IJSPT

ORIGINAL RESEARCH

RELIABILITY OF ANKLE-FOOT MORPHOLOGY, MOBILITY, STRENGTH, AND MOTOR PERFORMANCE MEASURES

John J. Fraser, PT, PhD, OCS[1]
Rachel M. Koldenhoven, MEd, ATC[1]
Susan A. Saliba, PT, PhD, ATC[1]
Jay Hertel, PhD, ATC[1]

ABSTRACT

Background: Assessment of foot posture, morphology, intersegmental mobility, strength and motor control of the ankle-foot complex are commonly used clinically, but measurement properties of many assessments are unclear.

Purpose: To determine test-retest and inter-rater reliability, standard error of measurement, and minimal detectable change of morphology, joint excursion and play, strength, and motor control of the ankle-foot complex.

Design: Reliability study.

Methods: 24 healthy, recreationally-active young adults without history of ankle-foot injury were assessed by two clinicians on two occasions, three to ten days apart. Measurement properties were assessed for foot morphology (foot posture index, total and truncated length, width, arch height), joint excursion (weight-bearing dorsiflexion, rearfoot and hallux goniometry, forefoot inclinometry, 1st metatarsal displacement) and joint play, strength (handheld dynamometry), and motor control rating during intrinsic foot muscle (IFM) exercises. Clinician order was randomized using a Latin Square. The clinicians performed independent examinations and did not confer on the findings for the duration of the study. Test-retest and inter-tester reliability and agreement was assessed using intraclass correlation coefficients ($ICC_{2,k}$) and weighted kappa (K_w).

Results: Test-retest reliability ICC were as follows: morphology: .80-1.00, joint excursion: .58-.97, joint play: -.67-.84, strength: .67-.92, IFM motor rating: K_w -.01-.71. Inter-rater reliability ICC were as follows: morphology: .81-1.00, joint excursion: .32-.97, joint play: -1.06-1.00, strength: .53-.90, and IFM motor rating: K_w .02-.56.

Conclusion: Measures of ankle-foot posture, morphology, joint excursion, and strength demonstrated fair to excellent test-retest and inter-rater reliability. Test-retest reliability for rating of perceived difficulty and motor performance was good to excellent for short-foot, toe-spread-out, and hallux exercises and poor to fair for lesser toe extension. Joint play measures had poor to fair reliability overall. The findings of this study should be considered when choosing methods of clinical assessment and outcome measures in practice and research.

Level of evidence: 3

Key Words: Assessment, examination, intrinsic foot muscles, manual therapy, repeatability

[1] Department of Kinesiology, University of Virginia, 210 Emmet Street South, Charlottesville, VA 22904-4407, USA

Conflicts of Interest: None

Presented in part at the American Physical Therapy Association, Combined Sections Meeting, February 18, 2017, San Antonio, TX and is archived at 10.7490/f1000research.1114545.1

CORRESPONDING AUTHOR
John J. Fraser,
Warfighter Performance Department, Naval Health Research Center, 140 Sylvester Road
San Diego, CA 92106
Phone: 757-438-0390
E-mail: john.j.fraser8.mil@mail.mil

INTRODUCTION

Traumatic and overuse injuries of the ankle and foot are frequently incurred in sport,[1,2] physical training,[3] and the workplace.[4] The most common lower extremity injury treated is the lateral ankle sprain,[5] with more than two million individuals injuring their lateral ankle annually in the United States.[6] Similarly, plantar fasciitis is a frequently occurring overuse injury in the foot that is responsible for more than one million ambulatory care visits in the United States per annum.[7] Clinical practice guidelines recommend physical examination of the ankle-foot complex to include observation of foot morphology and posture,[8–10] palpation,[8–11] range of motion,[8–10] test of joint play to assess ligamentous integrity,[9,11] and strength[9] in the assessment of these patients.

While test-retest and inter-rater reliability of some commonly utilized assessment measures of the ankle-foot complex have been studied, measurement properties for many others have yet to be established. The authors of a systematic review of ankle-foot examination articles published from 1966-2006 found that only a few studies rigorously assessed measurement properties such as reliability of ankle-foot posture, morphology, multisegmental joint mobility, strength, and motor function.[12] A ramification of imprecise physical examination measurements is the inability to distinguish actual clinical change from change resulting from random error. Consequences of measurement uncertainty are decreased sensitivity, specificity, and prognostic accuracy when assessing impairment throughout the disease or injury course or when tracking effectiveness of therapeutic interventions. Reliability of new, innovative examination measures of multisegmented ankle and foot motion, strength, and motor control[13,14] also need to be assessed. Novel measures of intersegmental mobility of the foot and neuromotor function of the intrinsic foot muscles (IFM) have been suggested to be clinically important in the assessment of the foot functions of shaping, force attenuation and transmission, and postural control.[14,15]

The purpose of this study was to determine test-retest and inter-rater reliability, standard error of measurement, and minimal detectable change of morphology, joint excursion and play, strength, and motor control of the ankle-foot complex in healthy, recreationally active individuals. Reliability of participant-reported task difficulty during short-foot and toe posture exercises were also assessed.

METHODS

Design

A reliability study was performed using a sample of convenience in which the independent variables were clinician (novice and experienced) and session (baseline and reassessment). Inter-rater and test-retest reliability, SEM, and MDC were assessed for each clinical measure of foot morphology, joint mobility, strength, and motor performance.

Participants

Data from 24 healthy, recreationally active adults aged 18-38 years were included (12 males, 12 females; mean age 21.5 ± 4.8 years; BMI 23.5 ± 2.9 kg/m^2). Participant demographic information and self-report measures are detailed in Table 1. "Recreationally active" was defined as participation in some form of physical activity for at least 20 minutes per day, at least three times a week. Individuals were excluded if they had any history of ankle or foot sprain, fracture in the leg or foot, disability secondary to lower extremity neuromuscular functional impairment, neurological or vestibular disorders that affected balance, diabetes mellitus, lumbosacral radiculopathy, a soft tissue disorder such as Marfan or Ehlers-Danlos syndrome, any absolute contraindication to manual therapy, or were pregnant. Participants who met inclusion criteria provided informed consent and the study was approved by the Institutional Review Board. Figure 1 is a flowsheet illustrating recruitment, retention, and time points for this study.

Assessors

All participants were evaluated by two clinicians. The first clinician (Tester 1) was an athletic trainer (height = 162.6-cm, mass = 59.0 kg, surface area of the palmar hand = 159.0-cm^2) with two years of clinical experience. The second clinician (Tester 2) was a physical therapist (height = 180.3-cm, mass = 88.5 kg, surface area of the palmar hand = 221.0-cm^2) with 14 years of clinical experience and was a board certified

Table 1. *Participant Demographic and Self-Reported Measures.*

		Mean (SD)			
		Male (n=12)		Female (n=13)	
Age (years)		23.3 (6.4)		19.8 (1.2)	
Height (cm)		176.0 (6.9)		162.6 (12.1)	
Mass (kg)		74.6 (9.4)		60.3 (9.0)	
BMI		24.0 (2.0)		23.0 (3.5)	
Godin Leisure Time Questionnaire		88.2 (24.6)		71.8 (26.7)	
Foot and Ankle Ability Measure	ADL	99.8 (0.6)		100 (0)	
	ADL SANE	99.8 (0.6)		100 (0)	
	Sport	99.8 (0.9)		99.5 (1.7)	
	Sport SANE	99.9 (0.3)		99.6 (1.4)	
Identification of Functional Ankle Instability (IdFAI)		0.6 (1.4)		1.6 (2.4)	
Tampa Scale of Kinesiophobia		16.0 (4.3)		15.4 (2.9)	
Veterans RAND 12-Item Health Survey	Physical Composite	56.6 (3.2)		58.0 (2.5)	
	Mental Composite	48.0 (0.8)		48.7 (1.2)	
	Physical Function	100 (0)		100 (0)	
	Role Physical	96.9 (7.8)		98.1 (6.9)	
	Bodily Pain	95.8 (9.7)		100 (0)	
	General Health	89.6 (16.7)		80.8 (11.0)	
	Vitality	68.8 (21.7)		59.6 (21.7)	
	Social Function	91.7 (16.3)		96.2 (9.4)	
	Emotional Role	99.0 (3.6)		94.2 (9.7)	
	Mental Health	79.2 (12.3)		67.3 (20.8)	
Time spent barefoot per day n (% sample)	0-29 minutes	5 (41.7)		7 (53.8)	
	30-59 minutes	5 (41.7)		3 (23.1)	
	> 60 minutes	2 (16.7)		3 (23.1)	
Foot Posture Index n (% sample)		Right	Left	Right	Left
	Highly Pronated	2 (16.7)	3 (25.0)	3 (23.1)	2 (15.4)
	Pronated	3 (25.0)	4 (33.3)	6 (46.2)	3 (23.1)
	Normal	4 (33.3)	3 (25.0)	2 (15.4)	6 (46.2)
	Supinated	3 (25.0)	2 (16.7)	2 (15.4)	2 (15.4)
	Highly Supinated	0 (0)	0 (0)	0 (0)	0 (0)

