Interexaminer Reliability of Seated Motion Palpation for the Stiffest Spinal Site

Kelly Holt, PhD,[a] David Russell, DC,[b] Robert Cooperstein, MA, DC,[c] Morgan Young, DC,[d] Matthew Sherson, DC,[a] and Heidi Haavik, DC, PhD[a]

Abstract

Objectives: The purpose of this study was to assess the interexaminer reliability of palpation for stiffness in the cervical, thoracic, and lumbar spinal regions.

Methods: In this secondary data analysis, data from 70 patients from a chiropractic college outpatient clinic were analyzed. Two doctors of chiropractic palpated for the stiffest site within each spinal region. Each were asked to select the stiffest segment and to rate their confidence in their palpation findings. Reliability between examiners was calculated as Median Absolute Examiner Differences (MedianAED) and data dispersion as Median Absolute Deviation (MAD). Interquartile analysis of the paired examiner differences was performed.

Results: In total, 210 paired observations were analyzed. Nonparametric data precluded reliability determination using intraclass correlation. Findings included lumbar MedianAED = 0.5 vertebral equivalents (VE), thoracic = 1.7 VE, and cervical = 1.4 VE. For the combined dataset, the findings were MedianAED = 1.1 VE; MAD was lowest in the lumbar spine (0.3 VE) and highest in thoracic spine (1.4 VE), and for the combined dataset, MAD = 1.1 VE. Examiners agreed on the segment or the motion segment containing the stiffest site in 54% of the observations.

Conclusions: Interexaminer reliability for palpation was good between 2 clinicians for the stiffest site in each region of the spine and in the combined dataset. This is consistent with previous studies of motion palpation using continuous analysis. (J Manipulative Physiol Ther 2018;41:571-579)

Key Indexing Terms: *Spine; Reproducibility of Results; Palpation; Observer Variation*

Introduction

A commonly used chiropractic method for assessment of spine function is motion palpation (MP).[1,2] This method has been found to be unreliable, with levels of agreement often being no better than those associated with chance.[3] This has led to criticism of chiropractic programs for teaching MP and for its use as an assessment in clinical practice.[4,5] It has been argued[6] that previous study designs may have had design flaws that could account for the poor agreement results. Among the design flaws that may have lowered concordance were forcing examiners to classify each segment as either hypomobile or not, regardless if the participant lacked a stiff spinal site, or forcing a response when examiners were unsure whether a specific level was hypomobile or not.[6]

A continuous measures analytic system, combined with stratification by examiner confidence, has demonstrated good interexaminer reliability in detecting areas of maximum spinal stiffness.[6-9] Level-by-level analysis of agreement at each spinal level, assessed using the κ statistic, addresses a different question and would not be expected to detect examiner agreement on locations of maximum spinal stiffness. Although this continuous measures approach seemed potentially valuable, the present authors thought it best that this prior work be replicated using similar methodology to substantiate it. The goal was to evaluate if seated motion assessment would obtain results comparable to those obtained using the prone, supine, and side-posture methods used in these prior studies.[6-9] Thus, the primary objective of this study was to assess interexaminer reliability of MP assessment for stiffness in the cervical, thoracic, and lumbar spinal regions.

[a] Center for Chiropractic Research, New Zealand College of Chiropractic, Aukland, New Zealand.
[b] Private Practice, Auckland, New Zealand.
[c] Palmer Center for Chiropractic Research, Palmer College of Chiropractic West, San Jose, California.
[d] Washington State Department of Labor and Industries, Olympia, Washington.
Corresponding author: Robert Cooperstein, MA, DC, Palmer College of Chiropractic San Jose, 90 East Tasman Drive, San Jose, CA 95134. Tel.: +1 408 944 6009.
(e-mail: *Cooperstein_r@palmer.edu*).
Paper submitted March 14, 2017; in revised form July 22, 2017; accepted August 23, 2017.
0161-4754

https://doi.org/10.1016/j.jmpt.2017.08.009

Methods

Design and Setting

This study was a secondary analysis of data gathered from a previous study.[10] Data for this secondary analysis included 70 patients from the sample. The trial was conducted at the Chiropractic Centre, a student training facility, of the New Zealand College of Chiropractic

(NZCC). The Health and Disability Ethics Committees, Ministry of Health in New Zealand (ethics.health.govt.nz), reviewed and exempted this secondary analysis. Data collection took place across 21 study sessions between October 2014 and March 2015.

Description of Previous Study

Participants were recruited from patients presenting to the Chiropractic Centre at NZCC. Inclusion criteria were all public patients attending the Chiropractic Centre during data collection sessions who were over the age of 18 and had verbally consented to participate in the study. Data collection took place during regular hours in the Chiropractic Centre when both examiners and research assistants (RAs) were available. All public patients attending the Chiropractic Centre during these shift times were invited to participate. When a patient over the age of 18 presented to the Chiropractic Centre during a data collection session, an RA explained the study and asked if he or she consented to participate. If the patient agreed to participate, he or she was escorted to a treatment room by the RA. Age and sex were recorded along with a recent history and whether the patient was currently experiencing any bodily pain.

