

Inter-examiner reliability of the interpretation of paraspinal thermographic pattern analysis

Barbara A. Mansholt, DC, MS¹

Robert D. Vining, DC²

Cynthia R. Long, PhD³

Christine M. Goertz, DC, PhD⁴

Introduction: A few spinal manipulation techniques use paraspinal surface thermography as an examination tool that informs clinical-decision making; however, inter-examiner reliability of this interpretation has not been reported. The purpose of this study was to report inter-examiner reliability for classifying cervical paraspinal thermographic findings.

Methods: Seventeen doctors of chiropractic self-reporting a minimum of 2 years of experience using thermography classified thermographic scans into categories (full pattern, partial +, partial, partial -, and adaptation). Kappa statistics (k) were calculated to determine inter-examiner reliability.

Results: Overall inter-examiner reliability was fair

Introduction : Quelques techniques de manipulation vertébrale utilisent la thermographie de la surface paravertébrale comme outil d'examen qui éclaire la prise de décision clinique; cependant, la fiabilité entre examinateurs de cette interprétation n'a pas été rapportée. L'objectif de cette étude était de rapporter la fiabilité entre les examinateurs pour classer les trouvailles thermographiques cervicales paravertébrales.

Méthodologie : Dix-sept chiropraticiens déclarant avoir un minimum de deux ans d'expérience de pratique de la thermographie ont classé les examens thermographiques en catégories (motif complet, partiel +, partiel - et adaptation). Les statistiques Kappa (k) ont été calculées pour déterminer la fiabilité entre les examinateurs.

Résultats : Dans l'ensemble, la fiabilité entre les

¹ Associate Professor, Clinic, Palmer College of Chiropractic barbara.mansholt@palmer.edu

² Associate Professor, Senior Research Clinician, Palmer Center for Chiropractic Research, Palmer College of Chiropractic robert.vining@palmer.edu

³ Professor, Director of Research, Palmer Center for Chiropractic Research, Palmer College of Chiropractic LONG_C@palmer.edu

⁴ Vice Chancellor for Research and Health Policy, Palmer College of Chiropractic christine.goertz@palmer.edu

Corresponding Author: Barbara A. Mansholt, DC, MS, Associate Professor, Clinic

Palmer College of Chiropractic, 1000 Brady St., Davenport, IA 52803

563-884-5542

barbara.mansholt@palmer.edu

Disclaimers: No conflicts of interest and no disclaimers to declare.

Funding: This study uses paraspinal thermography readings obtained in a project conducted under a developmental center award from the National Institutes of Health (NIH) 1U19AT004663-01 NIH/NCCAM.

© JCCA 2015

($k=0.43$). There was good agreement for identifying full pattern ($k=0.73$) and fair agreement for adaptation ($k=0.55$). Poor agreement was noted in partial categories ($k=0.05-0.22$).

Conclusion: Inter-examiner reliability demonstrated fair to good agreement for identifying comparable (full pattern) and disparate (adaptation) thermographic findings; agreement was poor for those with moderate similarity (partial). Further research is needed to determine whether thermographic findings should be used in clinical decision-making for spinal manipulation.

(JCCA 2015; 59(2):157-164)

KEY WORDS: chiropractic, technique, instrumentation, diagnosis

Introduction

Doctors of chiropractic (DCs) use complex clinical decision-making when determining where, when, and when not to perform spinal manipulation.¹ Factors considered may include the diagnosis, symptom severity, presence of co-morbid conditions, patient preferences, and other examination findings^{2,3} such as static or segmental motion palpation,^{4,5} posture analysis, leg length analysis,⁶ biomechanical interpretation of spinal radiographs, the presence of spinal/paraspinal tenderness, and abnormal muscle tone.⁷ Some chiropractic spinal manipulation techniques, particularly those focusing on upper cervical manipulation, use thermographic and other diagnostic instruments to provide primary information to determine whether treatment should or should not occur.⁸ The use of unique diagnostic instrumentation is not new to the chiropractic profession. B.J. Palmer, considered the “developer” of chiropractic, used an instrument called the electroencephaloneuromyotomograph and later, the neurocalometer.⁹ The neurocalometer was the predecessor of the current nervo-scope, which is still used by some practitioners using the Gonstead technique system.^{10,11}

A few studies indicate that there may be some potential

examineurs était passable ($k=0,43$). Il y avait une bonne entente dans l'identification des motifs complets ($k=0,73$) et une entente passable concernant l'adaptation ($k=0,55$). Une mauvaise entente a été notée dans les catégories partielles ($k=0,05-0,22$).

