10 DO St, trained and untrained | 10 Asx | Inter | 33 | $\kappa$ = 0.08 trained, 0.15 untrained | Slight |
| | SI | 10 DO St, trained and untrained | 10 Asx | Intra | 25 | $\kappa$ = 0.54 trained, 0.49 untrained | Moderate |
| Hart[57] | C1-C2 | 12 DC, NI | 31 Sx | Inter | 33 | $\kappa$ = −0.27 to 0.38<br>% = 11-58 | None to fair |
| Keating et al[37] | T12-S1 | 3 DC, >2.5 y | 46 (21 Sx, 25 Asx) | Inter | 67 | $\kappa$ = −0.16 to 0.22 | None to fair |
| Potter and Rothstein[58] | SI | 8 PT, >2 y | 17 Sx | Intra | 0 | % = 44-50 | Inconclusive |
| Spring and Tehan[59] | L1-L5 | 10 DO, St | 10 Asx | Inter | 83 | $\kappa$ = 0.04 | Slight |
| | L1-L5 | 10 DO, St | 10 Asx | Intra | 75 | $\kappa$ = 0.04 | Slight |

C indicates cervical; T, thoracic; L, lumbar; S, sacral; Sx, symptomatic; Asx, asymptomatic; Inter, interexaminer reliability; Intra, intraexaminer reliability; %, percent agreement; DC, doctor of chiropractic; DO, doctor of osteopathic medicine; PT, physical therapist; St, student; NI, no information presented.

[a] Percentage rounded off the nearest whole number.

**Table 4.** *Pairwise comparisons of the proportion of high-quality studies that reported good reliability between the 3 forms of static palpation*

| Comparison | $\chi^2$ | *P* |
|---|---|---|
| Location of painful or tender points vs position or alignment | 0.09 | .77 |
| Location of landmarks vs position or alignment | 0.00 | 1.0 |
| Location of painful or tender points vs location of landmarks | 0.4 | .53 |

high-quality studies that reported $\kappa$ values of 0.4 or greater in relation to the total number of studies for each of the 3 types of palpation (Table 4). Thus, no form of static palpation could be considered to be clearly superior. Figure 2 provides a graphical representation of the ranges of $\kappa$ presented by the various studies, arranged by type of static palpation.

### Levels of Evidence

The evidence supporting the interexaminer and intraexaminer reproducibility of the location of painful or tender points in the spine was conflicting, given that more than 2 high-quality studies reported dissimilar findings. There is also conflicting evidence with regard to the reproducibility of locating spinal landmarks. There is moderate evidence concerning the reproducibility of position or alignment of spinal structures pointing to poor reliability because only 1 low-quality study reported an acceptable level of intraexaminer reliability.

## Discussion

Overall, higher levels of inter- and intraexaminer reliability were found when palpation was used to elicit pain, which is consistent with the findings of the following 3 reviews. A review by Hollerwoger[1] that investigated manual cervical spine examination methods included several studies that assessed the reliability of palpation of tender or painful segmental levels. The author concluded that the detection of cervical segmental dysfunction based on manual assessment was uncertain, although studies that involved palpation for pain were generally more reliable than those that assessed segmental motion. Another review by Seffinger et al[2] assessed the reliability of back and neck palpation. They included a number of studies that dealt with palpation for the identification of landmarks and for pain provocation, as well as soft tissue tests. The authors concluded that pain provocation tests were the most reliable, as compared with the other palpation procedures, and that paraspinal soft tissue diagnostic tests lacked reliability. Stochkendahl et al[3] reviewed the literature relative to the reproducibility of studies that dealt with manual examination of the spine. They concluded that there was strong support for interexaminer reliability of osseous and soft tissue pain palpation as well as for intraexaminer reliability of soft tissue pain palpation and global assessment. There was little evidence pointing to the reliability of several other spinal evaluation procedures, including static and motion palpation.

