Intratester and Intertester Reliability of Clinical Measures of Lower Extremity Anatomic Characteristics: Implications for Multicenter Studies

By: Sandra J. Shultz, Anh-Dung Nguyen, Thomas C. Windley, Anthony S. Kulas, Timothy L. Botic, Bruce D. Beynnon

Shultz SJ, Nguyen A, Windley TC, Kulas AS, Botic TL, Beynnon BD. Intratester and Intertester Reliability of Clinical Measures of Lower Extremity Anatomical Alignment: Implications for Multi-center Studies. *Clinical Journal of Sports Medicine* 2006; 16(2):155-161.

Abstract:

Objective: To determine whether multiple examiners could be trained to measure lower extremity anatomic characteristics with acceptable reliability and precision, both within (intratester) and between (intertester) testers. We also determined whether testers trained 18 months apart could perform these measurements with good agreement.
Setting: University's Applied Neuromechanics Research Laboratory.
Participants: Sixteen, healthy participants (7 men, 9 women).
Assessment of Risk Factors: Six investigators measured 12 anatomic characteristics on the right lower extremity in the Fall of 2004. Four testers underwent training immediately preceding the study, and measured subjects on 2 separate days to examine intratester reliability. Two testers trained 18 months before the study (Spring 2002) measured each subject on day 1 to examine the consistency of intertester reliability when testers are trained at different times.
Main Outcome Measurements: Knee laxity, genu recurvatum, quadriceps angle, tibial torsion, tibiofemoral angle, hamstring extensibility, pelvic angle, navicular drop, femur length, tibial length, and hip anteversion.
Results: With few exceptions, all testers consistently measured each variable between test days (intraclass correlation coefficient>=0.80). Intraclass correlation coefficient values were lower for intertester reliability (0.48 to 0.97), and improved from day 1 to day 2. Intertester reliability was similar when comparing testers trained 18 months before those trained immediately before the study. Absolute measurement error varied considerably across individual testers.
Conclusions: Multiple investigators can be trained at different times to measure anatomic characteristics with good to excellent intratester reliability. Intratester reliability did not always

ensure acceptable intertester reliability or measurement precision, suggesting more training (or more experience) may be required to achieve acceptable measurement reliability and precision between multiple testers.

Keywords: lower extremity alignment | risk factor assessment | posture | measurement stability

Article:

Little is known about the factors that predispose an individual to anterior cruciate ligament (ACL) injury.[1,2] In part, this is because of the myriad of factors that have been proposed to explain the risk of ACL injury. Controlling and measuring multiple variables creates considerable challenges, as variables have the potential to interact with one another, and many of these variables are poorly defined or difficult to obtain reliably.[3] Many of the potential risk factors cannot be measured after ACL disruption, because the injury modifies the risk factors,[4–11] and bilateral asymmetry cannot always be assumed.[12,13]

Because of these limitations, large prospective studies are recommended to identify ACL injury risk factors.[14–17] However, prospective studies present their own challenges, as a large cohort of subjects is needed to yield sufficient ACL injuries. Uhorchak et al [16] prospectively followed 850 military cadets over 4 years and reported only 24 noncontact ACL injuries. Although a set of risk factors was found to be predictive of ACL injury, the authors concluded that 1000 subjects is far too small to achieve adequate statistical power for a wide selection of variables. Although the ultimate sample size depends on the expected injury rate of the population, the number of risk factors measured, and the power of the study on the basis of the measures of interest, this highlights the need for multiple centers and testers to make definitive conclusions on ACL injury risk factors.

Although risk factors are ideally measured by a single, experienced investigator,[18–20] this becomes impractical when multiple centers and testers are required to collect sufficient data. An added concern is the potential to lose examiners during the course of a multiyear study, requiring examiners to be replaced. Even with training and practice, some measurements lack the required precision to be useful in identifying those at increased risk of injury. Hence, before initiating multicenter studies, it is necessary to standardize the measurement technique, and demonstrate acceptable measurement consistency and precision, both within and between examiners.

Although there are many reports on the reliability of specific lower extremity anatomic measurements, little information exists on the reliability of a comprehensive selection of anatomic characteristics made by multiple investigators, or whether examiners can continue to make consistent measurements over extended time periods (eg, between seasons). We determine whether multiple examiners could be trained to measure lower extremity anatomic characteristics with an acceptable level of intratester and intertester reliability and precision. A secondary purpose was to determine intertester reliability between testers trained 18 months apart.

