VALIDITY OF COMPRESSIVE LEG CHECKING IN MEASURING ARTIFICIAL LEG-LENGTH INEQUALITY

Robert Cooperstein, DC,a Elaine Morschhauser, DC,b Anthony Lisi, DC,b and Todd G. Nick, PhD c

ABSTRACT

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

Design: Repeated measures (n = 26) on single subjects (n = 3).

Setting: Chiropractic college research clinic.

Methods: A pair of surgical boots were modified to permit continuous measurement of leg-length inequality (LLI). Multiple prone leg-check observations of a blinded examiner on 3 subjects were tested against artificial LLI that was created by randomly inserting 0 to 6 1.6-mm shims in either boot. Accuracy was assessed both within observations (observed versus artificial LLI) and between observations (observed versus artificial changes in LLI). The intraclass correlation coefficient (ICC), Lin’s concordance correlation coefficient (CCC), Bland-Altman limits of agreement, and linear regression statistics were obtained to determine the reliability and validity of compressive leg checking compared to a reference standard.

Results: For each shim condition, test-retest reliability was excellent (ICC = .85 and CCC = 0.95). The 95% confidence interval for the limits of agreement for observed versus artificial change in LLI was -5.44 to 5.67. The observed and artificial LLI shared 87% of their variation within observations (n = 78) and 88% between observations (n = 75). The mean examiner error was 1.72 mm and 2.01 mm, respectively.

Conclusion: Compressive leg checking seems highly accurate, detecting artificial changes in leg length ≤ 1.87 mm, and thus possesses concurrent validity assessed against artificial LLI. Pre–leg-check and post–leg-check differences should exceed 3.74 mm to be confident a real change has occurred. It is unknown whether compressive leg checking is clinically relevant. (J Manipulative Physiol Ther 2003;26:557-66)

Key Indexing Terms: Leg Length Inequality; Chiropractic; Validity

INTRODUCTION

Various chiropractic technique systems look to the “functional short leg” as both an indicator of subluxation and as an outcome measure and thus derive treatment strategies based on their leg-check findings.1 Many physical therapists and osteopaths also engage in leg-length assessment procedures of various kinds. Although somewhat controversial, there is evidence suggesting that anatomic leg-length inequality (LLI) can cause chronic low back pain5 and that correction of anatomic LLI with an appropriate heel lift may ameliorate low back pain.3

There have been many studies assessing the intraexaminer and interexaminer reliability of various leg-checking methods but far fewer that explore their validity. To be clinically useful, an examination method must be reliable; it must be possible for different examiners to obtain reproducible results. Moreover, the results obtained by that method must be valid, meaning the results are acceptably accurate when compared with the results obtained using an acknowledged gold standard. Although neither the reviews of Lawrence4 nor Mannello1 found much support for the reliability of leg checking, a few studies since then have been more encouraging for both prone5,7 and supine8 leg-length assessment procedures. Nonetheless, these newer studies cannot and do not address the following dilemma: When 2 examiners agree in their results, how do we know whether they are both right or both wrong? Without that knowledge, it would be clinically premature to place much confidence in their findings.

We are aware of very few other articles on the validity of leg-checking procedures. Rhodes et al,6.7 in the space of 2 articles, attempted to assess the validity of prone leg checking by comparing the results of prone and supine leg check-
ing with the radiological parameters of femoral head and iliac crest heights. Although these investigators obtained unusually high levels of intraexaminer reliability (they acknowledge the examiner was not adequately blinded between trials), they found very poor correlation between the leg-checking results and radiological parameters. This is not surprising, since in standing subjects, structural leg-length inequality may result in pelvic torsion that could affect the relative acetabular positions, reducing the correlation of radiological parameters with leg-checking results. Similarly, Venn et al, who also compared the results of prone leg checking with the radiographic measurements, did not find high correlation. Hoiriis et al correlated supine leg-check findings against anatometer findings, with somewhat equivocal results. Knutson investigated the possible correlation of supine leg-check findings with back pain, foot rotation, and pelvic crest unleveling. Russo has developed a tool called the Chiroslide Leg Discrepancy Analyzer (Chiroslide, Congers, NY) to quantify leg-length inequality, but we are not aware of any attempt to compare its results with any other measuring methods.