Conclusion : La fiabilité entre les examinateurs a indiqué une entente passable à bonne concernant l'identification de conclusions thermographiques comparables (motifs complets) et disparates (adaptation); l'entente était mauvaise pour ceux avec une similarité modérée (partiel). Des études plus poussées sont nécessaires pour déterminer si les conclusions thermographiques peuvent être utilisées dans la prise de décision clinique concernant la manipulation vertébrale.

(JCCA 2015; 59(2):157-164)

MOTS CLÉS : chiropratique, technique, appareillage, diagnostic

for thermography to provide information suggestive of underlying physiological processes¹² that may help inform spinal manipulation decisions.¹³ Roy reported changes in paraspinal surface temperature (comparing one side to the opposite side) using infrared thermographic methods following spinal manipulation.^{14,15} These findings suggest that paraspinal cutaneous and/or subcutaneous blood perfusion may be altered following spinal manipulation. However, without further study, it is unclear if these findings represent specific physiological mechanisms initiated by the manipulation or simply from normal changes over time or from tissue perturbation.

One application of thermography used by some practitioners, referred to as “pattern analysis,” is the interpretation of a series of skin surface temperature recordings obtained over the cervical spine. Similar thermographic findings obtained several hours apart are thought to suggest spinal dysfunction, which presumably contributes to a diminished autonomic response to external environmental cues resulting in muted adaptive changes in cutaneous and/or subcutaneous blood flow.⁸ The theory behind this interpretation is based on the following physiological principles: 1) skin temperature can serve

as an indirect gauge of autonomic function,¹⁶ 2) small variations in skin temperature over time suggest that the autonomic nervous system is appropriately functioning by adapting to an ever-changing environment,^{17,18} and 3) normal or abnormal environmental adaptation can be estimated by comparing sequential skin temperature measurements.

When an individual is “adapting” to their surroundings, it is assumed that autonomic mechanisms are functioning normally, resulting in subcutaneous blood flow change over time, detectable with thermography.¹⁹ Abnormal spinal function (vertebral subluxation complex) is thought to adversely impact spinal joints and neurological function, and initiate a compensatory neurovascular response(s) causing joint motion restriction, muscle contraction, vasomotor changes, and localized tenderness.^{20,21} Due to an impaired ability to maintain homeostasis over nearby spinal regions, these changes can potentially result in static thermographic findings.

A paraspinal thermographic “pattern” is determined when multiple scans, obtained over a period of several hours, reveal similar or identical temperature findings. When this occurs, autonomic malfunction is assumed to be caused by upper cervical spinal dysfunction and manipulative treatment is then considered appropriate. If few or no similarities exist between thermographic scans over several hours, a static pattern cannot be designated and treatment is usually considered unnecessary.²²

Establishing the reliability of a measurement tool is a necessary first step in determining whether information gained from its use can be used in a clinically meaningful way. Several authors have reported on the reliability of surface thermography in a chiropractic setting.^{18,19,23-26} Though thermography appears to reliably measure temperature, stable readings are dependent on strict environmental control^{27,28} and a single paraspinal measurement procedure has not yet been extensively tested, leading a recent systematic review to conclude that evidence is unfavorable for paraspinal skin temperature to be used to locate the site of manipulation.²⁹ However, the literature does not yet adequately address whether paraspinal skin temperature readings can inform a clinician regarding the need for spinal manipulation. Before that question can be logically answered, it is first necessary to determine if clinicians are able to interpret paraspinal thermographic findings consistently. In other words, what is

the inter-examiner reliability with respect to interpreting paraspinal thermographic findings?

The purpose of this study was to determine the inter-examiner reliability of interpreting paraspinal thermographic findings. Study findings are needed to help determine whether thermography can be a tool that informs clinical decision-making for spinal manipulation and to provide useful data to chiropractic educational institutions and practitioners seeking information that further informs evidence-based clinical practice.