METHODS

Subjects

Subject's sex, age, height, and mass were recorded, and then 12 anatomic characteristics were measured on the right lower extremity. Six testers (testers 1 to 4, trained in the month preceding the study; and testers 5 and 6, trained 18 months earlier) measured anatomic characteristics on day 1, and testers 1 to 4 repeated these measures on day 2, within 10 days of day 1 (Table 1 lists tester credentials). Because testers 5 and 6 had previously established intratester reliability, they measured only subjects on day 1 to examine how their measures compared with testers trained at a later date. Subjects and testers rotated among 5 measurement stations (order of subject station counterbalanced) of 2 to 3 anatomic characteristics each (Table 2). Measurements for each subject were obtained in the same order across days and testers. Results were manually recorded to the nearest degree or millimeter and later entered into a computer database. Testers were blinded to the other tester's results, and their own previous day results. All standing measures were taken with the subject's feet placed bi-acromial width apart, toes pointing forward, and in a relaxed stance. Three measures were taken by each tester on both test days.

TABLE 1. Credentials and Years of Clinical Experience for Investigators Measuring Anatomic Variables

Tester No.	Credentials	Years
1	Athletic trainer	2
2	Athletic trainer	2
3	Athletic trainer, occupational therapist	1, 6
4	Athletic trainer	6
5	Physical therapist	3
6	Athletic trainer	6

Procedures

Subject's sex, age, height, and mass were recorded, and then 12 anatomic characteristics were measured on the right lower extremity. Six testers (testers 1 to 4, trained in the month preceding the study; and testers 5 and 6, trained 18 months earlier) measured anatomic characteristics on day 1, and testers 1 to 4 repeated these measures on day 2, within 10 days of day 1 (Table 1 lists tester credentials). Because testers 5 and 6 had previously established intratester reliability, they measured only subjects on day 1 to examine how their measures compared with testers trained at a later date. Subjects and testers rotated among 5 measurement stations (order of subject station counterbalanced) of 2 to 3 anatomic characteristics each (Table 2). Measurements for each subject were obtained in the same order across days and testers. Results were manually recorded to the nearest degree or millimeter and later entered into a computer database. Testers were blinded to the other tester's results, and their own previous day results. All standing measures were taken with the subject's feet placed bi-acromial width apart, toes pointing forward, and in a relaxed stance. Three measures were taken by each tester on both test days.

Examiner Training

Four testers were trained by 1 instructor who had previously demonstrated day-to-day measurement consistency for the 12 anatomic characteristics, and had performed the measures extensively during the previous 2 years. Training consisted of twelve 2-hour practice sessions over 4 weeks. During each session, 1 to 3 anatomic measures were instructed and practiced.

After demonstrating the proper measurement methods, each tester practiced the measures with feedback from the instructor. Once comfortable with a measure, each tester measured a single subject and wrote their values for the 3 trials on a piece of paper (blinded to the other testers). Mean values were compared and if discrepancies were found, further practice and instruction was provided. Testers were also encouraged to practice each measure on their own until they felt proficient. As a final check, testers were observed as they performed each measure on 2 subjects to ensure all testers were identifying landmarks and performing measures as instructed.

TABLE 2. Order of Anatomic Variables Measured at Each Station

Stations	Anatomic Alignment Measures
1	HE, PA
2	StQA, TFA
3	FL, TL, ND
4	AKL, GR
5	SuQA, HA
6	TT

Outcome Measures

The following anatomic characteristics were measured on the right pelvis and lower extremity:

Pelvic Angle (PA)

With subject standing, the angle between the horizontal plane and a line from the anterior (ASIS) to posterior superior iliac spine was measured to the nearest degree, using an inclinometer (Performance Attainment Associates, St Paul, MN). A positive angle was defined as the ASIS positioned lower than the posterior superior iliac spine. (Modified from Gilliam et al.[21])

Hamstrings Extensibility (HE)

With the subject positioned supine and the right hip flexed to 120 degrees, a bar mounted on a steel frame affixed to the table served as a tactile cue to maintain this hip flexion angle. In this position, the subject actively extended the knee. After 5 practice trials, HE was recorded as the knee extension angle measured to the nearest degree, with a larger angle indicating greater extensibility. (Modified from Blackburn et al.[22])

Standing Quadriceps Angle (StQA)

With the axis of the goniometer over the center of the patella, the angle formed by a line from the ASIS to the center of the patella, and a line from the center of the patella to the center of the tibial tubercle was measured to the nearest degree.[23]

Tibiofemoral Angle (TFA)

With the subject standing and the goniometer axis positioned over the knee center in the frontal plane, the angle formed by a line from the knee center to a landmark midway between the ASIS and greater trochanter, and a line from the knee center to the ankle center (mid-malleolar distance) was measured to the nearest degree. (Modified from Chao et al.[24])

Femur Length (FL)
FL was defined as the distance from the superior aspect of the greater trochanter to the lateral joint line (LJL) of the knee, and was measured to the nearest millimeter by a sliding anthropometric caliper while standing.