A number of studies conducted outside of chiropractic should be noted, even though they do not concern the validity of prone leg checking per se. These studies show that the imposition of artificial leg-length inequality in standing subjects results in pelvic torsion, with posterior rotation on the lengthened leg side and anterior rotation on the contralateral side. Although Hanada et al were not interested in possible torsional effects, they also imposed artificial leg-length inequality on standing subjects to test for interexaminer reliability in using their method and to compare the results against a radiological gold standard. We call attention to these studies only because our study, which involves the production of artificial leg-length inequality in prone subjects, bears methodological similarity to these other studies on standing study subjects.

Although this article does not discuss the conceptual or clinical distinction between functional and anatomic LLI, a noninvasive and accurate means of clinically detecting undifferentiated LLI might influence treatment strategy and lead to improved outcomes in some types of low back pain patients. We assessed the concurrent validity (accuracy) of 1 examiner performing prone leg checks compared with a gold standard of our own making, artificial leg-length inequality of known magnitudes. Our null hypothesis was that the examiner would not be accurate in performing compressive leg checking when assessed against artificial leg-length inequality of known magnitudes.

**Methods**

In a chiropractic college research clinic, 3 asymptomatic subjects (2 males aged 37 and 28 and 1 female aged 34, all of average height and weight), none of whom had a history of serious foot or ankle trauma or pathology, were recruited for a prone leg-checking study. Each subject provided informed consent and the project received Institutional Review Board approval. Six “footprints” were cut from 1-inch (1.6-mm) poster board, to be used as shims inserted into a pair of modified surgical boots. The boots (Fig 1) were screwed to a ½-inch thick wood base, had a long heelless nail pointed medially from the heel that was used as an indicator for y-axis foot translation, and featured Velcro closures. The examination table had a metric ruler placed at its foot, between the subject’s legs, allowing measurement of absolute foot positions as indicated by the boot pointers.

In one experimental observation, a known number of shims (0 to 6) were inserted into either the left or the right surgical boot by a research assistant. A blinded examiner performed a leg check, assessing the location of each foot in absolute terms by observing where the medially directed pointer lay in relation to a metric ruler which was mounted between the legs of the prone subject (Fig 1). Our leg-checking method has been previously described as compressive leg checking because substantial cephalad force is applied by the examiner. Since there were 6 shims that could be inserted into either boot, there were 13 experimental conditions possible: left 1 to 6, right 1 to 6, and 0 shims. An experimental trial called for 2 each of these possible 13 shim conditions to occur, for a total of 26 observations. The magnitude, side, and order of shim conditions were determined by a random number generator. These 26 observations were obtained on each of 3 subjects, all by 1 examiner.

To maximize the independence of individual measurements, the examiner “reset” the subjects to their baseline condition between measurements. He inspected to make sure the subject’s pelvis was centered on the table and then tugged on the subject’s feet to undo any possible effect of the previous measurement’s cephalad pressure on the legs. Great care was taken by the examiner while obtaining the measurement to make sure both feet were perpendicular to the floor, by sighting directly down the long axis of the surgical boots to confirm symmetric placement. The subject remained prone during the entire procedure. There were approximately 3 minutes between measurements, the amount of time it took to change the shim condition. The
examiner reported his measurements to a data recorder, who also prepared the subject for each shim condition and observation.

**Statistical Methods**

Bland-Altman means versus difference plots and limits of agreement (LOA) were calculated to determine the appropriateness of having used linear regression to determine the correlation of compressive leg checking with artificial LLI and also to determine the accuracy of compressive leg checking. The concordance correlation coefficient (CCC) statistic determined the concurrent concordance and precision of the leg-checking procedure, and the intraclass correlation coefficient (ICC) determined the test-retest (intraexaminer) reliability of the leg-checking procedure.

**Results**

**Assessing Accuracy Within Observations and Between Observations**

In this experiment, there were 2 ways to determine the accuracy of the examiner, one assessed within observations and the other between observations.

