PARASPINAL SKIN TEMPERATURE PATTERNS: AN INTEREXAMINER AND INTRAEXAMINER RELIABILITY STUDY

Edward F. Owens, Jr, DC,[a] John F. Hart, DC,[b] Joseph J. Donofrio, DC,[c] Jason Haralambous, DC,[d] and Eric Mierzejewski, DC[e]

ABSTRACT

Background: Paraspinal thermography is used by chiropractors as an aid in assessing the presence of vertebral subluxation. Few reliability studies have been carried out, with mixed results. Digital infrared scanning equipment is now available with location tracking that may enhance reproducibility. Digitized scans enable a computer-aided interpretation of thermographic patterns.

Objective: To assess the ability of examiners to reproduce thermal patterns.

Study Design: Repeated measures with 2 examiners assessing the same patient on 2 occasions. Thirty asymptomatic students served as subjects.

Methods: A TyTron C-3000 handheld thermographic scanner interfaced to a Microsoft Windows compatible personal computer was used for all recordings. Each examiner recorded 2 scans on each patient. It took an average of 3 minutes to complete all 4 scans. Data were exported to a spreadsheet for initial analysis, then SPSS was used for calculation of intraclass correlation coefficients (ICC). Since the starting and stopping points of scans were not always the same, care was taken to align scans visually, using well-distinguished

peaks on the charts as guides. Scans were cropped to remove artifacts that might have occurred at the beginning and end of the scans. Intraexaminer and interexaminer ICCs were calculated.

Results: Skin temperatures ranged from 35.4'C to 30.0'C over all scans. The average temperatures changed little from the first to the last scans, indicating that subjects' overall skin temperatures were stable during the scanning procedure. Intraexaminer ICCs ranged from 0.953 to 0.984. The left and right channel data show slightly higher congruence than the Delta channel. The interexaminer reliability coefficients ranged from 0.918 to 0.975. Again, the Delta channel shows slightly less reliability, although the ICCs were quite high for all channels.

Conclusion: Intraexaminer and interexaminer reliability of paraspinal thermal scans using the TyTron C-3000 were found to be very high, with ICC values between 0.91 and 0.98. Changes seen in thermal scans when properly done are most likely due to actual physiological changes rather than equipment error. (J Manipulative Physiol Ther 2004;27:155-9)

Key Indexing Terms: Chiropractic; Skin Temperature; Thermography; Assessment; Vertebral Subluxation

INTRODUCTION

Several chiropractic theories predict that the nervous system can be influenced by alteration in vertebral joint kinematics.[1] Indirect measures of neural function, including paraspinal thermography, have been used to assess the impact of vertebral subluxation on the nervous system. Thermocouple devices were used in chiropractic as early as 1924 to measure the side-to-side skin temperature difference, with the information used as a clinical indicator of the need for vertebral adjustment.[2-4]

Today, there is a wide variety of methods for assessing skin temperature and likewise a wide variety of interpretations of the information. While infrared video and liquid crystal systems can assess temperature of the whole back or

155

limbs, chiropractic techniques tend to focus on the paraspinal area only. A variety of chiropractic techniques use paraspinal thermography, including upper cervical techniques and the Gonstead technique.[4-9] Equipment, scanning, and interpretive methods vary widely with different techniques. Some methods continue to use bilateral thermocouples that make direct contact with the skin to measure differential temperature, while others use noncontact infrared sensors.

There exists an extensive literature that describes the neural control of skin temperature, offering a good theoretical basis for using thermography as a clinical assessment.[10]

Still, the reliability and validity of the assessment as performed by chiropractic clinicians has not been established. Thermography, in general, was rated as investigational to equivocal by the 1993 Mercy Conference panel,[11] citing the dearth of studies that showed reliability or sensitivity of thermography to changes in health status.

There is controversy in the literature regarding the reliability of paraspinal thermography. One study of the Gonstead system by Plaugher et al[9] showed fair to good interexaminer reliability for the Nervoscope (Electronic Development/Labs Inc, Danville, Va) device as it is used to locate segmental side-to-side temperature differences (breaks), as well as moderate to excellent intraexaminer reliability. DeBoer et al[12] specifically tested interexaminer and intraexaminer reliability of an infrared system and found very high reliability. In that study, temperature profiles were digitized and compared with the intraclass correlation coefficient (ICC) statistic to determine reliability. A study by Keating et al[13] used a dermathermograph to detect segmental differences in temperature. Temperature difference was found to be 1 of the 4 most strongly reliable measures of lumbar segmental abnormality. A follow-up to that study, however, found that side-to-side temperature difference measurement was not reliable.[14]

The advent of computer-interfaced infrared devices has made it possible to generate more objective digital analysis methods of thermographs using mathematical algorithms.[15] Still, in most cases, a human operator performs the task of collecting the temperature data by guiding the temperature-sensing device along the patient's back. The purpose of this study was to test the repeatability of the thermographic scanning procedure between and among doctors using a particular computer-interfaced thermal scanning system. This study is the first step in a series of projects aimed at developing a valid tool for evaluation of the sympathetic nervous system response as related to vertebral subluxation.

