CROSS-SECTIONAL VALIDITY STUDY OF COMPRESSIVE LEG CHECKING IN MEASURING ARTIFICIALLY CREATED LEG LENGTH INEQUALITY

ROBERT COOPERSTEIN, MA, DC^a, ELAINE MORSCHHAUSER, DC^b,
ANTHONY J. LISI, DC^c

^aDirector of Technique and Research, Palmer College of Chiropractic West

^bPalmer Center for Chiropractic Research

^cUniversity of Bridgeport College of Chiropractic

Submit requests for reprints to: Dr. Robert Cooperstein, Palmer College of Chiropractic West, 90 E. Tasman Dr., San Jose, CA 95134.

Paper submitted April 14, 2004.

ABSTRACT

Objective: To determine the accuracy of instrumented, prone compressive leg checking.

Design: Point measures (n=29) on single participants.

Setting: Chiropractic college research clinic.

Methods: A pair of surgical boots was modified to permit continuous measurement of leg length inequality (LLI). The accuracy of prone leg checking for a masked examiner (n = 29) was determined, against the gold standard of artificial LLI that was created by randomly inserting zero to six 1.6 mm shims in either boot. Accuracy was defined as the examiner's ability to correctly assess the change in the number and side of shims inserted, in two consecutive observations per participant. Linear regression and Bland-Altman statistics were obtained to determine the concurrent validity of compressive leg checking compared to a reference standard.

Results: The observed and artificial LLI shared 86% of their variation (n = 29). The mean examiner error was 2.7 mm and the accuracy of dichotomous short leg determination for two shim insertions was 86.2%. The 95% confidence interval for the Bland-Altman limits-of-agreement for observed vs. artificial change in LLI was (-7.6, +5.2).

Conclusions: Instrumented, compressive leg checking seems highly accurate, detecting artificial changes in leg length of 2–3 mm, and thus possesses concurrent validity assessed against artificial LLI. Pre- and post leg check differences should exceed about 4–6 mm to be highly confident a real change has occurred. It is unknown whether compressive leg checking is clinically relevant. (*J Chiropr Med* 2004;3: 91–95)

Key Words: Leg Length Inequality; Chiropractic; Validity

INTRODUCTION

Various leg length assessment procedures are commonly used by practitioners within several professions: chiropractic, physical therapy, osteopathy, podiatry, and medical orthopedics (1,2). Although variations in leg length assessment procedures seem virtually endless, the crux of the matter involves determining the relative position of the distal lower extremities (the feet), where a difference would commonly be called leg length inequality (LLI). This commonly involves assessing Y-axis positional asymmetry of the distal lower extremities in the prone or supine position. Some of the methods are manual and visual, whereas others involve measuring tools ranging from tape measures to imaging procedures. There are leg checking protocols that require that the prone patient's head be turned sequentially left and right (3,4), that the knees be flexed (3–5), that the doctor or patient touch or press upon some part of the patient's body (5), or that there be some other type of "intervention" thought capable of producing an evoked response in apparent leg length.

In spite of the ubiquity of leg checking, surprisingly little is known about either its accuracy (6) or clinical utility, the extent to which leg check results correlate with or predict pathology of any kind. Moreover, it is not clear to what extent the various leg length checking procedures detect anatomical leg length inequality (actual differences in tibial or femoral lengths), or functional LLI (apparent but not actual difference in length). Opinions notwithstanding, very little quantitative or qualitative information is available as to what precisely is being measured, what creates the putative functional short leg (7), how stable it might be (8), how it may be distinguished from an anatomical short leg, or what pathological significance a functional or a structural short leg might have.

Although the reliability of manual leg checking procedures has to some extent been determined (6), their

accuracy has been largely unstudied. In research previously reported (6), we investigated the accuracy (concurrent validity) of a leg length assessment procedure we termed compressive leg checking (9). Compressive leg checking differs from conventional prone leg checking, as typically performed, only in that it uses somewhat more cephalad force, that we measured with a soft-tissue algometer to be about three kg per leg.