**Within-observations accuracy.** For each shim condition, the artificial LLI is the known or independent variable, and the examiner’s observed LLI is the dependent variable. For any 1 observation, we do not expect observed LLI to equal known artificial LLI, since the subject may have baseline LLI that would result in an “error” (it is not necessarily an examiner mistake) proportional to this baseline LLI. On the other hand, if we assume there is no systematic rater bias, either consistently overestimating or underestimating leg-length difference, we can correct each measurement by the baseline LLI or 26-observation mean error. Some of the rater’s calls will overestimate and some will underestimate LLI, but the 26-trial average mistake will be 0 if there is no systematic bias, so that the net mean error represents baseline LLI. As an example, suppose a subject is estimated to have a baseline 2-mm short right leg, the average difference between artificial and observed LLI across 26 observations. If 3 mm of shim are inserted in the left shoe, a perfectly accurate examiner would be expected to observe 3 + 2 = 5 mm of short right leg.

**Between-observations accuracy.** Although the within-observation differences cannot distinguish between a true baseline LLI and rater error, the between-observations can directly assess accuracy. An accurate observer should be able to accurately assess the change in shim condition from one observation to the next. For example, suppose 2 mm are inserted in the left shoe and the examiner perceives a 6-mm short right leg. If the next shim condition has 6 mm inserted in the left shoe, we would expect a perfectly accurate examiner to now perceive 10 mm of short right leg.

**Results for Within-Observations Observed Versus Artificial LLI**

In Figure 2, A, for each of the 3 subjects, there is a scatter plot for observed LLI versus artificial LLI (converting shims to millimeters, where 1 shim = 1.6 mm), and each plot includes a trendline. The aggregate data are plotted in Figure 2, B and C, which includes both regression and Bland-Altman differences versus means plots, and a histogram, Figure 2, D, confirming the normal distribution of the differences. (Although we produced Bland-Altman plots for each subject, we herein show only the aggregate plot.) As rationalized above, we corrected each observation by the arithmetically calculated mean error (arithmetical baseline LLI, Table 1). For each trendline, the slope represents the accuracy of the examiner in detecting the shim condition (the side and amount of artificial LLI created). The y-intercept represents the observed LLI when no shims are inserted, which amounts to the baseline LLI. A positive y-intercept denotes a short left leg, and a negative intercept denotes a short right leg. If the examiner were perfectly accurate and there were no confounding factors such as evoked responses in the subject, the slope of the trendline would be exactly 1.0. If the baseline LLI were equal to 0, the trendline would go through the origin. The $R^2$ values provide a measure of the amount of the examiner’s observed LLI explained by the shim insertion. Their values were .93, .85, .85 for subjects A, B, and C, respectively, and .87 for the aggregated data. The data are summarized in Figure 2, A, and Table 1.

The y-intercepts generated in the regressions are equivalent to the 26 trial mean “errors,” the arithmetically calculated baseline LLIs, as discussed previously. As can be seen in Table 1, the arithmetically and regression-obtained values were identical for all 3 subjects. For each of the shim conditions, we calculated the measurement error as the difference between artificial and observed LLI, after correcting for the arithmetical baseline LLI. The means of the absolute value of these measurement errors for subjects A, B, and C, respectively, were 1.55, 1.66, and 1.95 mm, and 1.72 for all 3 subjects. The standard errors (SEs) in Table 1 are calculated by the regression routine and are not expected to equal these measurement error absolute values (their relation is discussed below). In addition, for each of the 3 subjects, we counted the number of sign reversals, meaning the number of times a leg was made “longer” by the insertion of a shim and yet the examiner thought the leg was shorter than the baseline short leg. There were 3 such sign reversals (1, 0, and 2 for subjects A, B, and C, respectively). Finally, since each of the 13 shim conditions occurred twice, we were able to calculate an ICC for the first and second measurements, the test-retest reliability for each pair of point observations: these were .82, .85, and .82 for subjects A, B, and C, respectively, and .85 for all 3 subjects.