SD=standard deviation, cm=centimeters, kg=kilograms, BMI=body mass index, ADL=activities of daily living, SANE=single assessment numeric evaluation

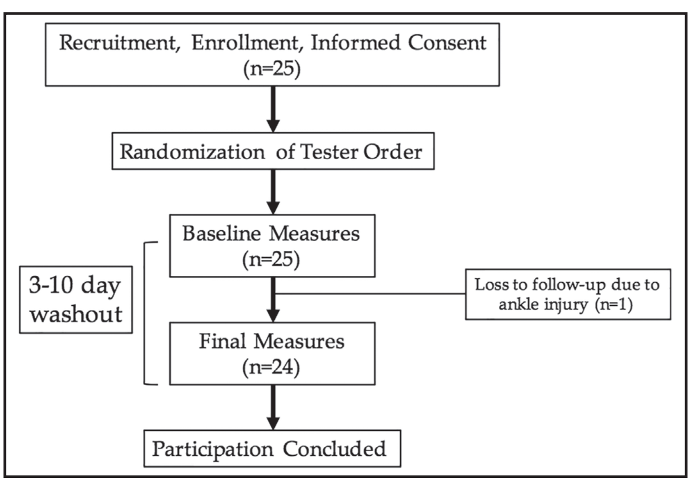

Figure 1. *Flowsheet illustrating recruitment, retention, and study time points.*

orthopaedic clinical specialist. Both clinicians were right-hand dominant and trained in morphologic assessment of the foot, goniometry, inclinometry, handheld dynamometry, gross motor assessment, and assessment of joint play and employed these skills regularly in practice.

Procedures

Prior to participant recruitment, both clinicians reviewed assessment procedures and performed collaborative trial assessments together to ensure agreement on examination technique and interpretation. Each clinician performed the examinations independently and did not confer on the findings. Clinician order was randomized using a Latin-square. Due to the potential influence on foot morphology and intrinsic foot muscle strength, participants were asked the amount of time spent barefoot daily. Participants provided demographic information, health and injury history, and completed the Foot and Ankle Ability Measure (FAAM) Activities of Daily Living[16] and Sport subscales,[17] Identification of Functional Ankle Instability (IdFAI),[18] Veterans Rand 12-item Health Survey (VR-12),[19] and the Godin Leisure-time Exercise Questionnaire.[20]

Morphologic Foot Assessment

Foot posture was assessed in standing using the Foot Posture Index–6 item version (FPI-6), a categorical measure of foot type that is based on five observations and one palpatory assessment.[21] Measurements of total and truncated foot length, arch height, and foot width were performed using the Arch Height Index Measurement System (JAKTOOL Corporation, Cranberry, NJ) in sitting and standing.

Joint Excursion Measures

Joint range of motion measures of rearfoot dorsiflexion, plantarflexion, inversion, and eversion were performed using a 30.5-cm transparent double arm plastic goniometer (Merck Corporation, Kenilworth, NJ). Forefoot inversion and eversion was measured using a digital inclinometer (Fabrication Enterprises, White Plains, NY) (Figure 2a) First metatarsal dorsiflexion and plantarflexion were measured utilizing a custom measuring device consisting of two metal rulers bent to 90° described by Greisberg and colleagues[22] (Figures 2b,c). The stationary arm of the device was constructed from a 16-cm metal ruler bent to 90-degrees. The moving arm of the device was cut to 10-cm (bent to 90-degrees at the 5-cm mark) and fastened to the stationary arm with two plastic zip ties and a rubber tensioner. First metatarsophalangeal flexion and extension were measured with a 17-cm double arm plastic goniometer with a semicircular scale (Upjohn Corporation, Kalamazoo, MI). The details to patient position and procedures for each joint excursion measure are outlined in Table 2. The total arc of motion within a plane was used for analysis of joint excursion.

Joint Play Motion

Proximal tibiofibular joint mobility was assessed for the presence or absence of hypomobility. Joint play was assessed using the 7-point Likert scale (0 = ankylosed, 1 = considerable hypomobility, 2 = slight hypomobility, 3 = normal, 4 = slight hypermobility, 5 = considerable hypermobility, 6 = unstable) developed for quantification of passive mobility intervertebral motion by Gonnella and colleagues.[23] Details to patient position and procedures for each joint play measure are outlined in Table 3.

Strength and Motor Function

Muscle strength was assessed with the MicroFET2 digital handheld dynamometer (Hoggan Health

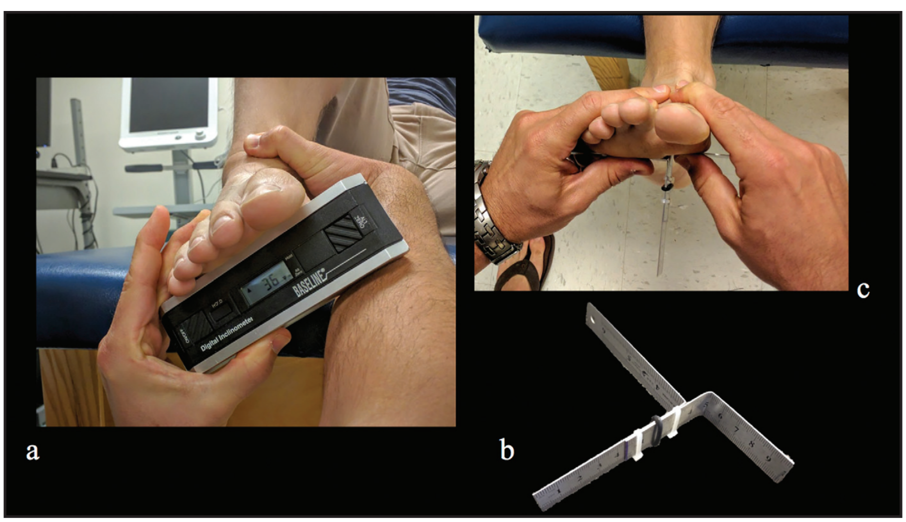

Figure 2. *a. Measure of forefoot on rearfoot inversion/eversion excursion. b. Custom measuring device and c. illustration of measurement of first metatarsal dorsiflexion and plantarflexion.*

Table 2. *Clinical Measures of Joint Excursion and Strength.*

		Measure	Instrument	Position	Description of Procedure
Joint Excursion	Rearfoot	DF/PF	30.5cm goniometer	Supine	Stationary arm: Midline lateral leg; Axis: Lateral malleolus; Moving arm: Lateral foot
		Inversion/ Eversion		Prone	Stationary arm: Midline posterior leg; Axis: Subtalar joint; Moving arm: Midline posterior calcaneus
		Standing DF	Paper tape measure	Standing Lunge	As described by Bennel (1998)
	Forefoot	Inversion/ Eversion	Inclinometer	Hook lying	The foot is cantilevered over the plinth edge. Rearfoot is manually stabilized by cupping and gripping the calcaneus. Inclinometer is aligned across the plantar metatarsal head. The forefoot is maximally inverted and everted on the rearfoot.
		First TMT DF/PF	Modified ruler	Supine	As described by Greisberg (2010)
	Hallux	Flexion/ Extension	17cm goniometer		Stationary arm: Midline lateral first metatarsal; Axis: Lateral metatarsal head Moving arm: Lateral proximal phalanx
Strength	Ankle	DF	Handheld Dynamometer	Supine	Tested in neutral DF/PF. Force measured at the metatarsal heads. Mobilization belt wrapped around assessor's pelvis for DF measurement and plinth leg during PF measurement to assist in providing counter force. The shank was manually stabilized
		PF		Prone	
		Inversion		Supine	Tested with foot and ankle in neutral. Force measured at the metatarsal heads. Shank was manually stabilized.
		Eversion			
		Hallux flexion		Hook lying	The foot was flat and positioned with the metatarsal heads at the plinth edge with the toe cantilevered off the end. The dorsal foot was manually stabilized. Force was measured at the pads of the toe.
		Lesser toe flexion			