METHODS

Subject Selection

Thirty subjects (8 female subjects, 22 male subjects) between the ages of 20 and 48 (SD = 6.5) were recruited from various clinical sciences classes taught at a chiropractic college. All subjects were informed of the nature of the procedure and the confidentiality of the records obtained and they indicated their verbal consent to participate. Any student was free to decline if they so wished. All subjects were also under routine spinal care in the college health center. Students were

excluded from participation if they were particularly hirsute or had visible skin blemishes in the paraspinal area of the back.

Data Collection

Paraspinal thermographs for this project were recorded using the TyTron C-3000 Infrared thermal scanner (Titronics Research & Development, Oxford, Iowa) interfaced to a Microsoft (Microsoft Corp, Redmond, Wash) Windows-based computer. The TyTron is a handheld dual probe scanner with wheels to keep it a uniform distance from the skin during data collection. One of the wheels is equipped with a position sensor that tracks its location along the spine as the temperature is recorded. TyTron scanners are calibrated by the manufacturer to be accurate to the nearest 0.5°C and can resolve temperature differences to the nearest 0.00625°C. The reliability of the sensors themselves is within 0.2°C. The sensors are mounted at the end of collimating barrels, producing a circular area of sensitivity 0.308 inch in diameter, which remains constant in size when the barrel tip is within 0.5 inch from the skin.

The thermal scans were carried out in the clinic area of the college, in a typical room used for patient evaluation. Care was taken to avoid any direct breezes or sunlight on the subject's back, but a temperature-controlled environment was not available. These conditions mimic those that are routinely used for thermal scanning in practitioners' offices. This test was performed in the springtime, and students were scanned in the mornings during a break between classes.

For scans, the subject exposed his or her back and immediately was seated on a special chair that supports only the outer lumbar area and leaves the paraspinal area clear. For the full-spine thermal scans, the investigator first located the second sacral tubercle by palpation, then placed the scanner with the wheels in contact with the skin and the scanner barrels straddling the second sacral tubercle. The trigger button on the scanner is depressed to initiate data collection, and the scanner is moved by hand up the spine. Care is taken to follow the contours of the spine, keeping the scanner barrels an equal distance from the spinous processes and perpendicular to the skin surface and a uniform distance from the skin. The scan continues into the hair in the upper neck. The investigator uses his or her free hand to hold the hair up and out of the way and moves the scanner barrels very close to the skin to help comb the hair out of the way. The scan is stopped at the shelf of the superior nuchal line. A typical full-spine scan can be collected in 15 seconds or less.
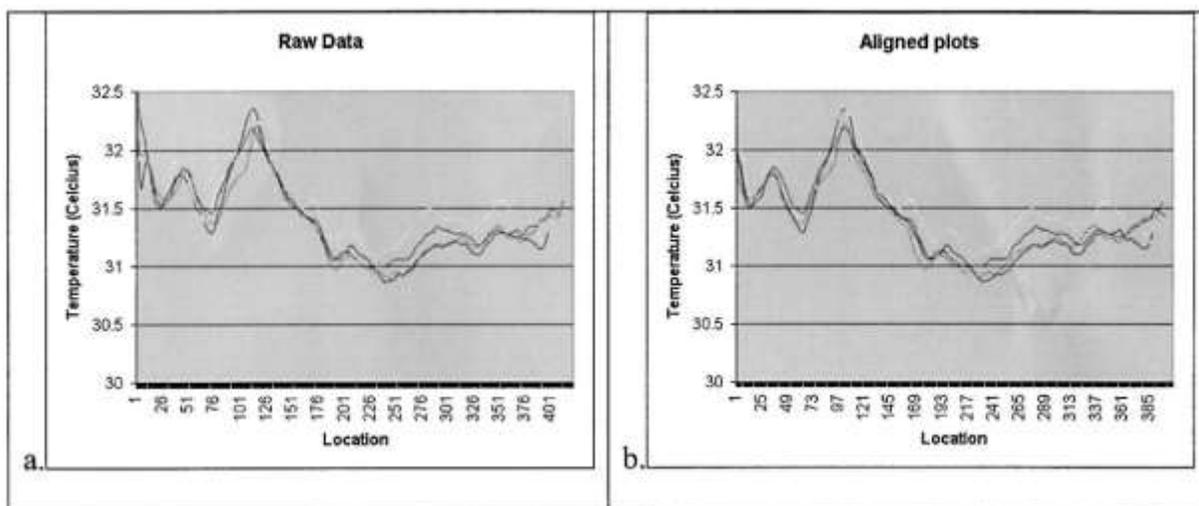
Fig 1. Line plots of temperature with respect to position for right side of one patient. All 4 scans are overlaid. Plots before (a) and after (b) cropping and alignment steps were performed.