**Results for Between-Observations Change in Observed Versus Artificial LLI**

Since there were 26 shim trials per subject, there were 25 changes in shim condition per subject. For each of the 3 subjects (Fig 3, A through F), we show a scatter plot, trendline, and an identity line (a 45° line through the origin)
for observed change in LLI versus artificial change in LLI and Bland-Altman plots. The aggregate data are plotted in Figure 4, which adds a histogram (Fig 4, C) confirming the normal distribution of the differences. For each of the observed versus artificial LLI trend lines, the slope represents the examiner’s accuracy in detecting the change in the side and amount of artificial LLI created. The y-intercept represents the observed change in LLI when no change in the shim condition occurred (this only occurred 3 times among the 75 deltas recorded for the 3 subjects). If the examiner...
Fig 3. A, Subject A, observed delta LLI versus artificial delta LLI, between observations. B, Subject A, Bland-Altman, between-observations. C, Subject B, observed delta LLI versus artificial LLI, between-observations. D, Subject B, Bland-Altman, between-observations. E, Subject C, observed delta LLI versus artificial LLI, between-observations. F, Subject C, Bland-Altman, between-observations.
were perfectly accurate, the trendline would be identical to the identity line that denotes perfect accuracy. The regression and other data are summarized in Table 2, and Table 3 summarizes the Bland-Altman LOA for each of the 3 subjects.

The $R^2$ values, which provide a measure of the amount of the observed change in LLI that can be explained by changes in the shim condition (the amount of shared variation), were .92, .90, and .89 for subjects A, B, and C, respectively, and .94 for all 3 subjects. The mean of the absolute value of the measurement errors were 1.89, 1.91, and 2.22 mm, and 2.01 for all 3 subjects. The SEs in Table 2 are calculated by the regression routine and are not expected to equal these measurement errors (their relation is discussed below). Finally, for each of the 3 subjects, we counted the number of sign reversals, meaning the number of times between observations that a leg was made “longer” by the insertion of a shim and yet the examiner thought it had been made shorter. There were 4 such sign reversals (1, 1, and 2 for subjects A, B, and C, respectively). Visual inspection of the Bland-Altman difference versus mean plots shows both accuracy (the scatters hug the x-axis) and the appropriateness of having used the linear regression statistic (random dispersion of points). Unlike in the case of the within-observations analysis, we were unable to compute ICC values, because the experimental design did not include a test-retest component. On the other hand, we were able to use Lin’s concordance correlation coefficient (CCC),$^{23}$ which takes into account parameters of both correlation and accuracy, to determine whether the aggregate between-observations data significantly differ from the identity line. The results are shown in Table 3.

**DISCUSSION**

This study addressed the question of whether compressive leg checking is an accurate procedure. Our null hypothesis, that the examiner would not be accurate in performing
compressive leg checking when assessed against artificial leg-length inequality of known magnitudes, was rejected.

Clinical Significance

Leg-length assessment procedures often call for the application of some cephalad pressure on the legs, as evidenced by the following selected phrases: the examiner may "apply a gentle constant headward pressure with the thumbs pushing through the long axis of the legs,"24 although one author advises not to "cram the legs into the acetabular joints or shake the legs."25 The compressive leg checking procedure that was used in our experiment differs from typical prone leg checking, such as the 6-point landing system described by Fuhr and Colloca,24 in that more cephalad force, measured with a soft tissue algometer to be about 3 kg/leg (somewhat more force than reported by Hartley and Charley26) was applied to the legs. We used compressive leg checking based on our prior clinical observation that it appeared to be more reproducible than other methods. Although we used an instrumented measuring method, we would expect clinicians who might adapt their methods based on our results to use a visual method, which in our opinion might be more than adequate.

Clinicians generally record the side of the short leg, if any, and less commonly the amount of LLI. These dichotomous left-right-even calls often lead to one adjustment versus another or a simple yes-no decision as to the presence or absence of some pathology, such as atlas misalignment or pelvic torsion. For that reason, we looked for both within-observation and between-observation sign reversals. That is, compressive leg checking would be expected to identify the short or shortened leg side, irrespective of magnitude, 95.4% of the time.