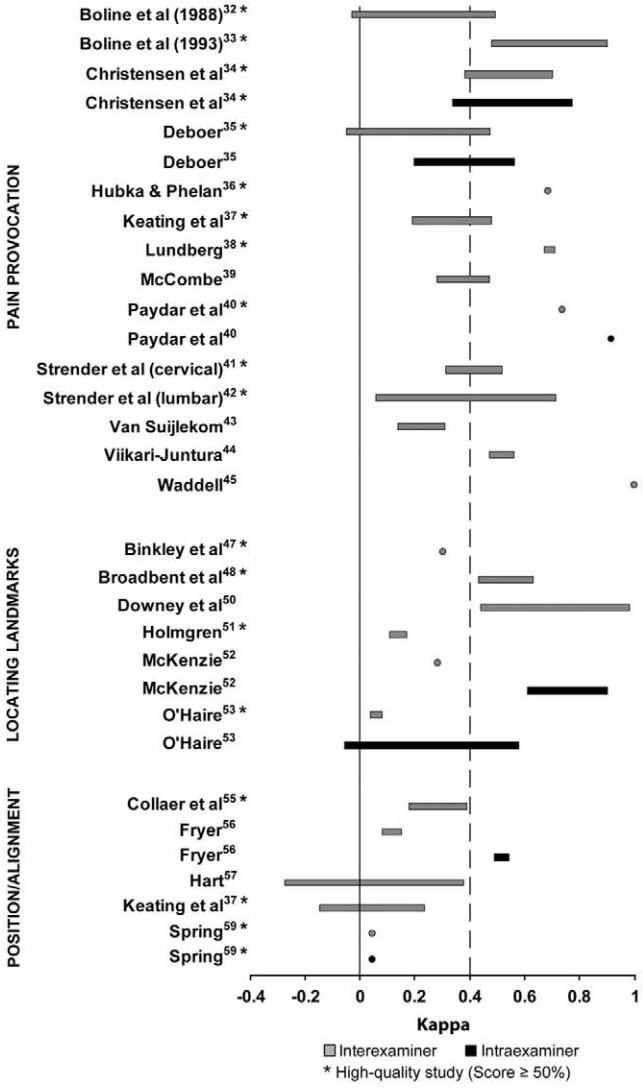


**Fig 2.** *Kappa ranges for studies that assessed the various forms of static spinal palpation.*

Although palpation of spinal tenderness has fared better than other procedures in inter- and intraexaminer reliability studies, there may be an inherent problem with this form of inquiry. Each patient is aware of the location of their own pain and/or tenderness and when examined repetitively are likely to direct subsequent examiners to the same area. Thus, these studies may not assess the examiners' reliability so much as they assess the patients' ability to recall the same site of pain. Issues of sensitization of an inflamed/tender area may further inflate the reproducibility of pain elicitation because, if prodded long enough, many soft tissues and bony prominences will become tender. With increased tenderness, agreement on finding a tender point would increase. On the other hand, when the examiner is not palpating for pain, a patient would not likely be aware of the locations of fixations or misalignments, leaving the decision process almost entirely up to the examiners.

Several studies have confirmed that the interexaminer reliability of naming the locations of spinal and SI landmarks is poor,[46,53] which may complicate palpation reliability studies. Hollerwoger[1] pointed out that the chance of errors in spinal palpation reliability studies is essentially doubled when examiners are required to name the specific level of involvement. This is because the particular level judged to be tender or misaligned might be reported as different due to misnaming the level, rather than true disagreement about the location of the problem. Thus, even when examiners are actually in agreement and call the same location dysfunctional, they might end up reporting different spinal levels due to level identification problems. Some researchers have compensated for this problem by having an independent person mark the bony landmarks before the examiners perform their palpations. This procedure limits the confounding variable of having the palpator determine the level because the levels are predetermined.

Christensen et al[34] noted that we should not be too harsh about criticizing the inability of examiners to agree on a specific level of spinal dysfunction because spinal manipulation is typically applied to a region. Moreover, even when the practitioner attempts to affect only 1 segment during manipulation, due to the relative broadness of hand contacts and smallness of the underlying segment, forces are distributed to the surrounding tissues resulting in the movement of adjacent segments.[60] Knowing the precise level of involvement among vertebral segments is therefore not as relevant in the clinical setting as one might think. Providing the general area of intended manipulation is correctly identified and forces are applied within that area, a positive clinical response would be expected.