The participants removed their shirts or changed into a gown to expose their backs. Marks were placed on the skin over their spine using a marking pen, over the inferior tip of the C7 and T12 spinous processes by the RA. The participant was instructed to straddle the foot of the chiropractic table in the seated position. The first examiner entered the room and performed MP of the lumbar, thoracic, and then cervical regions of the spine to determine the stiffest spinal site (SSS) in each spinal region.

The operational definition of the SSS for this study was the spinal location of maximal resistance to palpatory pressure, as perceived by the clinician, which is a force-displacement relationship.[11] Abbott et al[11] asked examiners to rate which was most important among 7 key parameters when assessing spinal segmental motion. "Quality of resistance" at 28.8% was second only to "pain response" at 31.8%. The participants in our study were not asked about possible pain on palpation.

While the participant was in a seated neutral position, the RA measured the distance of the examiner's designated SSS to the mark placed at C7 or T12, depending upon which region of the spine was being assessed. Table 1 describes the methods used[12] (Fig 1). If an examiner found more than 1 SSS in a given region, the more caudal finding was recorded. The examiner recorded his level of confidence in his palpation findings. The second examiner entered the room within 5 minutes of the first examiner and completed the palpation and repeated the assessment as described previously. The examiners were blinded to each other's selections, were not provided with any clinical information about the patients, alternated their order in assessing them, and did not converse with the participants. Both examiners were chiropractors with over 10 years of clinical experience, teachers within the technique program at the NZCC, and supervising clinicians in the Chiropractic Centre. Frequent consensus training sessions were held over a 3-month period prior to data collection to ensure the spinal assessments were performed consistently. The participants were assessed during the trial, with between 1 and 6 participants being assessed during each data collection session.

Table 1. *Motion Palpation Assessment Used in This Study*

Order	Spinal Region	Procedure and Definition of Positive Result
First	Lumbar	The patient was seated at the end of the table with arms crossed in front of them. The examiner used the left hand to hold the patient's crossed arms to direct motion and the right hand to assess for sagittal plane flexion and extension, as follows: the examiner contacted the spinous being and assessed with the index finger and the spinous of the inferior segment with the middle finger of the right hand. The examiner then moved the patient through flexion and extension, feeling for P-A motion of the spinous in relation to the spinous below. Each spinal level was individually assessed from L5 superiorly to L1 and was motioned only once by each examiner. A positive finding was recorded if there was restriction in either flexion or extension.
Second	Thoracic	The palpatory procedure and definition of a positive result was identical to that used in the lumbar spine, differing only in the spinal ranged assessed; each spinal level was individually assessed from T12 superiorly to T1 and was motioned only once by each examiner.
Third	Cervical	The patient was seated at the end of the table with arms uncrossed. The examiner used the left hand to hold the patient's head and guide its motion, using the right hand, into rotation and extension, as follows: the articular pillars of the level assessed were contacted bilaterally with the index finger and thumb of the right hand. The patient was then moved through left and right rotation, feeling for motion at the same spinal level. Each spinal level was individually assessed from C7 superiorly to C2 and was assessed only once by each examiner. Because the posterior tubercle is not considered palpable in this manner, to assess motion at C1, the examiner contacted the posterior aspect of the C1 TVP with the right hand while moving the patient into coupled extension and right or left rotation, motioned only in each direction by each examiner.

P-A, posteroanterior; *TVP*, transverse process.

Statistical Analysis

Data were analyzed using continuous measures analysis. The paired findings for the distance of the SSS from the C7 or T12 benchmark site were assessed for normality to

