

Intraobserver Reliability of Cranial Strain Patterns as Evaluated by Osteopathic Physicians: A Pilot Study

Kelly D. Halma, DO; Brian F. Degenhardt, DO; Karen T. Snider, DO; Jane C. Johnson, MA; M. Schaun Flaim, DO; and Danielle Bradshaw, OMS IV

Context: Few studies of inter- or intraobserver reliability have focused on evaluations of cranial strain patterns.

Objective: To determine whether substantial intraobserver reliability can be achieved by osteopathic physicians (DOs) using common palpatory tests to diagnose cranial dysfunction.

Methods: Forty-eight subjects were divided into three diagnostic groups, categorized as those with asthma, headaches, or neither asthma nor headaches (ie, healthy control group). Two blinded DO examiners separately evaluated approximately 8 subjects from each group (4 subjects per session), conducting diagnostic tests for cranial rhythmic impulse (CRI) rate, cranial strain patterns, and quadrants of restriction.

Results: Overall, among the three diagnostic procedures, cranial strain patterns showed the highest intraobserver reliability ($\kappa=0.67$). The highest intraobserver reliability was achieved in cranial strain patterns for the control group ($\kappa=0.82$), followed by the headache ($\kappa=0.67$) and asthma ($\kappa=0.52$) groups. Diagnoses of the left anterior quadrant of restriction also showed substantial intraobserver reliability for the headache and control groups ($\kappa=0.60$ and 0.61 , respectively). Diagnoses of three quadrants of restriction showed moderate overall intraobserver reliability ($\kappa=0.44-0.52$), while the left posterior quadrant had only fair overall intraobserver reliability ($\kappa=0.33$).

Conclusion: Osteopathic physicians can obtain substantial intraobserver reliability when diagnosing cranial strain patterns in healthy subjects as well as those with asthma or headache. However, results are less promising for diagnoses of CRI and quadrants of restriction.

J Am Osteopath Assoc. 2008;108:493-502

From the Kirksville (Mo) College of Osteopathic Medicine-A.T. Still University (Drs Halma, Snider, and Bradshaw), A.T. Still Research Institute (Dr Degenhardt and Ms Johnson), and Northeast Regional Medical Center (Dr Flaim).

This study was supported by a research fellowship grant from the American Osteopathic Association (Grant No. F03-08).

Address correspondence to Kelly D. Halma, DO, Department of Osteopathic Manipulative Medicine, Kirksville College of Osteopathic Medicine-A.T. Still University, 800 W Jefferson St, Kirksville, MO 63501-1443.

E-mail: khalma@atsu.edu

Submitted November 27, 2006; revision received June 7, 2007; accepted June 14, 2007.

Reliability is defined as the reproducibility of findings when a test is repeated to evaluate an unchanged attribute. When investigating the reliability of physical examination findings, two forms of reliability are commonly assessed—inter- and intraobserver reliability. Interobserver reliability is the degree to which multiple independent examiners reach the same conclusion, while intraobserver reliability describes the consistency in results when the same examiner performs the same test on two or more occasions.¹ Although interobserver reliability is more clinically significant than intraobserver reliability,¹ a well-performed assessment of intraobserver reliability can be an important step when evaluating subtle palpatory skills before testing for interobserver reliability.

In 2002, Hartman and Norton² reviewed six published studies of interobserver reliability in osteopathic technique within the cranial field. In these studies, the number of examiners ranged from two to ten, and the number of subjects from 9 to 40. Three of the studies they evaluated examined the interobserver reliability of osteopathic physicians (DOs) palpating the cranial rhythmic impulse (CRI) rate, using intraclass correlation coefficients (ICCs) to measure the interobserver reliability. The CRI, first proposed by John M. Woods, DO, and Rachel H. Woods, DO, in 1961,³ describes a physical manifestation that is routinely used to assess the primary respiratory mechanism, a concept articulated by William G. Sutherland, DO, decades earlier. The ICC values in these three studies² ranged from -0.009 to 0.59 . Low ICC values indicate minimal reliability, whereas high ICC values show greater reliability. Thus, the negative ICC values in these studies are an indication of poor reliability. Only one of the studies⁴ had ICC values that were statistically significant ($P<.001$). However, inconsistent methods in that investigation call into question the validity of the report's statistical significance.²

The remaining studies reviewed by Hartman and Norton² evaluated the interobserver reliability of palpation of the craniosacral rhythm determined by craniosacral therapists, primarily physical therapists and nurses. The ICC values in these three studies were also low, ranging from -0.005 to 0.22 .

None of the six studies² specifically examined intraobserver reliability. In an attempt to deduce intraobserver reliability, Hartman and Norton analyzed the raw data from the three osteopathic medical studies, noting differences in the variability within and between examiners. However, because

the study designs did not include repeated measures on the same subject by the same examiner, intraobserver reliability cannot be established.

