Reliability of a Diabetic Foot Evaluation

Jay E Diamond
Michael J Mueller
Anthony Delitto
David R Sinacore

The purpose of this study was to establish the interrater and intrarater reliability of various ankle and foot measures common to a diabetic evaluation. Bilateral biomechanical, sensory, and wound-size measurements were obtained in 31 subjects with diabetes mellitus. Twenty-five subjects were retested by the initial examiner to determine intratester reliability, and all subjects were retested by another examiner to determine intertester reliability. Both examiners participated in an extensive training period prior to the initiation of this study to minimize variability between and within measurers. Intraclass correlation coefficients for interrater and intrarater measurements ranged from .58 to .89 and from .74 to .99, respectively. The results of this study indicate that ankle and foot measurements common to a diabetic evaluation can be taken reliably between testers. We believe extensive examiner training in these clinically relevant measures can improve reliability between testers. [Diamond JE, Mueller MJ, Delitto A, et al: Reliability of a diabetic foot evaluation. Phys Ther 69: 797–802, 1989]

Key Words: *Diabetes; Diabetic neuropathies; Lower extremity, ankle and foot; Tests and measurements, range of motion.*

Plantar ulceration is a serious complication that often affects patients with diabetes mellitus. Neuropathy, a common sequela of diabetes mellitus, has been implicated as the primary cause of plantar ulceration.[1–3] Loss of sensation in the foot can allow unperceived trauma, contributing to plantar ulceration. Other contributing factors include limited joint mobility and foot deformities, both of which have been associated with the diabetic population.[4–9] These biomechanical factors may lead to excessive plantar pressure and subsequent ulceration in the presence of sensory loss. Because these factors may contribute to foot complications, it is imperative that biomechanical, sensory, and wound-size measures in a diabetic population be reliably established for use in research, evaluation, treatment, and prevention of these foot complications.

Various authors have published reports either describing or comparing devices that will standardize a biomechanical foot evaluation. Kaye and Sorto described a device called the "k-square" that measures subtalar joint (STJ) range of motion, forefoot-to-rearfoot (FF/RF) relationship, and ankle dorsiflexion (DF), but no reliability for the measurements was reported.[10] Baldwin and Graebner compared the use of the k-square and the tractograph when measuring STJ ROM and FF/RF relationship. Intraclass correlation coefficients (ICCs) revealed no significant agreement between testers. No intrarater reliability was assessed, and no mention of the population measured was reported.[11] Muwanga et al developed a device that measured total ankle ROM with the foot secured to a footplate.[12] They reported no significant difference between the same or independent observers using the device in a group of healthy volunteers.

Laftanza et al reported significant differences for measurements of STJ eversion in a closed versus an open

J Diamond, MHS, PT, is Physical Therapy Supervisor, Department of Physical Therapy, Irene Walter Johnson Rehabilitation Institute, Washington University Medical School, 509 S Euclid Ave, St Louis, MO 63110 (USA).

M Mueller, MHS, PT, is Instructor, Program in Physical Therapy, Washington University Medical School, and Physical Therapist, Department of Physical Therapy, Irene Walter Johnson Rehabilitation Institute.

A Delitto, MHS, PT, is Instructor, Program in Physical Therapy, Washington University Medical School, and Consulting Physical Therapist, Department of Physical Therapy, Irene Walter Johnson Rehabilitation Institute.

D Sinacore, MHS, PT, is a physical therapist and a doctoral candidate, Department of Anatomy, School of Medicine, West Virginia University, Morgantown, WV 26506.

This article was submitted October 13, 1988; was with the authors for revision for seven weeks; and was accepted March 21, 1989.

kinetic chain.[13] Their subjects were physical therapy students and physical therapists, and no reliability of the measurements was established. McPoil et al found high intrarater reliability (ICC = .90–.96) for three techniques used to measure a condition commonly referred to as "tibial varum," which the authors state is more accurately termed "tibiofibular varum."[14] No interrater reliability, however, was reported. Lohmann et al analyzed tibial varum measurements obtained by two raters while healthy subjects assumed three different positions.[15] Intrarater reliability coefficients were .46 and .83 for each of the raters, and interrater reliability coefficients by position were .41, .49, and .58. Boone et al used multiple testers to assess the reliability of various upper extremity and lower extremity motions.[16] The average interrater and intrarater ICCs for foot inversion were .69 and .80, respectively. This study was based on a small sample of healthy subjects (N = 12); thus, the results have little inference to the diabetic population.

