Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction

Andrea Reid, Trevor B Birmingham, Paul W Stratford, Greg K Alcock, J Robert Giffin

Background and Purpose

Although various hop tests have been proposed as performance-based outcome measures following anterior cruciate ligament (ACL) reconstruction, limited reports of their measurement properties exist. The purpose of this study was to investigate the reliability and longitudinal validity of data obtained from hop tests during rehabilitation after ACL reconstruction.

Subjects

Forty-two patients, 15 to 45 years of age, who had undergone ACL reconstruction participated in the study.

Methods and Measures

The study design was prospective and observational with repeated measures. The subjects performed a series of 4 hop tests on 3 separate occasions within the 16th week following surgery and on a fourth occasion 6 weeks later. The tests were a single hop for distance, a 6-m timed hop, a triple hop for distance, and crossover hops for distance. Performance on the ACL-reconstructed limb was expressed as a percentage of the performance on the nonoperative limb, termed the "limb symmetry index." Subjects also completed the Lower Extremity Functional Scale and a global rating of change questionnaire.

Results

Intraclass correlation coefficients for limb symmetry index values ranged from .82 to .93. Standard errors of measurement were 3.04% to 5.59%. Minimal detectable changes, at the 90% confidence level, were 7.05% to 12.96%. Changes in hop test scores on the operative limb were statistically greater than changes on the nonoperative limb. Pearson correlations (*r*) between change in hop performances and self-reported measures ranged from .26 to .58.

Discussion and Conclusion

The results show that the described series of hop tests provide a reliable and valid performance-based outcome measure for patients undergoing rehabilitation following ACL reconstruction. These findings support the use and facilitate the interpretation of hop tests for research and clinical practice.

A Reid, PT, MSc, was a physiotherapist at the Fowler Kennedy Sport Medicine Clinic, 3M Centre, University of Western Ontario, London, Ontario, Canada, during the completion of this project. She is currently a physiotherapist at the Allan McGavin Sports Medicine Centre, John Owen Pavilion, University of British Columbia, Vancouver, Canada.

TB Birmingham, PT, PhD, is Associate Professor and Tier 2 Canada Research Chair in Musculoskeletal Rehabilitation, School of Physical Therapy, Elborn College, University of Western Ontario, London, Ontario, Canada N6G 1H1, and Co-Director, Wolf Orthopaedic Biomechanics Laboratory, Fowler Kennedy Sport Medicine Clinic. Address all correspondence to Dr Birmingham at: tbirming@uwo.ca.

PW Stratford, PT, MSc, is Professor, School of Rehabilitation Science, and Associate Member, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, and a Scientific Affiliate in the Department of Surgery, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada.

GK Alcock, PT, MSc, is Physiotherapist, Fowler Kennedy Sport Medicine Clinic.

JR Giffin, MD, FRCS(C), is Assistant Professor, Department of Surgery, University of Western Ontario, and Co-Director, Wolf Orthopaedic Biomechanics Laboratory, Fowler Kennedy Sport Medicine Clinic.

[Reid A, Birmingham TB, Stratford PW, et al. Hop testing provides a reliable and valid outcome measure during rehabilitation after anterior cruciate ligament reconstruction. *Phys Ther*. 2007;87: 337–349.]

The importance of using standardized outcome measures in research and clinical practice has been described repeatedly in the orthopedic and physical therapy literature. For example, various outcome measures have been suggested for use when evaluating the effectiveness of different interventions being compared in clinical trials[1,2] and when making clinical decisions about individual patients.[3–5] Postoperative rehabilitation following anterior cruciate ligament (ACL) reconstruction is the focus of numerous research studies[6] and comprises a substantial portion of orthopedic physical therapist practice.[7] Accordingly, standardized outcome measures that are appropriate for assessing patients undergoing physical therapy following ACL reconstruction are required for comparing different postoperative rehabilitation strategies and for evaluating individual patient progress.

Standardized outcome measures can be described as measures with acceptable measurement properties that have been published with specific procedures for administration, scoring, and interpretation. Dissemination of this type of information has indeed occurred for a variety of self-report measures (questionnaires) and continues to progress. However, research reports focused on similar information for performance-based measures of physical function have not paralleled that for self-report measures. Specifically, although information about the measurement error and ability to detect change has been reported in a clinically interpretable way for many self-report measures, this often is not the case for performance-based measures.

Some authors[8–10] have suggested that self-report and performance-based measures quantify different aspects of function and that using one type of measure alone does not sufficiently capture the breadth of health concepts associated with the measurement of function. Researchers[8,9,11,12] investigating the relationship between self-report and performance-based measures have reported Pearson correlations (*r*) ranging from .02 to .59. Other authors[13] have emphasized that there are situations in which performance-based measures may be preferable and have suggested that these measures also be included in research and clinical practice. Owing to the increased emphasis on incorporating functional and sport-specific exercises into current ACL postoperative rehabilitation protocols, and the goal to have patients return to dynamic and potentially injurious activities, the inclusion of outcome measures that are performance-based may be especially important when evaluating these patients.

Hop testing has frequently been proposed as a practical, performance-based outcome measure that reflects the integrated effect of neuromuscular control, strength (force-generating capacity), and confidence in the limb and requires minimal equipment and time to administer.[14–17] Based on a review of the potential use of hop tests as measures of dynamic knee stability, Fitzgerald et al[8] suggested that hopping may be appropriate for use as a predictive tool for identifying patients who may have future problems as a result of knee injury or pathology and as an evaluative tool to reflect change in patient status in response to treatment.