Two previously published studies assessed intraobserver reliability for palpation of the CRI by DOs.^{5,6} In 2001, Moran and Gibbons⁵ analyzed intraobserver reliability in a study involving two DO examiners, with 4.5 and 6.5 years of experience, respectively, in osteopathy in the cranial field. In that study,⁵ examiners used momentary-action foot switches to record palpated CRI phases at either the head or the sacrum of 11 subjects. They found that the intraobserver reliability was fair to good with ICCs ranging from 0.52 to 0.72. To provide blinding against the possibility of auditory reference cues being exchanged between examiners, the foot switches were housed inside soundproof casings. In addition, a computer fan was used to produce white noise. In 1996, Norton⁶ found a significant correlation between CRI cycle lengths measured in 9 subjects by the same examiner at the sacrum and then at the cranium. The six DO examiners involved in the 1996 study were described as “experienced” in osteopathy in the cranial field. Limitations of this study were that ICCs were not used, nor were blinding protocols.

Two studies have been performed in the field of physical therapy regarding the intraobserver reliability of craniocervical palpation. In 1998, Rogers et al⁷ found intraobserver ICCs ranging from 0.18 to 0.30 for two physical therapist examiners who measured the craniocervical rate at the head or feet of 28 subjects. Those two examiners had 17 and 5 years’ experience, respectively, palpating the craniocervical rhythm. To prevent auditory cues from being exchanged between examiners, silent operation foot switches were used for blinding purposes during data collection. In another 1998 study of intraobserver reliability by Hanten et al,⁸ two physical therapist examiners, each with less than 1 year of experience palpating the craniocervical rhythm, examined 40 subjects. High ICC values, between 0.78 and 0.83, were reported. Although the examiners had no knowledge of the recorded data, no specific blinding procedure was used.

Few published studies have assessed the interobserver reliability of diagnoses of cranial strain patterns, and no published studies have assessed the intraobserver reliability of diagnoses of cranial strain patterns. In 1977, Upledger⁴ published data on interobserver reliability for assessment of cranial strain patterns, as well as for CRI rate, in a study in which three DOs and one osteopathic medical student evaluated 25 children for 19 different cranial strain parameters. He found the interobserver agreement for these parameters to range from 17% to 100% agreement with an aggregate of 71% agreement for all parameters for all DOs. However, inconsistencies in the study methods used limit the application of these findings.

In 1996, Fraval⁹ assessed the interobserver reliability of examinations by two DOs of cranial strain patterns in infants younger than 6 months. He reported 95.7% agreement on the

presence of dysfunction in a specific cranial bone and 90% agreement on the severity of the dysfunction with an overall Pearson correlation coefficient of 0.65 described as indicating good interobserver reliability.

A limitation of the Upledger⁴ and Fraval⁹ studies is that neither percent agreement nor Pearson correlation coefficients are rigorous statistical methods for measuring reliability. Either Cohen’s kappa (κ) coefficient or intraclass correlation coefficients should have been used in these analyses, depending on the measurement scale of the test findings. Cohen’s κ coefficient and related techniques have been identified as the optimal statistical method for quantifying intra- and interobserver reliability in most cases of palpatory examinations.^{10,11}

Previous manual medicine palpation studies have shown higher rates of intraobserver reliability than interobserver reliability.¹² However, the validity of analyses of intraobserver reliability has been questioned due to the difficulty of blinding examiners to their previous findings.¹ Furthermore, results of previous studies can be questioned for many other reasons, including examiners’ inexperience and small sample sizes.

The present study improves on previous research in several ways. Detailed blinding protocols were used for DO examiners. Both examiners are board-certified by the American Osteopathic Board of Neuromusculoskeletal Medicine in neuromusculoskeletal and osteopathic manipulative medicine. Both examiners likewise have completed at least seven accredited 40-hour courses in osteopathy in the cranial field. Each subject was evaluated three times by these examiners to reduce diagnostic results that would arise by chance alone. Subjects were selected for inclusion in an asthma or headache diagnostic group based on their potential to produce persistent cranial strain patterns, while healthy subjects in a control group were used to diversify the data collected. An improved statistical analysis, using Cohen’s κ coefficient, was used to assess agreement.

The objective of the present pilot study was to determine whether substantial intraobserver reliability ($\kappa > 0.60$) could be achieved by DOs using common palpatory tests to diagnose the cranial mechanisms in healthy subjects and those with one of two specified medical conditions. The palpatory tests that were assessed for intraobserver reliability included diagnoses of CRI rate, cranial strain patterns, and quadrants of restriction. We hypothesized that well-blinded, board-certified DOs specializing in neuromusculoskeletal medicine could establish substantial intraobserver reliability for these three diagnostic procedures.

Methods

The traditional teachings of osteopathy in the cranial field hypothesize that individuals with asthma as well as those with recurrent headaches have distinct cranial strain patterns.¹³ Therefore, 48 subjects were included in this study and were allocated to one of the three diagnostic study groups: asthma, headache, or healthy control.

Subjects were recruited from the local (Kirksville, Mo) community via solicitation by electronic mail and word of mouth. To participate in the study, subjects had to be between the ages of 18 and 75 years. In addition, subjects must have been diagnosed with asthma *or* had recurrent headaches at least twice per month for more than 3 months *or* had no symptoms or diseases. Potential participants were excluded from the study if they had asthma *and* recurrent headaches.