DF = dorsiflexion PF = plantarflexion TMT = Tarsometatarsal

Table 3. *Clinical Measures of Joint Play Motion.*

	Measure	Instrument	Position	Description of Procedure
Shank	Proximal tibiofibular	+/- hypomobility	Seated	The clinician palpates the joint line of the proximal tibiofibular joint as the patient actively cycles through DF/PF of the foot.
	Distal tibiofibular posterior glide	Passive Mobility Scale (Gonella 1982)	Supine	The heel is cupped in the treating clinicians stabilizing hand. The clinician contacts the anterior lateral malleolus using the thenar eminence of the mobilizing hand and applies an anterior-posterior force.
Rearfoot	Talar anterior and posterior glide			With the foot cantilevered over the plinth edge, the clinician stabilizes the shank and cups the calcaneus and talus using a C-grip. An anterior force is applied through the calcaneus for assessment of anterior glide; a posterior force is applied through the anterior talus to assess posterior glide.
	Inversion and eversion			With the foot cantilevered over the plinth edge, the clinician stabilizes the shank and cups the calcaneus and talus using a C-grip. A medial directed rotatory force is applied through the calcaneus for assessment of inversion; a lateral directed rotatory force is applied through the calcaneus for assessment of eversion
	Medial and lateral glide		Side-lying	With the foot cantilevered over the plinth edge, the clinician stabilizes the shank and cups the calcaneus and talus using a C-grip. A medial or lateral directed force is applied through the calcaneus.
Midfoot	Inversion and eversion		Supine	The foot is cantilevered over the plinth edge. Rearfoot is manually stabilized by using a C-grip around the calcaneus and talus. An inversion or eversion force is applied at the distal metatarsals.
	Abduction and Adduction			The foot is cantilevered over the plinth edge. Rearfoot is manually stabilized by using a C-grip around the calcaneus. A medial or lateral directed force is applied at the distal metatarsals.
Forefoot	First TMT dorsal and plantar glide		Hook-lying	With the foot flat on the plinth, the first cuneiform is manually stabilized. The base of the first metatarsal is gripped with the mobilizing hand and a dorsal force is applied. The procedure is repeated with a plantar directed force.
Hallux	First MTP dorsal and plantar glide			The great toe is cantilevered over the plinth edge. The first metatarsal is manually stabilized. The base of the first proximal phalanx is gripped with the mobilizing hand; a distraction force is applied followed by a dorsal directed glide. The procedure is repeated with a distraction and plantar directed force.

TMT = Tarsometatarsal MTP = Metatarsophalangeal DF = Dorsiflexion PF = Plantarflexion

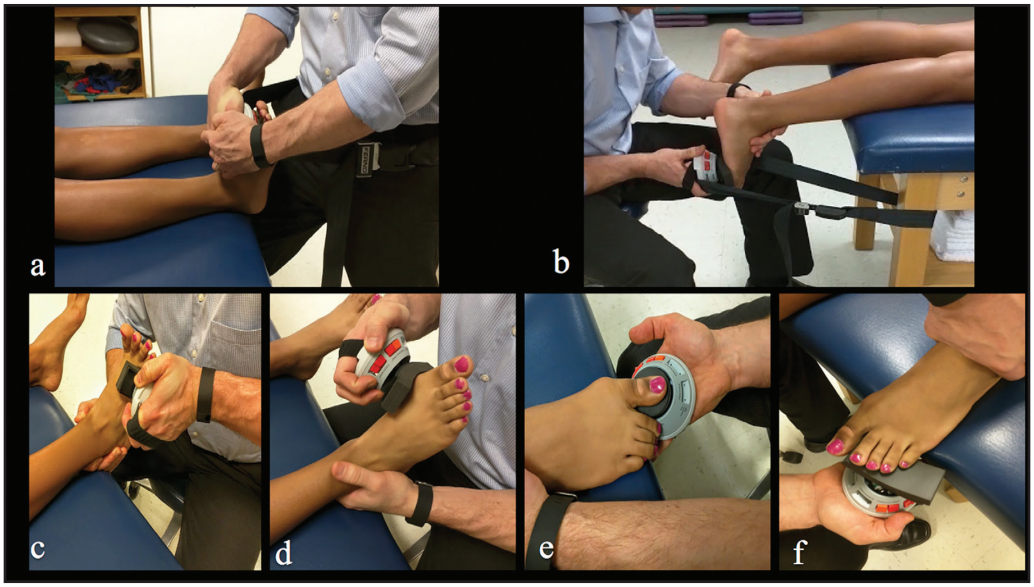

Figure 3. *Testing of a. dorsiflexion, b. plantarflexion, c. inversion, d. eversion, e. hallux flexion, and f. lesser toe flexion strength using a handheld dynamometer.*

Industries, West Jordan, UT). Details to patient position and procedures for each joint strength measure are outlined in Table 2 and illustrated in Figure 3. Strength measures were based on a single trial of a "make test" and reported in Newtons (N). In the case of an invalid trial (due to equipment difficulty, deviation from test position, or substitution motion), the participant rested prior to retesting to mitigate effects from fatigue. The IFM test was performed and graded using the scale (1 = Poor; 2 = Fair; 3 = Satisfactory) described by Jam.[13]

Motor performance and participant's rating of perceived difficulty were assessed during the medial longitudinal arch draw up maneuver (short-foot exercise), the toe-spread-out exercise, hallux extension, and lesser toe extension exercise. These exercises are employed clinically and have scant evidence to support their use in treatment of conditions of the foot and ankle.[24–26] The measurement properties of novel assessments of motor performance and task difficulty during IFM exercises need to be established before these interventions can be tested empirically. Motor performance was assessed using a scale adapted from the gross motor assessment developed by Bérard and colleagues[27] (0 = does not initiate movement or starting position cannot be maintained; 1 = partially completes the exercise; 2 = completes the exercise with compensations, slowness or obvious clumsiness; 3 = completes the exercise with a standard pattern). Perceived difficulty was assessed by asking the participant to rate the task using a 5-point Likert scale (1 = Very easy; 2 = Somewhat easy; 3 = Neutral; 4 = Somewhat Difficult; 5 = Very Difficult). The exercises were performed sitting barefoot, with the foot in contact with the floor. During the short-foot exercise, the participant was instructed to draw the medial longitudinal arch up while maintaining the metatarsal heads, toes, and heel in contact with the ground. This maneuver was performed correctly if there was an approximation of the calcaneus and first metatarsal head resulting in a shortening of the foot. The toe-spread-out exercise was performed sequentially by extending all the toes, followed by abduction, hallux flexion, and little toe flexion. Hallux extension was performed by extending the first metatarsophalangeal joint while maintaining the lesser toes (2-5) in contact with the floor. Lesser toe extension was performed by extending toes 2-5 while maintaining the hallux in contact with the ground. Video of the exercises can be accessed at https://goo.gl/ugffZ8. Patients were verbally instructed in the maneuvers and guided through a practice trial before assessment. Following instruction, the participant performed the exercise and motor performance was assessed. The participant was allowed a second attempt if motor performance was sub-optimal (rated < 3) on the first trial. Motor performance and rating of perceived difficulty were recorded immediately following each task.

Statistical Analysis

The level of significance was set *a priori* at p ≤0.05 for all analyses. *A priori* sample size estimation of 14 participants were needed based on two clinician measurements per variable, a reliability of ≥.70 considered desirable, an $\alpha = .05$, and $\beta = .20$.[28] Group descriptive statistics were calculated for participant demographic information and self-reported measures. Test-retest and inter-rater reliability of variables measured on a continuous scale or an ordinal scale with at least five items[29,30] were assessed with intraclass correlation coefficients ($ICC_{2,k}$), with > .75 interpreted as being excellent, .40-.75 as fair to good, and < .40 as poor.[31] Measures with negative ICCs were interpreted as having systematic disagreement.[32]

Linear weighted *Kw* statistics were used to assess test-retest and intertester agreement for measures of hypomobility in the proximal tibiofibular joint, motor performance during intrinsic foot exercises, and the intrinsic foot muscle test with agreement interpreted as almost perfect from 0.81-1.00; 0.61-0.80 as substantial; 0.41-0.60 as moderate; 0.21-0.40 as fair; 0.00-0.20 as slight, and < 0.00 as poor.[33] Descriptive statistics, ICC, and *Kw* estimates were computed using Statistical Package for Social Sciences (SPSS) Version 23.0 (SPSS, Inc., Chicago, IL). SEM and MDC were calculated from the mean variance of bilateral measures and both clinicians from visit one and the mean of ICC values for test-retest reliability for both limbs measured by both assessors using Microsoft Excel for Mac Version 15 (Microsoft Corp., Redmond, WA).

RESULTS

Morphologic Foot Measures

Test-retest and inter-rater reliability was found to be excellent (.81-1.00) for the FPI and morphologic measures of foot length, truncated foot length, foot width, and dorsal arch height (Table 4). The mean FPI scores were consistent between assessors and between visits (within 1 point on a 25-point scale). SEM was 2 points (rounded to the next integer) and MDC was 5 points (rounded). Group means were consistent between assessors and laboratory