Methods

Institutional review board approval for this project occurred in June 2011 through Palmer College of Chiropractic, IRB Assurance # X2011-6-15-M. The use of de-identified study data was determined exempt according to 45 CFR 46.101(b)(4); informed consent was obtained from the DCs who participated. The study was conducted from August of 2011 through January of 2012. This study complies with reporting standards as recommended by the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).³⁰

This study used thermographic scans obtained in a separate clinical trial conducted to determine the effectiveness of upper cervical chiropractic manipulation on stage 1 hypertensive patients during February through June of 2010, NCT 01020435.³¹ Paraspinal cervical scans were performed using the Tytron C-3000 (Titronics Research & Development, Oxford, Iowa), as follows: 1) participants were instructed to avoid caffeine for 2 hours and tobacco for 4 hours prior to assessment; 2) upon arrival to the study site, participants acclimated in a room maintained at 70-75 degrees for approximately 15 minutes; 3) during the scan, participants sat with their head flexed slightly to allow exposure of the cervical area with the feet flat and hands resting on the thighs; and 4) the examiner moved hair away from the posterior neck (when present) with one hand, held the paraspinal thermographic scanning instrument with the other hand, and obtained measurements between the vertebral prominence (T1 area) and the base of the occiput. The entire procedure lasted approximately 30 seconds. The resulting scan image appeared on a computer screen and consisted of 3 lines. The left line (or channel) represented the temperature gradient on the left paraspinal region from T1 to occiput; the right line (or channel) represented the temperature gradient on the right paraspinal

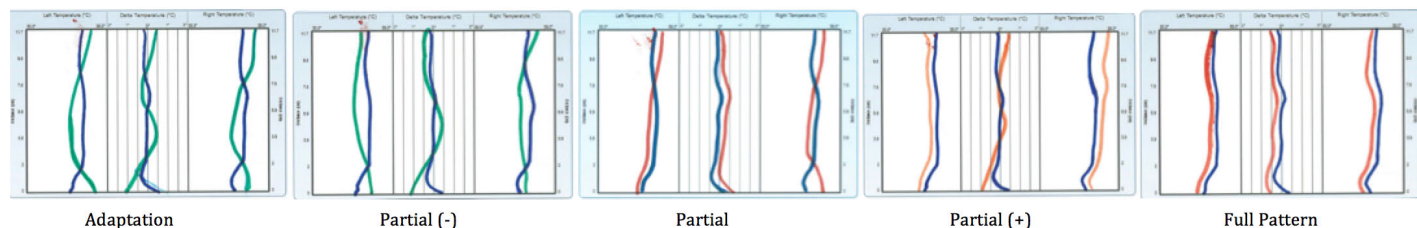


Figure 1.

Thermographic scans.

Lines in each column represent temperature readings over the cervical spine. Left column = left cervical spine region, Right column = right cervical spine region, Center column = average of left and right readings. Blue lines represent a static thermographic reading “pattern” obtained over the cervical spine (established by more than 1 reading over a ≥ 24 hour period) and overlaid with a current reading represented by green, red, or orange lines. Categories are based on subjectively comparing a patient’s designated “pattern” (blue lines) with current findings (green, red, or orange lines). Examples of scans representing categories used in this study are displayed:

Adaptation = completely dissimilar, Partial (-) = modest similarity, Partial = moderate similarity, Partial (+) = mostly similar, Full Pattern = virtually identical.

al region; and the center line (Delta) graphically displayed the difference between the right and left readings.

Prior to recruiting DC participants, de-identified scan pairings (2 scans from a single participant with at least 24 hours between scans), viewable on a computer monitor, were randomly selected. The final set included 17 scan pairings, which DC participants reviewed and classified.

Participant recruitment and eligibility

DCs self-reporting a minimum of 2 years of experience working with the Tytron software and using pattern analysis as a primary treatment indicator on a majority of their patients were eligible for this study. DCs were recruited at a chiropractic college event during a technique review class – a class that emphasizes the theories and application of thermography and pattern analysis. Knowledge of the study spread by word of mouth, and additional DCs volunteered over a period of six (6) months. Chiropractic college faculty DCs involved in teaching or research of pattern analysis or Tytron software were eligible if they met the above criteria. Basic demographic information was collected to determine eligibility.