Interpreting the Data

In our study, the accuracy of compressive leg checking was assessed for 3 subjects in 2 ways, by looking at (1) how well the examiner fared against the known amount of artificial LLI; and (2) how well the examiner could identify a change in artificial LLI. The Bland-Altman (also called Altman-Bland) bias plots27,28 we used graph the difference between the results of 2 measuring methods against their mean for each observation and obtain LOAs that will likely contain 95% of examiner errors. These limits are derived by calculating the mean difference (ie, bias) and its SD and then calculating the mean difference ± 1.96 SDs. In addition to the differences versus means and LOA plots, histograms, as recommended by Bland27 and Bland and Altman,28 may be provided (we did so for the aggregate analyses) to confirm that the mean values and SDs are reasonably constant throughout the ranges of measurement and are drawn from an approximately normal distribution.

To complement the Bland-Altman statistic, we computed the CCC. Lin introduced the CCC, a new reproducibility index, to evaluate agreement when the responses are measured on a continuous scale. It measures agreement like the statistic but for continuous rather than nominal or categorical data. The CCC combines measures of both precision and accuracy to determine whether the observed data significantly deviate from the identity line that represents perfect concordance (ie, 45° line). It increases in value as a

| Table 2. Data summary, accuracy of compressive leg checking between observations* |
|---------------------------------|---------|---------|---------|---------|
| Subjects | A (n = 25) | B (n = 25) | C (n = 25) | Aggregate (n = 75) |
| R²      | 0.92     | 0.90    | 0.89    | 0.88    |
| Standard error | 2.82    | 2.75    | 2.88    | 3.01    |
| Mean measurement error, absolute value | 1.89    | 1.91    | 2.22    | 2.01    |
| Intercept (95% CI) | .21 (−0.96, 1.38) | 0.16 (−0.98, 1.30) | −0.02 (−1.21, 1.18) | −0.08 (−0.77, 0.61) |
| Slope (95% CI) | 1.06 (0.92, 1.20) | .94 (0.80, 1.07) | .88 (0.74, 1.01) | .91 (0.83, 0.99) |

*Three subjects, 25 observations per subject.

| Table 3. Bland-Altman limits of agreement of Lin’s CCC |
|---------------------------------|---------|---------|---------|
| Subjects | A (n = 25) | B (n = 25) | C (n = 25) | Aggregate (n = 75) |
| Bias | 0.224 | 0.136 | 0.008 | 0.117 |
| 95% limits of agreement | −5.28, 5.73 | −5.24, 5.51 | −5.99, 5.97 | −5.44, 5.67 |
| CCC (95% CI) | 0.95 (0.90, 0.98) | 0.95 (0.88, 0.98) | 0.94 (0.87, 0.97) | 0.95 (0.92, 0.97) |

CCC, Concordance correlation coefficient.
function of both the correlation and accuracy of the observations with the gold standard (in this case, artificial LLI). We were also able to use the ICC to determine the test-retest reliability of compressive leg checking in this experiment, since each of the 13 shim conditions occurred twice.

In all 8 regressions summarized in Tables 1 and 2, the scatter plots clearly demonstrate linear correlation. Confidence intervals (CI) for the regression line slopes are provided in Tables 1 and 2, which contain the identity lines with 2 mild exceptions (subject C, within-observations, and for the aggregate between-observations) where the trendlines barely miss the CI. The P values for all 8 slopes are <.001. The Bland-Altman difference plots in Figure 2, C; Figure 3, A through C; and Figure 4 all show tight clustering around the x-axis, magnitudes of differences not proportional to shim size, horizontal trending, and relatively few outliers. This suggests (1) that the mean and SD of the differences are constant throughout the range of measurement; (2) these differences are from an approximately normal distribution; (3) it was appropriate to have used linear regression within the range of measures studied; and (4) compressive leg checking is not only reliable but accurate.