Tibial Length (TL)
TL was defined as the distance from the medial joint line to the inferior medial malleolus, and was measured to the nearest millimeter by a sliding anthropometric caliper when standing.

Navicular Drop (ND)
A straight edge ruler measured the change in navicular height from a standing neutral to a standing relaxed stance. Subtalar joint neutral was defined as the position where the medial and lateral aspects of the talar head were equally palpable with the thumb and index finger.[25]

Genu Recurvatum (GR)
With the subject supine, a 4-inch bolster was placed under the distal tibia. The knee was passively extended until a firm, soft tissue end feel was noted. With the axis over the LJL, the angle formed by a line from the LJL to the greater trochanter, and a line from the LJL to the lateral malleolus was measured to the nearest degree.[26]

Anterior Knee Laxity (AKL)
Subjects were positioned supine as per the manufacturer's guidelines, and AKL was measured as the amount of anterior displacement of the tibia on the femur at 133 N, using the KT 2000 knee arthrometer (MEDmetric Corp; San Diego, CA).

Hip Anteversion (HA)[27]
With the subject prone and knee flexed to 90 degrees, the hip was passively rotated until the greater trochanter was palpated to be in its most lateral position. The angle between the true vertical and the shaft of the tibia was measured to the nearest degree, using a goniometer with a bubble level attached. (Anteversion=positive angle).

Supine Quadriceps Angle (SuQA)
With the subject positioned supine, the feet positioned bi-acromial width apart, and the toes pointing vertically toward the ceiling, the same measurement methods described for StQA were used.

Tibial Torsion (TT)
The subject was positioned supine, and the femur was passively positioned so that a line between the epicondyles was parallel to the horizontal plane. In this position, the tester palpated the most prominent aspects of the medial and lateral malleoli. The angle formed between true vertical and a line bisecting the bi-malleolar axis was measured to the nearest degree. A bubble level ensured true vertical. (Modified from Stuberg et al.[28]).

Data Reduction and Statistical Analyses

For each test day and tester, 3 measurements for each anatomic characteristic were averaged for analyses. To examine intratester reliability, a repeated measure analysis of variance (ANOVA) with 1 within-subject variable (test day) was used to calculate intraclass correlation coefficients ($ICC_{2,k}$) and standard error of measurement $(SEM = sd \times \sqrt{(1 - ICC)})$ for Testers 1 to 4. We chose the more conservative $ICC_{2,k}$ over $ICC_{3,k}$ to generalize our findings to the greater population of testers in the interest of multicenter studies. Further, $ICC_{3,k}$ does not include the differences across time in the total variance, thus ignores the error because of the systematic changes in how a tester measures from 1 day to the next. To examine intertester reliability between testers 1 to 4 (newly trained), a repeated measure ANOVA with 1 within-subject variable (testers at 4 levels) was used to calculate ICCs ($ICC_{2,1}$) and SEMs for day 1 and day 2. We chose the more conservative $ICC_{2,1}$ over $ICC_{2,K}$,[29] as this formula is more sensitive to inconsistencies in tester measurement fluctuations between subject from one measurement time point to the next. Similar analyses examine intertester reliability between testers trained at different times points on day 1. Measurement error was also assessed using 95% limits of agreement (LOA),[30,31] calculated as the mean difference between measures±1.96 (standard deviation of the mean difference) after confirming first that the mean differences were normally distributed.

RESULTS

Table 3 lists means±standard deviations for each measure by tester and day. Table 4 lists intratester reliability estimates for testers 1 to 4. Generally, all testers consistently measured each variable across test days with good to excellent repeatability (ICC>=0.75). Tester 4 demonstrated excellent consistency on all measures. Only PA (testers 1 and 3), HA (tester 2), and TT (tester 3) demonstrated ICCs <0.80. ICC values were somewhat lower for intertester reliability, and improved from day 1 to day 2 (Table 5). Variables measured with good to excellent consistency (ICC>=0.75) were HE, StQA, FL, and TL. Measurement consistency was moderate (ICC range: 0.60 to 0.75) for ND, AKL, SuQA, and TT, and lower (ICC range: 0.48 to 0.59) for PA, TFA, GR, and HA. Intertester reliability was similar when comparing testers trained at different time points (Table 5). Table 6 provides the 95% LOA between day 1 and day 2 measures for testers 1 to 4. Tester 4 consistently demonstrated the highest measurement agreement, with tester 3 generally having the lowest agreement.