In that previous study, we tested the accuracy of a single blinded examiner in assessing the side and magnitude of artificially induced LLI created by the insertion of a number of shims into the left or right shoe of the experimental participants. The examiner obtained 26 measurements on each of three participants. The results of that study suggested that prone compressive leg checking, instrumented by means of modified surgical boots that enabled millimetric measurements, is quite accurate as performed by the examiner in that study. However, the longitudinal study design did not attempt to address whether the accuracy seen with these three participants would extend to a larger and more representative study participant pool. Moreover, it is not clinically feasible to perform a leg checking procedure 26 times on an individual patient. Therefore, the goal of this study was to assess the accuracy of the examiner in making point measurements on multiple participants.

Methods

In a chiropractic college research clinic, 29 asymptomatic participants (averaging 24 years in age, 17 males and 12 females), none of whom had a history of serious foot or ankle trauma or pathology, were recruited for a prone leg checking study. Each participant provided informed consent and the project received Institutional Review Board approval. As described in our previous article, six "footprints" cut from 1/16" (1.6 mm) poster board were used as shims to be inserted into a pair of surgical boots (figure 1) that were screwed to a 1/2" thick wood base. A long screw pointed medially from the heel as an indicator for Y-axis leg translation. The examination table had a metric ruler placed between the participant's feet, allowing measurement of absolute foot positions as indicated by the boot pointers.

For each of the 29 participants, two leg length measurements were obtained. In each measurement, anywhere from zero to six shims, each 1.6 mm thick, were inserted in either of the participant's surgical boots, thus lengthening either the left or right leg from zero to 9.6 mm. A research assistant randomly determined the magnitude by pulling a numbered index card from a paper bag and the side of insertion by flipping a coin.

Figure 1. Modified surgical boot used for instrumented leg checking.

The examiner, blinded to the side and amount of shim insertion, performed the compressive leg check, and read the boot pointers, as they intersected the ruler mounted between the participants' feet, to the nearest millimeter. Subtracting the right measurement from the left measurement provided the amount of left short leg for positive differences, or right short leg for negative differences. For each participant, we calculated the *actual* (known) change in the amount of artificial LLI between the first and second leg checks, as well as the change in the examiner's *observed* LLI from the first to the second observation. There was no need to determine a baseline (with no shims inserted), because our only intent was to determine if the examiner could accurately the change in shim condition between the two observations.

Suppose, for example, that two shims were added to the left boot for the first observation, and then removed while three shims were added to the right boot for the second observation. A perfectly accurate observer would have detected a difference of five shims between the two observations. This is irrespective of whether the non-measured baseline might have shown some degree of leg length inequality, be it functional or structural.

RESULTS

We produced a scatter diagram (figure 2) by plotting the examiner's observed change in LLI as a function of the actual change in LLI. This allowed a quick visual check to determine not only the examiner's accuracy in assessing changes in LLI, but also the appropriateness of using linear regression methods to quantify the results. Ultimately, a trendline was added to the scatter plot, the equation of which provided a numerical barometer of the compressive leg checking procedure. Moreover, the

Figure 2. Scatter diagram representing observed versus actual delta shim condition.

Bland-Altman 95% Limits-of-Agreement and histogram were obtained (figure 3), providing yet another measure of the examiner's accuracy. These assume that the mean and standard deviation of the differences between the independent and dependent variables (in this case, actual and observed change in shim condition) are constant throughout the range of measurement, and that these differences are from an approximately normal distribution.

The scatter plot of observed vs. actual change in LLI (figure 2) suggests a linear relationship, represented by

Figure 3. Bland-Altman 95% Limits-of-Agreement data.

a trendline with an equation representation of $y = 0.92x - 1.11$; $R^2 = 0.86$. The mean of the absolute value of the examiner errors equals 2.7 mm, and regression analysis computes a standard error (SE) of 3.2 mm. In five of 29 measurements (13.8%), the observed and actual change in LLI had opposite signs, or sign reversals. In these cases, a leg was made relatively longer from the first to the second observation, whereas the observer thought just the opposite had occurred. The Bland-Altman limits-of-agreement (LOA) and histogram are shown in figures 3 and 4 respectively, and were calculated to be (-7.6, 5.2). Various ways of assessing the accuracy of the compressive leg length assessment procedure are shown in table 1, and discussed below.