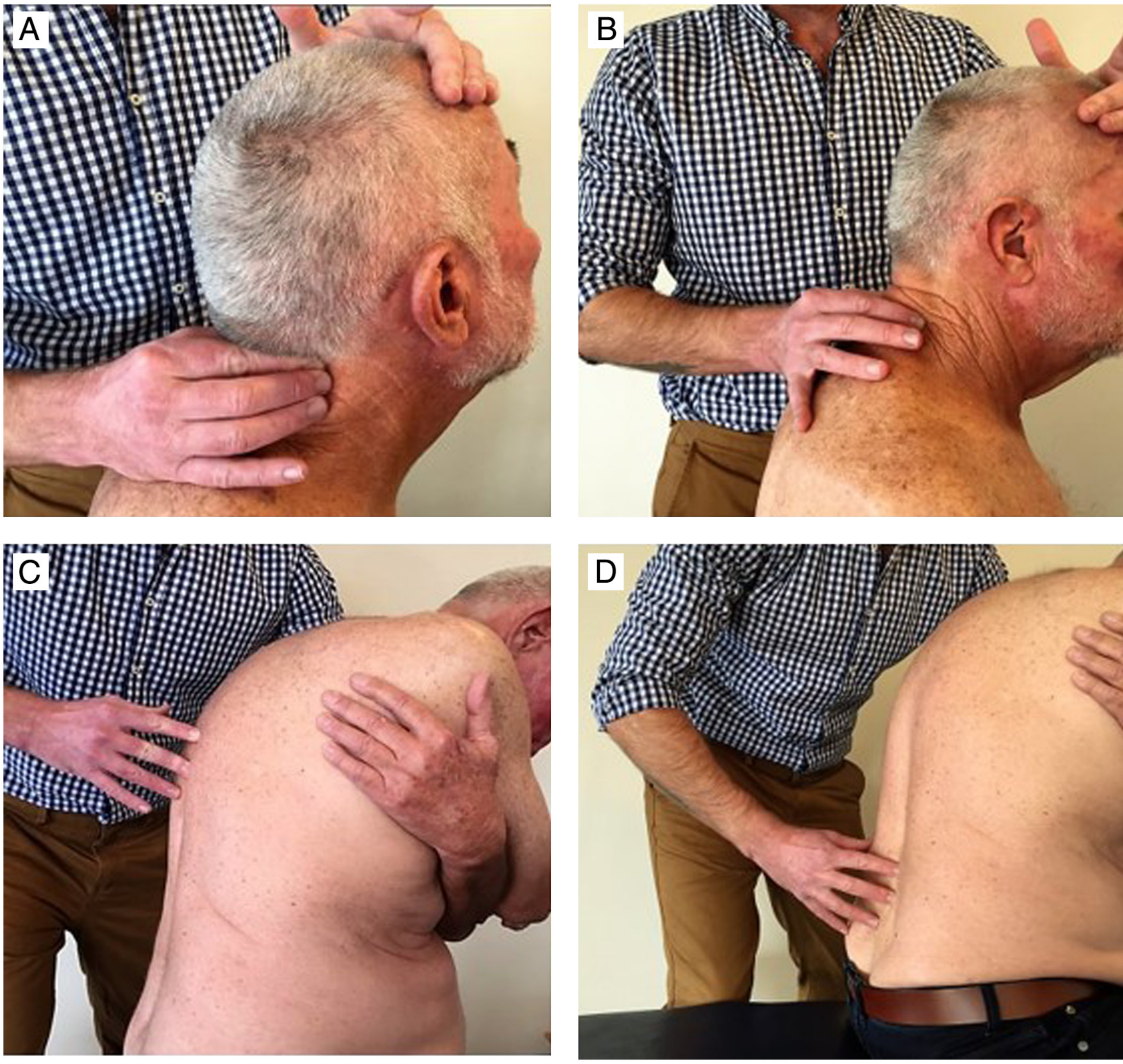

Fig 1. *Palpation procedures: upper cervical (A), lower cervical (B), thoracic (C), lumbar (D).*

determine the appropriate statistical function(s) to be used in calculating interexaminer reliability. Based on this, it was deemed most appropriate to determine interexaminer reliability by calculating the Median Absolute Examiner Differences (MedianAED) for each spinal region. Data dispersion was determined by calculating values for the Median Absolute Deviation (MAD).[13] In addition to being provided in centimeter units, MedianAED and MAD were transformed into, and presented as, vertebral equivalents (VE), where VE is based on the percentage of the height of a typical vertebra. For example, 1 VE = 100% of a vertebral height (Fig 2). This permitted combining the data for each spinal region to perform analysis on the combined dataset. To convert centimeters to VE, the following heuristic weighting factors were used: 2.3 cm for a typical thoracic segment,[14] 1.8 cm for a typical cervical segment,[15] and 4 cm for a typical lumbar segment.[16] A box and whisker plot was constructed to allow interquartile analysis of the paired examiner differences and to detect outliers (Fig 3).

Results

Table 2 summarizes the participant characteristics. Figure 4 shows the distributions of the examiners' determination of the stiffest segments for each spinal region: L3, T6, and C7.

Shapiro-Wilk testing for normality established that the paired data examiner differences did not form normal distributions, neither among the 3 subsets based on spinal region (the cervical, thoracic, and lumbar spines), nor for the combined dataset. As an alternative to intraclass correlation coefficient (ICC), the investigators determined interexaminer agreement as the median of the absolute