Subjects were required to undergo three head/cranial examinations and remain in a supine position for 45 to 60 minutes. Subjects were excluded from study participation if examination protocols would have caused major discomfort or exacerbation of symptoms from preexisting conditions.

Individuals interested in study participation responded by telephone, at which time they were screened using a series of questions used to determine subject eligibility for inclusion in one of the three cohorts, their hair length and hairstyle, and their availability on the designated testing dates. Screening information was recorded on forms for subjects meeting the inclusion criteria. Forms were later sorted based on subject availability, hair length, and diagnostic group.

Subjects selected for study participation were asked not to use cologne, perfume, or hair styling products (eg, gel, spray, mousse) on their designated testing day.

All subjects signed informed consent forms approved by the Institutional Review Board of Kirksville (Mo) College of Osteopathic Medicine-A.T. Still University (KCOM-ATSU) and completed medical history questionnaires before the physical examination process.

Examiners

Two DO examiners, each certified by the American Osteopathic Board of Neuromusculoskeletal Medicine in neuromusculoskeletal and osteopathic manipulative medicine, were recruited from the KCOM-ATSU Department of Osteopathic Manipulative Medicine. The first examiner (B.F.D.) had more than 14 years of clinical experience in osteopathic manipulative treatment (OMT) and had completed eight accredited 40-hour courses in osteopathy in the cranial field. The second examiner (K.T.S.) had more than 6 years of clinical experience in OMT and had completed seven accredited 40-hour courses in osteopathy in the cranial field.

Variables

As previously indicated, cranial examinations for each subject consisted of evaluating CRI rate, cranial strain pattern, and quadrants of restriction.

The CRI rate was measured in cycles per minute (cpm). One CRI cycle was defined as starting just as the flexion phase began (ie, after the completion of the extension phase) and ending with the completion of the extension phase.⁶ The CRI rate was measured using the following procedure:

1. The individual serving as the data recorder started a 60-second timer and stated, "Start."
2. After 60 seconds, the data recorder stated, "Stop."
3. The examiner verbally indicated the CRI rate.
4. The data recorder categorized the value into the currently accepted norms of low rate (0-7 cpm), normal rate (8-14 cpm), or high rate (>15 cpm).¹³⁻¹⁵

Diagnoses of cranial strain patterns consisted of palpatory tests for the following patterns:

- flexion
- extension
- torsion (left or right)
- sidebending rotation (left or right)
- lateral strain (left or right)
- vertical strain (superior or inferior)
- compression
- no strain^{13,14,16,17}

These palpatory patterns are commonly found in cranial osteopathic examinations. Examiners were instructed to identify the single most significant strain pattern found in each examination.

Quadrants of restriction are defined by the intersection of the cranium's sagittal and coronal planes. Transected by these planes, the cranium can be viewed as consisting of left and right anterior and posterior quadrants.¹³ Examiners were instructed to identify the quadrant(s) associated with any observed restricted motion.

Because the examiners each had their own preferred method of palpation, they were allowed discretion as to whether they kept their eyes open or closed during subject evaluations. The integrity of examiner blinding was not affected by this variable because a physical barrier (opaque sheets hung from the ceiling) obstructed both examiners' view of the subjects.

Procedure

Of the 48 subjects enrolled, each examiner evaluated 24 subjects, approximately 8 subjects per diagnostic group, with each subject evaluated three times. Because examiner blinding was an essential requirement of the present study, the 24 subjects for each examiner were subdivided into groups based on hair length and hairstyle. Hair length was one of the few features that blinded examiners could easily identify with their hands. Thus, subjects with long hair were examined separately as a group—as were subjects with moderate and short hair lengths. This division of subjects limited each examiner's ability to recall previous findings of a subject based on this physical (or palpatory) cue. Eliminating other identifiers, such as perfumes, hair styling products, and jewelry improved examiner blinding by removing other potentially identifying characteristics.

In the examination room, four identical treatment tables were arranged in a square so that the heads of the tables faced toward each other. This arrangement allowed for extra space for the examiner and a data recorder to move easily from one table to the next. A booth-like enclosure around the examination area was made of opaque sheets hung from the ceiling. Slits were cut horizontally in the sheets at the level of the treatment tables to allow examiners to reach subjects. A strip of tape was placed on each table to mark the spot where the subject's head was to be positioned.

A standard rolling office chair with an adjustable height setting was used to roll the examiner from one table to the next in the examination room. Between examinations, the examiner (while keeping his or her eyes closed) was rolled to the center of the booth by the data recorder before being moved to a different subject. This procedure was used to blind the examiners to their physical environment and to decrease the possibility of examiners inadvertently unblinding themselves. Examiner blinding was further enhanced by using overhead music and balanced room lighting to eliminate other potential environmental reference cues.