A combination of 4 different hop tests originally described by Noyes et al[18] may be particularly suitable as a performance-based outcome measure for patients who are undergoing rehabilitation after ACL reconstruction. The tests incorporate a variety of movement principles (ie, direction change, speed, acceleration-deceleration, rebound) that mimic the demands of dynamic knee stability during sporting activities and are suggested to prepare the patient for return to such activities.[7,19–22] This series of hop tests involves a single hop for distance, a 6-m timed hop, a triple hop for distance, and crossover hops for distance. Measurements are obtained on both extremities so that test performance on the operative limb can be expressed as a percentage of test performance on the opposite limb, termed the "limb symmetry index."

Based on performance on these 4 hop tests, the limb symmetry index has been used to help differentiate individuals with and without dynamic knee stability[18,23–27] and to compare different rehabilitation strategies following ACL reconstruction.[19] Some authors[7,20,21] also have advocated the use of these hop tests when monitoring progress in individual patients who are undergoing rehabilitation following ACL reconstruction. Various clinical practice guidelines include specific scores on the limb symmetry index that must be met in order for a patient to progress through phases of rehabilitation, to return to sports, or to be discharged from physical therapy.[7,20,21]

Bolgla and Keskula[28] evaluated the relative reliability of scores on the limb symmetry index based on the described series hop tests in subjects who were healthy and suggested that it is a reliable measure of lower-extremity performance (intraclass correlation coefficient [ICC]=.95–.96). Intraclass correlation coefficients also have been reported for individual hop tests in patients following ACL reconstruction (ICC=.76–.97 for the single hop for distance test,[11,29,30] ICC=.88–.97 for the 6-m timed hop test,[11,31] and ICC=.94–.98 for the crossover hops for distance test[31]). However, we are unaware of any previous reports providing estimates of the measurement error and minimal detectable change for the series of hop tests in patients

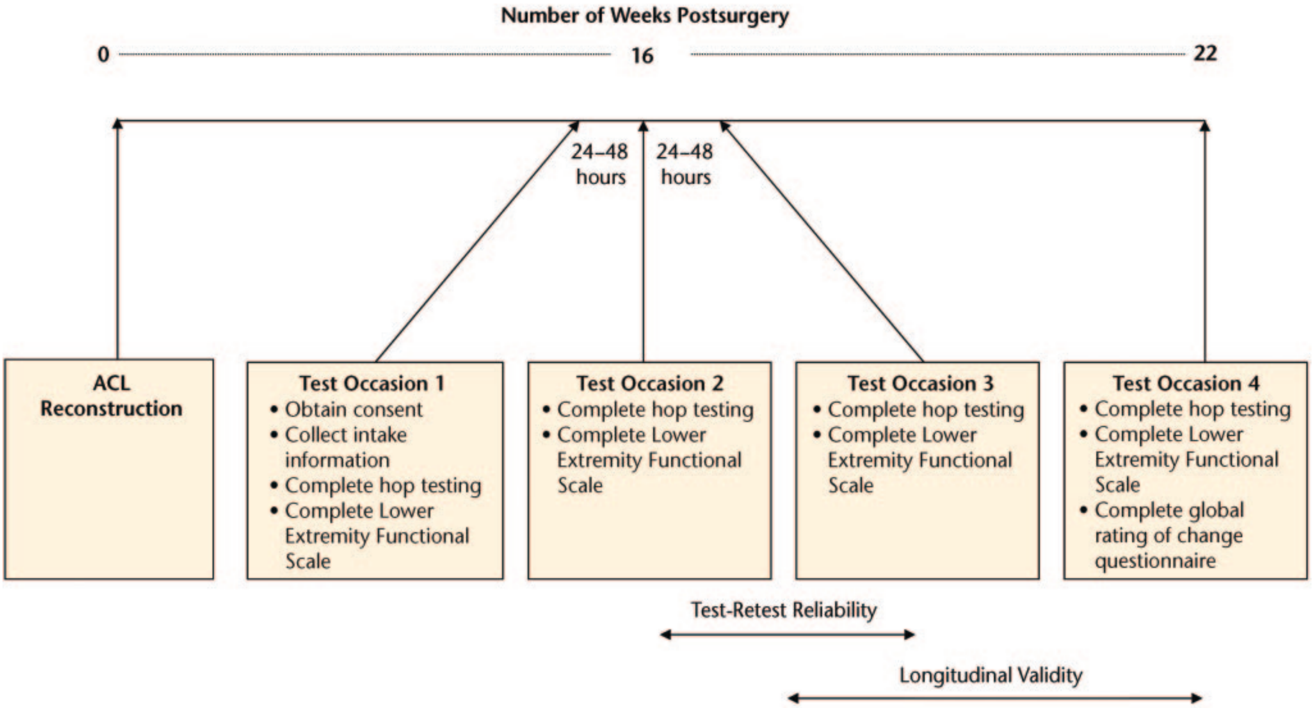

Figure 1.
Schematic diagram of study design. Subjects attended 3 test occasions within the 16th week following anterior cruciate ligament (ACL) reconstruction and a final test occasion 6 weeks later.

following ACL reconstruction, or the ability of this performance-based measure to detect change during postoperative rehabilitation.

In order to facilitate the use of the described series of hop tests as a standardized performance-based outcome measure for patients who are undergoing rehabilitation following ACL reconstruction, further information regarding its measurement properties should be provided. Specifically, further information regarding the reliability and longitudinal construct validity of data obtained from these hop tests is necessary to more accurately plan future clinical trials and to more confidently make clinical decisions about individual patients. Therefore, the objective of the present study was to investigate the reliability and longitudinal validity of data from these hop tests during rehabilitation after ACL reconstruction.