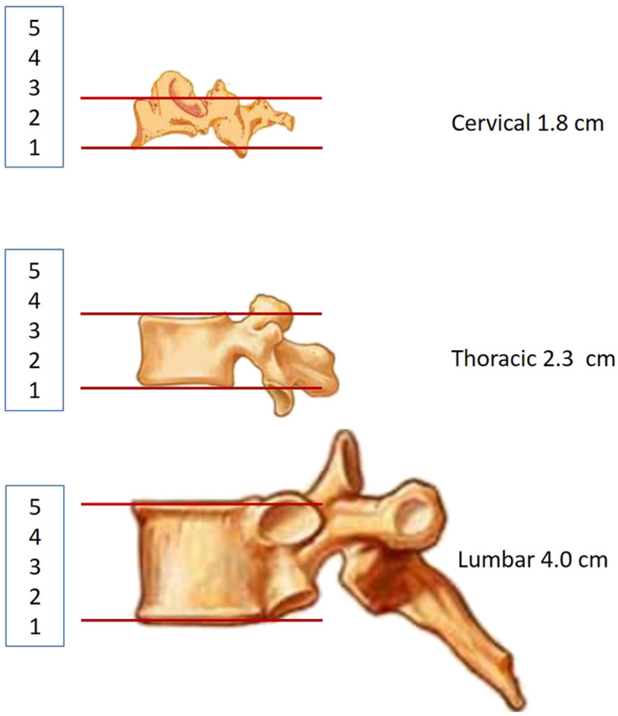

Fig 2. *Computation of vertebral equivalent.*

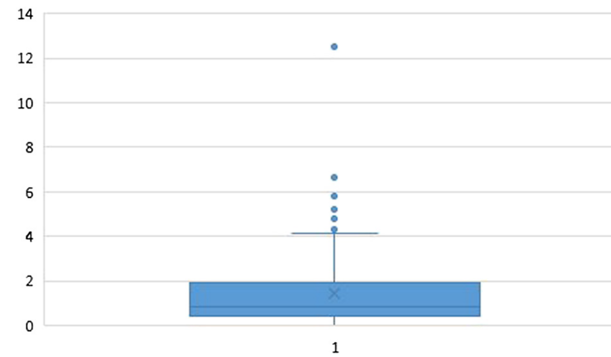

Fig 3. *Box and whisker plot, paired examiner differences.*

Table 2. *Patient Characteristics*

Characteristic	Value
Age [a]	46 (18)
Age range	19-87
Female [b]	40 (57)

[a] Mean (standard deviation).
[b] n (%).

values of examiner differences: MedianAED.[13] The dispersion of the examiner differences was found by subtracting each examiner difference from the median of such differences, converting this derived series of numbers to absolute values, and then finding their median. This procedure is given by the following equation: for each value x_i, MAD = median ($|x_i - \text{median}[x_i]|$). Although data were collected for the examiners' confidence levels, there were too few examiner ratings (14%) where 1 or both examiners lacked confidence to perform meaningful analysis.

Median Absolute Examiner Differences values and MAD calculations (in units of both centimeters and VE) are presented in Table 3. Given the different dimensions of typical vertebrae in the 3 spinal regions, VE units allow meaningful interregional comparisons of reliability and data dispersion and analysis of the combined dataset. The Median Absolute Examiner Difference in the lumbar region was 0.5 VE, in the thoracic region was 1.7 VE, and in the cervical region was 1.4 VE. For the combined dataset, MedianAED = 1.1 VE. Median Absolute Deviation, representing the dispersion of examiner differences, was lowest in the lumbar spine (0.3 VE) and highest in the thoracic spine (1.4 VE). For the combined dataset, MAD = 1.0 VE. Using examiner differences ≤1.25 VE as the cutoff point for defining agreement on the motion segment that included the SSS, examiners agreed 54.3% of the time for the combined dataset. In the box and whisker analysis, it could be seen that 75% of examiner differences among 210 observations were ≤2.0 VE, signifying agreement on a motion segment or a segment subjacent to it.

Discussion

This study found seated manual MP for SSS to have good reliability between 2 doctors of chiropractic. However, to be clinically useful, this procedure must also be valid when compared with a reference standard. The reference standard could be objective measurements of spinal stiffness, but that should ultimately be buttressed by clinical studies showing that treatment based on the identification of spinal stiffness results in enhanced clinical outcomes.

Few studies have reported on the validity for manual palpation of spinal stiffness. A systematic review[17] reported equivocal results, with variable results and generally low sensitivity. It included 5 studies; in 3 of the studies, the palpators assessed spinal stiffness in mannequins featuring variable segmental stiffness, whereas the 2 other studies used pain as a reference standard. The relevance of using mannequins to simulate in vivo spinal stiffness is unclear. In another validity study not included in the review, palpators examining the sacroiliac joint were unable to identify known cases of ankylosing spondylitis as fixated.[18] Another study with a similar design involving the cervical spine[19] demonstrated that palpators were able to detect fixation at the site of congenital block vertebrae. A comprehensive review by Snodgrass et al[20] discussed not only the measurement (manual and instrumented) of spinal stiffness,