To our knowledge, only two reliability studies on foot and ankle measures have used patients as subjects.[17,18] Elveru et al measured patients with neurological or orthopedic disorders.[17] The goniometric measurements were made by 14 physical therapists with limited or no experience in measuring the subtalar joint neutral (STJN) position or STJ passive ROM with a goniometer. They also investigated passive rearfoot inversion, eversion, and passive ankle DF and plantar flexion (PF). Intrarater reliability was adequate (ICC = .6–.9), but interrater reliability was low (ICC = .1–.5). The researchers concluded that clinical measurement of the STJ and ankle may be performed somewhat reliably if the same tester takes the measurements over a short time period; however, these same measurements taken by different therapists cannot be considered reliable. Pandya et al determined intrarater and interrater reliability of goniometric measurements of seven common upper extremity and lower extremity joints in children with Duchenne muscular dystrophy.[18] Their ICC values for passive ankle DF intrarater and interrater reliability were .90 and .73, respectively. The authors concluded that goniometric measurements taken by the same examiner are highly reliable.

Table 1. *Age and Years Insulin Dependent for Patients with Diabetes Who Participated in Interrater and Intrarater Trials*

	Interrater (n = 31)			Intrarater (n = 25)		
Variable	$\bar{X}$	*s*	Range	$\bar{X}$	*s*	Range
Age (yr)	59	12	(34–77)	59	11	(34–77)
Years insulin dependent	16	7	(2–31)	14	6	(2–24)

Decreased plantar sensation attributable to neuropathy is commonly seen in patients with diabetes. Several studies have documented the use of Semmes-Weinstein (SW) monofilaments for screening patients at risk for plantar ulcerations secondary to Hansen's disease and diabetes.[8,19,20] These studies document the 5.07 monofilament as the best discriminator for feet susceptible to ulceration.

The purpose of this study was to determine the interrater and intrarater reliability of various foot and ankle measurements, taken by experienced physical therapists, in a group of diabetic patients vulnerable to serious foot complications. Specifically, we assessed the following measures: passive calcaneal inversion and eversion, palpated STJN, FF/RF relationship in STJN, passive talocrural DF, sensation level, ulcer area and depth, calcaneal stance position, and tibial varum. We expected all measures would be reliable between and within raters because of extensive examiner training.

Method

Subjects

Thirty-one volunteers participated in this study. Subjects were recruited from patients referred to physical therapy from the diabetic foot center at Washington University Medical School and by private physicians. The purpose and procedure of the study were explained to each prospective subject, and informed consent was obtained. Subject characteristics are shown in Table 1.

Materials

A standard plastic goniometer with the scale marked in 2-degree increments was used for all biomechanical measurements. Semmes-Weinstein monofilaments 4.17, 5.07, and 6.10 were used for sensation measurements. An indelible ink marker was used to trace the ulcer area on a piece of sterilized x-ray film,[21] and a modified slide rule was used to measure ulcer depth.

Procedure

We used a test-retest design, with the same subject assessed three times during each session. To assess interrater reliability, subjects were randomly tested by one of the two testers (JED, MJM) and next by the other tester. To assess intrarater reliability, the subject was measured a third time by one of the testers, again determined randomly. Each evaluation took about 15 minutes to complete. Testers were in separate rooms when the measurements were taken. Interrater measurements were obtained for 31 subjects (17 men, 14 women), and intrarater measurements were obtained for 25 subjects (14 men, 11 women). Six subjects were not assessed a third time because of time constraints. Not all subjects had plantar ulceration. Ulcer size and depth were measured by the same tester on 14 subjects and by different testers on 9 subjects.

Both examiners underwent a lengthy training period with the measures used in this evaluation. Training sessions consisted of the evaluation of patients by each therapist independent of the other. Each patient's measurements were discussed and remeasured jointly if discrepancies existed. Approximately 20 of these training sessions were conducted over an 18-month period prior to this experiment. Both examiners commonly (ie, daily) used these measures in their clinical practice.