DISCUSSION

Before conducting a study to identify subjects at increased risk of injury, it is important to fully understand the reliability and precision of the potential risk factor measurements. Our findings within tester demonstrated good to excellent reliability, with most ICCs exceeding 0.80.31 However, there were clearly some inconsistencies when comparing measurements across testers. Time of training did not seem to have a significant impact on intertester reliability, indicating it is possible to add testers to the study at a later time without adversely affecting data collection.

TABLE 3. Means (SD) for Each Anatomic Measure by Tester and day

		Tester 1	Tester 2	Tester 3	Tester 4	Tester 5	Tester 6
Anatomic Measure	**Day**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
PA (deg)	1	13.5 (4.6)	14.5 (3.4)	8.7 (4.1)	12.3 (3.7)	12.4 (3.9)	13.6 (3.9)
	2	12.7 (4.0)	15.3 (4.1)	10.8 (4.6)	12.6 (3.4)		
HE (deg)	1	127.9 (20.3)	123.5 (20.2)	129.4 (20.5)	127.0 (19.5)	126.5 (21.6)	127.8 (21.0)
	2	124.6 (21.0)	124.7 (21.2)	125.4 (22.6)	126.3 (20.5)		
StQA (deg)	1	16.1 (7.5)	10.8 (6.2)	14.0 (6.0)	13.8 (5.4)	15.9 (6.3)	13.4 (6.4)
	2	18.0 (7.0)	11.8 (6.9)	14.1 (6.2)	14.8 (5.5)		
TFA (deg)	1	11.3 (2.0)	9.1 (3.0)	12.0 (2.4)	10.3 (2.0)	10.9 (2.3)	9.6 (2.0)
	2	11.9 (2.8)	10.4 (2.3)	11.9 (2.6)	11.3 (1.8)		
FL (cm)	1	42.6 (2.5)	42.0 (3.2)	42.0 (3.2)	41.8 (3.4)	41.4 (3.2)	41.3 (2.9)
	2	42.0 (2.8)	42.1 (3.2)	41.9 (3.6)	41.3 (3.5)		
TL (cm)	1	38.0 (4.4)	36.9 (3.6)	37.2 (3.3)	37.4 (3.1)	37.2 (3.6)	37.5 (3.4)
	2	37.6 (3.3)	37.4 (3.6)	36.5 (3.7)	37.2 (3.0)		
ND (mm)	1	7.1 (2.5)	7.0 (3.0)	5.4 (2.4)	6.1 (2.2)	6.9 (2.8)	7.1 (2.6)
	2	7.5 (2.8)	7.6 (3.6)	4.7 (2.2)	6.4 (2.1)		
GR (deg)	1	2.6 (4.0)	0.3 (3.3)	5.6 (4.1)	2.4 (2.7)	2.0 (4.2)	1.5 (3.0)
	2	1.9 (4.6)	0.5 (3.6)	5.7 (5.3)	2.0 (2.9)		
AKL (mm)	1	8.4 (1.9)	7.9 (1.9)	8.0 (1.8)	7.0 (1.7)	7.5 (2.0)	6.3 (1.8)
	2	8.3 (1.9)	8.0 (1.7)	7.5 (2.1)	7.1 (1.6)		
HA (deg)	1	10.8 (5.3)	8.1 (3.1)	4.9 (2.7)	9.8 (6.6)	10.0 (6.3)	11.2 (4.2)
	2	10.5 (4.5)	9.6 (4.1)	5.6 (3.2)	10.6 (5.8)		
SuQA (deg)	1	14.5 (4.5)	12.2 (4.6)	10.9 (4.2)	11.4 (5.7)	14.2 (4.7)	12.1 (5.4)
	2	15.9 (4.6)	11.2 (4.5)	12.9 (4.9)	12.3 (5.4)		
TT (deg)	1	15.5 (7.4)	15.4 (8.8)	10.9 (5.5)	15.8 (7.5)	16.7 (10.0)	14.9 (4.2)
	2	19.1 (7.0)	17.4 (10.0)	14.8 (6.4)	15.3 (7.5)		

TABLE 4. $ICC_{2,k}$ and SEM for Day-to-day Intratester Reliability (Testers 1 to 4)