Method

Study Design

The study design was prospective and observational with repeated measures (Fig. 1). Subjects performed the 4 hop tests and then completed self-report questionnaires on 4 different test occasions. The subjects were blinded to their hop test scores. The testing procedures were identical on each test occasion and were administered by the same investigator. The initial 3 test occasions occurred within the 16th week following ACL reconstruction, with a minimum of 24 hours between any 2 test occasions. The first test occasion was intended to allow motor learning. The second and third test occasions were used to evaluate test-retest reliability. The fourth and final test occasion took place 6 weeks later and was used to evaluate longitudinal validity.

A construct validation process was based on 2 theories of change. First, validity was evaluated based on the construct that changes in the hop performances on the operative limb should be significantly greater than changes in the hop performances on the nonoperative limb. We considered this comparison of limbs within individuals to be a form of known-groups validity, although it should be recognized that known-groups validity traditionally has involved comparisons among individuals. Second, convergent validity was evaluated based on the construct that change in limb symmetry index scores should be at least moderately correlated to changes in scores on self-report measures.

Participants

Forty-two patients between the ages of 15 and 45 years participated in this study (Tab. 1). All patients had

Table 1.
Patient Characteristics[a]

	Female Subjects	Male Subjects	Total
Sample size (n)	19	23	42
Age (y)[a]	23.1±8.2 (15–40)	27.7±9.7 (15–45)	25.6±9.2 (15–45)
Height (cm)[a]	165.3±6.2 (155.0–175.0)	177.2±8.4 (165.0–192.5)	171.8±9.5 (155.0–192.5)
Weight (kg)[a]	64.5±10.6 (47.7–81.8)	84.4±17.1 (54.5–115.9)	75.4±17.5 (47.7–115.9)
Body mass index[a]	23.1±3.2 (19–29)	26.7±5.3 (19–40)	25.2±4.8 (19–40)
Operative limb (right/left)	11/8	9/14	20/22
Dominant limb (right/left)	18/1	23/0	41/1
Meniscal repair (yes/no)	12/7	8/15	20/22
Self-rated activity level			
Sedentary	0	0	0
Recreationally active	12	15	27
Competitive athlete	7	8	15

[a] Mean ± standard deviation (minimum–maximum).

undergone primary unilateral ACL reconstruction at the Fowler Kennedy Sport Medicine Clinic using a semitendinosus and gracilis tendon autograft and were following the postoperative rehabilitation protocol used at that center. All patients had a stable contralateral knee (no injury or surgical interventions in the past 2 years), had full range of motion in the operative limb when compared with the nonoperative limb (flexion within 5°), and had only trace or no effusion. Patients with concomitant meniscal injury that required repair were included in the study, provided that they were permitted to undergo typical rehabilitation after ACL reconstruction involving immediate full weight-bearing gait and unrestricted non–weight-bearing range of motion.

Patients were excluded if they had concomitant posterior cruciate ligament or medial collateral ligament injury requiring treatment, had any concurrent musculoskeletal condition (eg, back, hip, or ankle injury) rendering them unable to hop on either extremity, had advanced degenerative changes (ie, Kellgren and Lawrence[32] grade of III or greater based on the preoperative radiograph or noted intraoperatively), or were unable to speak, read, write, or understand English. All participants provided informed consent prior to participation.

Sample size was based on parameter estimation of the reliability coefficient for overall limb symmetry index, with a lower confidence interval (CI) width of 0.1, an expected ICC of at least .85, and a one-tailed CI set to $1 - \alpha$ (α=.05).[33] Using these parameters, the estimated sample size required was 36 subjects. Given that the study design involved 4 repeated test occasions over a 6-week period, we conservatively recruited 50 subjects to account for a dropout rate of up to 25%.

One hundred seventeen patients were approached as potential participants. Those who did not enter the study were injured on their nonoperative side (n=5), had undergone revision surgery (n=4), had experienced a superficial wound infection (n=2), had an associated fracture (n=2), had nontypical ACL reconstruction (n=3), were away from home either traveling or attending university (n=20), were outside of a reasonable driving distance (n=23), were unwilling to participate (n=6), or failed to attend the scheduled appointment (n=4). Forty-eight patients were entered into the study.

During the course of the study, 6 patients withdrew from the study for the following reasons: 1 patient moved out of the area, 1 patient was diagnosed with pneumonia, 2 patients had scheduling difficulties, and 2 patients had complaints of thigh pain after 2 consecutive days of testing. Of the remaining 42 patients, 8 patients could attend only 2 of the 3 sessions completed within 1 week. Three patients did not complete the final test day (1 patient had a back injury rendering her unable to hop, 1 patient had hernia surgery, and 1 patient developed a knee effusion after playing ice hockey the previous day). As a result, the final sample consisted of 42 patients who attended either 3 or 4 test occasions and contributed data for summary statistics. Thirty-five patients contrib-

