Interexaminer Reliability and Validity for Diagnosis of Temporomandibular Disorders of Visual Leg Measurements Used in Dental Kinesiology

Mauro Farella, DDS, PhD
Assistant Professor

Ambra Michelotti, DDS, BDS
Associate Professor

Gioacchino Pellegrino, DDS
Clinical Instructor

Department of Dental and Maxillo-Facial Sciences
Section of Orthodontics and Clinical Gnathology
University of Naples Federico II
Naples, Italy

Umberto Giani, MD
Associate Professor
Department of Preventive Medicine
Section of Statistics in Medicine
University of Naples Federico II
Naples, Italy

Roberto Martina, MD
Professor and Chairman
Department of Dental and Maxillo-Facial Sciences
Section of Orthodontics and Clinical Gnathology
University of Naples Federico II
Naples, Italy

Correspondence to:
Dr Mauro Farella
Department of Dental and Maxillo-Facial Sciences
Section of Orthodontics and Clinical Gnathology
University of Naples Federico II
Via Pansini, 5
1-80131 Napoli, Italy
Fax: +39081 7462197
E-mail: farella@unina.it

Aims: *To determine the reliability and the validity of visual leg measurements used in dental kinesiology, which suggests that a masticatory dysfunction, such as occurs in temporomandibular disorders (TMD), can influence the length and the internal rotation of lower limbs.* ***Methods:*** *The leg-length inequality test and the internal foot-rotation test were performed independently by 3 different examiners on 41 subjects who were also screened for TMD. Data were analyzed by means of kappa statistics and by calculation of sensitivity and specificity values.* ***Results:*** *Chance-corrected reliability was generally poor for both the leg-length inequality test ($0.33 \leq k \leq 0.39$) and the internal foot-rotation test ($0.15 \leq k \leq 0.27$). Sensitivity and specificity values of the tests to differentiate TMD and healthy subjects were below acceptable thresholds; they ranged from 0.41 to 0.57.* ***Conclusion:*** *Visual evaluations of leg-length inequality and internal foot rotation were unreliable and are not valid for TMD diagnoses. The results of chiropractic visual leg measurement procedures in dentistry should be interpreted with caution, particularly when clinical decisions may lead to nonreversible dental treatment.* J OROFAC PAIN 2005;19:285–290

Key words: chiropractic, diagnostic tests, leg-length inequality, reliability, temporomandibular disorders

In the past 2 decades, it has been suggested that dental occlusion may influence the whole body posture.[1–6] According to this hypothesis, functional disturbances (eg, in chewing and swallowing) of the masticatory muscles can be transmitted to distal musculature along the so-called "muscle chains."[7,8] Masticatory disorders may therefore result in postural asymmetries and/or pain conditions that most often affect the craniocervical musculature, the shoulder musculature, the low back musculature, and the leg musculature.[9] It has also been suggested that this etiological chain of events may be reversed. Hence, a difference in length of the legs may be considered either a consequence of a disorder affecting the masticatory apparatus or a risk factor for a masticatory disorder.[10,11]

Although the current scientific evidence of these causal relationships is weak or nonexistent,[12] the dental kinesiologic approach is followed not only by numerous chiropractors,[10,11,13,14] but also by many dental practitioners. In Italy as well as in other developed countries (eg, the United States, France, Spain), these beliefs have also been widely spread by the Internet and by local magazines and television programs, yielding a large number of patients seeking concomitant treatment for occlusal and postural disorders. As a result, many dentists have introduced in their clinical practice diagnostic tests aiming to evaluate inequalities or "imbalances" in various part of the body, particularly the lower limbs, and their potential relation to dental occlusion and to masticatory function and dysfunction (eg, temporomandibular disorders [TMD]).[11–14] The outcome of these tests often influences the treatment modality suggested to the patient, which may consist of invasive and nonreversible occlusal therapy.