A standard biomechanical examination procedure described by McPoil and Brocato was used.[22] Briefly, patients were positioned prone with the leg not being examined flexed at the knee, laterally rotated, and abducted at the hip. The lower third of the tested leg and the calcaneus were bisected. For passive calcaneal inversion and eversion, the calcaneus was secured with one hand, maximally inverted or everted, and measured from anatomical zero. The goniometer arms were aligned with the bisection of the calcaneus and the bisection of the lower third of the leg. Palpated STJN is the position of inversion or eversion the calcaneus assumes, in relationship to the tibia, when the talus position is congruent between palpating fingers. The foot is passively pronated and supinated until the medial and lateral sides of the talar head are neither protruding nor depressed. The fourth and fifth metatarsal heads are then passively dorsiflexed, locking the midtarsal joints, and the calcaneal position is then measured as described previously. To determine the FF/RF relationship, first the STJN position is located. The goniometer is then placed with one arm parallel to the plane of the metatarsal heads and the other arm aligned perpendicular to the bisection of the calcaneus. Passive talocrural DF is also measured by first finding the STJN position. The axis of the goniometer is then placed over the lateral malleolus, and one arm is aligned with the lateral border of the foot while the other arm is aligned with the lateral border of the lower leg. The examiner then passively dorsiflexes the talocrural joint, and the angle is measured.

Sensation at the plantar surface of the foot was measured using 4.17, 5.07, or 6.10 SW monofilament similar to the method described by Birke and Sims.[19] Sites for testing were the first, third, and fifth toes and their associated metatarsal head; the medial (at the base of the first metatarsal) and lateral midfoot (at the base of the fifth metatarsal); and the heel. The monofilament was pressed perpendicular to the surface of the skin with enough pressure to bend the monofilament. The higher the value of the monofilament, the more difficult it is to bend. Five to 10 trials were performed at each site, and the subject needed to perceive 80% of the trials to be graded a given value at that site. If the subject was unable to sense at least 80% of the trials, he or she would be tested with the next highest monofilament. If the subject was unable to sense 80% of the trials using the 6.10 monofilament, he or she was graded the value ">6.10." The highest perceived monofilament value at any of the sites was taken as the overall grade of the foot.

If an ulcer was present, area measurements were obtained by tracing the circumference of the wound onto sterilized x-ray film.[21] Area was computed with a sonic digitizer.* Ulcer depth was obtained by placing the edge of a modified slide rule at the ulcer's deepest point.

The subject was then asked to stand, and calcaneal stance was measured by aligning one arm of the goniometer with the calcaneal bisect and the other arm parallel with the floor. Tibial varum was measured with the subject in relaxed single leg stance and with one arm of the goniometer aligned with the bisect of the lower third of the leg and the other arm parallel with the floor. All marks on the leg were then washed off, and the subject was measured by the next examiner.

Data Analysis

Reliability was assessed using ICCs (formula 2,1)[23] for all measures except sensation testing. Reliability for sensation measurement was determined by Cohen's Kappa statistic.[24]

Criteria for acceptable reliability were partially based on the work of Lahey et al[23] and included a minimal acceptable ICC of .70. Acceptable criteria for the Kappa statistic were based on the work of Landis and Koch,[25] who stated that a Kappa value of ≥ .61 infers a "moderate" strength of agreement.

In addition, we calculated the standard error of measurement (SEM) for all data except sensation using the formula[26]:

$$SEM = SD \times (1 - r)^{1/2}$$

where SD is the standard deviation and r is the correlation coefficient. The SEM is used to estimate the "reasonable limits" of the true score for a person with any obtained score; therefore, it is useful in the interpretation of the reliability of a single score.

Results

The means and standard deviations for all goniometric measurements are summarized in Table 2. Results of the interrater and intrarater ICCs and SEMs for goniometric measurements are summarized in Table 3. Intrarater reliability coefficients ranged from .74 to .99, and interrater reliability coefficients ranged from .58 to .89. All intrarater reliability coefficients were acceptable. Kappa values for interrater and intrarater reliability for sensation testing ranged from .72 to .83. Distri-

*Sciences Accessories, Inc, 970 Kings Hwy W, Southport, CT 06490.