Before each testing session, subjects were gathered outside the examination area. Consent forms were signed, and medical history questionnaires were completed. All subjects were instructed to remove any glasses, earrings, and necklaces they were wearing before entering the examination room. Each subject's hairstyle was inspected to ensure that it was uniform so no physical clues were provided to the examiner. Subjects were instructed to enter the examination room quietly and to remain as still as possible throughout the testing session. Subjects were given a 5-minute rest period in the supine position before testing began to allow their bodies to reach a state of equilibrium.

The sequence of the testing procedure was as follows. The data recorder accompanied the examiner into the booth and sealed the entrance. The examiner was seated in a rolling office chair and placed in the center of the booth. The subjects then entered the room and were asked to lie in the supine position on a treatment table. Each subject's position was checked to ensure that his or her head was at the standard distance from the end of the table. Next, the data recorder selected a subject and rolled the examiner to that subject's treatment table. The examiner then placed his or her hands through the slit in the sheet to make contact with the subject's head. The examiner conducted the evaluation and reported the results to the data recorder, who then rolled the examiner to a different treatment table. This process was repeated until each subject was evaluated three times, so that the diagnostic procedures for CRI rate, cranial strain pattern, and quadrants of restriction were conducted during a period of 45 to 60 minutes. The examiners had no access to the subjects' result forms at any time.

Statistical Analysis

Generalized κ coefficients were used to quantify the intraobserver reliability obtained for CRI rate, cranial strain patterns, and quadrants of restriction over and above chance agreement. The nomenclature for describing the level of reliability associated with a specific value of κ , as presented by Landis and Koch,¹⁸ is as follows:

- less than 0.00, poor
- 0.00 to 0.20, slight
- 0.21 to 0.40, fair
- 0.41 to 0.60, moderate
- 0.61 to 0.80, substantial
- 0.81 to 1.00, almost perfect

A κ coefficient greater than 0.60 (ie, at least in the "substantial" category) was the desired outcome to establish acceptable reliability. Logistic regression models were fit to test for differences between diagnostic groups in the probability of agreement between the findings for two evaluations. Statistical significance was defined as $P < .05$.

Results

Of the 48 subjects recruited to participate in the present study, 37 (77%) were women with a mean (SD) age of 37 (12) years. The majority of subjects (44 [92%]) were Caucasian with the remaining 4 subjects (8%) being Asian. By design, subjects were almost evenly distributed between the three diagnostic groups: asthma (16 [33%]), headache (17 [35%]), and healthy control (15 [31%]). Regarding hair length, there were 19 subjects (40%) with long hair, 14 subjects (29%) with medium-length hair, and 15 subjects (31%) with short hair. Most subjects (41 [85%]) had straight hair, while the remaining 7 subjects (15%) had wavy hair.

Table 1 provides a summary of the prevalence of the overall findings for CRI, cranial strain patterns, and quadrants of restriction as well as the overall κ values for intraobserver reliability and associated percent agreements.

A majority of subjects (31.3 [65%]) had CRI rates within the normal range (8-14 cpm). The overall κ coefficient for evaluations of subjects' CRI was 0.23, indicating fair intraobserver reliability.

For cranial strain patterns, 14.7 (31%) of subjects had no observed cranial strain pattern. The most commonly observed cranial strain patterns in subjects were sidebending rotation-right (13.3 [28%]), torsion-right (8 [17%]), and torsion-left (5 [10%]). The overall κ coefficient for measurements of cranial strain patterns was 0.67, indicating substantial intraobserver reliability in identifying cranial strain patterns.

Restriction was most prevalent in the right anterior quadrant (25.7 [54%]), followed by the right posterior quadrant (21.7 [45%]), the left posterior quadrant (16.7 [35%]), and the left anterior quadrant (10.7 [27%]). The overall κ coefficient was 0.52 for measurements of the left anterior quadrant of restriction,

Table 1
Cranial Strain Patterns: Overall Intraobserver Reliability of Palpatory Diagnosis (N=48)*

Diagnosis	Prevalence, No. (%)	κ Statistic (95% CI)	Reliability Level	Percent Agreement
■ Cranial rhythmic impulse		0.23 (0.04-0.43)	Fair	64
□ Low (0-7 cpm)	15.3 (32)			
□ Normal (8-14 cpm)	31.3 (65)			
□ High (≥ 15 cpm)	1.3 (3)			
■ Cranial strain patterns		0.67 (0.58-0.76)	Substantial	74
□ No strain pattern	14.7 (31)			
□ Flexion	0 (0)			
□ Extension	0 (0)			
□ Torsion				
– Left	5 (10)			
– Right	8 (17)			
□ Sidebending rotation				
– Left	2.7 (5)			
– Right	13.3 (28)			
□ Lateral strain				
– Left	1 (2)			
– Right	1.3 (3)			
□ Vertical strain				
– Superior	0 (0)			
– Inferior	0 (0)			
□ Compression	2 (4)			
■ Quadrants of restriction				
□ Anterior, restricted				
– Left	10.7 (22)	0.52 (0.25-0.79)	Moderate	83
– Right	25.7 (54)	0.44 (0.28-0.61)	Moderate	72
□ Posterior, restricted				
– Left	16.7 (35)	0.33 (0.13-0.52)	Fair	69
– Right	21.7 (45)	0.50 (0.33-0.66)	Moderate	75

* Prevalence estimates are calculated as the number of evaluations with a specific palpatory diagnosis divided by 3 (ie, the number of evaluations per subject) and are reported to one decimal place. In addition, the nomenclature for describing the level of reliability associated with a specific value of the kappa (κ) statistic, as presented by Landis and Koch,¹⁸ is as follows: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; 0.81-1.00, almost perfect.

