INTEREXAMINER RELIABILITY OF A MULTIDIMENSIONAL BATTERY OF TESTS USED TO ASSESS FOR VERTEBRAL SUBLUXATIONS

Kelly Holt[1], David Russell[2], Robert Cooperstein[3], Morgan Young[3], Matthew Sherson[4], Heidi Haavik[1]

[1] Centre for Chiropractic Research, New Zealand College of Chiropractic, Auckland, New Zealand, [2] Private Practice, Auckland, New Zealand, [3] Palmer Center for Chiropractic Research, Palmer College of Chiropractic, [4] Department Head (Technique), New Zealand College of Chiropractic, Auckland, New Zealand

Corresponding Author: Kelly Holt, kelly.holt@nzchiro.co.nz

INTEREXAMINER RELIABILITY OF A MULTIDIMENSIONAL BATTERY OF TESTS USED TO ASSESS FOR VERTEBRAL SUBLUXATIONS

ABSTRACT

Objective: The purpose of this study was to investigate the interexaminer reliability of assessing for vertebral subluxations using a multidimensional battery of tests and continuous measures analysis approach.

Methods: 70 participants were assessed by 2 blinded examiners. Examiners used a multidimensional battery of tests to assess for vertebral subluxations in 3 regions (cervical, thoracic, lumbar) of the spine, and indicated which segment had the most positive test findings in each spinal region. The distance was measured from the segment to marks that had been placed on the spine. Interexaminer reliability was determined by calculating the median absolute examiner difference in vertebral equivalents (VEs), where a VE is the height of a typical vertebra in each region of the spine. If the median examiner difference was ≤ 1VE, there was definite agreement on the motion segment that had the most subluxation findings. Differences > 1VE but ≤2VE suggested agreement on the same motion segment, and differences >2VE precluded agreement on the same motion segment.

Results: Median absolute examiner differences were 0.5 vertebral equivalents in the lumbar region, 1.0 vertebral equivalent in the cervical and thoracic regions, and 0.6 vertebral equivalents when combined across all regions. In the combined dataset, definite agreement (≤1 vertebral equivalent) occurred 63.3% of the time, possible agreement 19.0% of the time, and definite disagreement 17.6% of the time.

Conclusion: A multidimensional approach to vertebral subluxation assessment was reliable between examiners for detecting the level of vertebral subluxation in all regions of the spine. Median absolute examiner differences indicated examiners agreed on the motion segment with the most positive vertebral subluxation test findings most of the time. Vertebral subluxation assessment agreement, when analyzed using continuous data, indicates much higher reliability than has previously been associated with assessing agreement using discrete data. (*Chiropr J Australia 46;1:100-117*)

Key Indexing Terms: Chiropractic; Diagnostic Testing; Spine; Reliability; Vertebral Subluxation

INTRODUCTION

The primary objective of the chiropractic profession is to improve (primarily) spinal function in order to either improve nervous system function and general health and/or prevent or manage neuromusculoskeletal conditions.(1-3) To do this, chiropractors identify, analyze and correct areas of vertebral subluxation (sometimes referred to as spinal dysfunction) using a variety of chiropractic adjustment techniques, which predominantly involve manual procedures. (1,3) However, there seems to be little

agreement on what constitutes vertebral subluxation or what to call it.(4,5) It has variously been termed subluxation, vertebral subluxation, the vertebral subluxation complex, the chiropractic subluxation, spinal dysfunction, biomechanical joint dysfunction, or a manipulable or functional spinal lesion.(4, 6-9) The term traditionally, or historically, used by the chiropractic profession to define this dysfunction is vertebral subluxation.(10, 11) Recently a group of chiropractic colleges, known as The Rubicon Group, released a definition of 'chiropractic subluxation' that provided a testable model for this clinical entity.(9) In their definition and position statement this group states that:

> *"We currently define a chiropractic subluxation as a self-perpetuating, central segmental motor control problem that involves a joint, such as a vertebral motion segment, that is not moving appropriately, resulting in ongoing maladaptive neural plastic changes that interfere with the central nervous system's ability to self-regulate, self-organize, adapt, repair and heal."*

This definition provides a model that includes joints outside of the spine, so uses the term 'chiropractic subluxation', instead of the more exclusive term vertebral subluxation. One of the reasons for the release of this definition was that there is currently little consensus regarding the nature of the vertebral subluxation or its associated neurological manifestations.(6,9) One issue that has led to this paradox is that the chiropractic profession has struggled to demonstrate that it can reliably identify vertebral subluxations.(7,12)

Vertebral subluxation assessment generally involves evaluating what have been described as the 'pathophysiological consequences of manipulable lesions'.(7) These have been loosely aggregated into overlapping categories that are often referred to as a PARTS evaluation.(13) The categories include; Pain, Asymmetry, changes in relative Range of motion, changes in Tissue temperature/texture/tone, and other findings that can be identified using Special tests.(7, 13) Some methods of vertebral subluxation assessment, such as pain provocation at segmental levels, have been described as being reliable and valid.(7) However, many of the methods commonly used by chiropractors to functionally assess the spine have previously been found to have limited interexaminer reliability.(7)

In chiropractic practice a montage of examination tests, in combination with other aspects of the patient presentation, history, and preferences, is generally used to decide where to deliver a chiropractic adjustment, as opposed to a single evaluation method such as motion palpation alone.(14, 15) A number of studies have therefore used a combination of assessment methods to identify areas of vertebral subluxation with reliability results varying from poor to substantial.(7,12,16) When considering the results of these trials collectively, and taking study quality into account, it remains unclear whether multidimensional approaches contribute more than their component elements when deciding where to adjust the spine.(7,12)

A continuous measures system, combined with an assessment of examiner confidence, was found to lead to improved levels of interexaminer reliability for spinal motion palpation assessment.(17-21) The primary objective of this study was to use this continuous measures system to determine whether examiners agree on the vertebral

segment with the most indicators for adjustment based on the findings of a multidimensional battery of tests that can be used to assess for vertebral subluxations.