There were 2 ways of estimating the average measurement error—arithmetical and statistical. The arithmetical method involves calculating the average of the absolute values of the errors. (The average of the nominal errors, about half positive and half negative, approximates 0 and is of no significance.) Most of these were about 2 mm in this experiment (Tables 1 and 2) and their average was 1.87. This represents the sum of examiner error and subject-related variables, such as the patient shifting on the table or developing paraspinous muscle hypertonus during the course of the trial that hikes up one leg, confounding examiner calls. The statistically calculated SE, which derives from the algorithm that performs the regression analysis, more heavily weights larger misses than smaller misses. Therefore, the SEs in Tables 1 and 2 average 2.54 mm, exceeding the arithmetical estimate by .67 mm. From a clinician’s point of view, the arithmetical mean measurement error of less than 2 mm may be more relevant than the statistically calculated SE of around 2.5 mm.

Another statistical estimate of examiner error may be inferred from the LOA intervals in Figure 3, A through F, which are larger than the SEs, because these are not so much estimates of errors as intervals that will contain 95% of the observational misses. If differences within this range are of clinical significance, then the observational method may not be acceptable.

We were able to calculate the test-retest intraexaminer reliability of compressive leg checking in this experiment, since each of the 13 shim conditions occurred twice. The ICC between the first and second measurements was .92, .85, and .82 for subjects A, B, and C, respectively, and .95 for all 3 subjects (Table 3), suggesting that compressive leg checking is both accurate and precise.

Limitations

- In a rater/measurement validation process, the properties of a measurement are often characterized in the literature using linear regression analysis. There were 2 primary issues to consider in using linear regression to analyze the data: first, high correlation of a measure with a standard measure does not necessarily demonstrate accuracy; and second, it is somewhat unusual to apply linear regression methods to a repeated measures single-subject (SS) design. Bland27 discusses several pitfalls commonly encountered using correlation statistics when investigating a novel measuring method in comparison with an existing or gold standard or reference method. Of central significance is that r (the correlation coefficient) measures the strength of association of 2 variables, not their agreement. All the points may lie along a straight line, meaning there is perfect correlation, but this line may be very different from the 45° identity line that passes through the origin, denoting perfect concordance of the test method (compressive leg checking) with the standard or reference measure (artificial LLI). Bland and Altman28 suggest, nonetheless, that regression may be used to compare 2 methods of measurement provided supplemental analysis is provided (LOA plots and histograms).

- Again, it is not typical to apply linear regression in single-subject designs, when there are replicate measures for the subjects. Although correlation statistics are more often used in group analyses than in SS designs, any individual’s performance can be evaluated using a SS approach, provided sufficient data are obtained and his/her response remains stable during the observed interval. Bland27 and Bland and Altman28 also discuss the appropriateness of within-subject correlation. To do so, there must be reason to believe that the assumptions of normality and independence of measurements have not been violated. Clearly, it is easier to have confidence in the independence of measurements taken from different subjects rather than from one subject, as in our experiment. On the other hand, we carefully reset the subjects between observations and have no reason to suspect that a given observation had much impact on the next one. (Had we been studying the effect of a nutritional supplement taken daily on LLI, we would have had less confidence in our SS design, unless we had reason to believe that a given day’s vitamin supplement would not carry over into the next day.) The final potential problem with SS designs is the lack of generalizability to a population. It was for this reason that we applied the design to 3 individuals, although that is not enough to totally convince us that
our results with 3 asymptomatic young subjects can be safely extrapolated to different age brackets, symptomatic populations, etc.

- Whereas the Bland-Altman statistic is designed to test the accuracy of a proposed new measure compared with an existing measure which is not entirely accurate, Bland argues it may also be used to compare a new method against a validated gold standard, as in our experiment. He makes this argument against those who would plot differences against the gold standard, rather than the mean of the gold standard and observed values, and shows that this procedure is error prone and introduces spurious correlation.