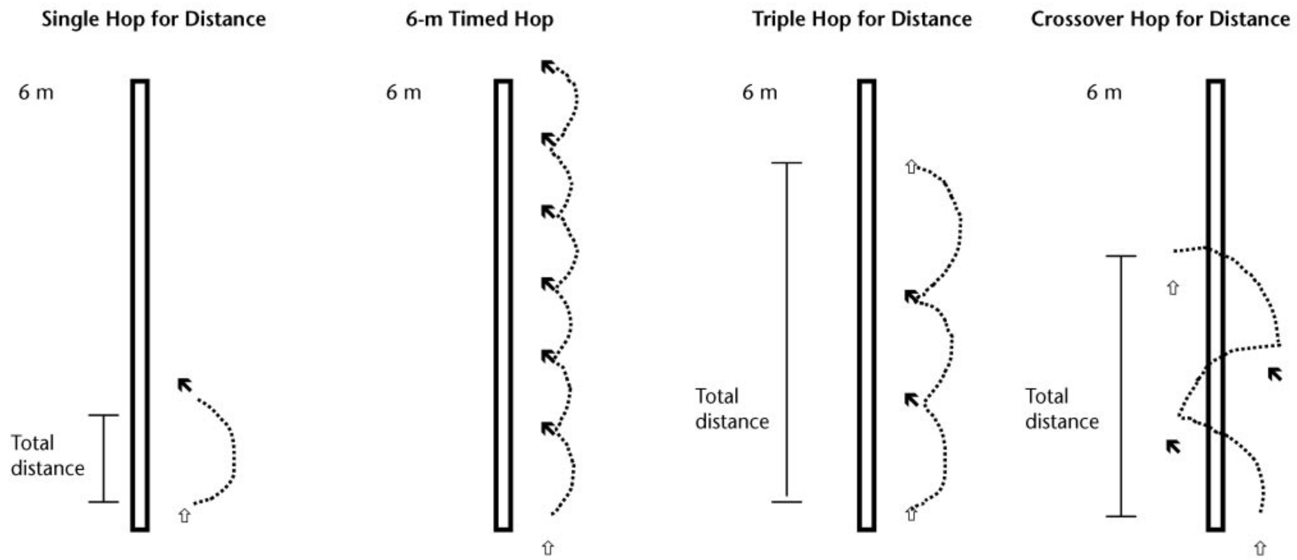

Figure 2.
Diagrammatic representation of the series of 4 hop tests: single hop for distance, 6-m timed hop, triple hop for distance, and crossover hop for distance. Adapted and reprinted by permission of Sage Publications Inc from: Noyes FR, Barber SD, Mangine RE. Abnormal lower limb symmetry determined by function hop tests after anterior cruciate ligament rupture. *Am J Sports Med.* 1991;19:513–518. Copyright 1991 by Sage Publications Inc.

uted data to the analysis of reliability, and 39 patients contributed data to the test of longitudinal validity.

Hop Testing Procedures

The series of 4 hop tests was administered in accordance with the protocols outlined by Noyes et al,[18] Barber et al,[34] and Daniel et al.[35] The tests were a single hop for distance, a 6-m timed hop, a triple hop for distance, and a crossover hop for distance (Fig. 2). In keeping with the original description,[18] the tests were administered in that order on each test occasion, followed by the administration of the self-report measures. The hop testing course was constructed on low-pile, rubber-backed carpet glued over concrete floor. The course consisted of a 6-m-long × 15-cm-wide marking placed on the floor.

For each hop test, the subjects performed one practice trial for each limb, followed by 2 measured and recorded trials. Consistent with the original description of the 4 hop tests, no additional warm-up activity was performed. For each set of tests, the subjects were instructed to begin with the nonoperative limb. To minimize fatigue, a rest period was offered between types of hop tests (up to 2 minutes) and between individual hop test trials if needed (typically less than 30 seconds was sufficient). Subjects started each test with the lead toe behind a clearly marked starting line. No restrictions were placed on arm movement during testing, and no instructions were provided regarding where to look. Subjects were encouraged to wear the footwear they would normally wear during their rehabilitation sessions.

For the hops for distance (single, triple, and crossover) to be deemed successful, the landing must have been maintained for 2 seconds. An unsuccessful hop was classified by any of the following: touching down of the contralateral lower extremity, touching down of either upper extremity, loss of balance, or an additional hop on landing. If the hop was unsuccessful, the subject was reminded of the requirement to maintain the landing, and the hop was repeated. No further instructions were provided to the subjects. Typically, 1 or 2 extra trials were required.

The single hop for distance was performed as outlined by Daniel et al.[35] The subjects stood on the leg to be tested, hopped, and landed on the same limb. The distance hopped, measured at the level of the great toe, was measured and recorded to the nearest centimeter from a standard tape measure that was permanently affixed to the floor. The timed 6-m hop was performed as outlined by Barber et al.[34] Subjects were instructed to perform large one-legged hops in series over the total distance. A standard stopwatch was used to record time. The stopwatch was started when a subject's heel lifted from the starting position and was stopped the moment that the tested foot passed the finish line. Measurements were recorded to the nearest 10th of a second.

The triple hop for distance was performed as outlined by Noyes et al.[18] Subjects were instructed to stand on one leg and perform 3 consecutive hops as far as possible, landing on the same leg. The total distance for 3 consecutive hops was recorded. Finally, the crossover hop for distance[18] was performed over a 15-cm strip on the floor. The subjects hopped forward 3 times while alternately crossing over a marking. The total distance hopped forward was recorded. Subjects were instructed to position themselves such that the first of the 3 hops was lateral with respect to the direction of crossover. The series of hop tests took approximately 10 minutes to administer.

Self-Report Measures

The Lower Extremity Functional Scale (LEFS) is a region-specific, self-report functional status measure.[36] Individuals' scores on this 20-item questionnaire range from 0 to 80, with higher scores indicating better functional status. Previous research[37] has determined the measurement properties of the LEFS, including its standard error of measurement (SEM) (3.4–3.9 LEFS points), 90% CI for a given score (±6 LEFS points), minimal detectable change at the 90% confidence level (9 LEFS points), and minimal clinically important difference (9 LEFS points).