METHODS

Design and Setting

This interexaminer reliability trial was conducted at the Chiropractic Centre (student training facility) of the New Zealand College of Chiropractic (NZCC) during regular operating hours. The trial was approved by the NZCC Research Committee and was given exemption from formal external ethical review by the local Ministry of Health ethics committee as it was deemed to be an evaluation of an existing practice against a standard that did not significantly differ from standard practice and/or quality assurance.

Participants

A convenience sample was recruited from patients presenting to the Chiropractic Centre. Potential participant's eligible for inclusion were all public patients attending the Chiropractic Centre during data collection sessions who were over the age of 18 and verbally consented to participate in the study. Data collection took place during regular shift times in the Chiropractic Centre when study examiners and research assistants were available. All public patients attending the Chiropractic Centre during these shift times were asked to participate as long as it did not interfere with the logistical operation of the Chiropractic Centre (e.g. scheduling clashes or room bookings). No incentives were given for participation and participation was only at the agreement of the participant.

Examiners

Two chiropractors, each with over 10 years of clinical experience, were the examiners in this study. Both chiropractors were involved in teaching within the technique program at the NZCC and regularly mentored interns as supervising clinicians in the Chiropractic Centre. Frequent consensus training sessions were held over a 3-month period prior to data collection in order to ensure the multidimensional spinal assessments were performed consistently.

Measurement/rating Process

When a patient over the age of 18 presented to the Chiropractic Centre during a data collection session, a research assistant (RA) assessed whether their participation in the trial would interfere with the logistical flow of operations in the Chiropractic Centre. If not, the RA explained the study to the patient and asked them if they consented to participate. If they agreed to participate they were escorted to an assessment room by the RA. Their age and gender were recorded along with the date of their most recent chiropractic care session and whether they were currently experiencing any bodily pain or not. If they were experiencing symptoms they were asked to describe the location and severity of any symptoms using a numeric pain rating scale ranging from 0 to 10, with 0 described as no pain and 10 being the worst pain imaginable. They were then

asked to remove their shirt or change into a gown and marks were placed on their spine over the inferior tip of the C7 and T12 spinous processes while the patient was seated.

The first examiner then entered the room, accompanied by an RA, and performed a multidimensional battery of vertebral subluxation assessment tests (Table 1). The battery of tests included motion palpation, leg length checks, soft tissue palpation, and joint play/end feel assessment. These assessments are all part of the routine spinal assessment package taught at the NZCC. When the examination was complete, the segment in each area of the spine that the examiner believed to have the most positive vertebral subluxation test indicators was identified and the RA measured the distance from the applicable skin marking (C7 for cervical and thoracic regions, and T12 for the lumbar region) to the position indicated by the examiner. All measurements were performed in the same seated position that was used for placing the marks on the skin. If the examiners found 2 levels in a region that had an equal amount of motion, soft tissue and joint play findings (with no additional finding to help make a decision (e.g. leg length inequality, Derifield tests or Cervical Syndrome), the lowest level was recorded as the level of vertebral subluxation. The examiner was also asked to indicate whether they were confident or not confident about their findings. The second examiner then entered the room within 5 minutes with a second RA, and while remaining blind to the findings of the first examiner, repeated the assessment. The examiners were not provided with any clinical information about the participants, they alternated their order in assessing the participants, and they did not converse with the participants.

Table 1: Battery of tests used in vertebral subluxation determination

Order	Assessment	Test Procedure	Positive test indicates
1	Motion - Lumbar	The patient was seated at the end of the table with their arms crossed in front of them. The examiner used their left hand hold the patients crossed arms to direct their motion and their right hand to assess for the following motions: Posterior/anterior (P-A) – The examiner contacted the spinous being assessed with their index finger and the spinous of the inferior segment with their middle finger of their right hand. The examiner 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.	Restriction is indicated when patient is in extension and/or flexion.
2	Motion - Thoracic	The patient was seated at the end of the table with their arms crossed in front of them. The examiner used their left hand to hold the patients	Restriction is indicated when patient is in extension and/or flexion