Abbreviation: CI, confidence interval; cpm, cycles per minute.

0.44 for the right anterior quadrant of restriction, and 0.50 for the right posterior quadrant of restriction—all indicating moderate intraobserver reliability. For the left posterior quadrant of restriction, the overall κ coefficient was 0.33, indicating fair intraobserver reliability.

The remaining tables provide summaries of the prevalence of the findings in individual diagnostic groups for CRI (Table 2), cranial strain patterns (Table 3), and quadrants of restriction (Table 4) along with κ coefficients in each diagnostic group for intraobserver reliability and associated percent agreements.

For measurements of CRI rate, the κ coefficients ranged from 0.10 (asthma group) to 0.29 (control group), indicating slight to fair intraobserver reliability (Table 2). No statistically significant difference was found in intraobserver reliability between individual diagnostic groups.

For diagnoses of cranial strain patterns, the κ coefficients ranged from 0.52 (asthma group) to 0.82 (control group), indicating moderate to almost perfect intraobserver reliability (Table 3). A significant difference ($P=.04$) was found for intraobserver reliability between the individual diagnostic groups, with diagnoses of cranial strain patterns for the control group having greater reliability than those for the asthma group.

For assessment of the left anterior quadrant of restriction, the κ coefficients ranged from 0.37 (asthma group) to 0.61 (control group), indicating fair to substantial intraobserver reliability (Table 4). For assessment of the right anterior quadrant of restriction, the κ coefficients ranged from 0.33 (asthma group) to 0.52 (headache group), indicating fair to moderate intraobserver reliability. For assessment of the left posterior quadrant of restriction, the κ coefficients ranged from 0.29

Table 2
Rate of Cranial Rhythmic Impulse: Intraobserver Reliability of Palpatory Diagnosis by Study Group (N=48)*

Study Group†	Prevalence, No. (%)	κ Statistic (95% CI)	Reliability Level	Percent Agreement
■ Asthma (n=16)		0.10 (-0.27 to 0.46)	Slight	58
□ Low (0-7 cpm)	4 (25)			
□ Normal (8-14 cpm)	11 (69)			
□ High (≥15 cpm)	1 (6)			
■ Headache (n=17)		0.23 (-0.16 to 0.62)	Fair	69
□ Low (0-7 cpm)	4.3 (25)			
□ Normal (8-14 cpm)	12.3 (72)			
□ High (≥15 cpm)	0.3 (3)			
■ Control (n=15)		0.29 (-0.01 to 0.58)	Fair	64
□ Low (0-7 cpm)	7 (47)			
□ Normal (8-14 cpm)	8 (53)			
□ High (≥15 cpm)	0 (0)			

* Prevalence estimates are calculated as the number of evaluations with a specific palpatory diagnosis divided by 3 (ie, the number of evaluations per subject) and are reported to one decimal place. In addition, the nomenclature for describing the level of reliability associated with a specific value of the kappa (κ) statistic, as presented by Landis and Koch,¹⁸ is as follows: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; 0.81-1.00, almost perfect.
 † P=.57, obtained from a logistic regression model comparing diagnostic groups on the probability of agreement between two evaluations of the cranial rhythmic impulse rate.

Abbreviation: CI, confidence interval; cpm, cycles per minute.

(asthma group) to 0.34 (headache and control groups), indicating fair intraobserver reliability. For assessment of the right posterior quadrant of restriction, the κ coefficients ranged from 0.43 (control group) to 0.52 (headache group), indicating moderate intraobserver reliability. No significant difference in intraobserver reliability was observed between individual diagnostic groups for any of the quadrants of restriction.

Comment
Cranial Strain Patterns

The overall results for evaluations of intraobserver reliability of cranial strain patterns for all three diagnostic groups indicate that our goal of substantial intraobserver reliability (κ>0.60) can be obtained. Since higher levels of intraobserver reliability are assumed to be more easily achieved than interobserver reliability,¹ we raised the standard level for acceptable reliability in this study from κ>0.40 to κ>0.60. However, because no standard for an acceptable κ coefficient for intraobserver reliability has been established, the apparently positive results of the present study should be interpreted with caution.

When cranial strain pattern results were analyzed by diagnostic group, the intraobserver reliability achieved for the headache and control groups met our goal, but those of the asthma group did not. The difference in intraobserver reliability between the control and asthma groups was statistically significant (P=.04). This finding may be the result of type I error, or it may be related to the inclusion in the asthma

group of participants who had mild, exercise-induced cases of asthma—inclusions that may have limited the diversity of somatic dysfunction within that cohort. We selected asthma for one of our diagnostic criteria because it is characterized by a distinct cranial strain pattern of chronic extension of sphenobasilar symphysis (SBS).¹³ However, we were unable to recruit a sufficient number of subjects with severe asthma.