Scans were taken in rapid succession in an attempt to capture the scans before any changes might occur on the part of the subject. The average time for completion of all 4 scans was 3 minutes; the longest scanning session took 11 minutes. In addition, the examiners avoided making any contact with the subject's skin other than the original palpation to locate the starting point and the light contact through the scanner's rollers.

Over the course of the study, 2 students and 1 faculty doctor performed the scans. For each subject, however, the faculty doctor and only 1 of the students performed the scans. The doctor and student each performed 2 scans, one immediately after the other. The order of the examiners was randomized; sometimes the doctor performed the scans first, and sometimes the student performed the scans first.

Since the temperature recording process is linked to the movement of the scanner roller, the number of data points in a scan depends on the length of the patient's spine. The scanner records data at a fixed rate as the roller tracks along the skin, 6 points per centimeter of roller travel.

In most cases, it was possible to identify characteristic peaks on the charts. Individual columns of data were shifted vertically when necessary and recharted to produce the best registration of the first major peak. Figure 1 shows a representative chart before the shifting and after.

After all channels had been shifted vertically to produce the best visible overlap, the files were cropped so that they were all the same length. In most cases, shifting the right channel by the same amount as the left channel produced good registration of the right channel as well. In one third of the files, however, right channel charts did not line up well, and in those cases, the right channel data were shifted separately from the left.

Once the left and right files had been shifted and cropped, a third channel was calculated as the difference between the right and left channel. This channel is called the Delta channel in future discussions. Descriptive statistics were performed on the data in the spreadsheet and then the data were exported to SPSS for Windows (Version 8.0.0; SPSS, Inc, Chicago, Ill.) for calculation of ICCs.

Data Analysis

Thermal data were stored in a patient database for later viewing and interpretation. For the purposes of this study, Titronics Research & Development provided us with a special export routine from their software that saves the centigrade temperature data in a comma delimited text

file. The temperature file consists of 2 columns of data, one representing the temperature of points on the left side of the spine and the other for the points on the right side of the spine.

For analysis, the 4 temperature files for each patient were imported into an Excel spreadsheet (Microsoft Corp) and the left channel data for the 4 scans was graphed in a chart.

RESULTS

Data collection resulted in 120 individual data files, which ranged in length from 451 points to 280 points (average = 372, SD = 36.7). The number of points in a data file depends on the length of the subject's spine, but there was also some variation in the length of individual patient's files, indicating variation in the starting and stopping points of the scans. Registration of the scans, by moving the data vertically within the spreadsheet, was necessary in all but 5 of the data sets. The amount of sliding varied from 0 to 23 steps (average = 5.5, SD = 5.12). This indicates that the starting points of the scans did vary from

Table 1. Intraclass correlation coefficients comparing sets of thermal scans after cropping and shifting of the data to achieve best visual alignment

	Scan sets compared	Left channel	Right channel	Delta channel
Intraexaminer	DrA1 vs DrA2	.983	.977	.953
	DrB1 vs DrB2	.982	.984	.956
Interexaminer	DrB1 vs DrA1	.970	.958	.925
	DrB1 vs DrA2	.975	.962	.918
	DrB2 vs DrA2	.972	.960	.919
	DrB2 vs DrA1	.964	.959	.920

Four scans were taken of each of 30 patients by 2 examiners. The examiners are denoted by DrA and DrB in the table, with individual scans labeled DrA1, DrA2, DrB1, and DrB2. Hence, scan sets labeled DrA1 vs DrA2 would denote the first scans performed by Dr A compared with second scans performed by DrA.

one scan to the next on the same patient. The registration and cropping step performed on the raw data eliminated differences in scan length within patients.