		crossed arms to direct their motion and their right hand to assess for the following motions: P-A – The examiner contacted the spinous of the level being examined with their index finger and contacted the spinous of the inferior segment with the middle finger of their right hand. The patient was moved 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 T12 superiorly to T1 and was motioned only once.	
3	Motion - Cervical	The patient was seated at the end of the table with their arms uncrossed. The examiner used their left hand hold the patients head to direct their motion and with their right hand performing the following motions: Rotation and extension – The examiner contacts the articular pillars bilaterally with the index finger and thumb of your right hand. The patient was moved through left and right rotation and extension, feeling for motion of the spinal level. Each spinal level was individually assessed from C7 superiorly to C2 and was motioned only once each side. Coupled motion of C1 – the examiner contacts the posterior aspect of the C1 TVP with your right hand. The patient was moved through extension and right and left rotation and was motioned only once each side. Coupled motion of C1 was used instead of flexion / extension as the lack of a C1 spinous process makes a flexion extension assessment of C1 movement challenging.	C2 - C7 – restriction is indicated when patient is in rotation and extension on either the left or right side, or both C1 – restriction indicated at end of coupled rotation and extension
4	LLI	Patient prone. The examiner was standing at the foot of the table, centred to the patient with their index and middle finger either side of malleolae, thumb over the cuboid placing firm cephalad pressure to the patients feet	Left short → left leg appears shorter than the right Right short → right leg appears shorter than the left
5	Derifield	The examiner was standing at the foot of the table, centred to the patient with thumbs on the ball of the foot and index finger to the side of the feet raise the legs to 90° and	Left long → left leg appears longer than the right in the flexed knee position indicating potential Left sided L5 subluxation

		view the relative leg length by looking from the centre of the sacrum up to the EOP. The examiner raised the legs only 1 time	Right long → right leg appears longer than the right in the flexed knee position indicating potential Right sided L5 subluxation
6	Cervical Syndrome	While the patient was prone they were instructed to tuck their chin and turn their head to the left and rest their right ear on the table. The examiner was standing at the foot of the table, centred to the patient, looking at the patients feet for any changes during this process. The patient was then instructed to tuck their chin and turn their head to the right and rest their left ear on the table. The patient only turned their head either side 1 time	Balance → leg lengths balance indicating a C1 subluxation if balanced quickly or lower cervical if balanced slowly None → leg lengths do not balance
7	ST tension - lumbar	The examiner used their index, middle and ring fingers, lightly palpating the soft tissues either side of the lumbar spine. Palpation was performed from L1 inferiorly to L5	Note which side has increased muscle tension
8	Joint play/end feel - lumbar	The examiner used the heel of their hand to assess for joint play in a primarily P-A motion. The examiner contacted the spinous of each spinal level, following the lordosis of the spine. Palpation was performed from L5 superiorly to L1 with pressure being applied to each level only once	Decreased P-A movement indicates restricted joint play/end feel
9	ST tension - thoracic	The examiner used their index, middle and ring fingers, lightly palpating the soft tissues either side of the thoracic spine. Palpation was performed from T1 inferiorly to T12.	Note which side has increased muscle tension
10	Joint play/end feel - thoracic	The examiner used the heel of their hand to assess for joint play in a P-A and I-S motion. The examiner contacted the TVP's of each spinal level, following the kyphosis of the spine. Palpation was performed from T12 superiorly to T1 with pressure being applied to each level only once	Decreased P-A and I-S movement indicates restricted joint play/end feel
11	ST tension - cervical	The examiner used their index, middle and ring fingers, lightly palpating the soft tissues either side of the cervical spine. Palpation was performed from C1 inferiorly to C7	Note which side has increased muscle tension

Statistical Analysis

The paired findings for examiner differences on vertebral subluxation were assessed for normality to determine the appropriate statistical function(s) to be used to assess interexaminer reliability. Following this, interexaminer reliability in this study was determined by calculating Median Absolute Examiner Differences (MedAED). Data dispersion was determined by calculating the Median Absolute Deviation (MAD).(22) Since standard deviation cannot be calculated in the usual manner when working with absolute values, data dispersion was characterized by MAD, the median of the absolute deviations of examiners differences from the median of such differences. MAD) is calculated as the median of the absolute value of each value, xi, minus the median: MAD = median (|xi - median(xi)|). In addition to being provided in “cm” units, MedAED and MAD were transformed into and presented as vertebral equivalents (VEs), where VE is defined as the height of a typical vertebra. Since the height of a typical vertebra varies according to the spinal region, examiner differences reported in cm would misleadingly imply different degrees of examiner reliability depending on the spinal region. For example, a MedAED of 4 cm constitutes a median difference of 1 vertebral body height in the lumbar spine, but in the cervical spine, where the vertebrae are shorter, would constitute median examiner differences of over 2 vertebral body heights. Reporting the data as VEs allows immediate comparisons of examiner reliability, irrespective of spinal region. To convert cm to VE, the following heuristic weighting factors were used: 2.3cm for a typical thoracic segment,(23) 1.8cm for a typical cervical segment,(24) and 4cm for a typical lumbar segment.(25) Calculations were also performed to determine the degree of examiner agreement on the level of vertebral subluxation with the following possible defined outcomes for identifying the most subluxated vertebra or the motion segment including it:

MedAED ≤1.0VE: definite agreement
MedAED >1.0VE and ≤2.0VE: indeterminate agreement
MedAED ≤1.5 VE: acceptable reliability
MedAED >2.0 VE: definite disagreement

Figure 1 illustrates why MedAED values were interpreted in this way. If MedAED is less than 1 VE, it may be stated there was definite agreement on the vertebral subluxation or at least the motion segment containing it. A spinal motion segment is the smallest spinal function unit that is comprised of 2 adjacent vertebrae and their accompanying ligaments and intervertebral disc.(26) At the other extreme, if MedAED is greater than 2 VE’s, there was definite disagreement as the examiners could not have agreed on the same vertebra, let alone the motion segment including it. In the range where MedAED is between 1 and 2 VE’s, there was indeterminate agreement, depending on whether an examiner happened to identify the vertebral subluxation near the center of a spinal segment, or rather identified the vertebral subluxation close to the top or bottom of a spinal segment (Figure 1). We thought it reasonable to identify the midpoint of this range, MedAED ≤1.5 VE, as the boundary of acceptable interexaminer reliability, wherein with great likelihood the examiners at least agreed on the motion segment including the vertebral subluxation. It would be very difficult, if not impossible, to reduce the size of the indeterminate zone. Doing so, would require untenable assumptions as to exactly where the spinal locations the examiners judged most subluxated were situated in relation to the actual center of the vertebrae.