Further, the most common cranial strain pattern found in asthma subjects in this study was a right sidebending rotation—a diagnosis that does not fit with the proposed model of asthma as associated with SBS extension.¹³ Relaxing our criteria for inclusion of participants in the asthma group may account for the increased percentage of subjects with no observed cranial strain pattern and for increased inconclusive findings (ie, all three evaluations had different cranial strain patterns). Additional complications in recruiting participants with asthma arose from the hair-length and availability protocol requirements of the present study. Future studies with increased sample sizes and more stringent criteria for inclusion in the asthma diagnosis group would be important for elucidating the proposed link between SBS extension and asthma.

Evaluations of subjects in the control group showed almost perfect intraobserver reliability for cranial strain patterns. One possible explanation for this high level of intraobserver reliability may be related to the stability of findings within the diagnostic group. Among all three study groups, the control group had the highest percentage of subjects with

Table 3
Cranial Strain Patterns: Intraobserver Reliability of Palpatory Diagnosis by Study Group (N=48)*

Study Group†	Prevalence, No. (%)	κ Statistic (95% CI)	Reliability Level	Percent Agreement
■ Asthma (n=16)		0.52 (0.37-0.68)	Moderate	63
□ No strain pattern	4 (25)			
□ Torsion				
– Left	2.7 (16)			
– Right	2.3 (15)			
□ Sidebending rotation				
– Left	0.7 (5)			
– Right	5 (31)			
□ Lateral strain				
– Left	0.7 (4)			
– Right	0 (0)			
□ Compression	0.7 (4)			
■ Headache (n=17)		0.76 (0.52-0.83)	Substantial	75
□ No strain pattern	4.7 (28)			
□ Torsion				
– Left	2.3 (13)			
– Right	2.3 (13)			
□ Sidebending rotation				
– Left	1 (6)			
– Right	5.3 (31)			
□ Lateral strain				
– Left	0.3 (3)			
– Right	1 (6)			
□ Compression	0 (0)			
■ Control (n=15)		0.82 (0.63-1.00)	Almost perfect	87
□ No strain pattern	6 (40)			
□ Torsion				
– Left	0 (0)			
– Right	3.3 (22)			
□ Sidebending rotation				
– Left	1 (7)			
– Right	3 (20)			
□ Lateral strain				
– Left	0 (0)			
– Right	0.3 (2)			
□ Compression	1.3 (9)			

* Prevalence estimates are calculated as the number of evaluations with a specific palpatory diagnosis divided by 3 (ie, the number of evaluations per subject) and are reported to one decimal place. In addition, the nomenclature for describing the level of reliability associated with a specific value of the kappa (κ) statistic, as presented by Landis and Koch,¹⁸ is as follows: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; 0.81-1.00, almost perfect.

† P=.04, obtained from a logistic regression model comparing diagnostic groups on the probability of agreement between two evaluations of cranial strain pattern. The probability of agreement for the control group was greater than that for the asthma group.

Abbreviation: CI, confidence interval.

no observed cranial strain pattern. In addition, right torsion and right sidebending rotation were frequently diagnosed in subjects in the control group. This clustering of findings into three predominant outcomes suggests that intraobserver reliability of cranial strain pattern diagnoses can be achieved for subjects with no known major medical diagnoses or histories of trauma. Being able to diagnose subjects with no cranial strain pattern is just as important as diagnosing subjects who do have cranial strain patterns.

Subjects in the headache group had fewer discernible cranial strain patterns, with no clear dominant pattern, compared with subjects in the other two study groups. This finding may be the result of the broad inclusion criteria used for recurrent headaches, which we defined as headaches occurring at least twice per month for more than 3 months. The severity, frequency, and type of headache (eg, cluster, migraine, or tension) were not specified in the inclusion criteria, which may have contributed to a lack of clear findings in the present study. It

Table 4
Cranial Quadrants of Restriction: Intraobserver Reliability of Palpatory Diagnosis by Study Group (N=48)*

Study Group	Prevalence, No. (%)	κ Statistic (95% CI)	Reliability Level	Percent Agreement
■ Asthma (n=16)				
□ Anterior, restricted				
- Left†	4.3 (27)	0.37 (-0.04 to 0.77)	Fair	75
- Right‡	8.3 (52)	0.33 (0.05 to 0.62)	Fair	67
□ Posterior, restricted				
- Left§	6 (37)	0.29 (-0.03 to 0.61)	Fair	67
- Right//	6.7 (42)	0.49 (0.19 to 0.78)	Moderate	75
■ Headache (n=17)				
□ Anterior, restricted				
- Left†	3 (18)	0.60 (0.05 to 1.00)	Substantial	88
- Right‡	7.3 (43)	0.52 (0.24 to 0.81)	Moderate	76
□ Posterior, restricted				
- Left§	5 (29)	0.34 (-0.03 to 0.71)	Fair	73
- Right//	9.3 (55)	0.52 (0.25 to 0.80)	Moderate	76
■ Control (n=15)				
□ Anterior, restricted				
- Left†	3.3 (22)	0.61 (0.13 to 1.00)	Substantial	87
- Right‡	10 (67)	0.40 (0.04 to 0.76)	Moderate	73
□ Posterior, restricted				
- Left§	5.7 (38)	0.34 (0.01 to 0.67)	Fair	69
- Right//	5.7 (38)	0.43 (0.11 to 0.76)	Moderate	73