	Tester 1	Tester 2	Tester 3	Tester 4
Measure	**$ICC_{2,k}$ (SEM)**	**$ICC_{2,k}$ (SEM)**	**$ICC_{2,k}$ (SEM)**	**$ICC_{2,k}$ (SEM)**
PA (deg)	0.77 (2.2)	0.90 (1.3)	0.64 (2.8)	0.98 (0.5)
HE (deg)	0.97 (3.7)	0.97 (3.6)	0.91 (6.8)	0.98 (2.8)
StQa (deg)	0.89 (5.9)	0.97 (1.1)	0.94 (1.6)	0.98 (0.8)
TFA (deg)	0.85 (1.1)	0.82 (1.2)	0.85 (1.0)	0.87 (0.7)
FL (cm)	0.95 (0.6)	0.97 (0.5)	0.97 (0.7)	0.99 (0.4)
TL (cm)	0.91 (1.3)	0.98 (0.5)	0.96 (0.7)	0.98 (0.4)
ND (mm)	0.95 (0.6)	0.95 (0.8)	0.91 (0.7)	0.97 (0.4)
GR (deg)	0.93 (1.2)	0.89 (1.2)	0.88 (1.8)	0.97 (0.5)
AKL (mm)	0.94 (0.5)	0.85 (0.7)	0.81 (0.9)	0.96 (0.3)
HA (deg)	0.97 (0.9)	0.77 (2.0)	0.90 (1.0)	0.97 (1.1)
SuQA (deg)	0.90 (1.5)	0.89 (1.5)	0.88 (1.7)	0.98 (0.7)
TT (deg)	0.91 (2.3)	0.94 (2.5)	0.69 (3.6)	0.99 (0.8)

TABLE 5. $ICC_{2,1}$ and SEMs for Intertester Reliability Between Newly Trained (Testers 1 to 4), Previously Trained (Testers 5 and 6), and all (Testers 1 to 6) on day 1, and Newly Trained Testers on day 2

	Day 1			Day 2
Measure	**Testers 1–4**	**Testers 5 and 6**	**Testers 1–6**	**Testers 1–4**
PA (deg)	0.48 (3.3)	0.68 (2.2)	0.56 (3.1)	0.53 (3.1)
HE (deg)	0.89 (6.7)	0.93 (5.7)	0.91 (6.6)	0.97 (4.2)
StQA (deg)	0.79 (3.5)	0.75 (3.2)	0.78 (3.5)	0.72 (3.7)
TFA (deg)	0.51 (2.1)	0.46 (1.7)	0.56 (2.0)	0.58 (1.8)
FL (cm)	0.90 (1.4)	0.80 (1.4)	0.88 (1.2)	0.88 (1.3)
TL (cm)	0.86 (1.6)	0.98 (.52)	0.90 (1.2)	0.92 (1.0)
ND (mm)	0.67 (1.7)	0.76 (1.4)	0.72 (1.6)	0.56 (2.3)
GR (deg)	0.48 (2.9)	0.75 (2.1)	0.57 (2.7)	0.56 (3.5)
AKL (mm)	0.59 (1.2)	0.56 (1.3)	0.57 (1.3)	0.77 (1.0)
HA (deg)	0.48 (3.8)	0.74 (3.2)	0.58 (4.2)	0.56 (3.8)
SuQA (deg)	0.62 (3.5)	0.75 (2.7)	0.69 (3.1)	0.74 (2.5)
TT (deg)	0.72 (4.6)	0.48 (7.2)	0.68 (5.7)	0.75 (5.0)

All values are $ICC_{2,1}$ (SEM).

TABLE 6. Mean Difference ± 95% LOA for Test-retest Measurements of Newly Trained Testers