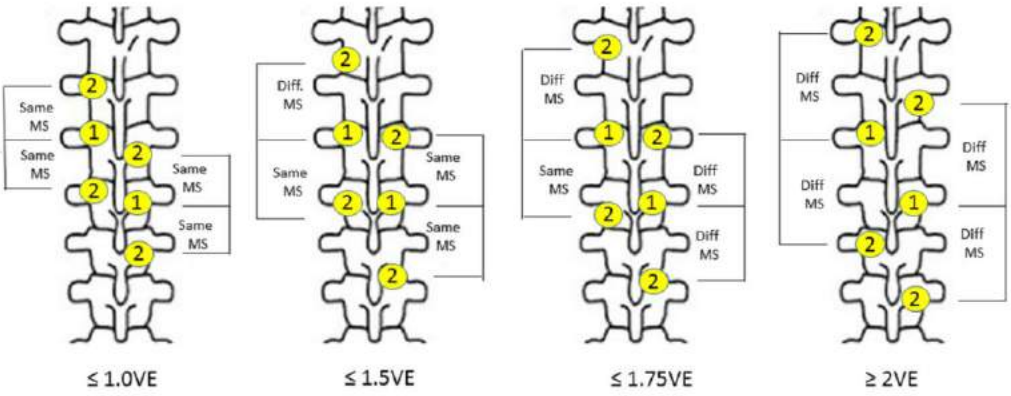

Figure 1: Whether examiners 1 and 2 agree on motion segment including vertebral subluxation depends on whether examiner 1 locates vertebral subluxation closer to top (or bottom) of vertebra, as on the left in illustrations, or closer to vertebral center, as on the right in illustrations. Agreement/disagreement on motion segment including vertebral subluxation is shown for 4 possible VE cut-points. At VE≤1.0, agreement is 100%. At VE ≤1.5, there is mostly agreement. At VE≤1.75 there is mostly disagreement. At VE≥2.0 100% disagreement. In this study, VE ≤1.5 was judged to reflect "adequate" examiner agreement.

RESULTS

Data collection took place across 21 study sessions between October 2014 and March 2015. Seventy patients were assessed during the trial with between 1 and 6 patients being assessed during each data collection session. All patients who were asked to participate agreed to do so. Fifty-one percent of participants reported the presence of bodily pain with the mean pain level being 4.3/10 (SD = 2.1) if present. A summary of patient characteristics is provided in Table 2.

Table 2: Patient characteristics

Characteristic	*Value*
Age*	46 (18)
Age range	19-87
Female[#]	40 (57%)
Bodily pain present[#]	36 (51%)
Pain severity (if present, out of 10)*	4.3 (2.1)
Pain severity, range (if present, out of 10)	1-8
Days since last chiropractic visit*	10 (10)
Days since last chiropractic visit, range	1-60
*Mean (SD), [#] n (%)	

Figure 2 shows the distributions of the examiners' determination of the most subluxated segments for each spinal region (calculated using the distance in VE's from the standardized measurement point). The most common levels identified in each region were L2, T7, and C3.