A critical issue when using a diagnostic test in clinical practice is the evaluation of its reliability and validity. A test is considered reliable when the variation of random fluctuation due to errors in measurement is relatively small; a test is valid when it identifies correctly people with or without a certain disease. The aim of the present study was to determine the interexaminer reliability and validity of 2 diagnostic tests for masticatory dysfunction that are frequently used by dentists: the leg-length inequality test and the internal foot-rotation test.

Materials and Methods

Subjects and Examiners

Dental students from the University of Naples were recruited for the study. The sample (n = 41) consisted of 23 male subjects and 18 female subjects ranging from 21 to 34 years of age (mean ± SD, 24.5 ± 3.4 years).

The subjects underwent a stomatognathic functional examination for TMD that was performed by a calibrated dentist (AM) according to Axis I of the Research Diagnostic Criteria for TMD (RDC/TMD).[15] This examination included palpation and auscultation of the temporomandibular joints (TMJs), palpation of the masticatory muscles, and measurement of the unassisted mandibular range of motion.

No exclusion or inclusion criteria were used to restrict the selection of the subjects to be included in the study. Informed consent was obtained from all the subjects participating in the study. Care was taken to provide the subjects with only general information about the study to avoid influencing the results.

Three examiners were selected for the study. The first examiner (rater A) was an expert chiropractor who had been practicing for 10 years. The second examiner (rater B) was a dentist with 10 years of experience in chiropractic. The third examiner (rater C) was a general dentist. Rater A had been extensively trained and calibrated during his 5-year educational program in chiropractic in France. Rater B had worked with rater A for about 10 years in a private practice and had been extensively trained by rater A. Rater C was acquainted with and trained in all the test procedures during a single-day session by rater A.

Diagnostic Tests and Procedures

All the tests were performed during a single-day session in a single room. The subjects were examined in sequence in alphabetical order by surname. Each subject was asked to remove his or her shoes and was invited to lie supine and relaxed on an examining table, keeping the arms straight. The subject was asked to breathe deeply and slowly, to swallow, and to maintain the mouth closed with the mandibular and maxillary teeth in slight contact. Thereafter, the examiner performed the first test, comparing the position of medial malleolus and identifying the side of leg-length inequality. Comparison between sides was performed by visual observation.[16] Differences in leg length up to approximately 0.5 cm, and differences of internal foot rotation less than approximately 15 degrees, were considered normal values. The possible outcomes of the test were normal, right leg longer, or left leg longer. In the second test, the examiner manually exerted a slightly forced internal rotation of the foot in order to detect potential differences in the amount of rotation and end-play. The possible outcomes of the test were: normal, rotation lower at the right leg, or rotation lower at the left leg. The order of examiners testing each subject was randomly determined. Examinations were performed independently, with each examiner blinded to the findings of the other examiners. The subject was repositioned at each evaluation. Subjects were allowed a few minutes' rest between examinations.

Statistical Analysis

Data for agreement were summarized in 3 × 3 tables. Interexaminer reliability was determined by means of proportion agreement (%) and chance-corrected kappa statistics for categorial scales.[17,18] p_{ij} denoted the observed proportion of subjects classified as category *i* by the first examiner and category *j* by the other. p_i represented the marginal probability of assignment to category *i* by 1 exam-

Tables 1a to 1c Agreement Among Raters for the Leg-Length Inequality Test

Table 1a Rater A Versus Rater B

	Rater B			
Rater A	Equal	R longer	L longer	Total
Equal	11	6	1	18
R longer	3	9	0	12
L longer	5	1	5	11
Total	19	16	6	41

R = right leg; L = left leg.

Table 1b Rater A Versus Rater C

	Rater C			
Rater A	Equal	R longer	L longer	Total
Equal	9	5	4	18
R longer	6	6	0	12
L longer	2	0	9	11
Total	17	11	13	41

R = right leg; L = left leg.

Table 1c Rater B Versus Rater C

	Rater C			
Rater B	Equal	R longer	L longer	Total
Equal	9	2	8	19
R longer	7	9	0	16
L longer	1	0	5	6
Total	17	11	13	41

R = right leg; L = left leg.