Anatomic Measure	Tester 1	Tester 2	Tester 3	Tester 4
PA (deg)	0.8 ± 7.4	−0.7 ± 4.4	−2.1 ± 8.5	−0.3 ± 2.0
HE (deg)	3.3 ± 12.9	−1.2 ± 13.7	4.0 ± 23.9	0.7 ± 10.8
StQA (deg)	−1.9 ± 8.5	−1.0 ± 3.8	−0.2 ± 6.1	−1.0 ± 2.7
TFA (deg)	−0.5 ± 3.4	−1.3 ± 3.4	0.1 ± 3.5	−1.1 ± 1.9
FL (cm)	0.6 ± 2.1	−0.1 ± 2.1	0.1 ± 2.5	0.5 ± 1.0
TL (cm)	−0.1 ± 2.4	−0.5 ± 1.9	1.0 ± 3.6	0.2 ± 1.5
ND (mm)	−0.4 ± 2.1	−0.5 ± 2.6	0.8 ± 2.2	−0.2 ± 1.4
GR (deg)	0.7 ± 4.2	−0.2 ± 4.4	0.0 ± 6.4	0.3 ± 1.8
AKL (mm)	0.0 ± 2.1	−0.2 ± 2.6	0.5 ± 3.0	−0.2 ± 1.1
HA (deg)	0.3 ± 3.2	−1.6 ± 5.8	−0.8 ± 3.3	−0.8 ± 3.9
SuQA (deg)	−1.4 ± 4.9	1.0 ± 5.5	−2.0 ± 4.8	−0.9 ± 2.3
TT (deg)	−3.6 ± 5.4	−2.0 ± 8.5	−3.8 ± 10.1	0.5 ± 3.0

Interpretation of Intertester Reliability

Lower intertester reliability in the presence of strong intratester reliability suggests systematic error may be the cause for the lower ICC values. Evaluation of the ANOVA results revealed a significant mean difference between testers for all but HE, AKL, and FL, and TL. Examination of the mean values obtained by each tester (Table 3) indicates measures for tester 3 were systematically different from the other testers on PA, ND, GR, HA, and TT. Tester 3 also demonstrated greater measurement error (Tables 4 and 6). When this tester was removed from the analyses, ICC values improved (PA=0.66; ND=0.77; GR=0.76; HA=75; SuQA=0.70; and TT=0.74). Hence, intertester reliability was largely affected by a single tester. However, some systematic differences were still apparent among the remaining testers ($P<0.05$). These observations suggest that testers differed somewhat in their measurement techniques, and that further training may be required. Further, it should be noted that 4 of 6 testers in this study had relatively few years of clinical experience (Table 1), which has been shown to impact reliability on some measures.[21,32–34] Perhaps, more experienced clinicians would have achieved a higher degree of reliability with the level of training provided. Although not a purpose of this study, post-hoc comparisons of testers 4 and 6 (each with 6 years of clinical experience) support this conclusion.

Although systematic error explains the majority of low intertester ICC measures, this was not the case for AKL. Evaluation of the raw data indicates there was little variation between subjects, with 8 of 16 subjects having values within 1.5 mm of one another. This lack of between-subject variability left little room for measurement error, and would naturally inflate the proportion of variance because of the systematic and random error. This was also reflected by somewhat lower intratester reliability coefficients (Table 4). Hence, the lower reliability on this measure seems to be because of the sample characteristics rather than simply an inability of testers to obtain an accurate measure.

Measurement Precision

The SEM provides a unit value of measurement precision that is based on the distribution of measurement error.[31] In the case of intertester reliability, the SEM indicates there is a 68% and 95% chance that the subject's true score falls within ±1 or ±2 SEMs, respectively, of the value obtained. In some cases (PA, TFA, AKL, HA), the expected measurement error was almost as large as the standard deviation of the sample, suggesting the resolution of the measure may not

be adequate to draw meaningful conclusions in this population. Our sample was relatively small (N=16), however, and the standard deviations may not be reflective of the larger population.

Because of these limitations, we also calculated the 95% LOA (Table 6), which is not dependent on the distribution of scores in the sample.[30,35] The 95% LOA indicates that the expected difference in day-to-day measures will be within 2Sd of the mean difference for 95% of the cases. This value can be useful in making clinical decisions as to whether a tester's measurement error is acceptable. Consider GR, where all testers demonstrated acceptable intratester reliability but poor intertester reliability. The 95% LOA for measures taken by tester 4 were within ±2 degrees for 95% of the subjects, but were within ±4 degrees for testers 1 and 2, and within ±6 degrees for tester 3. While considering that the clinical range of GR is relatively small, the measurement error for testers 1 to 3 seems problematic. Similar concerns are noted for testers 1 to 3 on other anatomic measures identified already as having lower reliability on the basis of their intraclass correlations. The fact that tester 4 consistently demonstrated substantially lower absolute measurement error suggests that the tester, rather than the measurement itself, may be the limiting factor in achieving acceptable intertester reliability on the majority of these measures.

These observations highlight the importance of understanding the variability in the measure, and the expected variability in the target population when selecting potential risk factor measurements to be included in a large scale study. The 95% LOA further emphasizes the need to examine absolute measurement error within each tester, and to make clinical judgments as to whether the magnitude of measurement error is acceptable for the intended study.