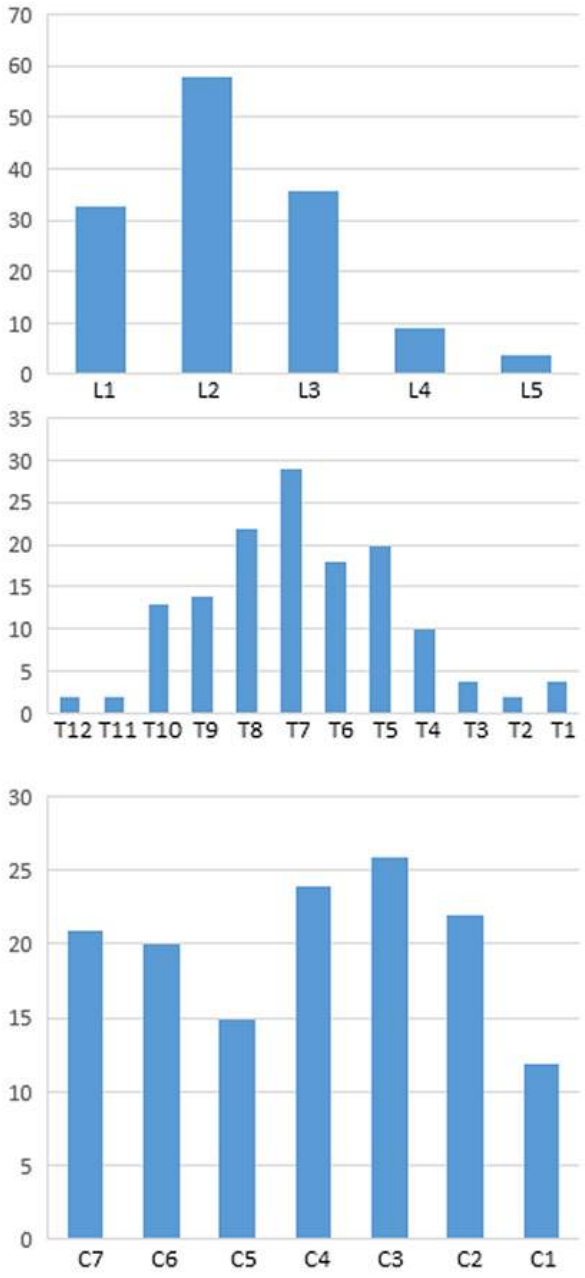

Figure 2: Distribution of the examiners' determination of the most subluxated segment for each spinal region. Results for the lumbar region are at the top, thoracic region are in the middle, and cervical region at the bottom. The y axis of each graph is the number of times either of the examiners identified each segment as being the most subluxated in each region.

Shapiro-Wilk testing for the cervical, thoracic, and lumbar spines demonstrated that in none of these regions were the pair examiners' ratings for vertebral subluxation normally distributed. This result technically precluded analysis using Intraclass Correlation (ICC) for parametric data, as well as calculating Bland-Altman Limits of Agreement. Given how commonly ICC is used to estimate reliability for continuous data, we thought it reasonable, despite this limitation, to provide their values: ICC values for interexaminer agreement were 0.55 in the cervical region, 0.57 in the thoracic region, and 0.61 in the lumbar region. Although these values are considered to be fair to good, (27) they should be interpreted with caution due to the non-normal distribution of examiner differences. Furthermore, there is another reason to be cautious in interpreting these values. ICC values are misleadingly depressed when subject variability is relatively low; i.e., the subjects are relatively homogeneous.(28) This is because ICC is a ratio of the variance within subjects to the total variance (the sum of within and between-subject variance). When within-subjects variance is small, as in the present study, where the most subluxated spinal locations were not randomly distributed in the thoracic and lumber spines, ICC values can be surprisingly low even when the examiners tend to agree.

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.

MedAED and MAD values, in both cm and VE's are provided (Table 3). Given the differing vertical dimensions of typical vertebrae in the 3 spinal regions, the authors believe it more meaningful to compare the results for the different regions as expressed in VE units. Median examiner differences for vertebral subluxation assessment were smallest in the lumbar region (0.5VE), and equal in the thoracic region and cervical regions (1.0VE). For the combined data, including all 3 spinal regions, MedAED was 0.6VE. MAD values, which represent data dispersion, ranged from a low of 0.3VE in the lumbar spine, to 0.8VE in the thoracic spine. Figure 3 summarizes the results of examiner agreement using a box-and-whisker plot, identifying 12 outliers, defined as examiner differences outliers out of the box by more than 1.5 times the interquartile range. Subgroup analyses did not indicate a significant effect of patient symptomatology on reliability results.

Table 3: Interexaminer reliability of a multidimensional vertebral subluxation assessment

DISTANCE		Delta Vertebral Subluxation, cm		Vertebral Subluxation, VE	
N	Spinal region	MedAED	MAD	MedAED	MAD
70	Cervical	1.8	1.3	1.0	0.7
70	Thoracic	2.3	1.8	1.0	0.8
70	Lumbar	2.0	1.0	0.5	0.3
210	Combined	Left blank deliberately		0.6	0.5
Abbreviations: MedAED=Median absolute examiner differences; MAD=Median absolute deviation; VE=Vertebral equivalent					