* All *P* values reported were obtained from a logistic regression model comparing diagnostic study groups on the probability of agreement between two evaluations. Prevalence estimates are calculated as the number of evaluations with a specific palpatory diagnosis divided by 3 (ie, the number of evaluations per subject) and are reported to one decimal place. The nomenclature for describing the level of reliability associated with a specific value of the kappa (κ) statistic, as presented by Landis and Koch,¹⁸ is as follows: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; 0.81-1.00, almost perfect.

† *P*=.17
‡ *P*=.54
§ *P*=.81
// *P*=.94

Abbreviation: CI, confidence interval.

might be valuable for future studies to evaluate cranial strain patterns in various subcategories of patients with headaches, including those with varying frequency rates and severity levels.

Cranial Rhythmic Impulse

The low κ values found for diagnoses of CRI rate in subjects—both in the overall results and by study group—were quite surprising. Although passive palpatory testing techniques (ie, light-touch monitoring of the head instead of active testing for each possible strain pattern) were used to minimize the potential influence of the diagnostic process on the cranial movement characteristics, changes may still have occurred due to the forces associated with repeated evaluations. The idea that palpation is an interaction between the observer and the subject—an interaction that can affect various findings in the subject—is an idea discussed in several studies of inter- and intraobserver reliability.^{6,8,19}

Another possible explanation for the low intraobserver reli-

ability in CRI diagnoses is that CRI may not be a constant physiologic phenomenon—but rather a dynamic phenomenon, like heart rate. It is likely that the heart rates of subjects decreased during the examination period because they were recumbent and resting. Likewise, CRI rates may have changed during this time because of changes in subjects' levels of consciousness. For example, a subject may have been in an awake or alert state during the first evaluation but in a resting or sleeping state during the second or third evaluations of the examination session. It was observed that at least 1 subject in every session fell asleep and began to snore. Because the snoring could have provided reference cues to the examiners and impacted blinding protocols, these subjects were gently awakened. Thus, a subject might have had three changes in his or her level of consciousness during the examination process—awake for the first evaluation, sleeping for the second evaluation, and reawakened for the third evaluation.

We are aware of no studies that have examined changes in CRI rate associated with sleep. Therefore, the existence of

CRI rate changes associated with levels of consciousness presents an interesting topic for future study. Such studies may derive benefit from keeping subjects at a constant, awake level of consciousness throughout the examination process. Moreover, such inherent problems in studies of manual medicine could be addressed by using electronically or mechanically simulated patient models.

Quadrants of Restriction

Quadrants of restriction were identified in the present study because they are a simpler, more direct method for assessing cranial somatic dysfunction rather than extrapolating a cranial strain pattern from a global assessment of cranial motion. For this reason, it was thought that intraobserver reliability might be greater for quadrants of restriction than for cranial strain patterns. However, overall and for each study group, the intraobserver reliability for identifying restricted quadrants showed mostly moderate reliability in contrast to the substantial reliability found with cranial strain patterns. One possible reason for this discrepancy might be that examiners are highly experienced at diagnosing cranial strain patterns but rarely diagnose quadrants of restriction in everyday clinical practice. Future studies that analyze the relationship between cranial strain patterns and quadrants of restriction may be better able to address this issue.

Blinding

A challenge in this line of research is the ability of researchers to maintain blinding protocols for examiners. To achieve an adequate level of scientific rigor, blinding techniques were carefully considered and adopted in the current study that would result in an environment devoid of stimuli, allowing no frame of reference for the examiner that could potentially cause unblinding. More rigorous methods of removing environmental stimuli, such as having the examiners wear surgical gloves and having the subjects wear surgical bonnets, were considered during the study design period. We concluded, however, that such methods would be too invasive for the natural diagnostic process and would likely impede the clinical observations of examiners. We also concluded that alternate forms of blinding, such as moving the subjects to different positions after each complete set of examinations (ie, after each subject in the group was examined once), would confound results by altering the state of equilibrium for subjects (ie, increasing heart rate and blood pressure).

Although the rigor of the blinding procedures in the current study was maintained between the examiners, future intraobserver reliability studies could investigate other means to further ensure examiner blinding protocols. For example, the use of silent operation, momentary-action foot switches as described in previous studies,^{5,7} when used in combination with the blinding procedures established in the present study, could further protect examiner blinding by reducing the likelihood of an examiner remembering previous diag-

nostic results during multiple data-collection intervals.

In future studies, it is important to consider increasing the number of examiners and subjects to better evaluate intraobserver reliability. Success at this level of testing for intraobserver reliability should then lead to studies of interobserver reliability, using the same examiners and palpation parameters, to determine the extent to which reliability of cranial mechanism diagnoses could be generalized in the clinical setting.