Table 4. *Reliability of Foot Morphologic Measures.*

		Group Means (SD)																	
		Baseline				Reassessment						Inter-rater Reliability				Test-Retest Reliability			
		Tester 1		Tester 2		Tester 1		Tester 2				Baseline		Reassessment		Tester 1		Tester 2	
		Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt	SEM	MDC	Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt
Unloaded	Foot Length (cm)	25.3 (2.0)	25.2 (2.0)	25.2 (2.0)	25.1 (2.1)	25.3 (2.0)	25.3 (2.0)	25.2 (2.1)	25.2 (2.0)	.1	.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Truncated Foot Length (cm)	18.6 (1.5)	18.3 (1.4)	18.6 (1.5)	18.5 (1.5)	18.6 (1.5)	18.4 (1.5)	18.5 (1.6)	18.4 (1.4)	.2	.4	.98	.98	.99	.98	.99	.99	.98	.99
	Foot Width (mm)	92.4 (6.7)	91.7 (5.7)	90.8 (6.7)	89.8 (6.0)	92 (6.7)	91.6 (5.9)	91.0 (6.7)	89.9 (6.3)	.9	2.5	.98	.97	.96	.95	.99	.96	.98	.98
	Dorsal Arch Height (cm)	6.5 (0.6)	6.5 (0.6)	6.6 (0.6)	6.6 (0.6)	6.6 (0.6)	6.6 (0.6)	6.7 (0.6)	6.7 (0.6)	.1	.2	.98	.98	.98	.99	.99	.98	.97	.97
Loaded	Foot Length (cm)	25.8 (2.0)	25.7 (2.0)	25.6 (2.0)	25.6 (2.0)	25.8 (2.0)	25.8 (2.1)	25.7 (2.0)	25.6 (2.0)	.1	.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Truncated Foot Length (cm)	19.1 (1.5)	18.7 (1.5)	18.9 (1.4)	18.8 (1.5)	19.1 (1.5)	18.8 (1.5)	18.9 (1.4)	18.8 (1.4)	.2	.4	.98	.99	.99	.98	.99	.99	.98	.99
	Foot Width (mm)	94.9 (6.9)	94.1 (5.9)	92.8 (6.9)	92.3 (6.3)	94.8 (6.4)	949.4 (6.2)	93.0 (6.7)	92.7 (6.0)	.9	2.5	.99	.98	.99	.99	.99	.99	.99	.98
	Dorsal Arch Height (cm)	6.0 (0.6)	6.0 (0.6)	6.1 (0.7)	6.1 (0.6)	6.1 (0.7)	6.1 (0.6)	6.2 (0.7)	6.1 (0.6)	.1	.2	.99	.98	.99	1.00	.99	.99	.99	.98
	Foot Posture Index*	4.8 (3.3)	5.0 (3.3)	5.3 (4.6)	4.7 (4.5)	5.6 (2.7)	5.6 (2.9)	6.4 (3.9)	5.3 (3.5)	1.6	4.5	.83	.81	.86	.84	.86	.82	.83	.81

* The Foot Posture Index scale ranges from -12 indicating highly supinated to +12 indicating highly pronated. A score of 0-5 is considered a "normal" foot. SD=standard deviation, SEM=standard error of measurement, MDC=minimal detectable change, Rt=right, Lt=left

Poor Fair Good Excellent

Table 5. *Reliability of Joint Excursion Measures of the Ankle-Foot Complex.*

		Group Means (SD)										Inter-rater Reliability				Test-Retest Reliability			
		Baseline				Reassessment						Baseline		Reassessment		Tester 1		Tester 2	
		Tester 1		Tester 2		Tester 1		Tester 2											
		Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt	SEM	MDC	Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt
Rearfoot	Standing DF (cm)	13.0 (3.9)	13.2 (3.6)	12.8 (3.9)	12.8 (4.0)	13.2 (3.8)	13.5 (3.5)	13.0 (3.5)	12.2 (3.7)	.7	2.0	.96	.96	.97	.96	.97	.96	.97	.96
	Talocrural DF/PF (deg)	81.0 (7.4)	80.8 (8.7)	73.8 (9.9)	79.2 (12.6)	79.8 (9.2)	79.8 (8.5)	73.0 (11.8)	77.2 (12.9)	4	12	.81	.84	.85	.76	.81	.81	.88	.88
	Subtalar Inv/Ev (deg)	52.3 (11.2)	53.8 (8.3)	36.2 (13.1)	37.4 (10.5)	57.5 (12.1)	57.7 (11.1)	37.6 (10.5)	37.2 (12.9)	5	15	.69	.53	.69	.61	.71	.65	.73	.58
Forefoot	Forefoot Inv/Ev (deg)	70.1 (16.7)	70.6 (12.9)	49.4 (11.3)	48.2 (12.4)	69.4 (15.0)	69.8 (14.2)	52.2 (10.8)	50.7 (10.5)	5	15	.83	.86	.77	.66	.81	.72	.86	.86
	First MT DF/PF (mm)	15.9 (2.9)	16.8 (2.4)	15.5 (4.1)	14.0 (3.9)	15.4 (2.6)	16.0 (2.4)	15.5 (4.9)	13.5 (5.1)	1	4	.47	.32	.42	.53	.85	.62	.86	.90
	Hallux Ext/ Flex (deg)	144.0 (16.0)	144.6 (15.7)	140.2 (19.3)	146.7 (20.1)	144.2 (15.4)	147.3 (13.3)	137.1 (22.1)	142.8 (22.2)	6	15	.91	.85	.88	.89	.82	.90	.95	.93

SD=standard deviation, SEM=standard error of measurement, MDC=minimal detectable change, Rt=right, Lt=left, DF=dorsiflexion, PF=plantarflexion, Inv=inversion, Ev=eversion, MT=metatarsal, Ext=extension, Flex=flexion

Poor | Fair | Good | Excellent

visits for measures performed using the Arch Height Index instrument, with SEM ranging from 1-2 mm and MDC ranging from 2-4 mm.

Joint Excursion Measures

Measures of joint excursion had excellent reliability (.81-.97) in 67% of the measures performed by Tester 1 and in 83% of measures performed by Tester 2. Table 5 details the reliability of joint excursion measures of the ankle-foot complex. Group means for standing dorsiflexion were consistent between assessors and laboratory visits. The measure had low variability, resulting in a SEM of 7-mm and a 2-cm MDC. Goniometric and inclinometric measures were consistent between visits for each clinician, but group means were higher for Tester 1 for subtalar and forefoot frontal plane excursion as compared to Tester 2. Standard error for goniometric and inclinometric measures ranged from 4-6°, with MDC ranging from 12-15°. Mean first metatarsal excursion measures were relatively consistent between assessors and visits. There was greater variability observed in measurements performed by Tester 2 on both visits. Standard error for first metatarsal excursion measures was 1-cm, with a MDC of 4-cm.

Joint Play Measures

Test-retest reliability of joint play motion (-.67-.1.00) varied widely between clinicians; with the more experienced clinician (Tester 2) demonstrating greater consistency (53% of measures good to excellent) compared the novice clinician (Tester 1) (28% of measures good to excellent). Inter-rater reliability was poor (-1.06-.39) in 73% of joint play measures (Tables 6 and 7). Group means were relatively consistent between sessions. When comparing clinicians, group means for Tester 2 were lower in many joint play measures. SEM ranged from 0.4-0.6, with a MDC of a full grade for all but two joint play measures.

Strength Measures

Reliability of strength measures of the ankle-foot complex were found to have excellent test-retest reliability (.76-.88) in 58% of the measures performed by Tester 1 and 92% performed by Tester 2. Table 8 details the reliability of strength measures of the ankle-foot complex. SEM and MDC for all measures sans ankle plantarflexion ranged from 18.0-23.6 N with a MDC of 49.8-65.5 N. Ankle plantarflexion was found to have a SEM of 41.6 and a MDC of 115.2 N.

Rating of Perceived Difficulty and Measures of Motor Performance during IFM Exercises

Repeatability of reported task difficulty was highest for the short-foot (.75-.90) and hallux extension exercises (.87-.96) and more variable during toe-spread-out exercise (.61-.82) and the lesser toe extension (.16-.50) (Table 9). Participants reported substantial decreases in perceived difficulty during the short-foot (2.1-2.4 to 1.9-2.0), toe-spread-out (3.1-3.9 to