Average Temperatures

Skin temperatures ranged from 35.4°C to 30.0°C over all scans. The average temperature changed little from the first to the last scans, indicating that subjects' overall skin temperature was stable during the scanning procedure. Some subjects had slight increases (0.5°C) in average temperature from the first to the last scan, while others showed slight decreases.

Intraclass Correlation Coefficient Calculations

The cropped scan data from each subject were stacked into 1 large spreadsheet consisting of 12 columns (4 scans x [Left, Right, and Delta channels]) and 10,330 rows and imported into SPSS for Windows for further analysis. The intraexaminer ICC (2-way, mixed-effect model, consistency agreement) was calculated by comparing scans done by the same operator in sets of 2 (Table 1).

Similarly, interexaminer ICCs compared scans performed by different operators (Table 1). Intraexaminer ICCs ranged from 0.953 to 0.984. The left and right channel data show slightly higher congruence than the Delta channel. The interexaminer reliability coefficients ranged from 0.918 to 0.975. Again, the Delta channel shows slightly less reliability, although the ICCs are quite high for all channels.

DISCUSSION

Thermography has been used since the early days of chiropractic as an assessment for the neurological component of the vertebral subluxation, and there have been many systems of measurement and interpretation developed. There also are many claims for the clinical usefulness of the methods. However, we could find no reports in the peer-reviewed literature on the validity of thermography as a tool for subluxation analysis.

The particular method of interpretation we intend to use for thermographic analysis is the "pattern analysis"developed by B.J. Palmer and others.[3,4] Temperature pattern analysis is 1 component of the subluxation assessment system in use at Sherman College. An article by Owens and Pennacchio[16] describes the system more fully.

In pattern analysis, successive thermal plots are compared, looking for certain constant features of the temperature profile. When enough constant features are found, the patient is considered "in pattern"and most likely in a subluxated state. Hence, it is crucial to know whether changes are due to errors in data collection or actual changes in the patient's skin temperature profile. This work is a step toward developing an evidence base that might satisfy that need. The first step is reliability testing of the equipment as clinicians use it.

We have attempted to remove the operator as much as possible from the interpretation of the similarity of thermal scans. DeBoer et al[12] were the first to suggest the use of the ICC in assessing reliability of thermal scans, such as those used in this study. Generally, a diagnostic test produces only 1 or 2 measures that are fairly easy to compare for congruence. Digital thermal scanning, on the other hand, produces a set of 300 or more data points per scan that makes comparison more difficult. The scan can be thought of as a waveform with amplitude and slope features. The ICC is useful because it provides one index of reliability that takes into account both aspects of the waveform and also can test the congruence between several different operators.

In this study, we took care to align and crop data sets to provide a good visual "fit"of the scans. The rationale used is that the temperature peaks represent real hot or cold spots on the patient's back that do not move between scans. The starting and stopping points of the scan were not marked, however, and did vary from scan to scan. The peaks on the temperature plots were used as comparison points to "register"the scans to each other. In future studies, reflective or insulating markers might be affixed to the patient's back to provide more definitive locations for registration.

The aligning and cropping step of data analysis was labor intensive, somewhat subjective, and perhaps prone to operator error. Alignment was produced by sliding files so that their plots began at different points, until the most obvious first peaks in the chart lined up. Using the waveform analogy again, the phase of the plot was shifted, but the amplitude was not changed. The reliability coefficients calculated were quite high, all above 0.90. Using this method, we found the reliability of thermal scanning with the TyTron to be better than that reported by Plaugher et al[9] but in the range of the work of DeBoer et al.[12] This work, especially regarding the statistical analysis of reliability, is most similar to DeBoer et al,[12] perhaps accounting for the

similarity of findings. The results of Plaugher et al[9] were more different from ours, perhaps because they used a temperature sensor that contacts the skin and also used a different approach, based on the "break"analysis, to determine reliability. Keating et al[13] and Boline et al[14] had conflicting results in their reliability studies, presumably using the same equipment and methods, but their methods are not well enough described to judge the similarity to the current study.

We have provided data on absolute temperatures as well as side-to-side differentials. Future studies might benefit from an automated process using a software application to search for the best fit by sliding the scans in increments.

One unexpected finding was the need to align the left and right side temperature scans separately. In two thirds of the files, aligning the left channel produced good alignment of the right channel as well, but in the rest, further alignment was needed. Apparently, there was some technical error in one third of our scans that caused either the left or right channel data collection to lag behind. Such an artifact could be due to holding the scanning head at an improper angle, where 1 probe leads the other. Care should be taken during scans to avoid twisting the scanning gun if precision results are desired.