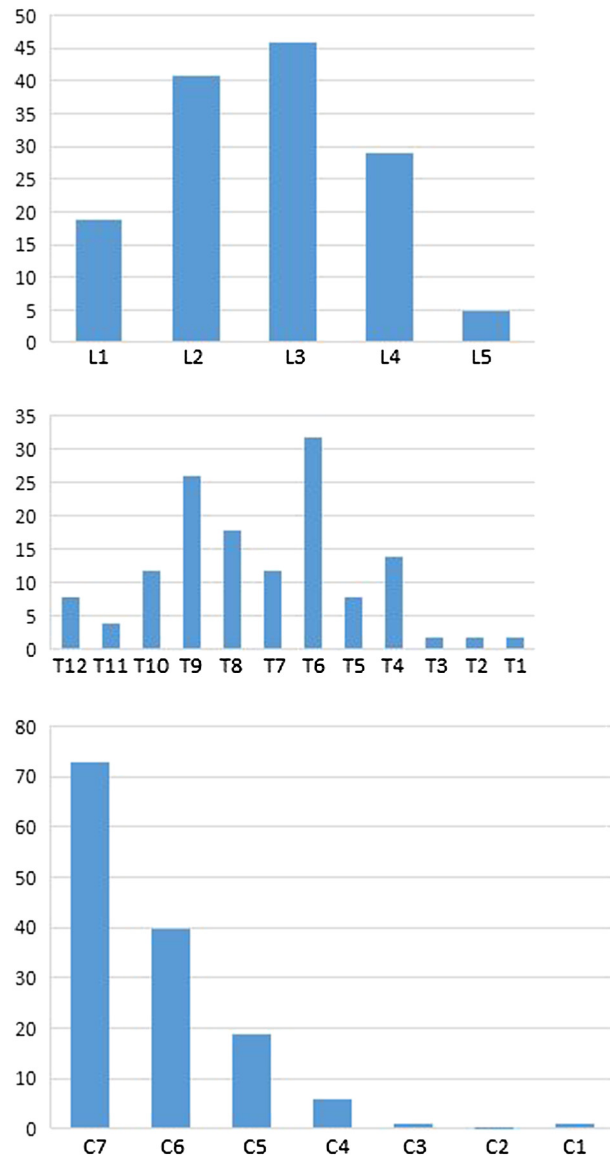

Fig 4. *Distributions of stiffest spinal segments.*

but also its utility in diagnosis, prognosis, and treatment decision-making. Both the reliability and validity of manual assessment of spinal stiffness were found equivocal at best.

Among the studies included by Snodgrass et al,[20] only that of Campbell and Snodgrass[21] used a stiffest spinal site method similar to the one in the present study, rather than more typical assessment of segmental stiffness. In their study, an experienced physiotherapist manually identified the thoracic segment perceived to be the stiffest by applying posteroanterior pressure on the spinous processes of the thoracic spine. An instrument was used both before and after spinal manipulation to measure stiffness at not only the segment perceived to be stiffest, but also the 2 levels caudal and cephalad to it. Stiffness at spinal locations judged to be very stiff tended to decrease with spinal manipulation.

Table 3. *Interexaminer Reliability of MP for SSS*

		Delta SSS, cm		Delta SSS, VE	
N	Spinal Region	MedianAED	MAD	MedianAED	MAD
70	Cervical	2.5	1.5	1.4	0.8
70	Thoracic	4.0	2.5	1.7	1.1
70	Lumbar	2.0	1.3	0.5	0.3
210	Combined	—	—	1.1	1.0

MAD, median absolute deviation; *MedianAED*, median absolute examiner differences; *MP*, motion palpation; *SSS*, stiffest spinal site; *VE*, vertebral equivalent.

The poor reliability of manually assessing individual spinal segments for stiffness[22-24] does not preclude reliability in determining the stiffest site within a defined spinal region. A continuous measures analytic system, combined with stratification by examiner confidence, has demonstrated good interexaminer reliability in detecting areas of maximum spinal stiffness.[6-9] When a clinical variable can be measured using either discrete or continuous analysis, there are good reasons to expect finding greater reliability when using continuous data. Two meta-analyses used continuous and categorical measures to assess psychopathology.[25] These studies found a 15% increase in reliability and 37% increase in validity using continuous measures, allowing a 50% reduction in sample size for any given power analysis of participant requirements. Another study[26] showed low reliability in assessing patients may result from the use of discrete diagnostic criteria that fail to recognize continuous variation in patients' presentations.

One explanation for why segmental palpation is less likely to detect agreement than using a stiffest site paradigm may be that when findings are measured on a continuum, information is lost. It is proposed that if stiffness were distributed over a spinal region spanning more than 1 spinal motion segment, rating adjacent motion segments as moving or not moving might fail to identify larger fields of spinal stiffness and, thus, miss the overlap of those fields among examiners' perceptions. Some investigators have attempted to address this issue by liberalizing their definition of agreement. For example, Christensen et al[14] considered motion palpators to agree on a site of hypomobility when their findings were ±1 spinal segment, as did Harlick et al[15] in a study of the accuracy of static spinal palpation. Although these efforts to transcend the requirement that examiners agree on the same segment seem reasonable, the SSS paradigm goes further by liberalizing the definition of agreement by extending it to fields of spinal stiffness that are centered around the SSS. Addressing the SSS would affect the contacted vertebra and the 2 motion segments including it, each of which, in turn, will influence the subjacent segment. This "field of stiffness" analysis thus leads directly to a "field of manipulative impact" analysis, possibly explaining clinical outcomes for spinal manipulation despite frequently low levels of examiner agreement on exact sites for spine care.[16]

Assessing the stiffest spinal sites in a defined region may more closely match the practice of chiropractic clinicians than formulating judgements for each spinal vertebra as "moving or not moving," which is the examination method that has been used in typical research settings. It is hypothesized that level-by-level analysis may find multiple levels within a spinal region, although a clinician would probably not adjust each hypomobile segment. Since spinal manipulation often results in multiple cavitations,[27,28] it may be less specific than usually intended. However, since manipulation necessarily impacts at least 2 motion segments, treating what seems to be the most hypomobile segment within a defined region may impact adjacent spinal levels. This study's finding that 75% of examiner differences among 210 observations were ≤2 VE apart implies this SSS protocol would usually "capture" the site of the patient's complaint.