DISCUSSION

Many clinicians use a leg checking procedure – with some procedures being quite elaborate – in an attempt to garner information about the pelvis or more remote parts of the body, such as the atlanto-occipital articulation. Nonetheless, it is surprising how little attention has been paid to the most obvious suggestion that an apparent leg length difference may simply reflect an anatomic leg length inequality, as might be ascertained by tape measure, scanogram x-ray, sonogram, or some other direct leg length measuring method. Procedures that consider any observed LLI to automatically be related to spinal-pelvic dysfunction, without considering that said LLI may be structural, seem inherently flawed. Furthermore, efforts to discriminate functional LLI from

Figure 4. Histogram relating Bland Altman Limits-of-Agreement.

TABLE 1
ACCURACY GAUGES FOR PRONE COMPRESSIVE LEG CHECKING

ACCURACY GAUGE	CROSS-SECTIONAL STUDY (CURRENT)	LONGITUDINAL STUDY (PREVIOUSLY PUBLISHED)
SLOPE OF REGRESSION	.92	.95
R ²	.86	.90
Y INTERCEPT OF REGRESSION	-1.1 MM	.1 MM
MEAN ABSOLUTE VALUE OF MEASUREMENT ERROR	2.7 MM	2.0 MM
STATISTICAL STANDARD ERROR (SE)	3.2 MM	2.8 MM
LOA (95% CI)	(-7.6, +5.2)	(-5.4, +5.7)
SIGN CONCORDANCE	86.2%	94.7%

structural LLI, and to associate functional LLI (a phenomenon barely shown to exist) with pathology are not likely to be convincing so long as the accuracy of the leg length measuring method is in unknown.

There are several ways of estimating the accuracy of the observer, listed in table 1.

- *Linear regression:* The 45° identity line represents what would have been obtained with perfect accuracy, corresponding to a slope of 1.0. The .92 slope of the trend line through the scatter plot indicates that the observer's perceptions of changes in LLI were highly correlated with the induced LLI changes. Moreover, the R² value of .86 suggests that 86% of the observed change in LLI were accounted for by the experimentally produced change in LLI from one observation to the next.
- *Y intercept:* The Y intercept of -1.1 mm, which would equal zero with perfect accuracy, is the change in observed LLI when in fact the actual LLI does not change between observations. The 95% confidence interval around the Y intercept (-2.4, .13) includes the origin.
- *Standard error and mean error of measurement:* The SE of 3.2 mm is statistically calculated. It is larger than the mean error of measurement (the mean of the absolute values for each measurement error) of 2.7 mm, since the computation of SE more heavily weights larger errors than smaller errors.
- *Bland-Altman statistic:* This plots the differences between the observed and produced changes in LLI against their average. The horizontal lines in figure 3 represent 95% confidence intervals for these examiner errors. This means we can expect 95% of examiner errors to fall between these limits. In this study, these limits were -7.6 to 5.2 mm. The Bland-Altman histogram (figure 4) shows the frequency of different amounts of examiner errors. The approximately normal distribution suggests the data are not skewed, that there are few outlying data points, that the magnitude of examiner errors

is not related to the amount of artificially produced change in LLI, and finally that the use of linear regression is statistically appropriate.

- *Sign concordance:* Since in five of 29 measurements (13.8%), the observed and actual change in LLI had opposite signs, or sign reversals, looking at it dichotomously (right or wrong result), compressive leg checking was correct 86.2% of the time.

Table 1 also includes data from our previous longitudinal study. Every single measure showed greater accuracy in the earlier longitudinal study, just as one might have supposed: a participant being measured 26 times over a one hour period eventually becomes acclimated to the prone test environment, less likely to engage in confounding leg and trunk movements. Such extraneous movements hypothetically are more likely to occur during point observations of participants who are prone for just a few minutes, and presumably less adapted to that position. The accuracy seen in this current experiment is probably more relevant to typical practice settings.

Since a mean of repeated measures would be more accurate than the point observations used in this study (10), the use of such repeated measures in this study might have achieved a level of accuracy closer to that achieved in the previous study. Figure 5 (based on data from the longitudinal study) shows the examiner errors

Figure 5. Observed vs. actual delta shim condition.

for both single and three repeated measures observations; the standard deviation of 2.4 for single measures is reduced to 1.7 for repeated measures. Thus, the accuracy of prone instrumented compressive leg checking, using the mean of 3 measurements, would be about 2 mm, and pre-post measures differing by 4 mm or more would be significant. The optimal number of LLI measurements to take must lie somewhere between the 26 measurements performed per participant in the initial longitudinal study, and the two measurements per participant used in this cross-sectional study. Cooperstein (10) suggests that three compressive leg check measurements constitutes a reasonable compromise between the goals of being accurate and efficient in the use of the clinician's time.