Table 2. *Range-of-Motion Measurements (in Degrees) from Right and Left Sides of Patients with Diabetes (N = 31)*

Measurement	Right		Left	
	$\bar{X}$	*s*	$\bar{X}$	*s*
Calcaneal inversion	22	7	22	5
Calcaneal eversion	5	4	5	3
STJN[a]	3	2	5	3
FF/RF[b]	1	3	1	3
DF[c]	5	4	4	4
Calcaneal stance	0	2	0	2
Tibial varum	9	2	8	2

[a]STJN = subtalar joint neutral (varus).

[b]FF/RF = forefoot-to-rearfoot relationship (varus).

[c]DF = dorsiflexion.

Table 3. *Interrater and Intrarater Intraclass Correlation Coefficients (ICCs) and Standard Errors of Measurement (SEMs) for Seven Biomechanical Measurements Obtained from Patients with Diabetes*

	Interrater (n = 31)				Intrarater (n = 25)			
	Left		Right		Left		Right	
Measurement	ICC	SEM	ICC	SEM	ICC	SEM	ICC	SEM
Calcaneal inversion	.89	3	.86	3	.96	2	.92	2
Calcaneal eversion	.78	4	.79	2	.96	1	.96	1
STJN[a]	.79	2	.62	3	.96	1	.74	0
FF/RF[b]	.58	3	.77	2	.93	1	.91	1
DF[c]	.87	2	.74	3	.96	1	.89	3
Calcaneal stance	.75	1	.84	1	.92	1	.91	2
Tibial varum	.66	1	.62	1	.84	1	.86	1

[a]STJN = subtalar joint neutral.

[b]FF/RF = forefoot-to-rearfoot relationship.

[c]DF = dorsiflexion.

bution of patients according to sensation measurements is summarized in Table 4. The mean ulcer area and depth were 1.0 ± 0.9 cm^2 and 3.4 ± 2.4 mm, respectively. The ICC value of interrater and intrarater reliability for ulcer area was .99. The ICC values for interrater and intrarater reliability of ulcer depth measurements were .92 and .99, respectively. The SEM for ulcer depth and ulcer area was zero.

Discussion

Two previous studies examined the reliability of foot and ankle goniometric measurements in patient populations.[17,18] Our study differs from those of Elveru et al[17] and Pandya et al[18] in two basic ways. Our study involved only two testers (Elveru et al's study[17] involved 14 testers, and Pandya et al's study[18] involved 5 testers), and the testers had undergone extensive training with the measures involved in our evaluation. Elveru et al's examiners had limited or no experience in measuring the STJN position or passive STJ ROM. Pandya et al's examiners were "five experienced therapists," but there is no mention of examiner training with the measurements used. Elveru et al's study reported poor interrater reliability for STJN and passive STJ ROM measurements, but this finding is not surprising because the therapists had limited or no experience in these measures.

We attempted to minimize our measurement variability by undergoing extensive training in the measures being used. Our results indicated acceptable interrater reliability for all except four measurements (Tab. 3). Our results indicate that the same rater may reliably take foot and ankle measurements in this patient population. Different measurers, with training, may be reliable with certain foot and ankle measurements in this population. The extensive training our examiners have undergone, however, infers the generalization of our results only to other examiners who have undertaken similar training. We believe the most important aspects of these training sessions were defining and agreeing on common techniques of measurement. Although clinicians may believe they use similar measurement techniques, comparing outcome measurements often proves otherwise. The training sessions offered an opportunity to discover and correct these differences, which decreased variability in the measure. Although our training encompassed an 18-month time period, a shorter time period with more frequent sessions may achieve similar results.

Our acceptable reliability findings using the SW monofilaments to measure sensation agree with the results of Birke and Sims[19] and Holewski et al.[20] Sensation measurements obtained with SW monofilaments are reliable, require minimal training to perform, and appear to be a valid indicator of sensory loss.[8,19,20] We also found that measuring wound area by tracing the circumference of the ulcer onto sterilized x-ray film[21] was extremely reliable and offered a good comparison of wound size for future

Table 4. *Number of Patients with Diabetes Who Demonstrated Sensitivity to Specific Sizes of Semmes-Weinstein Monofilaments on Plantar Surfaces of Their Right and Left Feet (N = 31)*

Monofilament Size	Right	Left
4.17	0	0
5.07	9	9
6.10	11	9
>6.10	11	13

office visits. Unlike the biomechanical measures, taking reliable wound measurements appears to require minimal training.