We suggest that the form of static spinal palpation that showed the highest level of reliability, the location of painful or tender points, may not be so much of an assessment of the reliability of the examiners as it is an assessment of the ability of the examined subjects to consistently recall the same site of pain from palpation to palpation. Assuming suitable training on the palpation procedures, the examiners should be applying reasonably similar forces to similar tissues from subject to subject and then asking the subjects for a response. Thus, what is actually being tested is the consistency of the examiners' palpation procedures, as well as the consistency of the subjects' ability to recall the same site of pain.

### Limitations

We searched 4 biomedical databases using assorted search terms and also harvested several references from the included articles. Nevertheless, it is possible that some relevant citations may have been missed. The studies were generally not amenable to direct comparisons because of heterogeneity of their methods, statistical analyses, and regions investigated. Palpation reliability studies are

inherently prone to error because the examiners may name the incorrect level, even when they agree on their palpation findings. This is because of the difficulties associated with examiners simply attempting to agree on the identification of a given spinal level.

## Conclusion

The indices of agreement that were reported in the included studies were generally low. More of the pain-related palpation studies reported acceptable $\kappa$ levels than those that investigated landmark location or alignment. However, when we compared the proportions of high-quality studies that pointed to acceptable levels of reliability based on the type of palpation, although the graphs seem to show that pain palpation is superior, the differences were not statistically significant. Consequently, 1 method of static palpation cannot be deemed clearly superior to the others based on this review.

**Practical Applications**

- Static spinal palpation studies have focused on the assessment of the reliability of locating painful or tender points, the location of landmarks, as well as the position or alignment of bone structure.
- The reported reliability for each of the 3 types of static spinal palpation is generally low.
- The proportion of studies that reported good reliability favored palpation to locate painful or tender points over the other types, although the differences were not statistically significant.

## References

1. Hollerwoger D. Methodological quality and outcomes of studies addressing manual cervical spine examinations: a review. Man Ther 2006;11:93-8.
2. Seffinger MA, Najm WI, Mishra SI, et al. Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. Spine 2004;29:E413-25.
3. Stochkendahl MJ, Christensen HW, Hartvigsen J, et al. Manual examination of the spine: a systematic critical literature review of reproducibility. J Manipulative Physiol Ther 2006;29:475-85, 85 e1-10.
4. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. Am J Epidemiol 1987;126:161-9.
5. Rosner B. Fundamentals of biostatistics. 4th ed. Belmont (Calif): Duxbury Press; 1995. p. 518.
6. Haas M. Statistical methodology for reliability studies. J Manipulative Physiol Ther 1991;14:119-32.
7. Portney LG, Watkins MP. Foundations of clinical research: applications to practice, 2nd ed. Upper Saddle River (NJ): Prentice Hall; 2000. p. xiv, 768.
8. Chakraverty R, Pynsent P, Isaacs K. Which spinal levels are identified by palpation of the iliac crests and the posterior superior iliac spines? J Anat 2007;210:232-6.
9. Atula TS, Grenman R, Varpula MJ, et al. Palpation, ultrasound, and ultrasound-guided fine-needle aspiration cytology in the assessment of cervical lymph node status in head and neck cancer patients. Head Neck 1996;18:545-51.
10. Brantingham J, Hubka M, Snyder W, Diballa S, Wilke P, Biedebach D. Interexaminer reliability of palpation for foot and ankle joint tenderness: a pilot study. J Neuromusculoskelet Syst 1995;3:188-91.
11. Gemmell H, Paydar D, Thiel H. Intra- and interexaminer reliability of certain pelvic palpatory procedures and the sitting flexion test for sacroiliac joint mobility. J Neuromusculoskelet Syst 1994;2:65-9.
12. Jende A, Peterson CK. Validity of static palpation as an indicator of atlas transverse process asymmetry. Eur J Chiropr 1997;45:35-42.
13. Hubka MJ. Palpation for spinal tenderness: a reliable and accurate method for identifying the target of spinal manipulation. J Chiropr Tech 1994;6:5-8.
14. Cooperstein R, Lisi A. Pelvic torsion: anatomic considerations, construct validity, and chiropractic examination procedures. Top Clin Chiropr 2000;7:38-49.
15. Hart J. Comparison of x-ray listings and palpation listings of the upper cervical spine. J Vertebral Subluxation Res 2000;4.
16. Jull G, Bogduk N, Marsland A. The accuracy of manual diagnosis for cervical zygapophysial joint pain syndromes. Med J Aust 1988;148:233-6.
17. Downey B, Taylor N, Niere K. Can manipulative physiotherapists agree on which lumbar level to treat based on palpation? Physiotherapy 2003;89:74-81.
18. Phillips DR, Twomey LT. A comparison of manual diagnosis with a diagnosis established by a uni-level lumbar spinal block procedure. Man Ther 1996;1:82-7.
19. Sebastian D, Chovvath R. Reliability of palpation assessment in non-neutral dysfunctions of the lumbar spine. Orthop Phys Ther Pract 2004;16:23-6.
20. Tong HC, Heyman OG, Lado DA, et al. Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position. J Am Osteopath Assoc 2006;106:464-8.
21. French SD, Green S, Forbes A. Reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low-back pain. J Manipulative Physiol Ther 2000;23:231-8.
22. Hawk C, Phongphua C, Bleecker J, et al. Preliminary study of the reliability of assessment procedures for indications for chiropractic adjustments of the lumbar spine. J Manipulative Physiol Ther 1999;22:382-9.
23. McPartland JM, Goodridge JP. Counterstrain and traditional osteopathic examination of the cervical spine compared. J Bodywork Move Ther 1997;1:173-8.
24. Tuchin P, Hart C, Johnson C, et al. Interexaminer reliability of chiropractic evaluation for cervical spine problems—a pilot study. Australasian Chiropr Osteopat 1996;5:23-9.
25. Haig AJ, Moffroid M, Henry S, et al. A technique for needle localization in paraspinal muscles with cadaveric confirmation. Muscle Nerve 1991;14:521-6.
26. Braun B, Schiffman EL. The validity and predictive value of four assessment instruments for evaluation of the cervical and stomatognathic systems. J Craniomandib Disord 1991;5:239-44.
27. Drerup B, Hierholzer E. Objective determination of anatomical landmarks on the body surface: measurement of the vertebra prominens from surface curvature. J Biomech 1985; 18:467-74.