Participant interpretation of scans

Interested DCs completed the basic demographic survey to determine eligibility. When eligibility was confirmed,

interested DC participants signed an informed consent document. DC participants were instructed to classify each scan pairing (left channel readings, delta readings, and right channel readings) into one of the following categories (see Figure 1):

1. Pattern: 3 lines are the same
2. Partial (+): 2 lines are the same and the 3rd line is similar
3. Partial: 2 lines are the same
4. Partial (-): 1 line is the same
5. Adaptation: 3 lines are different

Participating DCs either met in person or corresponded via e-mail and telephone with the lead author (BAM). All were provided the study objectives, instructions for participants and categorical classification definitions. If participants completed the study in person, they were guided through the Tytron software, viewed the scan pairings on the Tytron software, and categorized the scan pairings on the data collection form. The remainder of DCs received an Adobe® PDF file of scan pairings with written instructions and the scan analysis data collection form. DC participants designated each scan into one of five categories, and returned the data collection form via e-mail. Each participant viewed 17 unique scan pairings.

Data Entry and Analysis

Both the scan pairings and the DC raters were samples of convenience. The data were double key-entered and exported to and analyzed in SPSS for Windows (Version 17.0.0, SPSS, Inc. Somers, NY). The multi-rater un-weighted Kappa statistic³² and 95% confidence intervals based on Fleiss' corrected standard error³³ were calculated overall and for each of the 5 categories. Because SPSS does not calculate Kappa and associated confidence intervals for the multi-rater case, we used a publicly available SPSS macro.³⁴ Kappa statistics (k) were interpreted according to Fleiss: $k > 0.75$ was considered excellent, $0.40 \leq k \leq 0.75$ was fair to good agreement, and $k < 0.40$ was poor or less than expected by chance.³⁵

Results

Seventeen DCs participated in the study, reporting use of the Tytron software a mean of 7.7 years (SD 4.5). Five DCs viewed scan pairings in person; 12 viewed scan pairings and returned the scan analysis form via e-mail. DCs reported using Tytron analysis as a primary clinical decision-making indicator on a mean of 82% of patients. While practicing DCs used various spinal manipulative techniques, 14 primarily focused their treatment on the upper cervical region, 7 of whom reported using upper cervical manipulative procedures exclusively. Five DCs held chiropractic college faculty positions (Table 1).

Overall inter-examiner reliability was fair, $k=0.43$ (95% CI 0.38, 0.47) (Table 2). Reliability coefficients were highest for the individual categories of full pattern ($k=0.73$) and adaptation ($k=0.55$), and lowest for partial pattern ($k=0.05$).

Discussion

To our knowledge, this is the first study investigating inter-examiner reliability of interpreting thermographic pattern scans as taught and practiced by a few chiropractic techniques (e.g., Toggle Recoil or Blair) focused exclusively on the cervical spine. Though paraspinal thermography has been studied in chiropractic settings, strong evidence demonstrating how it can be best used clinically is currently lacking, in part because of wide variations in how these findings are interpreted to relate to abnormal physiological states.

One method of interpretation compares paraspinal skin temperature at single vertebral levels from the occiput to

Table 1:

Demographics of doctors of chiropractic interpreting thermographic scan pairings (n=17).

Female/Male	3/14
Chiropractic College Faculty Status	5
Chiropractic Education	
<i>Logan College of Chiropractic</i>	2
<i>Palmer College of Chiropractic</i>	14
<i>Sherman College of Chiropractic</i>	1
Av Pts/Week †	85 (55.9)
# Techniques Used	
<i>Only 1</i>	7
Primary Technique	
<i>Upper Cervical</i>	14
<i>Palmer Package</i>	3
Other Techniques Used	
<i>Activator</i>	4
<i>Diversified</i>	5
<i>Gonstead</i>	4
<i>Palmer Package</i>	3
<i>Pierce Stillwagon</i>	2
<i>SOT</i>	4
<i>Thompson</i>	3
<i>Upper Cervical</i>	5
<i>Other</i>	2
Use of Tytron Years †	7.73 (4.5)
1° Adj Indicator	
% patients ‡	82.2 (33.5)

*counts unless otherwise noted

† mean (SD)

‡ median (IQR)

Table 2:

Kappa (k) statistics measuring inter-reliability [kappa] statistics of thermographic pattern identification.