Tables 2a to 2c Agreement Among Raters for the Internal Foot-Rotation Test

Table 2a Rater A Versus Rater B

	Rater B			
Rater A	Equal	R longer	L longer	Total
Equal	14	4	2	20
R lower	2	5	0	7
L lower	12	1	1	14
Total	28	10	3	41

R = right leg; L = left leg.

Table 2b Rater A Versus Rater C

	Rater C			
Rater A	Equal	R longer	L longer	Total
Equal	11	4	5	20
R lower	1	5	1	7
L lower	7	1	6	14
Total	19	10	12	41

R = right leg; L = left leg.

Table 2c Rater B Versus Rater C

	Rater C			
Rater B	Equal	R longer	L longer	Total
Equal	15	4	9	28
R lower	3	5	2	10
L lower	1	1	1	3
Total	19	10	12	41

R = right leg; L = left leg.

iner; p_j represented the marginal probability of assignment to category *j* by the other examiner. The chance-corrected agreement by the kappa coefficient for a nominal scale with *n* categories may be estimated as follows:

$\kappa = \frac{p_o - p_e}{1 - p_e}$ where $p_o = \sum_{i=1}^{n} p_{ii}$ is the proportion of observed agreement, and $P_e = \sum_{i=1}^{n} p_{i}\cdot p_{\cdot i}$ is the expected agreement.

The standard error (SE) and the overall values of kappa for multiple examiners were calculated according to Fleiss et al.[19,20] A 1-tailed statistical test by reference to the standard normal distribution (z test) was carried out in order to test the null hypothesis that agreement was better than chance. Kappa values below 0.4 indicated poor agreement, between 0.4 and 0.75 fair agreement, and values from 0.75 to the maximum of 1 excellent agreement.[17] Diagnostic validity of the tests performed by the chiropractor to differentiate between TMD patients and healthy subjects was assessed with sensitivity and specificity values. The "gold standard" for the TMD diagnosis was the examiner (AM) calibrated according to the TMD/RDC criteria.[15] Sensitivity is the proportion of subjects with a disease who obtain a positive test result, whereas specificity is the proportion of those without a disease who get a negative test result. Statistical analyses were carried out by means of commercial software (Microsoft Excel 97).

Results

Seven subjects (17%) were diagnosed as having myofascial pain of the jaw muscle without limited opening, and 12 subjects (29%) as having anterior disc displacement of the TMJ. Twenty-eight subjects (68%) were free of any TMD. Leg-length inequality was found to be highly prevalent in the subjects investigated. Indeed, the percentages of

Table 3 Computational Procedure for Kappa Statistics

	p_o	p_e	κ	SE(κ)	z	*P*
Leg-length inequality test						
Rater A – Rater B	0.61	0.36	0.39	0.11	3.56	.00018
Rater A – Rater C	0.58	0.34	0.37	0.11	3.30	.00048
Rater B – Rater C	0.56	0.35	0.33	0.11	3.04	.00120
Internal foot-rotation test						
Rater A – Rater B	0.49	0.40	0.15	0.10	1.50	.067
Rater A – Rater C	0.54	0.37	0.27	0.11	2.38	.008
Rater B – Rater C	0.51	0.40	0.19	0.10	1.86	.031

P_o = proportion of observed agreement; P_e = proportion of expected agreement; SE(k) = standard error of kappa. *P* values were 1-tailed.

Table 4 Sensitivity and Specificity of Leg-Length and Internal Rotation Tests for Myofascial Pain of the Jaw Muscle and Anterior Disc Displacement (RDC Ia and IIa)

	Sensitivity	Specificity
Leg-length inequality test		
Myofascial pain of the jaw muscles	0.43	0.41
Anterior TMJ disc displacement	0.50	0.41
Internal foot-rotation test		
Myofascial pain of the jaw muscles	0.43	0.47
Anterior TMJ disc displacement	0.57	0.52

Only data obtained from rater A (the chiropractor) were used for calculations.

length inequalities found by the 3 examiners were 56.1%, 53.6%, and 58.5%, respectively, for the raters A, B, and C. A different internal foot rotation was found in 51.2% of the subjects by rater A, in 31.7% of the subjects by rater B, and in 53.6% of the subjects by rater C.