Clinical Relevance

The influence of anatomic factors on ACL injury risk remains unknown.[1,2] The measurements examined in the current study were intended to characterize lower extremity posture,[36,37] and were chosen on the basis of proposed injury risk factors [1,14,16,38–40] and known sex differences.[12,16,22,41] Although sex differences in lower extremity anatomic characteristics have been hypothesized to be related to ACL injury risk,[16,38,39,42,43] others do not support this.[44] However, all but one 16 of these investigations were retrospective and limited to small sample sizes. Further, the anatomic factors examined varied considerably between studies, and measurement reliability and precision were often inadequately reported. To understand how static postural abnormalities influence ACL injury risk, acceptable measurement reliability and precision must be determined a priori, and sufficient subjects and risk factors must be evaluated to draw meaningful conclusions. Further, a combination of anatomic characteristics may be more likely to predict ACL injury than a single characteristic, given compensations that occur in the lower extremity with postural malalignments, and their interrelationships with one another.[36,37,39]

Our findings are limited to measurements on the right side from subjects who were relatively lean (average BMI<24 kg/m^2). Research has demonstrated that side-to-side symmetry of anatomic characteristics cannot always be assumed,[12,13] making it necessary to measure both the left and right side in prospective injury risk studies. Because some measures require the examiner to change position and hand placements when measuring side-to-side, investigators should establish measurement reliability for both sides. Measurement reliability is also

dependent on accurately identifying bony landmarks, which may be more challenging in subjects with a high body mass index.

ACKNOWLEDGMENTS

The authors thank Jatin Ambegaonkar MS, ATC, OT, CSCS and Yohei Shimokochi MA, ATC for their valuable assistance with this project.

REFERENCES

1. Griffin LY, Agel J, Albohm MJ, et al. Noncontact anterior cruciate ligament injuries: risk factors and prevention strategies. J Am Acad Orthop Surg. 2000;8:141–150.

2. McClay-Davis I, Ireland ML. ACL research retreat: the gender bias April 6-7, 2001. Clin Biomech. 2001;16:937–939.

3. Beynnon BD, Renström PA, Alosa DM, et al. Ankle ligament injury risk factors: a prospective study of college athletes. J Orthop Res. 2001;19:213–220.

4. Beard DJ, Kyberd PJ, Fergusson CM, et al. Proprioception after rupture of the anterior cruciate ligament: an objective indication for the need for surgery? J Bone Joint Surg. 1993;75-B:311–315.

5. Beynnon BD, Fleming BC, Labovitch R, et al. Chronic anterior cruciate ligament deficiency is associated with increased anterior translation of the tibia during the transition from non-weightbearing to weightbearing. J Orthop Res. 2002;20:332–337.

6. Branch TP, Hunter R, Donath M. Dynamic EMG analysis of anterior cruciate deficient legs with and without bracing during cutting. Am J Sports Med. 1989;17:35–41.

7. Daniel DM, Stone ML, Sachs R, et al. Instrumented measurement of anterior knee laxity in patients with acute anterior cruciate ligament disruption. Am J Sports Med. 1985;13:401–407.

8. Gauffin H, Tropp H. Altered movement and muscular-activation patterns during the one-legged jump in patients with an old anterior cruciate ligament rupture. Am J Sports Med. 1992;20:182–192.

9. McNair PJ, Marshall RN. Landing characteristics in subjects with normal and anterior cruciate ligament deficient knee joints. Arch Phys Med Rehabil. 1994;75:584–589.

10. Mizuta H, Shiraishi M, Kubota K, et al. A stabilometric technique for evaluation of functional instability in anterior cruciate ligament-deficient knee. Clin J Sports Med. 1992;2:235–239.

11. Wojtys EM, Huston LJ. Neuromuscular performance in normal and anterior cruciate ligament-deficient lower extremities. Am J Sports Med. 1994;22:89–104.

12. Livingston LA, Mandigo JL. Bilateral within-subject Q angle assymetry in young adult females and males. Biom Sci Instrument. 1997;33:112–117.

13. Rosene JM, Fogarty TD. Anterior tibial translation in collegiate athletes with normal anterior cruciate ligament integrity. J Athl Train. 1999;34:93–98.

14. Bonci CM. Assessment and evaluation of predisposing factors to anterior cruciate ligament injury. J Athl Train. 1999;34:155–164.

15. Soderman K, Alfredson H, Pietila T, et al. Risk factors for leg injuries in female soccer players: a prospective investigation during an out-door season. Knee Surg Sports Traumatol Arthrosc. 2001;9:313–321.