Table 6. *Reliability of Joint Play Motion Measures of the Shank and Rearfoot.*

		Group Means (SD)										Inter-rater Reliability				Test-Retest Reliability			
		Baseline				Reassessment						Baseline		Reassessment		Tester 1		Tester 2	
		Tester 1		Tester 2		Tester 1		Tester 2											
		Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt	SEM	MDC	Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt
Tibiofibular	Proximal Superior Translation (%)*	8.0	20.0	8.0	20.0	0	8.3	0	37.5	-	-	-.09	.25	1.00	-.16	1.00	-.14	1.00	.24
Tibiofibular	Distal Post Glide†	3.0 (0.5)	3.0 (0.4)	3.0 (0.4)	2.5 (0.7)	3.0 (0.3)	3.0 (0.4)	3.1 (0.5)	2.5 (0.6)	.4	1.1	.00	-.80	.00	.51	.40	-.67	.64	.75
Talocrural	Anterior Glide†	3.2 (0.6)	3.3 (0.6)	2.8 (0.7)	2.7 (0.5)	3.2 (0.5)	3.3 (0.6)	2.8 (0.5)	2.8 (0.7)	.5	1.3	.71	.26	.67	.36	.44	.62	.64	.43
Talocrural	Posterior Glide†	2.9 (0.4)	2.9 (0.5)	2.5 (0.8)	2.5 (0.6)	2.9 (0.3)	2.9 (0.3)	2.7 (0.7)	2.8 (0.4)	.4	1.2	.01	-.77	.36	-.53	.46	.30	.61	.49
Talocrural	External Rotation†	3.1 (0.5)	3.0 (0.5)	2.8 (0.6)	2.9 (0.4)	3.1 (0.4)	3.2 (0.5)	3.0 (0.6)	3.1 (0.4)	.5	1.4	.09	.05	.04	-.19	.24	-.19	.21	.11
Talocrural	Internal Rotation†	2.8 (0.6)	2.8 (0.4)	2.9 (0.5)	3.1 (0.6)	2.9 (0.3)	3.2 (0.2)	3.1 (0.6)	3.3 (0.6)	.5	1.3	-1.06	-.48	.09	-.30	.18	-.19	.69	.27
Subtalar	Inversion†	3.6 (0.6)	3.7 (0.5)	3.0 (0.5)	3.3 (0.7)	3.5 (0.6)	3.5 (0.5)	3.1 (0.4)	3.3 (0.6)	.4	1.0	.56	.43	.52	-.09	.62	.68	.40	.42
Subtalar	Eversion†	2.9 (0.3)	2.8 (0.4)	2.5 (0.8)	2.9 (0.4)	3.1 (0.3)	3.0 (0.4)	2.6 (0.7)	2.6 (0.5)	.5	1.3	.29	-.32	-.08	-.21	.20	-.13	.58	.25
Subtalar	Medial Glide†	3.0 (0.4)	3.0 (0.4)	2.8 (0.6)	2.6 (0.8)	3.0 (0.2)	3.0 (0)	2.8 (0.7)	2.7 (0.6)	.4	1.0	.00	-.47	.36	.66	.00	.84	.68	.62
Subtalar	Lateral Glide†	3.0 (0.4)	3.0 (0.3)	2.8 (0.8)	2.9 (0.6)	3.0 (0.2)	3.0 (0.4)	2.8 (0.5)	3.0 (0.8)	.5	1.3	.34	.34	-.50	-.30	.69	.00	.38	.31

* Proportion of sample judged to have hypomobility during active dorsiflexion.
† As measured by the 0-7 joint mobility scale described by Gonnella.[23]
SD=standard deviation, SEM=standard error of measurement, MDC=minimal detectable change, Rt=right, Lt=left, Post=posterior

ICC†: Poor, Fair, Good, Excellent
Kw*: Poor, Slight, Fair, Moderate, Sub-stantial, Almost Perfect

Table 7. *Reliability of Joint Play Motion Measures of the Midfoot and Forefoot.*

		Group Means (SD)*										Inter-rater Reliability				Test-Retest Reliability			
		Baseline				Reassessment						Baseline		Reassessment		Tester 1		Tester 2	
		Tester 1		Tester 2		Tester 1		Tester 2											
		Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt	SEM	MDC	Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt
Forefoot on Rearfoot	Inversion	3.0 (0.7)	2.8 (0.7)	3.2 (0.8)	3.5 (0.7)	3.1 (0.7)	3.2 (0.7)	3.2 (0.6)	3.3 (0.7)	.5	1.3	.58	.60	.67	.17	.70	.55	.68	.63
Forefoot on Rearfoot	Eversion	2.9 (0.5)	2.8 (0.4)	2.7 (0.8)	3.2 (0.6)	3.0 (0.2)	3.1 (0.3)	2.9 (0.5)	3.3 (0.6)	.5	1.5	.65	.50	.48	.33	.06	.27	.39	.72
Forefoot on Rearfoot	Abduction	3.0 (0.5)	2.9 (0.4)	3.0 (0.4)	2.8 (0.6)	3.0 (0.4)	2.8 (0.5)	3.0 (0.2)	2.9 (0.7)	.4	1.2	.00	.36	.04	-.13	.39	-.22	.04	.08
Forefoot on Rearfoot	Adduction	3.0 (0.4)	3.1 (0.4)	2.9 (0.6)	3.2 (0.6)	3.1 (0.3)	3.0 (0.2)	3.0 (0.5)	3.1 (0.7)	.4	1.1	.33	.39	.05	-.03	-.07	-.07	.63	.68
First TMT	Dorsal Glide	3.2 (0.6)	3.2 (0.6)	3.1 (0.7)	3.2 (0.6)	3.2 (0.6)	3.3 (0.4)	3.1 (0.7)	3.2 (0.7)	.4	1.2	.40	-.20	.57	.35	.54	.22	.70	.65
First TMT	Plantar Glide	3.0 (0.4)	3.0 (0.4)	2.5 (1.0)	3.0 (0.9)	3.0 (0.4)	2.9 (0.3)	2.8 (0.7)	3.2 (0.9)	.6	1.6	.19	.31	.22	-.15	.00	-.07	.70	.60
First MTP	Dorsal Glide	3.3 (0.6)	3.4 (0.8)	3.0 (0.7)	3.2 (0.6)	3.3 (0.8)	3.2 (0.7)	3.0 (0.7)	3.5 (0.7)	.5	1.4	.28	.21	.39	.73	.26	.51	.43	.37
First MTP	Plantar Glide	2.8 (0.6)	2.8 (0.7)	2.8 (0.6)	2.9 (0.7)	2.8 (0.6)	2.7 (0.6)	3.0 (0.7)	3.3 (1.0)	.4	1.1	.23	-.31	.57	.46	.50	.38	.69	.49

* As measured by the 0-7 joint mobility scale described by Gonnella.[23]
SD=standard deviation, SEM=standard error of measurement, MDC=minimal detectable change, Rt=right, Lt=left, TMT=tarsometatarsal, MTP=metatarsophalangeal

Poor, Fair, Good, Excellent

2.9-3.2), hallux extension (2.6-3.0 to 2.2-2.6), and lesser toe extension (2.0 reported for both limbs to each assessor to 1.6-1.8) exercises from baseline to reassessment. The subjective rating reported to the clinicians was consistent (.76-.97) within session for the short-foot, toe-spread-out, and hallux extension exercises and highly variable for the lesser toe extension during baseline assessment (-.08-.00). Consistency of the patient reported difficulty for lesser toe extension improved during reassessment (.71-.93).

Table 8. *Reliability of Strength Measures of the Ankle-Foot Complex.*

	Group Means (SD)										Inter-rater Reliability				Test-Retest Reliability			
	Baseline				Reassessment													
	Tester 1		Tester 2		Tester 1		Tester 2				Baseline		Reassessment		Tester 1		Tester 2	
	Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt	SEM	MDC	Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt
Ankle DF (N)	267.1 (52.9)	257.4 (59.3)	259.2 (53.0)	251.2 (43.8)	275.7 (45.8)	283.0 (52.2)	274.0 (54.7)	259.3 (46.5)	20.4	56.6	.60	.61	.76	.67	.82	.88	.83	.88
Ankle PF (N)	365.8 (79.3)	360.2 (89.6)	453.3 (118.1)	455.6 (119.1)	411.1 (119.8)	433.3 (131.3)	484.4 (115.9)	467.2 (119.0)	41.6	115.2	.77	.88	.83	.89	.68	.86	.84	.92
Ankle Inversion (N)	196.6 (46.1)	197.0 (49.7)	237.6 (41.4)	251.6 (50.0)	199.7 (35.9)	202.8 (41.8)	243.7 (44.5)	265.2 (40.6)	18.8	52.2	.69	.90	.53	.85	.73	.87	.83	.83
Ankle Eversion (N)	194.0 (55.0)	191.5 (42.7)	234.3 (47.5)	222.5 (37.7)	207.6 (43.6)	198.1 (34.3)	233.9 (39.0)	228.7 (37.3)	23.6	65.5	.74	.79	.71	.65	.85	.79	.78	.74
Hallux Flexion (N)	112.3 (38.5)	111.8 (41.0)	142.7 (44.9)	144.7 (48.0)	117.1 (38.0)	119.8 (43.3)	154.6 (52.3)	155.2 (42.7)	18.5	51.4	.75	.87	.82	.87	.68	.76	.85	.92
Lesser Toe Flexion (N)	103.9 (35.0)	110.4 (36.8)	121.5 (35.0)	135.5 (45.1)	117.1 (38.0)	108.4 (29.8)	129.1 (36.4)	144.1 (34.6)	18.0	49.8	.66	.77	.87	.82	.67	.74	.77	.77

DF = dorsiflexion, PF = plantarflexion, N = Newtons, SD=standard deviation, SEM=standard error of measurement, MDC=minimal detectable change, Rt=right, Lt=left

Poor | Fair | Good | Excellent

Table 9. *Reliability of Participant-Reported Task Difficulty and Measures of Motor Performance of Short-foot and Toe Exercises.*