On the final test occasion, subjects also completed a global rating of change questionnaire that asked them how much they had changed over the last 6 weeks (ie, since first performing the hop tests).[38] This tool was used to provide an indication of the subjects' perception of the size of the change experienced. The questionnaire asks patients to indicate whether they are better, worse, or the same, and, if appropriate, how much they have changed on a 15-point scale (−7 to 7) that includes descriptors ranging from "a tiny bit, almost same" to "a very great deal."[39]

Data Analysis

On each test occasion, all hop test scores were recorded as absolute distance (in centimeters) or time (in seconds) and were calculated as the mean of the 2 recorded trials. Also using the mean of 2 trials, the limb symmetry index was calculated such that the score on the ACL-reconstructed limb was expressed as a percentage of the score on the nonoperative limb. Limb symmetry index scores were calculated for each of the 4 hop tests and for the overall combination of hop tests. Although the limb symmetry index scores were the outcome measures of most interest, absolute scores on each limb also were presented to better understand the behavior of the calculated index scores upon repeated assessments.

Hop test scores on each of the 4 test occasions were compared using repeated-measures analyses of variance (ANOVAs). Separate ANOVAs were completed for the operative and nonoperative limbs using data from all subjects. Following a significant main effect, Scheffé *post hoc* tests were used to compare scores for each test occasion.

Reliability. Test-retest reliability was assessed using the hop values obtained from test occasions 2 and 3. Reliability first was estimated using ICC(2,1).[40] The ICC is a ratio of the variance between patients to the total variance, it provides an indication of how well a measure can distinguish among patients, and it therefore can be considered a measure of relative reliability. Reliability then was estimated using the SEM.[41] The SEM provided an expression of an individual subject's hop test measurement error in the original test units (eg centimeters, seconds, percentage), and therefore can be considered an absolute measure of reliability. An upper one-sided 95% CI for the point estimate of the SEM was constructed using the method described by Stratford and Goldsmith.[5] The point estimate of the SEM then was used to estimate the error in an individual subject's score at a given point in time, at the 90% confidence level, by multiplying the SEM by the z value for 90% confidence (1.64).

The point estimate of the SEM also was used to calculate an estimate of the minimal detectable change at the 90% confidence level by multiplying the SEM by the square root of 2 (this accounts for measurement error at 2 testing occasions) and the z value for 90% confidence (1.64).[42] We used a different level of confidence when creating CIs for point estimates (95%) than when describing the interpretation of an individual's score (90%), partly to emphasize that these concepts are indeed different and because we believed that clinical interpretations based on a single subject's score should be interpreted more liberally than estimates based on our study's sample of subjects (n=35). We felt that the 90% level represented that sentiment while still being quite conservative.

Longitudinal validity. Change scores were calculated as the difference between scores obtained on test occasion 4 and the mean of test occasions 2 and 3 (n=35). For the subjects without occasion 3 data, the values for test occasion 2 were used (n=4). For known-groups validity, we compared change scores on the absolute hop scores between limbs on each of the 4 hop tests using paired t tests. For convergent validity, we evaluated the correlation between change in limb symmetry index scores and: (1) change in the LEFS and (2) the global rating of

Table 2.
Mean ± Standard Deviation (Minimum–Maximum) for All Subjects for Hop Test Absolute Scores on the Operative and Nonoperative Limbs, the Limb Symmetry Index (Operative Limb Expressed as a Percentage of Nonoperative Limb), and the Lower Extremity Functional Scale Scores on 4 Separate Test Occasions

Test	Day 1 (16 wk Postoperatively)	Day 2 (+24–48 hr)	Day 3 (+24–48 hr)	Day 4 (22 wk Postoperatively)
n	42	42	35	39
Single hop				
Operative limb (cm)	112.0±32.5 (39.0–179.5)	127.4±32.3 (41.5–187.5)	128.9±32.4 (61.5–192.5)	141.4±28.1 (74.0–187.5)
Nonoperative limb (cm)	135.3±31.2 (71.5–204.0)	154.4±30.0 (77.0–213.5)	158.4±28.3 (92.5–215.0)	160.0±26.0 (100.5–212.0)
Limb symmetry index (%)	82.9±15.4 (33.8–110.1)	82.2±12.3 (47.2–103.2)	81.0±12.1 (51.6–103.7)	88.2±9.5 (63.8–103.2)
6-m timed hop				
Operative limb (s)	3.4±2.1 (1.7–12.8)	2.9±1.2 (1.8–7.7)	2.9±1.2 (1.7–6.4)	2.6±0.8 (1.6–5.9)
Nonoperative limb (s)	2.5±0.71 (1.6–5.1)	2.3±0.5 (1.5–3.5)	2.3±0.6 (1.5–3.8)	2.3±0.5 (1.5–3.9)
Limb symmetry index (%)	81.7±16.3 (33.8–109.5)	81.8±13.4 (45.4–102.8)	83.2±12.7 (50.2–100.3)	89.6±9.5 (66.0–102.1)
Triple hop				
Operative limb (cm)	344.8±91.4 (124.0–532.5)	363.5±89.0 (159.0–570.0)	371.7±96.5 (173.0–553.5)	393.2±88.9 (193.5–618.0)
Nonoperative limb (cm)	416.1±84.1 (247.0–576.5)	440.1±81.4 (271.5–606.5)	452.3±91.9 (249.0–633.5)	450.6±99.4 (239.0–666.5)
Limb symmetry index (%)	82.6±13.3 (45.1–99.6)	82.4±11.7 (48.4–99.7)	82.1±13.2 (54.4–102.7)	87.7±10.2 (68.0–102.3)
Crossover hop				
Operative limb (cm)	303.3±90.7 (68.5–514.0)	328.0±92.3 (128.5–552.5)	330.9±98.7 (136.0–544.5)	358.6±89.3 (152.0–589.0)
Nonoperative limb (cm)	362.6±93.2 (140.0–534.0)	387.3±84.8 (204.5–602.0)	399.1±89.5 (220.5–604.5)	405.6±89.8 (194.0–618.5)
Limb symmetry index (%)	83.1±13.0 (48.9–106.1)	84.4±14.1 (46.0–112.5)	82.2±13.3 (47.5–103.4)	88.3±9.6 (68.2–105.7)
Overall combination of hops: limb symmetry index (%)	82.6±13.0 (41.8–99.6)	82.7±11.9 (47.3–100.8)	82.1±11.6 (55.4–102.1)	88.5±8.5 (70.0–101.7)
Lower Extremity Functional Scale	66.0±9.9 (24–79)	66.0±9.1 (28–79)	65.5±8.9 (26–78)	69.3±8.3 (30–80)