Interestingly, although several ICC values were unacceptable, we found relatively low SEMs. The SEM is useful in the interpretation of the reliability of a single score. Thus, the SEM is a useful statistic in clinical settings for estimating the range of values within which a patient's "true" value may lie. Caution must be exercised when interpreting the SEM because this measure is "end sensitive." That is, as the obtained value approaches the extremes of the distribution, the SEM will increase dramatically. Nevertheless, if values obtained in applied settings are near the means obtained in this study, the low SEMs can be interpreted to mean that a patient's true value may lie within a very small range of the obtained value. For example, if a clinician measures calcaneal inversion at 22 degrees (the mean obtained in this study) in a diabetic patient, the true value can be interpreted to fall within a range of 18 to 26 degrees (using the obtained value of ± 2 SEMs).

The magnitude of the SEM is directly related to the standard deviation and indirectly related to the correlation coefficient between the test-retest values. As the correlation coefficient approaches 1, the SEM approaches zero. As the standard deviation decreases, however, so does the SEM. The range of values for most of the measurements obtained in this study is relatively low. As Lahey et al[23] have pointed out, restricted ranges have a deleterious effect on obtained reliability coefficients. Thus, we are left with contradictory results: although the reliability coefficients at times appear to be unacceptable, the obtained SEMs are certainly within acceptable and useful ranges for the clinician. For example, the measure of tibial varum had low interrater reliability coefficients (ICC = .66 and .62 for left and right sides, respectively), although it had acceptable intrarater reliability (ICC = .84 and .86 for left and right sides, respectively). The SEM for both intrarater and interrater tibial varum measurements, however, was 1 degree (Tab. 3). Even with a conservative estimate of ± 2 SEMs, a clinician obtaining a tibial varum measurement of 9 degrees should expect the "true" value to fall within a range of 7 to 11 degrees for both interrater and intratester situations. Consequently, we believe that judging the reliability of some measurements obtained in this study needs further clarification beyond failing to meet the predetermined ICC acceptability level of ≥ .70.

Alternative explanations for obtaining low reliability coefficients should be ruled out before a measure is judged unreliable. Lahey et al suggest ruling out the following explanations before concluding that a measure is unreliable: 1) sampling from a restricted range of values, thus decreasing between intrarater and interrater variability and artificially decreasing the obtained ICC value; 2) interrater interactions, usually resulting when one rater misinterprets instructions (eg, consistently misreads the goniometer in the opposite direction); and 3) low intertrial correlations.[23] Accordingly, we have attributed the low ICC values obtained in this study to the small range of measurements sampled and not to unreliability of the measurement. As mentioned previously, Lahey et al noted that data with a restricted range of values have a deleterious effect on the obtained reliability coefficients.[23] Furthermore, they point out that low reliability coefficients obtained from values in a restricted range should not necessarily be interpreted as unreliable. Instead, they suggest stratifying the sample to increase the range of values. This strategy was not practical in our study. Instead, we can only point out that although some reliability coefficients did not meet our predetermined level of acceptability, this fact alone does not infer that the measure is unreliable. We have outlined this situation with interrater assessment of tibial varum measurements, and we point out that analogous situations exist in all other measures with ICC values below .70.

Conclusion

A foot and ankle evaluation is an important consideration for diabetic patients because of the frequency of injury or disability of this area of the body. We have shown that in the test-retest situation of this foot and ankle evaluation, the SEMs of all measures were sufficiently low for them to be clinically useful. These results indicate that therapists who train extensively can take many clinical foot and ankle measurements in the diabetic population reliably.