28. Geelhoed MA, McGaugh J, Brewer PA, et al. A new model to facilitate palpation of the level of the transverse processes of the thoracic spine. J Orthop Sports Phys Ther 2006;36:876-81.
29. Karim A, Mukherjee D, Gonzalez-Cruz J, et al. Accuracy of pedicle screw placement for lumbar fusion using anatomic landmarks versus open laminectomy: a comparison of two surgical techniques in cadaveric specimens. Neurosurgery 2006;59:ONS13-9.
30. Burton AK, Edwards VA, Sykes DA. Invisible skin marking for testing palpatory reliability. J Man Med 1990;5:27-9.
31. Kim HW, Ko YJ, Rhee WI, et al. Interexaminer reliability and accuracy of posterior superior iliac spine and iliac crest palpation for spinal level estimations. J Manipulative Physiol Ther 2007;30:386-9.
32. Boline P, Keating J, Brist J, et al. Interexaminer reliability of palpatory evaluations of the lumbar spine. Am J Chiropr Med 1988;1:5-11.
33. Boline PD, Haas M, Meyer JJ, et al. Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality: part II. J Manipulative Physiol Ther 1993;16:363-74.
34. Christensen HW, Vach W, Vach K, et al. Palpation of the upper thoracic spine: an observer reliability study. J Manipulative Physiol Ther 2002;25:285-92.
35. Deboer KF, Harmon Jr R, Tuttle CD, et al. Reliability study of detection of somatic dysfunctions in the cervical spine. J Manipulative Physiol Ther 1985;8:9-16.
36. Hubka MJ, Phelan SP. Interexaminer reliability of palpation for cervical spine tenderness. J Manipulative Physiol Ther 1994; 17:591-5.
37. Keating J, Bergmann T, Jacobs G, et al. Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. J Manipulative Physiol Ther 1990;13:463-70.
38. Lundberg G, Gerdle 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.
39. McCombe PF, Fairbank JC, Cockersole BC, et al. Volvo Award in clinical sciences. Reproducibility of physical signs in low-back pain. Spine 1989;14:908-18.
40. Paydar D, Thiel H, Gemmell H. Intra- and interexaminer reliability of certain pelvic palpatory procedures and the sitting flexion test for sacroiliac joint mobility and dysfunction. J Neuromusculoskelet Syst 1994;2:65-9.
41. Strender LE, Lundin M, Nell K. Interexaminer reliability in physical examination of the neck. J Manipulative Physiol Ther 1997;20:516-20.
42. Strender LE, Sjoblom A, Sundell K, et al. Interexaminer reliability in physical examination of patients with low back pain. Spine 1997;22:814-20.
43. Van Suijlekom HA, De Vet HC, Van Den Berg SG, et al. Interobserver reliability in physical examination of the cervical spine in patients with headache. Headache 2000;40:581-6.
44. Viikari-Juntura E. Interexaminer reliability of observations in physical examinations of the neck. Phys Ther 1987;67: 1526-32.
45. Waddell G, Main CJ, Morris EW, et al. Normality and reliability in the clinical assessment of backache. Br Med J (Clin Res Ed) 1982;284:1519-23.
46. Billis EV, Foster NE, Wright CC. Reproducibility and repeatability: errors of three groups of physiotherapists in locating spinal levels by palpation. Man Ther 2003;8:223-32.