Category	k	CI
Overall	0.43	(0.38, 0.47)
Full Pattern	0.73	(0.59, 0.87)
Partial +	0.22	(0.11, 0.33)
Partial	0.05	(-0.05, 0.15)
Partial -	0.22	(0.08, 0.37)
Adaptation	0.55	(0.34, 0.76)

the sacrum, i.e., segmental analysis. Findings potentially indicate subsurface hyperemia from abnormal physiology such as unilateral hypertonic muscle contraction or local inflammation;^{12,23} another method compares the temperature of the right and left mastoid fossa (slightly anterior and inferior to the mastoid process) as an indicator of general health.³⁶ Hart investigated paraspinal thermographic patterns and thermographic mastoid fossa temperature differences with patient health perceptions. However, no definitive conclusions were reached regarding a relationship between mastoid fossa temperature and health perceptions.³⁶⁻³⁸ Hart also recently proposed a statistical approach to tracking a patient's paraspinal thermographic mastoid fossa findings, which has not yet been validated.³⁹ Brown explored the association between mastoid fossa temperature findings and paraspinal thermographic patterns, concluding that mastoid fossa asymmetry does not necessarily co-exist with paraspinal thermographic patterns.⁴⁰ Roy identified statistically significant temperature changes at the L5 vertebral level after a lumbar side posture manipulation when compared to a sham treatment.¹⁵

Thermographic pattern interpretation differs significantly from "segmental" analysis because it assumes the ability to adapt to a changing environment (homeostasis) will result in disparate sequential time-delayed findings. According to this theory, these differences suggest normal physiological function and thus, no need for treatment. A patient's "pattern" is established when multiple scans, obtained over a period of several hours, reveal similar or identical temperature findings. Subsequent readings are compared to this "thumbprint" pattern to determine the need for additional treatment. If completely similar, the patient is considered to be non-adapting, and treatment to the upper cervical spine is indicated (see example "pattern," Figure 1). If completely dissimilar, no treatment is indicated (see example, "adaptation," Figure 1). Partial categories are defined to clarify readings on the continuum between the two clear readings – perhaps where many clinical presentations fall. When a "partial" reading appears closer to the patient's pattern (but not completely similar), a practitioner may rely on a few additional clinical findings (static or motion palpation findings, postural abnormalities, tenderness, or muscle tone) to determine the need for treatment. Conversely, when a "partial" reading appears closer to adaptation, a practitioner may deter-

mine that no treatment is needed unless significant other clinical findings are present. Note "partial +," "partial," and "partial -," in Figure 1. This study found inter-examiner reliability for identifying or interpreting "partial" patterns to be very poor. Thus we recommend reducing to three categories (pattern, partial, and adaptation).

If reliability regarding this interpretation classification system is established, further investigation is needed regarding its validity. With this pattern interpretation theory, a patient's adjustment is considered "successful" if the consistent static readings (pattern) begin to change after treatment. Future studies should focus on whether pattern readings do change after treatment, as well as whether pattern vs. adaptation readings correlate with patient outcomes.

Thermography provides relatively reliable and objective information compared with other measures used in a clinical exam such as motion palpation.⁴ However, the results of this study indicate that there is substantial subjectivity in interpreting thermographic findings creating a challenge with utilizing the information gained in a consistent and clinically meaningful manner. Thus, largely due to the need for additional evidence, there does not appear to be a consensus on how thermographic findings should influence clinical decisions regarding spinal manipulation.

This study identified "full" or "adaptation" reliability as good and fair, respectively. If the use of this instrument in education and practice will continue, research should focus on the validity of its use. Further, clinical outcomes based on this form of clinical decision-making have not yet been reported, and more research is needed to determine if inter-examiner reliability can be enhanced (by increasing the participation, providing more rigorous standardized training, and reducing the number of category classifications) or whether clinical decisions based on this technology are associated with clinical improvement.

Limitations

This study used a convenience sample consisting of self-reported experienced DCs in the use of thermographic pattern analysis. However, there are currently no criteria other than years of experience by which to determine relative expertise. The method by which each DC viewed the scans (i.e., consecutive on PDF v. guided through software) may have had an effect on the results of their

interpretation. Future studies may want to include DCs who use this method, regardless of how often, and those who do not. Study findings from a sample size of 17 also limits the generalizability of results. Further, as the scans were performed on patients being assessed for stage 1 hypertension, it may be argued that the scans were not representative of typical chiropractic patients.