The outcomes of the 2 tests for measuring the interexaminer agreement are detailed in Tables 1a to 1c and 2a to 2c. The proportion of observed agreement between examiners ranged from 0.56 to 0.61 for the leg-length test and from 0.49 to 0.54 for the internal rotation test. The computational procedure for calculation of kappa is summarized in Table 3. Kappa values were in general low, ranging from 0.33 to 0.39 for the leg-length inequality test and from 0.15 to 0.27 for the internal rotation test. The kappa values were statistically significant ($P < .001$) between all the examiners for the leg-length inequality test, and between raters A and C and raters B and C for the internal rotation test. The null hypothesis that agreement was better than chance could not be rejected for the internal foot-rotation test between rater A and rater B ($P > .05$). The proportion of full agreement among all the raters was 0.41 for the leg-length test and 0.31 for the internal rotation test. The overall values of kappa among the 3 examiners were 0.35 (SE = 0.065) for the leg-length test and 0.20 (SE = 0.002) for the internal rotation test. Sensitivity and specificity ranged from 0.41 to 0.57. These values are summarized in Table 4.

Discussion

The proportions of both leg-length inequality and different internal rotation resulting from the 2 diagnostic tests used in this study were surprisingly high. Differences in regard to leg length and rotation for the 2 legs were found in more than half the sample. Such high prevalence values, however, are in agreement with other chiropractic findings in studies where exclusion and inclusion criteria were not used.[21,22]

In spite of the widespread use of these "quick" visual leg measurement procedures, their reliability and biomechanical significance are still controversial; so far, mixed findings have been reported.[23] While several authors have reported good reliability (k = 0.7) for the supine leg length test when used by experienced clinicians,[24] others have reported very low agreement values for both expert and normal examiners.[25] A recent systematic review made by chiropractors themselves[23] concluded that the reliability of chiropractic tests for lumbo-pelvic spine had not been sufficiently evaluated, and that most current tests do not yield consistently acceptable results.

In the present study, the agreement between an experienced chiropractor and 2 dental examiners was evaluated. The experienced chiropractor was considered the reference examiner, and no attempt was made to test the measurement validity against a more objective procedure (eg, x-ray, sliding calipers, or mechanical identification devices). The aforementioned procedures have been used elsewhere for evaluation of leg-length inequality.[26,27] This is a limitation of the present study. since the experienced chiropractor cannot be considered an acceptable "gold standard" to determine a "true" leg asymmetry. In previous studies, the validity of visual measurements has been tested against the reference of radiographs taken in a standing position, with rather poor results.[28,29] The findings of these studies, however, have been questioned by chiropractors, because x-ray measurements reveal actual anatomic asymmetry, whereas supine leg measurements determine "abnormal activation of pelvic and suprapelvic muscles" (ie, physiologic/functional asymmetry).[22] Other measurements not related solely to anatomic asymmetries would probably be more appropriate, but unfortunately, robust gold standards for these purposes do not appear to exist. In the present study, the decision to test "visual examination" of leg length against a chiropractic test was intentional, as this represents what is commonly done by colleagues using this approach.