47. Binkley J, Stratford PW, Gill C. Interrater reliability of lumbar accessory motion mobility testing. Phys Ther 1995;75:786-92.
48. Broadbent CR, Maxwell WB, Ferrie R, et al. Ability of anaesthetists to identify a marked lumbar interspace. Anaesthesia 2000;55:1122-6.
49. Byfield D, Humphreys K. Intra- and inter-examiner reliability of bony landmark identification in the lumbar spine. Eur J Chiropr 1992;40:13-7.
50. Downey BJ, Taylor NF, Niere KR. Manipulative physiotherapists can reliably palpate nominated lumbar spinal levels. Man Ther 1999;4:151-6.
51. Holmgren U, Waling K. Inter-examiner reliability of four static palpation tests used for assessing pelvic dysfunction. Man Ther 2007.
52. McKenzie A, Taylor N. Can physiotherapists locate lumbar spinal levels by palpation? Physiotherapy 1997;83:235-9.
53. O'Haire C, Gibbons P. Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study. Man Ther 2000;5:13-20.
54. Simmonds M, Kumar S. Health care ergonomics. Part II: location of bony structures by palpation—a reliability study. 11 (2):145-151. Int J Ind Ergon 1993;11:145-51.
55. Collaer JW, McKeough DM, Boissonnault WG. Lumbar isthmic spondylolisthesis detection with palpation: interrater reliability and concurrent criterion-related validity. J Man Manipulative Ther 2006;14:22-9.
56. Fryer GM, McPherson H, O'Keefe P. The effect of training on the interexaminer and intra-examiner reliability of the seated flexion test and assessment of pelvic anatomical landmarks with palpation. Int J Osteopat Med 2005;8:131-8.
57. Hart J. Palpation and X-ray of the upper cervical spine: a reliability study. J Vertebral Subluxation Res 2006.
58. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. Phys Ther 1985;65: 1671-5.
59. Spring F, Gibbons P, Tehan P. Intra-examiner and inter-examiner reliability of a positional diagnostic screen for the lumbar spine. J Osteopat Med 2001;4:47-55.
60. Herzog W, Kats M, Symons B. The effective forces transmitted by high-speed, low-amplitude thoracic manipulation. Spine 2001;26:2105-10.

## Appendix A. Six-point scale used to assess the quality of studies. Questions 2, 4, 5, and 6 apply to intraexaminer studies

1. *Order of observers conducting the test(s) randomized (1 point)*
2. *Case mix (1 point)* (Both symptomatic and asymptomatic subjects. For studies with an intentional case mix, the case mix must represent a natural clinic population, that is, a population consisting of a variety of patients with regards to sex, age, and problems in different part of the spine. If the purpose is to look at subgroups, such as asymptomatic patients only, credit will be given.)
3. *Observers blind to other observers findings (1 point)*
4. *Observers blind to confounding information (1 point)* (such as patient history)
5. *Subjects blind to observer findings* (*1 point* for true/complete blinding, *0.5 point* for no blinding but subjects unable to affect results)
6. *Kappa* ($\kappa$, $\kappa_o$, $\kappa_g$, $\kappa_w$) *or ICC* (l-way ICC, 2-way ICC, or generalizability coefficients for single rating by each observer) used for analysis *(1 point)*