Since only one examiner was tested, this study was not designed to determine how well another examiner could be trained to achieve similar accuracy. Although this and the previous study (6) show compressive leg checking using a modified surgical boot is quite accurate, it remains unknown whether non-instrumented (i.e., purely visual) compressive leg checking would also be accurate, at least for determining the side of the short leg. This study was not designed to determine the clinical utility of compressive leg checking, nor any other measure of LLI.

CONCLUSIONS

Although we found compressive leg checking in our previous longitudinal study to be very accurate, since we used only three participants we could not rule out the possibility that there were special characteristics of these three participants that would make the results non-generalizable. In other words, this first study did not have external validity. The results of this cross-sectional data on 26 participants do lend external validity to our previous conclusion, that compressive, instrumented (i.e., surgical boot-assisted) prone leg checking is an accurate, valid examination procedure. Investigators and clinicians might be warranted in using it as a pre-post measure during clinical interventions, should additional studies find its results correlate with other examination findings and pathologies.

We doubt clinicians routinely apply as much cephalad pressure as we use in performing compressive leg checking. Our data suggest clinicians might consider employing more cephalad pressure in performing leg checks, or at least lend their existing leg checking protocols to a similar calibrating study, to determine their

accuracy as compared with the protocol used in this study. It remains unknown whether the results of compressive leg checking correlate with clinically relevant findings, or if pre-post changes in the results of compressive leg checking predict clinically relevant changes in patients.

Among the future studies that could be performed as follow-ups to the present study, probably the most pressing would be to determine the interexaminer reliability of prone, compressive leg checking. Although a procedure may prove accurate in one setting, say, as conducted by an individual examiner, it would be of little clinical utility if other examiners can not be trained to achieve a similar level of accuracy.

ACKNOWLEDGMENTS

This work was supported by the Palmer Center for Chiropractic Research, in its domain of technology assessment. We thank Dr. Todd G. Nick, of the University of Mississippi Medical Center, for serving as our co-investigator in the initial longitudinal study, the methodology for which served as a template for this cross-sectional study.

REFERENCES

1. Cooperstein R, Grunstein J. Interdisciplinary assay of literature pertaining to leg length inequality. In: Owens E, editor. 10th Annual Vertebral Subluxation Conference; 2002 Dec 7-8; Hayward, CA: Sherman College of Straight Chiropractic; 2002. <http://www.sherman.edu/research/subcon10.html#coop>; Access date August 18, 2004.
2. Mannello DM. Leg length inequality. *J Manipulative Physiol Ther* 1992;15(9):576-590.
3. Thompson C. Thompson technique reference manual. Elgin, Illinois: Thompson Educational Workshops, Williams Manufacturing; 1984. p. 36-7.
4. Stillwagon G, Stillwagon KL. The Pierce-Stillwagon technique. *American Chiropractor* 1983(July/August):20-23, 28-31, 65.
5. Fuhr A. Activator Methods: Basic manual, Activator Methods Chiropractic Technique seminars. Phoenix, Arizona: Activator Methods, Inc.; 1990. p. 5, 13.
6. Cooperstein R, Morschhauser E, Lisi A, Nick T. Validity of compressive leg checking in measuring artificial leg length inequality. *J Manipulative Physiol Ther* 2004;26(9):557-566.
7. Jansen RD, Cooperstein R. Measurement of soft tissue strain in response to consecutively increased compressive and distractive loads on a friction-based test bed. *J Manipulative Physiol Ther* 1998;21(1):19-26.
8. Cooperstein R, Bricker D, Jansen R. Confirmation of leg length inequality as a stable, objectively-measured entity: The construction and implementation of a friction-reduced, multi-segmented table. In: Cleveland III C, Haldeman S, editors. Conference Proceedings of the Chiropractic Centennial Foundation; 1995; Washington, DC: Chiropractic Centennial Foundation; 1995. p. 325-326.
9. Cooperstein R, Morschhauser E, Lisi A. Validity of prone, compressive leg checking in detecting artificial leg length inequality. In: 2002 International Conference on Spinal Manipulation; 2002; Toronto CA: FCER; 2002. p. 129.
10. Cooperstein R. The importance of repeated measures. *Journal of the American Chiropractic Association* 2003;40(6):30-31.