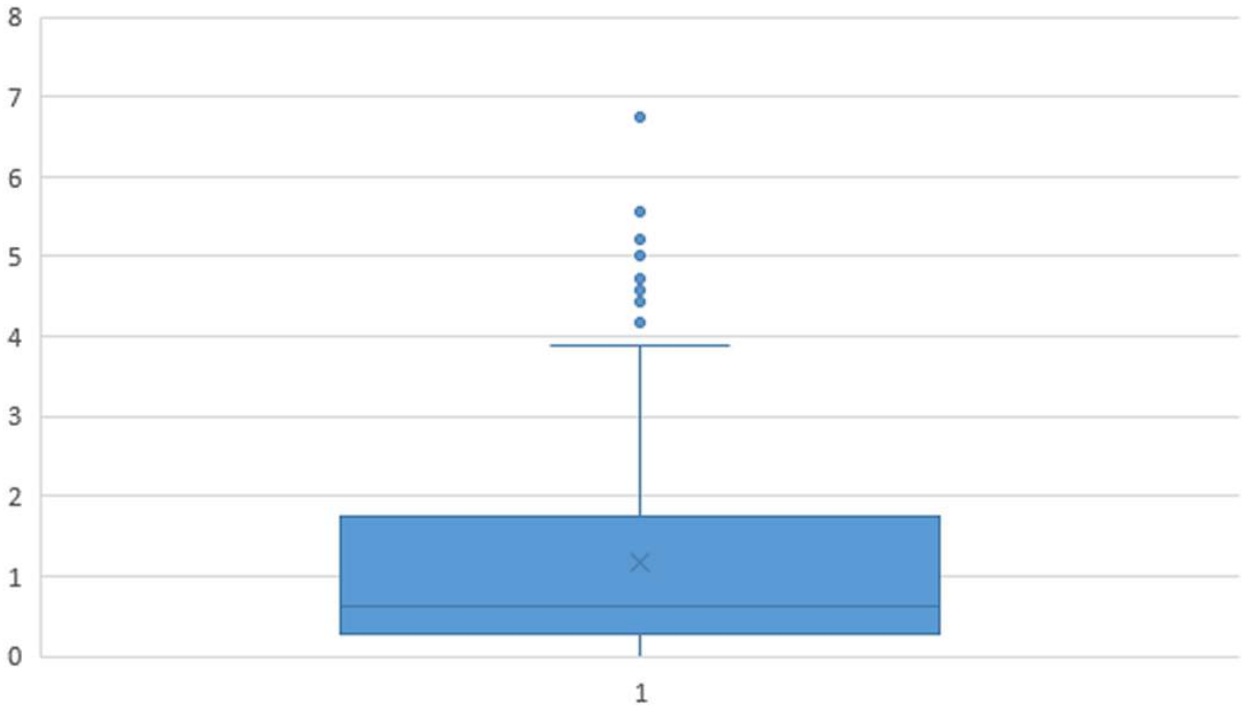

Figure 3: Box-and-whisker plot of level of agreement in vertebral equivalents results in the combined dataset. The low whisker represents the bottom 25% of examiner differences; the box represents the middle half of the examiner differences; and the upper whisker the top 25%. The 12 dots in the upper whisker represent outliers that are out of the box by more than 1.5 times the interquartile range.

In the regional analyses, examiner agreement on the vertebral subluxation, or the motion segment including it, ranged from a high of 90% in the lumbar spine to a low of 40% in the thoracic spine (Table 4). Definite disagreement on the vertebral subluxation, or the motion segment including it, ranged from a high of 40.0% in the thoracic spine to a low of 0% in the lumbar spine. Adequate agreement varied from a high of 97.1% in the lumbar spine to a low of 47.1% in the thoracic spine. In the combined dataset, definite agreement was 63.3%, definite disagreement was 17.6%, and acceptable agreement was 73.3%.

Table 4. Degrees of agreement on segmental vertebral subluxation or motion segment including it

	Regional Agreement	AED ≤1VE	1VE<AED≤2VE	AED>2 VE	Total	AED ≤1.5VE
N		Definite agreement	Indeterminate agreement	Definite disagreement		Adequate agreement
70	Cervical	48.6	31.4	20.0	100.0	62.9
70	Thoracic	40.0	20.0	40.0	100.0	47.1
70	Lumbar	90.0	10.0	0.0	100.0	97.1
210	Combined	63.3	19.0	17.6	100.0	73.3
Abbreviations: AED=Absolute examiner differences; VE=Vertebral equivalent						

DISCUSSION

The results of this study suggest that a multidimensional approach to vertebral subluxation assessment was reliable between examiners for detecting the level of vertebral subluxation in all regions of the spine. In at least 63% of assessments the examiners agreed on the same motion segment across all regions of the spine. It has been hypothesized that reliability would increase when the examiners were confident in their findings, as had been the case in prior studies by Cooperstein et al in the thoracic (19), cervical (20), and lumbar (17) regions. In the present study there were too few observations in which 1 or both examiners lacked confidence to test this hypothesis.

When a clinical variable can be measured using either discrete or continuous analysis, there are good reasons to expect to find greater reliability when using continuous data. Markon, Chmielewski (29) performed 2 meta-analyses including 58 studies in which both continuous and categorical measures were used to assess psychopathology. They 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 subject requirements. Baca-Garcia, Perez-Rodriguez (30) suggested that low reliability in assessing patients may result from the use of discrete diagnostic criteria that fail to recognize continuous variation in patients' presentations.

In assessing vertebral subluxation, the most obvious explanation why strict segmental assessment is less likely to detect agreement than using a most subluxated site paradigm is that when the finding actually lies on a continuum, and is then artificially discretized, information is lost. If vertebral subluxation were understood to involve at least one motion segment, rating individual segments as subluxated or not will fail to identify larger fields of vertebral subluxation and thus miss the overlap of those fields among examiners' assessments.

Due to the non-parametric nature of the data, rendering conventional ICC analysis suspect, the authors strongly emphasized understanding interexaminer reliability as the median of absolute examiner absolute differences, MedAED, which provides a measure of the typical difference between examiners. In fact, it is especially useful when examiner differences are not normally distributed.(31). MAD is a robust measure of dispersion (functionally similar to standard deviation) that is resilient to outliers and is suitable for datasets that are not normally distributed. Data points at the very extremes of the distribution of examiner differences do not impact the calculation of MedAED any more than less extreme values.

Figure 2 suggests that the examiners found the subjects relatively homogeneous in their most subluxated level in the thoracic and lumbar regions, but in the cervical region the examiners' findings for the most subluxated level were relatively dispersed throughout the range of the cervical spine.

Most likely the higher agreement (90.0%) seen in the lumbar spine compared to the thoracic spine (40.0%) reflects the fact that there are only 5 lumbar segments vs. 12 thoracic segments to choose among. Had the thoracic spine been divided into upper

and lower divisions, in all likelihood, examiner agreement would have increased. Even granting the limitation that the thoracic spine was not subdivided into sections comparable in numbers of segments to the cervical and lumbar spines, acceptable examiner agreement was 73.3% in the combined dataset.