Conclusion

Overall inter-examiner reliability of thermographic findings was fair. Although the reliability of those designated as “pattern” (completely similar to a reference scan) was good, reliability of those designated as “adaptation” (completely dissimilar to a reference scan) was fair, and there was poor agreement for scans with partial similarity. These findings indicate that other clinical findings should be relied upon to determine treatment necessity. Further research is needed to better understand if treatment decisions based on thermographic findings are related to clinical outcomes.

References

1. Murphy DR, Hurwitz EL, McGovern EE. A nonsurgical approach to the management of patients with lumbar radiculopathy secondary to herniated disk: A prospective observational cohort study with follow-up. *J Manip Physiol Ther.* 2009;32(9):723-733. doi: 10.1016/j.jmpt.2009.10.007; 10.1016/j.jmpt.2009.10.007.
2. Murphy DR, Hurwitz EL. A theoretical model for the development of a diagnosis-based clinical decision rule for the management of patients with spinal pain. *BMC Musculoskelet Disord.* 2007;8:75. doi: 10.1186/1471-2474-8-75.
3. Murphy DR, Hurwitz EL. Application of a diagnosis-based clinical decision guide in patients with low back pain. *Chiropr Man Therap.* 2011;19:26-709X-19-26. doi: 10.1186/2045-709X-19-26 [doi].
4. 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. doi: 10.1016/j.jcm.2010.06.004; 10.1016/j.jcm.2010.06.004.
5. Cooperstein R, Young M, Haneline M. Interexaminer reliability of cervical motion palpation using continuous measures and rater confidence levels. *J Can Chiropr Assoc.* 2013;57(2):156-164.
6. Cooperstein R. Heuristic exploration of how leg checking procedures may lead to inappropriate sacroiliac clinical interventions. *J Chiropr Med.* 2010;9(3):146-153. doi: 10.1016/j.jcm.2010.06.003; 10.1016/j.jcm.2010.06.003.
7. Centers for Medicare and Medicaid Services/NHIC, Inc. Chiropractic billing guide. <http://www.medicarenhic.com/providers/pubs/Chiropractic%20Billing%20Guide.pdf>. Accessed 06/03, 2013.
8. Eriksen K. Upper Cervical Subluxation Complex: A Review of the Chiropractic and Medical Literature. Baltimore, MD: Lippincott Williams and Wilkins; 2004.
9. Palmer B. Chiropractic Clinical Controlled Research. Hammond, IN: WB Conkey Company; 1951.
10. Herbst RW. Chapter 11, instrumentation. In: Shi-Chi Publications, ed. *Gonstead Chiropractic Science & Art.* ; 1980:157.
11. Bergman TF, Peterson DH. Galvanic skin resistance. In: Elsevier, ed. *Chiropractic Technique: Principles and Procedures.* 2011:80.
12. Roy RA, Boucher JP, Comtois AS. Consistency of cutaneous thermal scanning measures using prone and standing protocols: A pilot study. *J Manip Physiol Ther.* 2010;33(3):238-240. doi: 10.1016/j.jmpt.2010.01.008 [doi].
13. Wu CL, Yu KL, Chuang HY, Huang MH, Chen TW, Chen CH. The application of infrared thermography in the assessment of patients with coccygodynia before and after manual therapy combined with diathermy. *J Manip Physiol Ther.* 2009;32(4):287-293. doi: 10.1016/j.jmpt.2009.03.002 [doi].
14. Roy RA, Boucher JP, Comtois AS. Effects of a manually assisted mechanical force on cutaneous temperature. *J Manip Physiol Ther.* 2008;31(3):230-236. doi: 10.1016/j.jmpt.2008.02.006.
15. Roy RA, Boucher JP, Comtois AS. Paraspinal cutaneous temperature modification after spinal manipulation at L5. *J Manip Physiol Ther.* 2010;33(4):308-314. doi: 10.1016/j.jmpt.2010.03.001.
16. Guyton AC. *Textbook of Medical Physiology.* 8th ed. Philadelphia, PA: W.B. Saunders Company; 1991.
17. Owens EF, Pennacchio VS. Operational definitions of vertebral subluxation: A case study [procedures used at Sherman College of Straight Chiropractic. *Top Clin Chiropr.* 2001;8(1):40-48.
18. Owens EF Jr, Hart JF, Donofrio JJ, Haralambous J, Mierzejewski E. Paraspinal skin temperature patterns: An interexaminer and intraexaminer reliability study. *J Manip Physiol Ther.* 2004;27(3):155-159. doi: 10.1016/j.jmpt.2003.12.019.
19. Hart J, Owens EF Jr. Stability of paraspinal thermal patterns during acclimation. *J Manip Physiol Ther.* 2004;27(2):109-117.
20. Lantz C. The vertebral subluxation complex PART 1 An introduction to the model and kinesiological component. *Chiropractic Research Journal.* 1989;1(2):23-36.
21. Lantz C. The vertebral subluxation complex PART 2: The neuropathological and myopathological components. *Chiropractic Research Journal.* 1990;1(4):19-38.