Interexaminer reliability in this study was determined by calculating MedianAED and MAD; the nonparametric distribution of examiner differences precluded calculating ICC in the usual manner. It also precluded using Bland-Altman limits of agreement[29,30] for the same reason. Median Absolute Examiner Difference provides a useful measure of the typical examiner difference in localizing the SSS and is especially useful when examiner differences are not normally distributed.[13,31] Median Absolute Deviation is a robust measure of dispersion (similar to a standard deviation) that is resilient to outliers and is suitable for datasets that are not normally distributed.[13] It is mathematically defined as the median of the absolute deviations of all paired observations from the median absolute examiner difference.[13,32] It is resistant to extreme outliers at the minimum and maximum ends of the range of examiner differences because relatively large or small differences have no greater impact on the median value than smaller ones. Median Absolute Deviation quantifies the variability of the data, painting a more detailed picture than more typically reported range values, the maximum and minimum values. The magnitude of a range is strongly affected by extreme values. Because the height of a typical vertebra varies according to the spinal region, examiner differences reported in centimeters misleadingly imply different degrees of examiner reliability, depending on the spinal region. Reporting the data as VEs allowed immediate comparisons of examiner reliability, irrespective of spinal region, and is more clinically intuitive to understand than centimeters.

To assist in interpreting the findings for MedianAED (Fig 5), consider a case in which the first examiner judged the SSS to be at the middle of a given segment (right). If the second examiner identifies an SSS that is ≤1.5 VE segments distant, it must be concluded the examiners agreed that 1 of the 2 motion segments including the first examiner's SSS was hypomobile. This does not imply their findings were in a range spanning 2 motion segments, but only that agreement may have occurred on the motion segment above or below. In this study, this occurred 85.6% of the time in the lumbar spine, 47.1% of the time in the thoracic spine, and 61.4% of the time in the cervical spine.

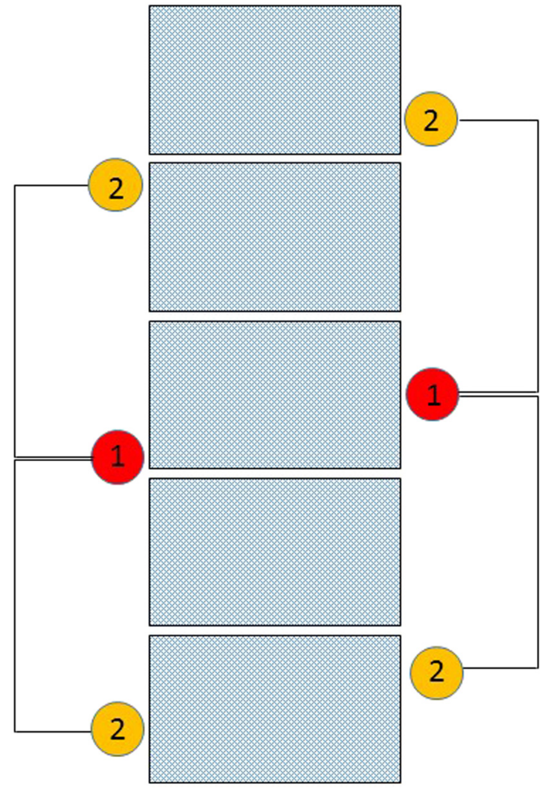

Fig 5. *Interpretation of Median Absolute Examiner Differences. Examiner 1's stiffest spinal site appears in the darker shade circles, and examiner 2's stiffest spinal site appears in the lighter shade circles.*

If 1 of the examiners had identified the SSS at either the very top or the very bottom of a segment (left), the MedianAED consistent with agreeing on the hypomobile motion segment, including the SSS, narrows to ≤1.0 VE. This is because a MedianAED > 1.0 VE would mean the second examiner might have found the SSS to lie more than 1 vertebra distant (Fig 5, lower left), thus within the subjacent motion segment. Examiner disagreement ≤1.0 VE occurred 75.8% of the time in the lumbar spine, 40.1% of the time in the thoracic spine, and 41.4% of the time in the cervical spine.

For the combined dataset, agreement on the SSS or the motion segment at the ≤1.5 VE cutoff point (ie, first examiner found the SSS at a vertebral center) was 56.2% and at the ≤1.0 VE cutoff point was 46.7% (ie, the first examiner found the SSS at the top or bottom of a vertebra). Under the reasonable assumption that an examiner was equally likely to locate the SSS nearest the middle or nearest the top or bottom of a vertebra, it may be most reasonable to use ≤1.25 VE as the cutoff point for stating the examiners agreed on the SSS or the motion segment including it. This occurred 54.3% of the time for the combined dataset. These figures show less agreement compared with data reported[9] in prior studies.[6-8] In the prior analysis, examiner differences were reported to be ≤1.5 VE 77.0% of the time in the combined dataset, exceeding the 56.2% seen in the present study. In their secondary analysis, unstratified MedianAED = 0.7 VE, compared with the 1.1 VE in the present study.

A box and whisker plot summarizes the results for the combined dataset (Fig 3). This plot divides the data into quartile groups: the low whisker represents the smallest 25% of examiner differences, the blue box represents the middle half of the examiner differences, and the upper whisker represents the largest 25% of examiner differences. The 13 dots above the line across the upper whisker represent data points defined as outliers because they exceed the third quartile of the data by more than 1.5 times the interquartile range or height of the box. The value for VE at the top of the box is 2.0 VE, including 75% of examiner differences among 210 observations.