A box-and-whisker plot is provided to summarize the results in the combined dataset (Figure 3). Analysis of the plot leads to the conclusion that 154/210 (73%) of examiner differences were ≤1.5VE, which the authors deem the boundary of acceptable reliability, and 12/210 (6%) of examiner differences were ≥1.5 times the interquartile range, extreme data points generally considered outliers (21).

Previous research has suggested that the interexaminer reliability of clusters of tests to identify vertebral subluxations in the spine is questionable, with reviews of the literature concluding that evidence for examination montages is either unclear or that no good quality studies exist that show a testing regimen is reliable.(7, 12) Two studies have suggested that by clustering the results of a number of tests for sacroiliac joint dysfunction substantial interexaminer reliability can be demonstrated.(32, 33) However, previous studies that have investigated multidimensional assessment methods across multiple spinal levels, that also met quality standards,(12) showed marginal interexaminer reliability.(15, 16, 34) French, Green (16) used a multidimensional spinal diagnostic method commonly used by chiropractors to assess interexaminer reliability in the lower thoracic spine, lumbar spine, and sacrum, and found fair agreement (kappa (κ) = 0.27) when averaged across all spinal joints tested. Hawk, Phongphua (34) also used a combination of commonly used chiropractic assessment procedures in their interexaminer reliability study of the lumbar spine and reported levels of agreement that averaged less than chance (κ = -0.08), with the maximum level of agreement across 42 comparisons barely reaching acceptable levels (κ = 0.44). Keating, Bergmann (15) also investigated interexaminer reliability of the lumbar spine using a multidimensional approach that included the 4 strongest tests from an 8 test regimen. They reported slightly better results with ICC's across the levels ranging from 0.34 to 0.62 with an average of 0.46. The average ICC for the multidimensional vertebral subluxation assessment across the spinal regions for the present study was 0.58 which exceeds the values from these previous studies, but must be interpreted with caution due to violation of normality assumptions.

Interestingly, a spinoff study from the present study assessed examiner agreement purely based on the motion palpation assessment included in the present study.(35) In this motion palpation study, the MedAED for examiner agreement in the combined dataset was 1.1 VE, almost twice as large as the MedAED examiner agreement in the combined dataset for the present study, which was 0.6VE. This suggests that using a multidimensional approach to assessing vertebral subluxations is more reliable than using motion palpation alone. Of more clinical relevance is the finding that the examiners in the present study agreed on the same motion segment 73.3% of the time across all spinal regions. This suggests that it is possible to create a multidimensional assessment of vertebral subluxation that is reliable. Interestingly enough, the assessment that was used in the present study did not include pain provocation at segmental levels, which currently has the most convincing favorable evidence for interexaminer reliability.(7) This suggests that the reliability of multidimensional testing

approaches may exceed that reported in this trial if more reliable component parts were to be included in the multidimensional approach.

Limitations of the study

As full-time chiropractic educators it could be argued that the examiners in the study were not representative of chiropractors in the field, though both examiners were still active in part-time private practice. Although examiners were blinded to any other prior findings during the study it is possible that they were familiar with some patient's prior findings or clinical or non-clinical cues based on previous visits to the Chiropractic Centre that they may have supervised; this is also a limitation. It has been hypothesized that reliability would increase when the examiners were confident in their findings, as had been the case in prior studies by Cooperstein et al in the thoracic (19), cervical (20), and lumbar (17) regions. There were too few observations in which 1 or both examiners lacked confidence to test this hypothesis. Data violated normality assumptions which meant neither the more traditional ICC analysis nor Bland-Altman Limits of Agreement could be used in this study. One of the strengths of MedAED for analyzing reliability data such as these, is that it is not influenced by the variability amongst possible responses.

We chose the relatively stringent criterion for "acceptable" agreement that the median examiner difference was ≤1.5VE, corresponding to apparent agreement on the motion segment including the vertebral subluxation. It is entirely possible the clinical field of impact for vertebral subluxation includes not only the motion segment including it, but the motion segments adjacent *to it*, presumably to a lesser extent. Pursuing that logic, clinically relevant examiner agreement in this study may have corresponded to a higher median examiner differences. For example, at the VE≤3VE cut point, agreement in this study occurred 90% of the time in the combined data.

This study showed high levels of interexaminer reliability for a multidimensional battery of tests for detecting vertebral subluxations, but it did not address the validity of these tests. A reliable test cannot be assumed to be useful for clinical decision-making if it has not been shown to be valid. Further research is required to assess the validity of the tests that were used in this study.

The results of this study indicate that chiropractors can agree on the vertebral level to be adjusted which is important when it comes to clinical practice and teaching examination techniques to chiropractic students. Future research is required to determine whether the findings of the multidimensional battery of tests change after an adjustment is provided at that level, and whether patient clinical outcomes are influenced by adjusting at the spinal level with the most positive test findings as opposed to some other means for determining the preferred adjustment site.

CONCLUSION

In this study, high levels of interexaminer reliability were observed in each region of the spine when a multidimensional approach to detect vertebral subluxations was used.

Since the combined MedAED for vertebral subluxations was 0.6VE, it can be stated with confidence that examiners usually agreed on at least the motion segment containing the most positive vertebral subluxation test indicators, and very frequently on the same segment. Vertebral subluxation assessment, when analyzed using continuous data, indicate much higher levels of agreement than has been heretofore associated with assessing agreement using discrete data and the Kappa statistic.