Conclusion

The overall objective of the present study was to test whether substantial intraobserver reliability ($\kappa > 0.60$) could be achieved by DOs using common palpatory tests to diagnose cranial mechanisms in subjects divided into three study groups—subjects with asthma, subjects with headaches, and healthy control subjects. Diagnoses of cranial strain patterns in the healthy control group showed almost perfect intraobserver reliability, even with two-thirds of the subjects (ie, 10 of 15) diagnosed with some type of cranial strain pattern in at least one of the three palpatory evaluations. Diagnoses of cranial strain patterns in the headache group showed substantial intraobserver reliability. Although diagnoses of cranial strain patterns in the asthma group showed only moderate reliability, the κ coefficient for this group was 0.08 from the substantial reliability goal. The moderate intraobserver reliability found in cranial strain pattern diagnoses for the asthma group may be associated with the low enrollment of subjects with severe asthma—a factor that may also be responsible for the observed statistical difference between the asthma and control groups.

The low κ coefficients observed for CRI rate and the quadrants of restriction may be related to physiologic changes in the subjects resulting from examiner-subject interaction associated with each manual examination and illustrating an inherent limitation of this type of manual medicine study.^{6,8,19} Despite this hypothesized interaction, the present study was successful in achieving substantial intraobserver reliability for diagnoses of cranial strain patterns in healthy subjects and in subjects with specified medical conditions.

Acknowledgment

The authors thank Deborah Goggin, MA, for her editorial assistance.

References

1. Haas M. The reliability of reliability [review]. *J Manipulative Physiol Ther.* 1991;14:199-208.
2. Hartman SE, Norton JM. Interexaminer reliability and cranial osteopathy [review]. *Sci Rev Altern Med.* 2002;6:23-34. Available at: <http://faculty.une.edu/com/shartman/sram.pdf>. Accessed August 25, 2008.
3. Woods JM, Woods RH. A physical finding related to psychiatric disorders. *J Am Osteopath Assoc.* 1961;60:988-993.
4. Upledger JE. The reproducibility of craniocervical examination findings: a statistical analysis. *J Am Osteopath Assoc.* 1977;76:890-899.

(continued)

ORIGINAL CONTRIBUTION

5. Moran RW, Gibbons P. Intraexaminer and interexaminer reliability for palpation of the cranial rhythmic impulse at the head and sacrum. *J Manipulative Physiol Ther.* 2001;24:183-190.
 6. Norton JM. A challenge to the concept of craniosacral interaction. *Am Acad Osteopath J.* 1996;6(4):15-21. Available at: <http://faculty.une.edu/com/jnorton/challenge.htm>. Accessed August 25, 2008.
 7. Rogers JS, Witt PL, Gross MT, Hacke JD, Genova PA. Simultaneous palpation of the craniosacral rate at the head and feet: intrarater and interrater reliability and rate comparisons. *Phys Ther.* 1998;78:1175-1185. Available at: <http://www.ptjournal.org/cgi/reprint/78/11/1175>. Accessed August 25, 2008.
 8. Hanten WP, Dawson DD, Iwata M, Seiden M, Whitten FG, Zink T. Craniosacral rhythm: reliability and relationships with cardiac and respiratory rates. *J Orthop Sports Phys Ther.* 1998;27:213-218.
 9. Fraval M. The reliability of examination findings of cranial motion. *Aust J Osteopath.* 1996;8:4-7.
 10. Fleiss JL. Measuring nominal agreement among many raters. *Psychol Bull.* 1971;76:378-382.
 11. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley & Sons; 1981.
 12. Beal MC, Goodridge JP, Johnston WL, McConnell DG. Interexaminer agreement on patient improvement after negotiated selection of tests. *J Am Osteopath Assoc.* 1980;79:432-440.
 13. Magoun HI. *Osteopathy in the Cranial Field.* 2nd ed. Fort Worth, Tex: Sutherland Cranial Teaching Foundation; 1966.
 14. Lay EM. Cranial field. In: Ward RC, ed. *Foundations for Osteopathic Medicine.* Philadelphia, Pa: Lippincott Williams & Wilkins; 1997:901-913.
 15. Lockwood MD, Degenhardt BF. Cycle-to-cycle variability attributed to the primary respiratory mechanism [published correction appears in *J Am Osteopath Assoc.* 1998;98:137]. *J Am Osteopath Assoc.* 1998;98:35-36,41-43.
 16. Rivera-Martinez S, Wells MR, Capobianco JD. A retrospective study of cranial strain patterns in patients with idiopathic Parkinson's disease. *J Am Osteopath Assoc.* 2002;102:417-422. Available at: <http://www.jaoa.org/cgi/reprint/102/8/417>. Accessed August 25, 2008.
 17. Lippincott RC, Lippincott HA. *A Manual of Cranial Technique.* Indianapolis, Ind: The Cranial Academy; 1995.
 18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.
 19. Nelson KE. The primary respiratory mechanism. *Am Acad Osteopath J.* 2002;12(4):24-33.
-