change. We calculated Pearson correlation coefficients (r) and lower one-sided 95% CI. Given that previously reported correlations between performance-based and self-report measures have typically ranged from approximately 0 to 0.6,[8,9,11,12] we decided on the following criteria for strength of evidence for longitudinal validity: good, $r > .5$; moderate, $r = .36-.5$; low, $r = .2-.35$; and no evidence, $r < .2$.

Results

Summary statistics for hop test and LEFS scores on all test occasions are presented in Table 2 for the entire sample and in Tables 3 and 4 for female and male subjects, respectively. For all of the absolute hop test scores on both the operative and nonoperative limbs (Tab. 2), the ANOVAs indicated a significant main effect for time ($P < .001$). For all tests completed on the operative limb, *post hoc* comparisons indicated that absolute hop scores on the first test occasion were significantly different from those on the second test occasion ($P < .01$). There was no signifi-

Table 3.
Mean ± Standard Deviation (Minimum–Maximum) for Female Subjects for Hop Test Absolute Scores on the Operative and Nonoperative Limbs, the Limb Symmetry Index (Operative Limb Expressed as a Percentage of Nonoperative Limb), and the Lower Extremity Functional Scale Scores on 4 Separate Test Occasions

Test	Day 1 (16 wk Postoperatively)	Day 2 (+24–48 hr)	Day 3 (+24–48 hr)	Day 4 (22 wk Postoperatively)
n	19	19	18	18
Single hop				
Operative limb (cm)	105.9±26.2 (39.0–139.0)	116.4±29.5 (41.5–154.5)	121.6±28.4 (61.5–164.0)	133.2±25.9 (74.0–170.5)
Nonoperative limb (cm)	129.8±23.0 (78.5–166.0)	141.6±29.1 (77.0–188.0)	146.4±24.8 (92.5–182.0)	151.6±25.0 (100.5–188.0)
Limb symmetry index (%)	81.4±13.8 (46.4–98.7)	82.2±13.9 (47.1–103.2)	82.8±12.5 (53.7–103.7)	88.0±10.4 (63.8–103.2)
6-m timed hop				
Operative limb (s)	3.7±2.4 (2.1–12.8)	3.2±1.3 (1.9–7.7)	3.0±1.1 (2.0–6.4)	2.8±0.9 (1.7–5.9)
Nonoperative limb (s)	2.7±0.7 (1.8–5.1)	2.4±0.5 (1.8–3.5)	2.5±0.6 (1.7–3.8)	2.5±0.6 (1.7–3.9)
Limb symmetry index (%)	79.9±16.2 (39.8–109.5)	81.1±14.7 (45.5–100.0)	84.4±11.2 (59.4–99.8)	89.8±10.1 (66.0–102.1)
Triple hop				
Operative limb (cm)	307.7±76.2 (124.0–411.5)	329.8±82.6 (159.0–488.0)	343.7±87.7 (173.0–489.0)	362.2±82.1 (193.5–493.0)
Nonoperative limb (cm)	388.6±74.9 (247.0–538.0)	408.1±68.2 (271.5–518.5)	411.0±79.4 (249.0–559.0)	412.3±88.2 (239.0–552.0)
Limb symmetry index (%)	79.0±13.2 (49.2–94.5)	80.4±12.6 (48.4–94.1)	83.6±13.9 (54.4–102.7)	88.2±10.4 (69.6–102.3)
Crossover hop				
Operative limb (cm)	265.7±81.3 (68.5–378.5)	301.4±85.3 (128.5–416.5)	305.1±87.7 (136.0–431.5)	336.9±87.9 (152.0–479.5)
Nonoperative limb (cm)	328.7±82.3 (140.0–469.5)	360.9±67.6 (237.0–461.0)	362.0±75.7 (220.5–472.0)	376.1±83.2 (194.0–500.0)
Limb symmetry index (%)	79.8±13.6 (48.9–97.7)	82.7±15.6 (46.0–99.4)	83.4±14.1 (47.5–103.4)	89.1±9.7 (68.2–105.7)
Overall combination of hops: limb symmetry index (%)	80.0±12.8 (46.1–99.6)	81.6±13.5 (47.3–99.0)	83.5±12.1 (55.9–102.1)	88.7±9.3 (70.0–101.7)
Lower Extremity Functional Scale	64.2±8.0 (45–76)	64.6±6.9 (53–76)	66.0±5.9 (55–77)	68.8±5.1 (61–78)

cant difference in absolute scores completed on the second and third test occasions ($P>.89$). With the exception of the timed hop ($P=.17$), there was a significant difference between absolute scores obtained on the second and fourth test occasions ($P<.001$).