There were several differences between the present study and the previous, otherwise similar, studies,[6-8] possibly accounting for the modest difference in outcomes. In the present study, MP was performed in the seated position, assessing the excursion of primarily forward flexion and extension movements; whereas the prior studies assessed prone end-feel (participants prone in the thoracic spine assessments, supine in the cervical spine, and in side-posture in the lumbar spine). Excursion MP assesses the motion of a vertebra in relation to an adjacent vertebra, whereas end-feel palpation assesses the stiffness to palpation of a single vertebra.[33] Unlike the present study, in the prior studies,[9] interexaminer reliability varied with examiner confidence; when both examiners were confident in their findings (53.4%), the median examiner difference decreased to 0.6 VE, increased to 1.0 VE when 1 lacked confidence, and increased to 1.8 VE when both lacked confidence. In the previous studies, paired examiner differences were normally distributed,[9] allowing the use of both ICC and Bland-Altman limits of agreement to assess interexaminer reliability in addition to the MedianAED calculations used in both the previous and present studies.

Limitations

Limitations of this study include the use of a convenience sample of participants based on logistics in the Chiropractic Centre and examiner availability, which likely resulted in sampling bias. Only 2 examiners, who may have had unique experience and skills, participated in the study. The examiners were full-time chiropractic educators; thus, it could be argued that they were not representative of chiropractors in the field. Therefore, these findings may not necessarily be extrapolated to other chiropractors.

Although examiners were blinded to any other prior findings during the study, it is possible that they were familiar with some patients' prior findings or clinical or nonclinical cues based on previous visits to the Chiropractic Centre that they may have supervised. Although there was no power analysis to determine sample size, the investigators thought 70 participants, each of whom were examined in each spinal region for a total of 210 observations, was ample. Although ICC could not be used because of the nonparametric nature of the data, the recommended number of participants for either a complete dataset or a subset in this kind of study is about 35 participants, to have 80% power at the 5% significance level to detect ICC ≥ 0.6.[34] It was likely reasonable to extrapolate this judgement on sample size to the MedianAED calculations used in the present study. Although MAD is robust to extreme values because a larger extreme value has no greater impact than a smaller extreme value, this is also an important weakness. So-called extreme outliers at the lower and upper quartiles of examiner differences may represent an important characteristic of the examination method under investigation.

The methodological choice to select the more caudal site for the SSS, when an examiner found at least 2 sites indistinguishably stiff, may have biased the results to some degree. Without there having been a reference standard, it cannot be confirmed there actually were areas of clinically significant spinal stiffness among the participants studied. Although clinician disagreement on the stiffest or otherwise defined optimal site of spine care may lead to suboptimal results or even harm patients, the authors are not aware of studies confirming or excluding that possibility. Snodgrass et al did conclude[20] that there is limited evidence that manually assessed spinal stiffness may be associated with radiographic findings of rotational hypomobility or hypermobility and that short-term spinal stiffness decreases immediately after high-velocity, low amplitude manipulation in symptomatic persons. In the end, the limited amount of information available on kinematic and clinical changes after manipulation of arbitrarily selected specific spinal segments may not be entirely relevant to pre-post changes in the properties of the SSS.

Motion palpation findings should not solely determine clinical interventions, but instead should consider additional parameters related to the patients and doctors. These include the patients' symptoms, comorbidities, prior response to care, preferences, values, goals, and specific diagnoses. It should also be noted that demonstrating an examination procedure to be reliable does not, by itself, demonstrate it is clinically useful. It must also be shown to obtain valid information, generally through comparison of its results with those of a reference standard. Future studies might address whether deploying this stiffest site paradigm results in improved clinical outcomes.

Conclusion

More than half the time (54.3%) the examiners in this study agreed on the exact segment or at least the motion segment that was stiff in a given spinal region. The MedianAED for the combined dataset was 01.1 VE. This information supports high levels of interexaminer reliability for the SSS in each region of the spine and in the combined dataset. Reliability estimates based on absolute examiner differences were relatively higher and seemed to be more trustworthy than prior estimates based on discrete analysis and analysis with κ because the assessment method more closely resembled that used by clinicians in clinical practice. The reliability seen in this study is broadly consistent with previous studies of MP using continuous analysis.

Funding Sources and Conflicts of Interest

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

Contributorship Information

Concept development (provided idea for the research): K.H., R.C., M.Y.
Design (planned the methods to generate the results): K.H., D.R.
Supervision (provided oversight, responsible for organization and implementation, writing of the manuscript): K.H., D.R., R.C., M.S., H.H.
Data collection/processing (responsible for experiments, patient management, organization, or reporting data): K.H., D.R.
Analysis/interpretation (responsible for statistical analysis, evaluation, and presentation of the results): K.H., D.R., R.C.
Literature search (performed the literature search): K.H., D.R.
Writing (responsible for writing a substantive part of the manuscript): K.H., D.R., R.C.
Critical review (revised manuscript for intellectual content, this does not relate to spelling and grammar checking): K.H., D.R., R.C.

Practical Applications

- In this study, spinal motion palpation using continuous data for determining the stiffest spinal site had good reliability.
- About half of examiner differences for the stiffest spinal site are smaller than 1.25 vertebral equivalents, the dimension of a typical vertebra.
- Examining an area of clinical interest for the stiffest spinal site, considering some degree of nonspecificity in spinal manipulation, may be a useful clinically relevant protocol.