		Group Means (SD)										Inter-rater Reliability				Test-Retest Reliability			
		Baseline				Reassessment													
		Tester 1		Tester 2		Tester 1		Tester 2				Baseline		Reassessment		Tester 1		Tester 2	
		Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt	SEM	MDC	Rt	Lt	Rt	Lt	Rt	Lt	Rt	Lt
Perceived Difficulty*	Short-foot Exercise	2.2 (1.3)	2.4 (1.3)	2.4 (1.4)	2.1 (1.3)	2.0 (1.3)	2.0 (1.5)	2.0 (1.4)	1.9 (1.3)	.4	1.2	.89	.84	.97	.94	.90	.75	.88	.91
	Toe-spread-out Exercise	3.1 (1.3)	3.5 (1.2)	3.8 (1.3)	3.9 (1.1)	3.1 (1.4)	3.2 (1.4)	2.9 (1.4)	3.0 (1.3)	.7	2.0	.77	.76	.87	.89	.61	.72	.64	.82
	Hallux Extension	2.8 (1.4)	2.6 (1.4)	3.0 (1.5)	2.7 (1.5)	2.5 (1.4)	2.2 (1.5)	2.5 (1.5)	2.6 (1.6)	.4	1.1	.90	.95	.94	.91	.87	.87	.91	.96
	Lesser Toe Extension	2.0 (1.0)	2.0 (1.1)	2.0 (1.0)	2.0 (0.5	1.7 (0.8)	1.8 (1.0)	1.7 (0.7)	1.6 (0.9)	.7	1.8	-.08	.00	.71	.93	.50	.16	.47	.15
Motor Performance	Short-foot Exercise†	2.4 (0.7)	2.3 (0.7)	2.0 (0.5)	2.1 (0.5)	2.6 (0.5)	2.4 (0.8)	2.1 (0.5)	2.2 (0.4)	-	-	.19	.15	.19	.26	.50	.22	.24	.31
	Toe-spread-out Exercise†	2.2 (0.9)	1.9 (0.9)	1.7 (0.7)	1.6 (0.7)	2.1 (0.9)	1.8 (0.9)	1.8 (0.5)	1.8 (0.6)	-	-	.40	.25	.27	.37	.51	.47	.26	.38
	Hallux Extension†	2.3 (0.8)	2.2 (0.9)	1.9 (0.7)	2.2 (0.7)	2.4 (0.9)	2.4 (0.9)	2.3 (0.7)	2.1 (0.7)	-	-	.40	.50	.56	.40	.71	.65	.62	.66
	Lesser Toe Extension†	2.6 (0.6)	2.7 (0.6)	2.2 (0.5)	2.4 (0.5)	2.9 (0.3)	2.8 (0.5)	2.3 (0.6)	2.6 (0.5)	-	-	.02	.20	.16	.35	-.01	.02	.68	.46
	Intrinsic Foot Muscle Test‡	2.1 (1.0)	1.9 (1.0)	2.0 (0.8)	2.0 (0.8)	2.3 (0.8)	2.3 (0.8)	1.8 (0.8)	1.8 (0.8)	-	-	.12	-.02	.31	.40	.35	.39	.44	.17

* Rated on a 1-5 Likert scale (1=very easy, 5=very difficult).
† Rated using the motor scale described by Bérard.[27]
‡ Rated by clinician using the scale described by Jam.[13]
SD=standard deviation, SEM=standard error of measurement, MDC=minimal detectable change, Rt=right, Lt=left

ICC*: Poor | Fair | Good | Excellent
K_W†,‡: Poor | Slight | Fair | Moderate | Substantial | Almost Perfect

Similarly, agreement of clinician's rating of motor performance was fair to moderate (.51-.86) for the short-foot and toe-spread-out exercises and substantial (.62-.71) for hallux extension exercise. Test-retest agreement of lesser toe extension varied between clinicians, with poor to slight (-.01-.02) observed in Tester 1 and moderate to substantial (.46-.68) in Tester 2.

The IFM test had slight to moderate (.17-.44) test-retest reliability. Similarly, inter-rater reliability was poor to moderate during baseline assessment (-.02-.50) that improved to slight to moderate agreement (.16-.56) on the second session. Rating of motor performance during hallux extension and lesser toe extension exercises improved from baseline to reassessment.

DISCUSSION

The principal findings of this study were that reliability was excellent in all morphologic foot measures and rearfoot and hallux sagittal excursion, fair to good in rearfoot frontal plane excursion, poor to excellent in joint play motion with disparity between clinicians, and good to excellent in strength measures. Participant-reported difficulty and motor performance during short-foot and toe exercises had fair to excellent agreement, except for lesser toe extension and the IFM test which had poor inter-rater agreement on the baseline assessment.

Morphologic Foot Assessment

Morphologic measures of the foot using the Arch Height Index Measurement System and the FPI were found to be expedient and having excellent reliability. These findings are consistent with other studies investigating the reliability of Arch Height Index Measurement System and FPI. Butler and colleagues[34] reported the test-retest reliability of the Arch Height Index Measurement System to be .96-.99 and inter-rater reliability to be .98-.99 for arch height, truncated foot length, and foot length. These measures may be utilized in isolation, or as part of a composite measures such as the arch height index[35] or foot mobility magnitude,[36] to quantify longitudinal or transverse deformation of the foot across loading. The MDC values were relatively small and ranged from 2-4 mm, which makes morphologic measurements suitable as a potential outcome measure in clinical practice or in research.

The FPI, which is comprised of five observations of foot morphology and one palpation of talar head position, had excellent test-retest and inter-rater reliability (.81-.86). While excellent test-retest reliability has previously been reported, the findings of excellent inter-rater reliability in this study were substantially higher than the moderate reliability found in prior study of the FPI.[37,38] MDC of the FPI was five points (rounded to the next integer) and reflective of a full categorical shift in foot morphotype (i.e. normal to pronated foot type). Despite having excellent repeatability, the large magnitude MDC in this measure is a product of high variability of foot morphotype observed in this sample (and likely reflective of the population in general). While this measure may not be ideal for assessing changes to the foot in the short term, it may have clinical and research utility when studying morphologic changes to the foot over the lifespan. It is recommended that morphologic foot measures be considered as a clinical or research outcome measure to capture deformation of the foot across loading.

Joint Excursion Measures

Excellent inter-rater and test-retest reliability (.96-.97) was observed during standing lunge dorsiflexion. These findings are consistent with data presented by Bennell and colleagues.[39] Reliability of goniometric measures of the rearfoot were the highest when anatomical landmarks used for alignment were readily identifiable and not obfuscated by soft tissue. Rearfoot sagittal plane excursion had excellent test-retest and inter-rater reliability (.81-.88). The midline of the lateral foot, lateral fibula, and lateral malleolus are easily identifiable and utilized to align the moving arm, stationary arm, and the axis of goniometer during sagittal measures of dorsiflexion. Reliability of rearfoot frontal plane excursion was fair to good (.53-.73) with variability of measurement likely attributed to error in identification and alignment with obscure landmarks such as the subtalar joint and vertical axis of the calcaneus.

Clinician dexterity likely contributed to joint excursion measurement error. Tester 1, who had smaller stature and hand size of the two clinicians, demonstrated side-to-side differences in test-retest reliability in forefoot frontal plane and first metatarsal sagittal measures (fair to good reliability in the left limb and excellent reliability on the right). Both measures require a degree of ambidexterity to simultaneously stabilize and move adjacent segments while holding the instrument. It is plausible that disparate inconsistency between left and right measures is attributed to the 28% less hand surface area in this assessor.

Differences in the magnitude of overpressure applied by the clinicians during first metatarsal excursion measurement likely affected inter-rater reliability. While test-retest reliability was good to excellent for first metatarsal excursion in both raters (.62-.90), inter-rater reliability was poor to fair (.32-.53). Disparities in the amount of applied overpressure and soft tissue deformation due to the pliability of the metatarsal pads are the most likely contributing factors to the inter-test reliability observed in this measure.

Standard error for goniometric and inclinometric measures were 4-6°, with a MDC of 12-15° for the total arc of motion. For sagittal plane talocrural and first metatarsophalangeal measures, this equates to 10-15% of the total excursion. In frontal plane rearfoot and forefoot measures, this equates to a 25% of the total excursion. MDC was 2-cm (15% total excursion) and 4-mm (25% total excursion) for weight-bearing dorsiflexion and first metatarsal excursion, respectively. The relatively higher MDC in frontal plane measures are likely attributed to measurement error (more so in the rearfoot) and normal variability in motion.

Joint excursion measures, to include novel measures of forefoot on rearfoot inversion, eversion and first metatarsal dorsiflexion excursion, had good to excellent repeatability and small MDC values. These measures may have utility when assessing effectiveness of intervention aimed at improving range of motion, such as manual therapy, in the clinical population.

Joint Play Motion

Test-retest reliability of joint play motion ranged from poor to excellent (-.67-.84). Inter-rater reliability was poor (-1.06-.39) in 73% of the joint play measures. The distal tibiofibular posterior glide, subtalar medial glide, forefoot inversion and adduction, and tarsometatarsal dorsal and plantar glides had the highest test-retest reliability. Joint play measures were highly subjective, relied on a clinician's perceived magnitude of displacement, and were likely susceptible to bias. MDC for joint play measures were found to be a full grade on the mobility scale for all but two measures.

The mobility scale proposed by Gonnella and colleagues[23] was developed as a means of quantifying and improving passive intervertebral joint play motion assessment reliability. This scale, which was based on the 4-point mobility scale (immobile, hypomobile, normal, and hypermobile) developed by Kaltenborn and Lindahl[40], added qualifiers that differentiated varying degrees of hypomobility and hypermobility in the assessment scheme. The utility of these qualifiers is questionable during the assessment of small magnitude joint play motions in the foot and ankle. Future study comparing the reliability of both scales is warranted for assessment of joint play motion in the ankle and foot.