For all tests completed on the nonoperative limb, *post hoc* comparisons indicated that absolute hop scores on the first test occasion were significantly different from those on the second test occasion ($P<.05$). There was no significant difference in absolute scores completed on the second and third test occasions ($P>.1$). Unlike the operative limb, there were no significant differences between absolute scores obtained on the second and fourth test occasions ($P>.1$), with the exception of the crossover hop ($P=.035$).

When scores were expressed as a percentage of the nonoperative limb (ie, limb symmetry index scores, Tab. 2), the ANOVAs also indicated a significant main effect for time ($P<.001$) for each of the hop tests and for the combination of tests (overall limb symmetry index). For all tests, *post hoc* comparisons indicated that the limb symmetry index on the final test occasion was significantly different from those on all other test occasions ($P<.005$), but there were no significant differences among the first, second, and third test occasions ($P>.40$).

Table 4.
Mean ± Standard Deviation (Minimum–Maximum) for Male Subjects for Hop Test Absolute Scores on the Operative and Nonoperative Limbs, the Limb Symmetry Index (Operative Limb Expressed as a Percentage of Nonoperative Limb), and the Lower Extremity Functional Scale Scores on 4 Separate Test Occasions

Test	Day 1 (16 wk Postoperatively)	Day 2 (+24–48 hr)	Day 3 (+24–48 hr)	Day 4 (22 wk Postoperatively)
n	23	23	17	21
Single hop				
Operative limb (cm)	117.0±36.8 (44.0–179.5)	136.4±32.4 (70.0–187.5)	136.7±35.4 (70.5–192.5)	148.5±28.5 (96.5–187.5)
Nonoperative limb (cm)	139.8±35.9 (71.5–204.0)	165.1±26.9 (115.5–213.5)	171.1±26.9 (123.0–215.0)	167.3±25.3 (122.0–212.0)
Limb symmetry index (%)	84.1±16.8 (33.8–110.1)	82.1±11.0 (50.5–99.7)	79.1±11.8 (51.6–92.9)	88.5±8.8 (71.5–102.7)
6-m timed hop				
Operative limb (s)	3.1±1.9 (1.7–9.1)	2.7±1.1 (1.8–6.4)	2.7±1.3 (1.7–6.0)	2.4±0.6 (1.6–4.0)
Nonoperative limb (s)	2.3±0.6 (1.6–4.5)	2.2±0.4 (1.5–3.5)	2.1±0.5 (1.5–3.1)	2.1±0.4 (1.5–2.9)
Limb symmetry index (%)	83.1±16.7 (33.8–99.6)	82.4±12.5 (47.5–102.8)	81.8±14.4 (50.2–100.3)	89.5±9.2 (70.4–100.7)
Triple hop				
Operative limb (cm)	375.4±93.1 (183.0–532.5)	391.3±86.0 (255.0–570.0)	401.3±99.0 (231.5–553.5)	419.8±87.7 (279.0–618.0)
Nonoperative limb (cm)	438.8±86.1 (265.5–576.5)	466.5±83.2 (317.5–606.5)	496.0±85.4 (302.5–633.5)	483.4±98.6 (310.5–666.5)
Limb symmetry index (%)	85.6±12.9 (45.1–99.6)	84.0±11.0 (55.5–99.7)	80.6±12.7 (57.2–96.2)	87.4±10.2 (68.0–101.3)
Crossover hop				
Operative limb (cm)	334.3±87.8 (157.0–514.0)	349.9±94.0 (216.5–552.5)	358.2±104.9 (206.5–544.5)	377.2±88.3 (238.0–589.0)
Nonoperative limb (cm)	390.6±91.1 (195.5–534.0)	409.0±92.5 (204.5–602.0)	438.3±88.1 (240.0–604.5)	431.0±89.4 (240.5–618.5)
Limb symmetry index (%)	85.8±12.1 (54.2–106.1)	85.8±12.9 (58.2–112.5)	80.9±12.6 (48.4–93.2)	87.7±9.7 (69.2–99.0)
Overall combination of hops: limb symmetry index (%)	84.7±13.1 (41.8–98.9)	83.6±10.6 (52.9–100.8)	80.6±11.3 (55.4–92.1)	88.2±7.9 (72.1–98.1)
Lower Extremity Functional Scale	67.4±11.2 (24–79)	67.1±10.6 (28–79)	64.9±11.4 (26–78)	69.6±10.4 (30–80)

In general, comparison of hop scores over the 4 test occasions indicated that substantial motor learning took place on both the operative and nonoperative limbs between the first and second test occasions, which then leveled off by the third test occasion. The significant increases in hop scores on the fourth test occasion on the operative limb, but not on the nonoperative limb, suggested that hop performance improved over the 6-week period.

Reliability

Reliability statistics for the hop test limb symmetry index scores are presented in Table 5. The ICCs ranged from .82 to .93 and can be described as indicating excellent relative reliability.[43] The single hop test and overall limb symmetry index scores demonstrated the highest relative reliability. The SEM was lowest for the single hop test and overall limb symmetry index scores, suggesting that these measures also demonstrated the highest absolute reliability. The error in an individual's limb symmetry index scores at one point in time and the minimal detectable changes upon reassessment, both at the 90% confidence level, also are presented in Table 5. An example of their interpretation is provided in the "Discussion" section.