References

1. Holt K, Kelly B, Taylor H. Practice characteristics of chiropractors in New Zealand. *Chiropr J Aust.* 2009;39(3):103-109.
2. Walker BF, 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(9):583-589.
3. Haneline M, Cooperstein R, Young M, Birkeland K. An annotated bibliography of spinal motion palpation reliability studies. *JCCA J Can Chiropr Assoc.* 2009;53(1):40-58.
4. Troyanovich SJ, Harrison DD. Motion palpation: it's time to accept the evidence. *J Manipulative Physiol Ther.* 1998;21(8):568-571.
5. Hestbaek L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. *J Manipulative Physiol Ther.* 2000;23(4):258-275.
6. Cooperstein R, Haneline M, Young M. Interexaminer reliability of thoracic motion palpation using confidence ratings and continuous analysis. *J Chiropr Med.* 2010;9(3):99-106.
7. Cooperstein R. Interexaminer reliability of cervical motion palpation using continuous measures and rater confidence levels. *J Can Chiropr Assoc.* 2013;57(2):156-164.
8. Cooperstein R, Young M. The reliability of lumbar motion palpation using continuous analysis and confidence ratings. *J Can Chiropr Assoc.* 2016;60(2):146-157.
9. Cooperstein R, Young M. The reliability of spinal motion palpation determination of the location of the stiffest spinal site is influenced by confidence ratings: a secondary analysis of three studies. *Chiropr Man Therap.* 2016;24:50.
10. Holt K, Russell D, Cooperstein R, Young M, Sherson M, Haavik H. Interexaminer reliability of the detection of vertebral subluxations using continuous measures and confidence levels. *Chiropr J Aust.* 2018;46(1):100-117.
11. Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: intent and validity. *Man Ther.* 2009;14(1):36-44.
12. Bergmann T, Peterson DH. *Chiropractic Technique*. 3rd ed. St. Louis, MO: Elsevier; 2011.
13. Leys CL, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J Exp Soc Psychol.* 2013;49(4):764-766.
14. 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):285-292.
15. Harlick JC, Milosavljevic S, Milburn PD. Palpation identification of spinous processes in the lumbar spine. *Man Ther.* 2007;12(1):56-62.
16. Triano JJ, Budgell B, Bagnulo A, et al. Review of methods used by chiropractors to determine the site for applying manipulation. *Chiropr Man Therap.* 2013;21(1):36.
17. Najm WI, Seffinger MA, Mishra SI, et al. Content validity of manual spinal palpatory exams - A systematic review. *BMC Complement Altern Med.* 2003;3:1.
18. Mior SA, McGregor M, Schut B. The role of experience in clinical accuracy. *J Manipulative Physiol Ther.* 1990;13(2):68-71.
19. 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;5:19.
20. Snodgrass SJ, Haskins R, Rivett DA. A structured review of spinal stiffness as a kinesiological outcome of manipulation: its measurement and utility in diagnosis, prognosis and

treatment decision-making. *J Electromyogr Kinesiol.* 2012;22 (5):708-723.
21. Campbell BD, Snodgrass SJ. The effects of thoracic manipulation on posteroanterior spinal stiffness. *J Orthop Sports Phys Ther.* 2010;40(11):685-693.
22. Matyas TA, Bach TM. The reliability of selected techniques in clinical arthrometrics. *Aust J Physiother.* 1985;31(5):175-199.
23. Maher C, Adams R. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. *Phys Ther.* 1994;74(9):801-809 discussion 809-811.
24. Binkley J, Stratford PW, Gill C. Interrater reliability of lumbar accessory motion mobility testing. *Phys Ther.* 1995;75(9): 786-792 discussion 793-795.
25. Markon KE, Chmielewski M, Miller CJ. The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. *Psychol Bull.* 2011;137(5):856-879.
26. Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, et al. Diagnostic stability of psychiatric disorders in clinical practice. *Br J Psychiatry.* 2007;190:210-216.
27. Ross JK, Bereznick DE, McGill SM. Determining cavitation location during lumbar and thoracic spinal manipulation: is spinal manipulation accurate and specific? *Spine.* 2004;29 (13):1452-1457.
28. Ross K, Bereznick DE, McGill SM. The accuracy and specificity of lumbar and thoracic spinal manipulation. *J Chiropr Educ.* 2004;18(1):26.
29. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-310.
30. Bland JM, Altman DG. Comparing methods of measurement: why plotting differences against standard method is misleading. *Lancet.* 1986;346(8982):1085-1087.
31. Rouse MW, Borsting E, Deland PN. Convergence Insufficiency and Reading Study (CIRS) Group. Reliability of binocular vision measurements used in the classification of convergence insufficiency. *Optom Vis Sci.* 2002;79(4): 254-264.
32. Huang S, Wang T, Yang M. The evaluation of statistical methods for estimating the lower limit of detection. *Assay Drug Dev Technol.* 2013;11(1):35-43.
33. Cooperstein R. Two types of motion palpation: the excursion and the end-feel methods. *JACA Online.* 2008;45(5):25-26.
34. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther.* 1994;74(8):777-788.