Lack of a clearly defined reference is likely to be deleterious to consistency in assessment of joint play motion. Gonnella and colleagues[23] described the reference for the joint mobility scale as the "expected normal for the patient when age, body type, and activity level are considered." Differences in patient phenotype and clinician interpretation of normality make this reference a moving target. In peripheral assessment of joint laxity, clinicians often use the contralateral joint as a reference for comparison. By doing so, a quandary arose during assessment that is likely to have impacted reliability. It was challenging to determine if the assessed joint was hypermobile with normal mobility in the referenced contralateral joint, or if the assessed joint was normal with hypomobility in the contralateral joint. This differentiation was difficult, even when a participant's morphotype and generalized joint laxity was considered. In the current study, measures were made independently without consideration for other clinical correlates that would otherwise provide context for the findings. Ambiguity due to a poorly defined reference made judging magnitude of joint play motion difficult and likely contributed to inconsistency in these measures.

Clinical experience and expertise is likely to influence test-retest reliability in joint play measures. While both clinicians had fair to good reliability at best, the experienced clinician demonstrating greater consistency (53% of measures good) compared the novice clinician (28% of measures good). The higher reliability in measures conducted by the experienced clinician may be a result of uniformity of examination technique or consistency in interpretation of examination findings. Joint play measurements have previously been shown to be influenced by clinician technique, such as grip and test style.[41] Magnitude[42] and direction[43] of force applied and joint position[44] has also been shown to influence joint displacement. Habitual motor patterns formed from years of practice are likely to be more consistent in an experienced clinician between multiple test sessions. Pattern recognition, improved analytical thinking, and intuition are components of clinical expertise that are formed with time and practice.[45] Greater consistency in interpretation of examination findings is also a likely consequence to clinical expertise.

While joint play motion assessment had primarily poor to fair reliability, it is not the standard of care to utilize these measures in isolation. While not ideal as a primary outcome measure, these assessments may have utility as a clinical correlate with other measures when developing a diagnosis and plan of care. When assessing response to treatment, a change of one full grade on the mobility scale when combined with increases in joint excursion measu res, should provide the clinician and researcher with ample information regarding changes in osteokinematic and arthrokinematic motion. The wider implications of these findings are that joint play measures, such as the talocrural anterior glide, are used clinically in surgical decision-making. The talocrural anterior glide (anterior drawer test) is used in the assessment of mechanical laxity and serves as a primary indication for surgical stabilization of the ankle.[46] To ensure diagnostic accuracy and avoidance of unnecessary treatment (and associated financial burden and risk), it is recommended that instrumented measures of joint laxity, specifically ankle arthrometry of joint play in all three planes of motion, be utilized in lieu of manual assessment if surgery is a consideration.

Strength

Test-retest reliability and inter-rater reliability of hand-held dynamometric measures were found to be good to excellent. These findings are similar to those observed in the same laboratory by Kelln and colleagues[47] While variability in motor performance is expected during a maximal isometric task[48] and with differences in strength between assessing clinicians[49] The authors posit that stabilization of the proximal segments, use of mobilization strap to assist in resisting force, and testing at a consistent joint angle likely contributed to the repeatability and inter-rater reliability observed for these measures in this study. Kelln and colleagues[47] were unable to assess the reliability of handheld dynamometry measure of plantarflexion due to the tester being overpowered during testing. Use of a mobilization strap with the handheld dynamometer allowed the assessors to test this muscle group with excellent consistency.

The SEM for all strength measures, sans plantarflexion, ranged from 18.0-23.6 N, with MDC ranging from 49.8-65.5N. Plantarflexion measures, which were substantially higher in magnitude, had a SEM of 41.6 N and a MDC of 115.2 N. The MDC observed during strength assessment, which was 20-25% of the measure, are likely attributed to the variability often observed during motor tasks.

These findings should be considered when assessing changes in strength in healthy individuals using handheld dynamometry. It is recommended that clinicians ensure consistency of test position, joint angle, and proper stabilization when using a handheld dynamometry. Use of a strap to support the dynamometer and assist in resistance may improve consistency by reducing dependency on a clinician's upper body strength.

Motor Performance during Short-foot, Toe Exercises, and IFM Test

Motor performance of the IFM short-foot and toe exercises in this study were assessed using a novel application of the scale developed by Bérard and colleagues.[27] In the original instrument, the motor scale was developed to assess gross motor function in individuals with neuromuscular disease.[27] Reliability of motor performance varied with task complexity, the number of joint segments moving, and the plane of motion. Excellent test-retest and inter-rater reliability was observed in the uniarticular, single plane motion of hallux extension. To contrast this finding, tasks that were more complex, multisegmented, or involved less discrete movement patterns (such as the short-foot exercise, toe-spread-out exercise, and the IFM test) had less consistency and demonstrated fair to good repeatability and inter-rater reliability. Motor performance during tasks involving drawing up of the medial longitudinal arch (short-foot and IFM test) were generally less reliable compared to tasks involving the toes. It is unclear if this a consequence of greater volitional control of the toes compared to the medial longitudinal arch or the subtlety of motion during a short-foot maneuver. Small subtle motions may be harder to be distinguish by the assessor and contribute to error.

There were data to suggest evidence of motor learning occurring within and between assessment visits. Inter-rater reliability for the lesser toe extension exercise was poor on baseline assessment and improved to good on the latter visit. This finding was likely attributed to the participant's unfamiliarity with the task and variable performance during baseline testing. By the time the second assessor evaluated motor performance during the baseline visit, the participant had performed the tasks multiple times and consequently adapted to the novel task. Motor performance ratings provided by both clinicians also improved between visits, with higher ratings of performance during reassessment.

Rating of Perceived Difficulty during Intrinsic Foot Muscle Exercises

Repeatability of perceived task difficulty was excellent for the short-foot and hallux extension exercises, good to excellent in the toe-spread-out, and poor to fair for lesser toe extension. Tasks which had the lowest test-retest reliability were also the most complex (toe-spread-out and lesser toe extension) and demonstrated the largest decrease in perceived difficulty from baseline to reassessment. A disparity in task difficulty between limbs during the lesser toe extension tasks was also observed, with the left limb demonstrating greater variability between sessions. Except for the short-foot exercise, reliability of perceived difficulty paralleled motor performance in three of the four tasks. There was a disconnect between perceived difficulty and motor performance during the short-foot exercise, with participants reporting low difficulty while performing the task inappropriately or with substitution.

Inter-rater reliability was excellent for perceived task difficulty during short-foot, toe-spread-out, and hallux extension exercises. The lesser toe extension task had poor inter-rater reliability during baseline testing, which was also observed during assessment of motor performance of the same task. The authors posit that these findings are attributed to a decrease in task difficulty resulting from repetition and improvement in motor strategies.

The MDC for rating of task difficulty was 1-point (rounded to the next integer) for the short-foot and hallux extension exercises and 2-points for the toe-spread-out and lesser toe extension exercises. The minimal 2-point change is attributed to the variability of difficulty with the toe-spread-out exercise and poor to fair test-retest reliability in the lesser toe extension exercise.

Rating of motor performance and perceived difficulty during short-foot and toe exercises may have clinical utility in assessing IFM function. When assessing baseline measures, it is recommended that patients are assessed immediately following instruction of the tasks and after they have had an opportunity to practice the task. This additional baseline measure may capture any immediate motor learning that may occur. Subsequent measures may be contrasted to both baseline measures to assess effect of intervention.

Limitations

The current study does present with some limitations. While the assessing clinicians had similar training, they had different levels of experience and anthropometric characteristics. It is unclear which of these factors was most contributory to the findings in this study and should be investigated in the future. While an equal number of healthy male and female participants were recruited in this study, foot morphotype was not controlled for and resulted in an unequal representation in the current study's sample. Previous research has suggested a link between foot posture and mobility.[50] Future reliability studies should consider foot morphotype as a potential delimitation. These findings should be interpreted with caution when determining the effect of treatment in the clinical population. Variability in joint mobility or neuromotor function resulting from a clinical condition may increase the SEM and MDC of the measures. Future research to establish reliability, SEM, and MDC in specific clinical populations is needed.

CONCLUSION

Measures of ankle-foot posture, morphology, joint excursion, and strength demonstrated fair to excellent test-retest and inter-rater reliability. Test-retest reliability for rating of perceived difficulty and motor performance was good to excellent for short-foot, toe-spread-out, and hallux exercises and poor to fair for lesser toe extension. Joint play measures had poor to fair reliability overall. The findings of this study should be considered when choosing methods of clinical assessment and outcome measures in practice and research.