- This study was designed to be clinically relevant, meaning it employed a very low-tech procedure. However, this low-tech design did leave several possible confounding factors uncontrolled: (1) the cephalad forces applied to each foot were not measured; (2) the feet were not carefully confirmed to be symmetrically dorsiflexed; (3) the subject’s pelvis was not proven to be midline on the table, nor the legs parallel to the long axis of the table; and (4) no effort was made to stabilize the pelvis while the legs were being pushed. It is not hard to conceive how using load receptors and additional rulers or the equivalent could increase the precision of the test system, which would then become more an instrument for basic science research rather than a vehicle for the immediately practical clinical research we set out to perform.

- Bias may have been introduced, because for each observation (except the 2.0-shim conditions), one surgical boot had no shims at all, whereas the other had at least 1. Since the boot has a somewhat contoured arch, pressure on the nonshimmed side may affect its relative travel compared with the shimmed shoe. This would be likely to change the slope of the regression lines, although probably not its intercept.

- If the examiner introduces systematic bias, say by pushing harder on one leg than the other most of the time, the estimate of the baseline LLI would be altered. Data analysis did not show evidence of this having occurred.

- There may be subject characteristics that introduce bias, such as differences in the left-right flexibility of the lumbar spine; even with equal cephalad pressures on the left and right legs, the pelvis may laterally flex on the lumbar spine under loading conditions confounding the baseline LLI. However, we believe the small forces we used were unlikely to have done so.

- It was not possible in this experiment to determine if baseline LLI represented anatomic, structural, or a composite LLI. It would be useful to now compare the baseline with a gold standard measure of anatomic LLI.

- This study was not designed to determine whether the results of compressive leg checking correlate with clinically relevant findings, such as patient symptoms or radiographic findings, or whether prechanges and postchanges in the results of compressive leg checking predict clinically relevant changes in signs and symptoms.

- Since this study tested only 1 examiner, the interexaminer reliability of compressive leg checking has not been carefully studied, although preliminary data (not herein reported) suggest it is reproducible and thus easily learned.

- In this experiment, we obtained multiple measurements on 3 individual experimental subjects, unlike the more typical leg-checking studies that obtained a very few measurements on many subjects. In so many words, we could not demonstrate external validity in this design. It will be necessary to repeat the study taking a pair of measurements, randomizing the shim condition, across a larger sample of subjects than the 3 used in this study, so that the results may be generalized. This would allow use of linear regression methods in a more traditional experimental setting.

- The same experimental protocol has not yet been applied to other leg-checking methods to test their accuracy (concurrent validity).

**Conclusion**

It seems safe to conclude that compressive leg checking is highly accurate, capable of detecting artificial changes in leg length ±1.87 mm. What this means clinically is that the difference between pre-leg-checks and post-leg-checks should exceed 3.74 mm if we are to be highly confident that a real change has occurred. Moreover, if pathologic significance is to be attributed to LLI (a matter of some controversy), say as detected by compressive prone leg checking, the magnitude should exceed 4 mm. Below that 4-mm threshold, the prevalues and postvalues and/or the inference of pathological significance may not be warranted.

We have no reason to suppose clinicians routinely apply as much cephalad pressure as we used in performing compressive leg checking. Since none of the more typical leg-checking methods has been shown to be more accurate when compared with a gold standard, our data suggest that clinicians might consider employing more cephalad pressure in performing leg checks, since that is now shown to be quite accurate in identifying intervention-related changes in leg length. Compressive leg checking is thus partially validated (lacking primarily external validity) in the same sense as treatment procedures that can become validated as described by Kaminski et al.

We do not know if the results of compressive leg checking correlate with clinically relevant findings or if prechanges and postchanges in the results of compressive leg checking predict clinically relevant changes in patients. These observations point toward other studies that might be...
done. However, in the meantime, we would suggest that if leg checking is to be done, whether to identify putative pathology, to monitor the outcome of care, or rationalize the use of a heel lift, prone compressive leg checking, instrumented in our study, appears to be a very accurate method to use.

ACKNOWLEDGMENTS

We thank the Research Council of the Palmer Center for Chiropractic Research for reviewing and endorsing this project prior to obtaining IRB approval and the Palmer University system for funding and the use of its facilities. We also acknowledge the contribution of our Research Assistants at Palmer West: E. Holmes, G. Saavedra, E. Crumm, and A. Morton.

REFERENCES