16. Uhorchak JM, Scoville CR, Williams GN, et al. Risk factors associated with non-contact injury of the anterior cruciate ligament. Am J Sports Med. 2003;31:831–842.

17. Trimble MH, Bishop MD, Buckley BD, et al. The relationship between clinical measurements of lower extremity posture and tibial translation. Clin Biomech. 2002;17:286–290.

18. Smith-Orrichio K, Harris BA. Interrater reliability of subtalar neutral, calcaneal inversion and eversion. J Orthop Sports Phys Ther. 1990;12:10–15.

19. Elveru RA, Rothstein JM, Lamb RL, et al. Methods of taking subtalar joint measurements. Phys Ther. 1988;68:678–682.

20. Baumhauer JF, Alosa D, Renström P, et al. Test-retest reliability of ankle injury risk factors. Am J Sports Med. 1995;23:571–574.

21. Gilliam J, Brunt D, MacMillan M, et al. Relationship of the pelvic angle to the sacral angle: measurement of clinical reliability and validity. J Orthop Sports Phys Ther. 1994;20:193–199.

22. Blackburn JT, Riemann BL, Padua DA, et al. Sex comparison of extensibility, passive, and active stiffness of the knee flexors. Clin Biomech. 2004;19:36–43.

23. Livingston LA, Mandigo JL. Bilateral Q angle asymmetry and anterior knee pain syndrome. Clin Biomech. 1999;14:7–13.

24. Chao EYS, Neluheni EVD, Hsu RWW, et al. Biomechanics of malalignment. Orthop Clin North Am. 1994;25:379–386.

25. Brody DM. Techniques in the evaluation and treatment of the injured runner. Orthop Clin North Am. 1982;13:541–558.

26. Loudon JK. Measurement of knee-joint-position sense in women with genu recurvatum. J Sport Rehabil. 2000;9:15–25.

27. Ruwe PA, Gage JR, Ozonoff MB, et al. Clinical determination of femoral anteversion. J Bone Joint Surg. 1992;74-A:820–830.

28. Stuberg W, Temme J, Kaplan P, et al. Measurement of tibial torsion and thigh-foot angle using goniometry and computed tomography. Clin Ortho Relat Res. 1991;272:208–212.

29. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psych Bull. 1979;86:420–428.

30. Bland M. An Introduction to Medical Statistics. 2nd ed. New York: Oxford University Press; 1995.

31. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2000.

32. Greene CC, Edwards TB, Wade MR, et al. Reliability of the quadriceps angle measurement. Am J Knee Surg. 2001;14:97–103.

33. Jonson SR, Gross MT. Intraexaminer reliability, interexaminer reliability, and mean values for nine lower extremity skeletal measures in healthy naval midshipment. J Orthop Sports Phys Ther. 1997;25:253–263.

34. Tomisch DA, Nitz AJ, Threlkeld AJ, et al. Patellofemoral alignment: reliability. J Orthop Sports Phys Ther. 1996;23:200–208.

35. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26:217–238.

36. Kendall FP, McCreary EK. Muscles: Testing and Function. 3rd ed. Baltimore: Williams & Wilkins; 1983.

37. Hruska R. Pelvic stability influences lower extremity kinematics. Biomechanics. 1998;6:23–29.

38. Hertel JN, Dorfman JH, Braham RA. Lower extremity malalignments and anterior cruciate ligament injury history. J Sports Sci Med. 2004;3:220–225.

39. Loudon JK, Jenkins W, Loudon KL. The relationship between static posture and ACL injury in female athletes. J Orthop Sports Phys Ther. 1996;24:91–97.

40. Murphy DF, Connolly DAJ, Beynnon BD. Risk factors for lower extremity injury: a review of the literature. Br J Sports Med. 2003;37:13–29.

41. Woodland LH, Francis RS. Parameters and comparisons of the quadriceps angle of college-aged men and women in the supine and standing positions. Am J Sports Med. 1992;20:208–211.

42. Woodford-Rogers B, Cyphert L, Denegar CR. Risk factors for anterior cruciate ligament injury in high school and college athletes. J Athl Train. 1994;29:343–346.

43. Beckett ME, Massie DL, Bowers KD, et al. Incidence of hyperpronation in the ACL injured knee: a clinical perspective. J Athl Train. 1992;27:58–60.

44. Smith J, Szczerba JE, Arnold BA, et al. Role of hyperpronation as a possible risk factor for anterior cruciate ligament injuries. J Athl Train. 1997;32:25–28.