The rationale for the choice of the 3 examiners represents an attempt to mimic what usually happens in the clinical setting: chiropractors identify a "general dysfunction" originating from the masticatory system, and thereafter they refer the patient to a dentist who is often unacquainted with the approach. Some dentists do attend short courses about dental kinesiology and learn the basic principles of this approach, and a few dentists embrace dental kinesiology completely and start to treat patients accordingly. As a first objective, the authors determined whether a general dental practitioner would be able to perform chiropractic tests correctly; secondly, the authors determined whether an experienced dentist would perform better than an inexperienced dentist. The results of the present study suggest that agreement between dentists and a chiropractor, although better than chance, was in general poor, particularly for the internal rotation test, and provide little support for the use of these clinical tests in dentistry. In addition, when the findings of an experienced dentist were compared with those of an ordinary dentist, the former did not produce better agreement with the chiropractor; in some cases, agreement was worse. It is therefore unlikely that the poor results may be ascribed only to the examiners' lack of adequate experience or to their educational background.[30]

The diagnostic validity of visual leg measurements as tests to differentiate between TMD patients and healthy subjects was tested by calculation of sensitivity and specificity using the chiropractor as the reference standard. The recommendations of the American Academy of Orofacial Pain[31] indicate that sensitivity and specificity of diagnostic tests for TMD should be higher than 0.7. In the present study, however, both the sensitivity and specificity values of the tests were far below 0.7; thus, the present findings gave little support to the diagnostic validity of visual leg measurements for TMD, as both the sensitivity and specificity values of the tests were far below 0.7. Hence, it may be argued that in most cases, a decision of "treatment need" should not be made based on these tests. It should also be stressed that visual leg measurements are often performed in apparently healthy subjects, and that these tests suggest the need for irreversible occlusal treatments (eg, orthodontics, prosthodontics).

The proportions of subjects with myofascial pain and anterior disc displacement found in the sample fit with previous estimates of the prevalence of TMD in the general population.[32] However, although TMD diagnoses were made in the sample investigated, the subjects could not be considered representative of a patient population seeking treatment for TMD complaints. Thus, it would be appropriate to consider the present data as preliminary findings that need to be replicated with larger sample sizes and multiple examiners.

In conclusion, leg-length inequality and asymmetric internal foot rotation are highly frequent in a student population. The interexaminer reliabilities of the leg-length inequality test and of the internal foot-rotation test were poor. Indeed, kappa values obtained from a chiropractor and 2 dentists were always below the critical threshold of 0.4, and in several cases the agreement between examiners was not better than chance. The diagnostic validity of quick visual leg measurements as tests to differentiate between TMD patients and healthy subjects was also very low. The outcome of these tests used in dentistry should therefore be interpreted with caution, particularly when clinical decisions may lead to nonreversible dental treatment.

Acknowledgments

A special thanks to the chiropractor who performed the clinical evaluations for the present research. The study was supported by the Italian Ministry for Scientific Research (grant PRIN Prot.n. 2001062735).