REFERENCES

1. World Health Organization. WHO guidelines on basic safety and training in chiropractic. Geneva: World Health Organization; 2005.
2. Association of Chiropractic Colleges. The Association of Chiropractic Colleges Position Paper # 1. July 1996. . ICA Rev 1996;November/December.
3. Chiropractic WFo. Definitions of Chiropractic 2015 [Available from: https://www.wfc.org/website/index.php?option=com_content&view=article&id=90&Itemid=110.
4. Gatterman ML. Foundations of chiropractic: subluxation. 1st ed. St Louis: Mosby-Year Book, Inc; 1995.
5. Ebrall P. Subluxation, what's in a name. Chiropr J Aust 2011;41(3):110-2.
6. Nelson C. The subluxation question. J Chiropr Humanit 1997;7(1):46-55.
7. Triano JJ, Budgell B, Bagnulo A, Roffey B, Bergmann T, Cooperstein R, et al. Review of methods used by chiropractors to determine the site for applying manipulation. Chiropr Man Ther 2013;21(1):36.
8. Ebrall P, Draper B, Repka A. Towards a 21 century paradigm of chiropractic: stage 1, redesigning clinical learning. J Chiropr Educ 2008;22(2):152-60.
9. Definition and Position Statement on the Chiropractic Subluxation [press release]. [Online] Available at: http://www.therubicongroup.org/#/policies/: The Rubicon Group, 22/5/2017 2017.
10. Gliedt JA, Hawk C, Anderson M, Ahmad K, Bunn D, Cambron J, et al. Chiropractic identity, role and future: a survey of North American chiropractic students. Chiropr Man Ther 2015;23(1):4.
11. 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):583-9.
12. Gemmell H, Miller P. Interexaminer reliability of multidimensional examination regimens used for detecting spinal manipulable lesions: A systematic review. Clin Chiropr 2005;8:199-204.
13. Bergmann TF. P.A.R.T.S. Joint assessment procedure. Chiropr Tech 1993;5(3):135-6.
14. Walker BF. Most common methods used in combination to detect spinal subluxation: A survey of chiropractors in Victoria. Australas Chiropr Osteop 1998;7(3):109-11.
15. Keating JC, Jr., Bergmann TF, Jacobs GE, Finer BA, Larson K. Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. J Manipulative Physiol Ther 1990;13(8):463-70.

16. 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(4):231-8.
17. Cooperstein R, Young M. The reliability of lumbar motion palpation using continuous analysis and confidence ratings. J Canadian Chiropr Assoc 2016; 60(2): 146-57.
18. 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.
19. 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.
20. Cooperstein R, Young M, Haneline M. Interexaminer reliability of cervical motion palpation using continuous measures and rater confidence levels. J Canadian Chiro Assoc 2013;57(2):156-64.
21. Hotmath.com. Box-and-Whisker Plots [Available from: http://hotmath.com/hotmath_help/topics/box-and-whisker-plots.html.
22. Leys C, 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):764-6.
23. Gray H. Anatomy of the human body 1918 [Available from: http://www.bartleby.com/107/25.html.
24. Gilad I, Nissan M. Sagittal evaluation of elemental geometrical dimensions of human vertebrae. J Anat 1985;143:115-20.
25. Terazawa K, Akabane H, Gotouda H, Mizukami K, Nagao M, Takatori T. Estimating stature from the length of the lumbar part of the spine in Japanese. Medicine, science, and the law 1990;30(4):354-7.
26. White AA, Panjabi MM. Clinical biomechanics of the spine. Philadelphia: Lippincott; 1990.
27. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 1994;64(4):284-90.
28. Lee KM, Lee J, Chung CY, Ahn S, Sung KH, Kim TW, et al. Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research. Clinics Orthop Surg 2012;4(2):149-55.
29. Markon KE, Chmielewski M, Miller CJ. The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. Psychological Bull 2011;137(5):856-79.
30. Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, Fernandez del Moral AL, Jimenez-Arriero MA, Gonzalez de Rivera JL, et al. Diagnostic stability of psychiatric disorders in clinical practice. Br J Psychiatry 2007;190:210-6.
31. Rouse MW, Borsting E, Deland PN. Reliability of binocular vision measurements used in the classification of convergence insufficiency. Optom Vis Sci 2002;79(4):254-64.

32. Cibulka MT, Delitto A, Koldehoff RM. Changes in innominate tilt after manipulation of the sacroiliac joint in patients with low back pain. An experimental study. Phys Ther 1988;68(9):1359-63.
33. Kokmeyer DJ, Van der Wurff P, Aufdemkampe G, Fickenscher TC. The reliability of multitest regimens with sacroiliac pain provocation tests. J Manipulative Physiol Ther 2002;25(1):42-8.
34. Hawk C, Phongphua C, Bleecker J, Swank L, Lopez D, Rubley T. Preliminary study of the reliability of assessment procedures for the indications for chiropractic adjustments of the lumbat spine. J Manipulative Physiol Ther 1999;22(6):382-9.
35. Cooperstein R, Holt K, Russell D, Young M, Sherson M, Haavik H. Interexaminer reliability of seated motion palpation in defined spinal regions for the stiffest spinal site using continuous measures analysis. J Manipulative Physiol Ther 2017;IN PRESS.