22. Strazewski J. *The Essentials of Toggle Recoil (HIO)*. Davenport, Iowa: Brandt Printing; 2010.
23. Roy R, Boucher JP, Comtois AS. Validity of infrared thermal measurements of segmental paraspinal skin surface temperature. *J Manip Physiol Ther*. 2006;29(2):150-155. doi: 10.1016/j.jmpt.2005.12.004.
24. Hart J, Omolo B, Boone WR, Brown C, Ashton A. Reliability of three methods of computer-aided thermal pattern analysis. *J Can Chiropr Assoc*. 2007;51(3):175-185.
25. Seay C, Gibbon C, Hart J. Intraexaminer and interexaminer reliability of mastoid fossa readings using a temporal artery thermometer. *J Chiropr Med*. 2007;6(2):66-69.
26. McCoy M, Campbell I, Stone P, Fedorchuk C, Wijayawardana S, Easley K. Intra-examiner and inter-examiner reproducibility of paraspinal thermography. *PLoS One*. 2011;6(2):e16535. doi: 10.1371/journal.pone.0016535.
27. Boone WR, Strange M, Trimpi J, Wills J, Hawkins C, Brickey P. Quality control in the chiropractic clinical setting utilizing thermography instrumentation as a model. *J Vert Sublux Res*. 2007;Oct(12):Online access only p. 1-6.
28. Roy RA, Boucher JP, Comtois AS. Digitized infrared segmental thermometry: Time requirements for stable recordings. *J Manip Physiol Ther*. 2006;29(6):468.e1-468.10. doi: S0161-4754(06)00160-6 [pii].
29. 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):36-709X-21-36. doi: 10.1186/2045-709X-21-36 [doi].
30. Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud*. 2011;48(6):661-671. doi: 10.1016/j.ijnurstu.2011.01.016 [doi].
31. United States National Institutes of Health. Clinical trials. gov. <https://clinicaltrials.gov/ct2/show/NCT01020435?term=01020435&rank=1>. Accessed 01/27, 2015.
32. Fleiss J. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378.
33. Fleiss J, Nee J, Landis J. Large sample variance of kappa in the case of different sets of raters. *Psychol Bul*. 1979;86:974.
34. <ftp://public.dhe.ibm.com/software/analytics/spss/support/Stats/Docs/Statistics/Macros/Mkappasc.htm>2013.
35. Fleiss J. The measurement of interrater agreement. statistical method for rates and proportions. In: New York: John Wiley and Sons, Inc., New York; 1981.
36. Hart J. Mastoid fossa temperature differentials & health perception. *J Vert Sublux Res*. 2010;Nov(14):Online access only p 1-6.
37. Hart J, Omolo B, Boone WR. Thermal patterns and health perceptions. *J Can Chiropr Assoc*. 2007;51(2):106-111.
38. Hart J. Six-minute acclimated thermal scans and health perception. *J Vert Sublux Res*. 2007;Jul(30):Online access only 5 p.
39. Hart J. Using basic statistics on the individual patient's own numeric data. *J Chiropr Med*. 2012;11(4):306-309. doi: 10.1016/j.jcm.2012.10.008 [doi].
40. Brown M, Coe A, DeBoard TD. Mastoid fossa temperature imbalances in the presence of interference patterns: A retrospective analysis of 253 cases. *J Vert Sublux Res*. 2010;Jul(15):Online access only 13 p.