Shifting the left and right channels with respect to each other has a dramatic effect on the Delta channel, which is the calculated side-to-side difference between the channels. In our study, we recalculated the Delta channel after the best alignment was achieved. Care should be taken in interpreting changes that are seen in the Delta channel as calculated by the TyTron C-3000 software. The changes could be due to changes in actual patient temperature patterns, but they could also be produced by tilting the scanner head during the scan. Changes seen in the individual left or right channel temperature profiles would not be prone to this error.

Digital thermal scanning with the TyTron C-3000 appears to be reliable enough for further clinical testing. Thermal scans are stable in the short-term (3 to 10 minutes), but we do not yet know how they fluctuate over longer time periods or in response to care. Other work has been done in this lab to look at changes that occur in thermal profiles over a 31-minute period.[17] The eventual goal is to develop a system of paraspinal skin temperature assessment and interpretation that will serve as a valid and reliable tool to detect the presence of neurological effects indicative of vertebral subluxation.

CONCLUSION

Intraexaminer and interexaminer reliability of paraspinal thermal scans using the TyTron C-3000 were found to be very high, with ICC values between 0.918 and 0.984. Intraexaminer reliability is slightly higher than interexa-
miner reliability. Changes seen in thermal scans when properly done are most likely due to actual physiological
changes rather than equipment error.

REFERENCES

1. Leach RA. Facilitation hypothesis in the chiropractic theories: principles and clinical applications. 3rd ed. Baltimore: Williams & Wilkins; 1994. p. 89-119.
2. Rademacher WJ. A premise for instrumentation. Chiropr Tech 1994;6:84-94.
3. Palmer BJ. Chiropractic clinical controlled research. Vol. XXV. Hammond (IN): W.B. Conkey Co; 1951. p. 587.
4. Duff SA. Chiropractic clinical research, interpretation of spinal bilateral skin temperature differentials. San Francisco: Paragon Printing; 1976. p. vi-vii.
5. Kessinger RC, Boneva DV. Bell's palsy and the upper cervical spine. Chiropr Res J 1999;6:47-56.
6. Kessinger RC, Boneva DV. Case study: acceleration/deceleration injury with angular kyposis. J Manipulative Physiol Ther 2000;23:279-87.

7.   Kessinger RC, Boneva DV. Vertigo, tinnitus and hearing loss in the geriatric patient. J Manipulative Physiol Ther 2000;23: 352-62.

8.   Kessinger RC, Boneva DV. A new approach to the upper cervical specific, knee-chest adjusting procedure: part 1. Chiropr Res J 2000;7:14-32.

9.   Plaugher G, Lopes MA, Melch PE, Cremate EE. The inter-and intra-examiner reliability of a paraspinal skin temperature differential instrument. J Manipulative Physiol Ther 1991;14: 361-7.

10.   Wallace H, Wallace J, Resh R. Advances in paraspinal thermographic analysis. Chiropr Res J 1993;2:39-54.

11.   Haldeman S, Chapman-Smith D, Peterson DM. Guidelines for chiropractic quality assurance and practice parameters. Gaithersburg (MD): Aspen Publishers; 1993. p. 146-7, 151-2.

12.   DeBoer KF, Harmon RO, Chambers R, Swank L. Inter- and intra-examiner reliability study of paraspinal infrared temperature measurements in normal students. Res Forum 1985;2: 4-12.

13.   Keating JC, Bergmann TF, Jacobs GE, Finer BA, Larson K. Interexminer reliability of eight evaluative dimensions of lumbar segmental abnormality. J Manipulative Physiol Ther 1990; 13:463-70.

14.   Boline PD, Haas M, Meyer JJ, Kassak K, Nelson C, Keating JC. Interexminer reliability of eight evaluative dimensions of lumbar segmental abnormality: part II. J Manipulative Physiol Ther 1993;16:363-74.

15.   Owens EF, Stein T. Computer-aided analysis of paraspinal thermographic patterns: a technical report. Chiropr Res J 2000;7:65-9.

16.   Owens EF, Pennacchio VS. Operational definitions of vertebral subluxation: a case study. Top Clin Chiropr 2001;8: 40-8.

17.   Hart J, Owens EF. Stability of paraspinal thermal patterns during acclimation. J Manipulative Physiol Ther 2004;27: 109-17.