References

1 Levin ME: The Diabetic Foot, ed 4. St Louis, MO, C V Mosby Co, 1988

2 Brand PW: The diabetic foot. In Ellenberg M, Rifkin H (eds): Diabetes Mellitus: Theory and Practice, ed 3. New Hyde Park, NY, Medical Examination Publishing Co Inc, 1983, pp 829–849

3 Boulton AJ, Betts RP, Franks CI, et al: The natural history of foot pressure abnormalities in neuropathic diabetic subjects. Diabetes Res 5:73–77, 1977

4 Grigic A, Rosenbloom AL, Weber TF, et al: Joint contractures common manifestation of childhood diabetes. J Pediatr 88:584–588, 1976

5 Rosenbloom AL, Silverstein JH, Lexotte DC, et al: Limited joint mobility in childhood diabetics indicates increased risk for microvascular diseases. N Engl J Med 305:191–194, 1982

6 Rosenbloom AL, Silverstein JH, Riley WJ, et al: Limited joint mobility in childhood diabetics: Family studies. Diabetes Care 6:370–373, 1983

7 Shinabarger NA: Limited joint mobility in adults with diabetes mellitus. Phys Ther 67:215–218, 1987

8 Mueller MJ, Diamond JE, Delitto A, et al: Insensitivity, limited joint mobility, and plantar ulcers in patients with diabetes mellitus. Phys Ther 69:453–459, 1989

9 Birke JA, Cornwall MW, Jackson M: Relationship between hallux limitus and ulceration of the great toe. Journal of Orthopaedic and Sports Physical Therapy 10:172–176, 1988

10 Kaye JM, Sorto LA: The k-square: A new biomechanical measuring device for the foot and ankle. J Am Podiatr Assoc 69:58–64, 1979

11 Baldwin EB, Graebner JE: A comparison of the k-square and tractograph. J Am Podiatr Assoc 72:629–633, 1982

12 Muwanga CL, Dove AF, Plant GR: The measurement of ankle movements: A new method. Injury 16:312–314, 1985

13 Laftanza L, Gray GW, Kantner RM: Closed versus open kinematic chain measurements of subtalar joint eversion: Implications for clinical practice. Journal of Orthopaedic and Sports Physical Therapy 9:310–314, 1988

14 McPoil TG, Schuit D, Knecht HG: A comparison of three positions used to evaluate tibial varum. J Am Podiatr Med Assoc 78:22–28, 1988

15 Lohmann KN, Rayhel HE, Schneiderwind WP, et al: Static measurement of tibia vara: Reliability and effect of lower extremity position. Phys Ther 67:196–202, 1987

16 Boone DC, Azen SP, Lin C-M, et al: Reliability of goniometric measurements. Phys Ther 58:1355–1360, 1978

17 Elveru RA, Rothstein JM, Lamb RL: Goniometric reliability in a clinical setting: Subtalar and ankle joint measurements. Phys Ther 68:672–677, 1988

18 Pandya S, Florence JM, King WM, et al: Reliability of goniometric measurements in patients with Duchenne muscular dystrophy. Phys Ther 65:1339–1342, 1985

19 Birke JA, Sims DS: Plantar sensory threshold in the Hansen's disease ulcerative foot. In: Proceedings of the International Conference on Biomechanics and Clinical Kinesiology of Hand and Foot. Madras, India, December 16–18, 1985, pp 137–138

20 Holewski JJ, Stress RM, Graf PM, et al: Aesthesiometry: Quantification of cutaneous pressure sensation in diabetic peripheral neuropathy. J Rehabil Res Dev 25:1–10, 1988

21 Bohannon RW, Pfaller BA: Documentation of wound surface area from tracings of wound perimeters: Clinical report of three techniques. Phys Ther 63:1622–1624, 1983

22 McPoil TG, Brocato RS: The foot and ankle: Biomechanical evaluation and treatment. In Gould JA, Davies GJ (eds): Orthopaedic and Sports Physical Therapy. St Louis, MO, C V Mosby Co, 1985, vol 2, pp 313–341

23 Lahey MA, Downey RG, Saal FE: Intraclass correlations: There's more there than meets the eye. Psychol Bull 93:586–595, 1983

24 Cohen J: A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:27–36, 1960

25 Landis RJ, Koch GG: The measurement of observer agreement for categorical data. Biometrics 33:159–174, 1977

26 Anastasi A: Psychological Testing, ed 6. New York, NY, Macmillan Publishing Co, 1988, pp 133–135