REFERENCES

1. Luciano A de P, Lara LCR. Epidemiological study of foot and ankle injuries in recreational sports. *Acta Ortop Bras*. 2012;20(6):339-342.
2. Sobhani S, Dekker R, Postema K, Dijkstra PU. Epidemiology of ankle and foot overuse injuries in sports: A systematic review. *Scand J Med Sci Sports*. 2013;23(6):669-686.
3. Almeida SA, Williams KM, Shaffer RA, Brodine SK. Epidemiological patterns of musculoskeletal injuries and physical training. [Miscellaneous Article]. *Med Sci Sports Exerc*. 1999;31(8):1176-1182.
4. Conti SF, Silverman L. Epidemiology of foot and ankle injuries in the workplace. *Foot Ankle Clin*. 2002;7(2):273–290.
5. Lambers K, Ootes D, Ring D. Incidence of patients with lower extremity injuries presenting to us emergency departments by anatomic region, disease category, and age. *Clin Orthop*. 2012;470(1):284-290.
6. Waterman BR. The epidemiology of ankle sprains in the United States. *J Bone Jt Surg Am*. 2010;92(13):2279.
7. Riddle DL, Schappert SM. Volume of ambulatory care visits and patterns of care for patients diagnosed with plantar fasciitis: a national study of medical doctors. *Foot Ankle Int*. 2004;25(5):303-310.
8. Martin RL, Davenport TE, Reischl SF, et al. Heel pain—plantar fasciitis: revision 2014. *J Orthop Sports Phys Ther*. 2014;44(11):A1-A33.
9. Martin RL, Davenport TE, Paulseth S, Wukich DK, Godges JJ. Ankle stability and movement coordination impairments: ankle ligament sprains: clinical practice guidelines linked to the international classification of functioning, disability and health from the orthopaedic section of the american physical therapy association. *J Orthop Sports Phys Ther*. 2013;43(9):A1-A40.
10. Thomas JL, Christensen JC, Kravitz SR, et al. The diagnosis and treatment of heel pain: a clinical practice guideline–revision 2010. *J Foot Ankle Surg*. 2010;49(3, Supplement):S1-S19.
11. Tiemstra JD. Update on acute ankle sprains. *Am Fam Physician*. 2012;85(12):1170–1176.
12. Wrobel JS, Armstrong DG. Reliability and validity of current physical examination techniques of the foot and ankle. *J Am Podiatr Med Assoc*. 2008;98(3):197-206.
13. Jam B. *Evaluation and retraining of the intrinsic foot muscles for pain syndromes related to abnormal control of pronation*. Advanced Physical Therapy Education Institute; 2006.
14. Fraser JJ, Feger MA, Hertel J. Clinical commentary on midfoot and forefoot involvement in lateral ankle sprains and chronic ankle instability. part 2: clinical considerations. *Int J Sports Phys Ther*. 2016;11(7):1191-1203.
15. Fraser JJ, Feger MA, Hertel J. Clinical commentary on midfoot and forefoot involvement in lateral ankle sprains and chronic ankle instability. part 1: anatomy and biomechanics. *Int J Sports Phys Ther*. 2016;11(6):992-1005.
16. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the foot and ankle ability measure (FAAM). *Foot Ankle Int*. 2005;26(11):968–983.
17. Carcia CR, Martin RL, Drouin JM. Validity of the foot and ankle ability measure in athletes with chronic ankle instability. *J Athl Train*. 2008;43(2):179-183.
18. Donahue M, Simon J, Docherty CL. Reliability and validity of a new questionnaire created to establish the presence of functional ankle instability: the IdFAI. *Athl Train Sports Health Care*. 2013;5(1):38-43.
19. Selim AJ, Rogers W, Fleishman JA, et al. Updated U.S. population standard for the veterans rand 12-item health survey (vr-12). *Qual Life Res*. 2009;18(1):43-52.
20. Godin G, Shephard RJ. Godin leisure-time exercise questionnaire. *Med Sci Sports Exerc*. 1997;29(6):36-38.
21. Redmond AC, Crosbie J, Ouvrier RA. Development and validation of a novel rating system for scoring standing foot posture: the foot posture index. *Clin Biomech*. 2006;21(1):89-98.
22. Greisberg J, Prince D, Sperber L. First ray mobility increase in patients with metatarsalgia. *Foot Ankle Int*. 2010;31(11):954-958.
23. Gonnella C, Paris SV, Kutner M. Reliability in evaluating passive intervertebral motion. *Phys Ther*. 1982;62(4):436–444.
24. Kim M-H, Kwon O-Y, Kim S-H, Jung D-Y. Comparison of muscle activities of abductor hallucis and adductor hallucis between the short foot and toe-spread-out exercises in subjects with mild hallux valgus. *J Back Musculoskelet Rehabil*. 2013;26(2):163-168.
25. Kim M-H, Yi C-H, Weon J-H, Cynn H-S, Jung D-Y, Kwon O-Y. Effect of toe-spread-out exercise on hallux valgus angle and cross-sectional area of abductor hallucis muscle in subjects with hallux valgus. *J Phys Ther Sci*. 2015;27(4):1019-1022.

26. Jung D-Y, Kim M-H, Koh E-K, Kwon O-Y, Cynn H-S, Lee W-H. A comparison in the muscle activity of the abductor hallucis and the medial longitudinal arch angle during toe curl and short foot exercises. *Phys Ther Sport*. 2011;12(1):30-35.
27. Bérard C, Payan C, Hodgkinson I, Fermanian J, Group TMCS. A motor function measure scale for neuromuscular diseases. Construction and validation study. *Neuromuscul Disord*. 2005;15(7):463-470.
28. Donner A, Eliasziw M. Sample size requirements for reliability studies. *Stat Med*. 1987;6(4):441-448.
29. Rhemtulla M, Brosseau-Liard PÉ, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods*. 2012;17(3):354-373.
30. Norman G. Likert scales, levels of measurement and the "laws" of statistics. *Adv Health Sci Educ*. 2010;15(5):625-632.
31. Fleiss JL. *Design and Analysis of Clinical Experiments*. John Wiley & Sons; 2011.
32. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34.
33. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159-174.
34. Butler RJ, Hillstrom H, Song J, Richards CJ, Davis IS. Arch height index measurement system: establishment of reliability and normative values. *J Am Podiatr Med Assoc*. 2008;98(2):102-106.
35. Williams DS, McClay IS. Measurements Used to Characterize the Foot and the Medial Longitudinal Arch: Reliability and Validity. *Phys Ther*. 2000;80(9):864-871.
36. McPoil TG, Vicenzino B, Cornwall MW, Collins N, Warren M. Reliability and normative values for the foot mobility magnitude: a composite measure of vertical and medial-lateral mobility of the midfoot. *J Foot Ankle Res*. 2009;2(1):6.
37. Cornwall MW, McPoil TG, Lebec M, Vicenzino B, Wilson J. Reliability of the modified foot posture index. *J Am Podiatr Med Assoc*. 2008;98(1):7–13.
38. Evans AM, Copper AW, Scharfbillig RW, Scutter SD, Williams MT. Reliability of the foot posture index and traditional measures of foot position. *J Am Podiatr Med Assoc*. 2003;93(3):203-213.
39. Bennell K, Talbot R, Wajswelner H, Techovanich W, Kelly D, Hall A. Intra-rater and inter-rater reliability of a weight-bearing lunge measure of ankle dorsiflexion. *Aust J Physiother*. 1998;44(3):175-180.
40. Kaltenborn F, Lindahl O. Reproducibility of the results of manual mobility testing of specific intervertebral segments. *Lakartidningen*. 1969;66:962-965.
41. Hurley WL, McGuire DT. Influences of clinician technique on performance and interpretation of the Lachman test. *J Athl Train*. 2003;38(1):34-43.
42. de Souza MVS, Venturini C, Teixeira LM, Chagas MH, de Resende MA. Force-displacement relationship during anteroposterior mobilization of the ankle joint. *J Manipulative Physiol Ther*. 2008;31(4):285-292.
43. Fujii M, Suzuki D, Uchiyama E, et al. Does distal tibiofibular joint mobilization decrease limitation of ankle dorsiflexion? *Man Ther*. 2010;15(1):117-121.
44. Kovaleski JE, Norrell PM, Heitman RJ, Hollis JM, Pearsall AW. Knee and ankle position, anterior drawer laxity, and stiffness of the ankle complex. *J Athl Train*. 2008;43(3):242-248.
45. Pelaccia T, Tardif J, Triby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online*. 2011;16.
46. Baumhauer JF, O'Brien T. Surgical considerations in the treatment of ankle instability. *J Athl Train*. 2002;37(4):458-462.
47. Kelln BM, McKeon PO, Gontkof LM, Hertel J. Hand-held dynamometry: reliability of lower extremity muscle testing in healthy, physically active, young adults. *J Sport Rehabil*. 2008;17(2):160.
48. Moraux A, Canal A, Ollivier G, et al. Ankle dorsi-and plantar-flexion torques measured by dynamometry in healthy subjects from 5 to 80 years. *BMC Musculoskelet Disord*. 2013;14(1):1.
49. Wikholm JB, Bohannon RW. Hand-held dynamometer measurements: tester strength makes a difference. *J Orthop Sports Phys Ther*. 1991;13(4):191-198.
50. Cornwall MW, McPoil TG. Relationship between static foot posture and foot mobility. *J Foot Ankle Res*. 2011;4:4.