References

1. Hellsing E, McWilliam J, Reigo T, Spangfort E. The relationship between craniofacial morphology, head posture and spinal curvature in 8, 11 and 15 years old children. Eur J Orthod 1987;9:254–264.
2. Huggare JA, Raustia AM. Head posture and craniovertebral and craniofacial morphology in patients with craniomandibular dysfunction. J Craniomandib Pract 1992;10:173–178.
3. Kritsineli M, Shim YS. Malocclusion, body posture, and temporomandibular disorder in children with primary and mixed dentition. J Clin Pediatr Dent 1992;16:86–93.
4. Darlow LA, Pesco J, Greenberg MS. The relationship of posture to myofascial pain dysfunction syndrome. J Am Dent Assoc 1987;114:73–75.
5. Hellsing E, Reigo T, Mc William J, Spangfort E. Cervical and lumbar lordosis and thoracic kyphosis in 8, 11 and 15-year-old children. Eur J Orthod 1987;9:129–138.
6. Huggare J, Pirttiniemi P, Serlo W. Head posture and dentofacial morphology in subjects treated for scoliosis. Proc Finn Dent Soc 1991;1:151–158.
7. Valentino B, Melito F. Functional relationship between the muscles of mastication and the muscles of the leg. An electromyographic study. Surg Radiol Anat 1991;13:33–37.
8. Valentino B, Fabozo A, Melito F. The functional relationship between the occlusal plane and the plantar arches: An EMG study. Surg Radiol Anat 1991;13:171–174.
9. Gole DR. A clinical observation: A relationship of occlusal contacts to distal musculature. Cranio 1993;11:55–62.
10. Chinappi AS, Getzoff H. The dental-chiropractic cotreatment of structural disorders of the jaw and temporomandibular joint dysfunction. J Manipulative Physiol Ther 1995;18:476–481.
11. Eversaul G. Dental Kinesiology. Las Vegas: Author, 1977.
12. Michelotti A, Manzo P, Farella M, Martina R. Occlusion and posture: Is there evidence of correlation? [in Italian]. Minerva Stomatol 1999; 48:525–534.
13. Walther DS. Applied Kinesiology. Vol 1: Basic procedure and muscle testing. Pueblo CO: Systems DC, 1981.
14. Jensen CR, Schultz GW, Bangerter BL. Applied Kinesiology, ed 3. Vol 2: Head, neck and jaw pain and dysfunction. The stomatognathic system. New York: McGraw-Hill, 1983.
15. Dworkin SF, LeResche L. Research Diagnostic Criteria for Temporomandibular Disorders: Review, criteria, examinations and specifications, critique. J Craniomandib Disord 1992;6:301–355.
16. Fuhr AW, Osterbauer PJ. Interexaminer reliability of relative leg length evaluations in the prone extended position. Chiropractic Technique 1989;1:13–18.
17. Cohen J. A coefficient of agreement for nominal scales. Educ Physiol Meas 1960;20:37–46.
18. Landis J, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
19. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. Psychol Bull 1969;72:323–377.
20. Fleiss JL. The measurement of interrater agreement. In: Statistical Methods for Rates and Proportions, ed 2. New York: John Wiley & Sons, 1981:212–236.
21. Specht DL, De Boer KF. Anatomical leg length inequality, scoliosis and lordotic curve in unselected clinic patients. J Manipulative Physiol Ther 1991;14:368–375.
22. Knutson GA. Incidence of foot rotation, pelvic crest unleveling, and supine leg length alignment asymmetry and their relationship to self-reported back pain. J Manipulative Physiol Ther 2002;25:110E [erratum 2002;25:425].
23. Hestbœk L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. J Manipulative Physiol Ther 2000;23:258–275.
24. Hinson R, Brown SH. Supine leg length differential estimation: An inter- and intra-examiner reliability study. Chiropractic Res J 1998;5:17–22.
25. Rhudy TR, Burk JM. Inter-examiner reliability of functional leg-length assessment. Am J Chiropractic Med 1990;2:63–66.
26. Grundy PF, Roberts CJ. Does unequal leg length cause back pain? A case-control study. Lancet 1984;4;2: 256–258.
27. Mannello DM. Leg length inequality. J Manipulative Physiol Ther 1992;15:576–590.
28. Friberg O, Nurminen M, Korhonen K, Soininen E, Manttari T. Accuracy and precision of clinical estimation of leg length inequality and lumbar scoliosis: Comparison of clinical and radiological measurements. Int Disabil Stud 1988:10:49–53.
29. Rhodes DW, Mansfield ER, Bishop PA, Smith JF. The validity of the prone leg check as an estimate of standing leg length inequality measured by x-ray. J Manipulative Physiol Ther 1995:18:343–346.
30. Boline PD, Keating JC, Brist J, Denver G. Inter-examiner reliability of palpatory evaluations of the lumbar spine. Am J Chiropractic Med 1988;1:5-11.
31. American Academy of Orofacial Pain. Okeson JP (ed). Orofacial Pain. Guidelines for Assessment, Diagnosis, and Management. Chicago: Quintessence, 1996.
32. Dworkin SF, Huggins KH, LeResche L, et al. Epidemiology of signs and symptoms in temporomandibular disorders: Clinical signs in cases and controls. J Am Dent Assoc 1990;120:273–281.