Longitudinal Validity

Limb symmetry index change scores were 6.5% (95% CI=4.5–8.5) for the

Table 5.
Reliability of Hop Test Limb Symmetry Index Scores (n=35): Intraclass Correlation Coefficients (ICC) With Lower One-Sided 95% Confidence Intervals (CI) in Parentheses; Standard Errors of Measurement (SEM) With Upper One-Sided 95% CIs in Parentheses; and Corresponding Estimates of the Error in an Individual's Score at One Point in Time and Minimal Detectable Change, Both Estimated Using the *z* Value for 90% Confidence (1.64)

Limb Symmetry Index	ICC (Lower 95% CI)	SEM (%) (Upper 95% CI)	Error in an Individual's Score (%)	Minimal Detectable Change (%)
Single hop test	.92 (0.87)	±3.49 (4.37)	±5.72	±8.09
6-m timed hop test	.82 (0.70)	±5.59 (7.01)	±9.17	±12.96
Triple hop test	.88 (0.80)	±4.32 (5.41)	±7.08	±10.02
Crossover hop test	.84 (0.74)	±5.28 (6.62)	±8.66	±12.25
Overall combination of hop tests	.93 (0.89)	±3.04 (3.81)	±4.99	±7.05

single hop test, 7.9% (95% CI=5.3–10.5) for the 6-m timed hop test, 5.3% (95% CI=2.8–7.8) for the triple hop test, 4.8% (95% CI=2.2–7.4) for the crossover hop test, and 6.1% (95% CI=4.2–8.0) for the overall combination of hop tests. The changes in absolute scores for hop tests on the operative limb were statistically greater than the changes on the nonoperative limb for the single hop test (paired $t_{38}=6.4$, $P<.001$), the 6-m timed hop test (paired $t_{38}=4.5$, $P<.001$), the triple hop test (paired $t_{38}=3.3$, $P=.002$), and the crossover hop test (paired $t_{38}=3.1$, $P=.004$). Correlations among hop test change scores, the global rating of change, and LEFS change scores are reported in Table 6. Correlations (*r*) between performance-based and self-report measures ranged from .26 to .58. The global rating of change was most highly correlated to the overall limb symmetry index.

Discussion

This study provides comparative hop scores in both absolute and limb symmetry index values for male and female subjects at the time during postoperative rehabilitation where training dynamic knee stability is emphasized (Tabs. 2, 3, and 4). Although we are unaware of previously published data describing the entire series of hop tests in patients undergoing rehabilitation after ACL reconstruction, the present values are similar to those previously reported for individual hop tests evaluated in these types of patients.[11,17,29–31] In general, comparison of hop scores over the 4 test occasions indicated that substantial motor learning took place on both the operative and nonoperative limbs from the first to second test occasions, which tended to level off by the third test occasion. There were substantial increases in hop scores on the fourth test occasion on the operative limb, but not on the nonoperative limb, suggesting that the functional status of the operative limb improved over the 6-week period.

Table 6.
Longitudinal Validity: Pearson *r* Values With Lower One-Sided 95% Confidence Intervals in Parentheses for Correlations Between Hop Test Limb Symmetry Index Change (Scores From Day 4 Versus the Averaged Score From Days 2 and 3), the Global Rating of Change, and Lower Extremity Functional Scale Change Scores (n=39)

Limb Symmetry Index Change	Global Rating of Change	Lower Extremity Functional Scale Change
Single hop test	.48 (.24)	.37 (.11)
6-m timed hop test	.46 (.22)	.28 (.01)
Triple hop test	.44 (.20)	.26 (.00)
Crossover hop test	.45 (.21)	.41 (.16)
Overall combination of hop tests	.58 (.37)	.41 (.16)

Limb symmetry index values provide important measures of performance on the operative limb in relation to the nonoperative limb. The fact that limb symmetry index values were relatively stable over the first 3 test occasions (ie, the limb symmetry index accounted for learning that occurred in both limbs) and were similar for male and female subjects also supports their use. However, examining absolute scores also is important. For example, although limb symmetry index values were similar for test occasions 1 and 2, the absolute scores were very different. Examining limb symmetry index in isolation would mask this change in performance.

The ICCs observed in the present study for limb symmetry index scores suggest excellent relative reliability[43] and indicate that these tests are appropriate for distinguishing among patients, such as is done when comparing groups of patients participating in randomized clinical trials of different postoperative protocols. The relative reliability of the single hop for distance test in patients 1 to 2 years following ACL reconstruction has previously been reported.[11,29,31] Intraclass correlation coefficients for the single hop for distance test reported in Table 5 were similar to those previously reported by Kramer et al[29] (ICC=.76–.96). The ICCs for limb symmetry scores on the 6-m timed hop and crossover hop tests (Tab. 5) were slightly lower than those reported by Hopper et al[31] (6-m timed hop test, ICC=.93–.96; crossover hop test, ICC=.94–.98). To our knowledge, the ICC for the triple hop for distance test has not been previously reported in patients following ACL reconstruction.

We are unaware of previous reports of the SEM for hop test scores in patients following ACL reconstruction. The present findings facilitate the clinical use of hop tests by providing estimates of measurement error and minimal detectable change (Tab. 5) that enable clinicians to determine how much confidence they can place in their assessment of an individual's hop test limb symmetry index. For example, based on an individual's performance on the overall combination of hops assessed at one point in time (Tab. 5), the limb symmetry index could vary ±4.99% simply due to measurement error (ie, ±SEM × z value for 90% confidence=±3.04×1.64=±4.99%). Additionally, based on the observed minimal detectable change for the overall limb symmetry index (Tab. 5), 90% of stable patients would change by less than 7.05% on repeated measures (ie, ±SEM × z value for 90% confidence×$\sqrt{2}$=±3.04×1.64×$\sqrt{2}$=±7.05%).

The following description provides an example of how a physical therapist might use these values in clinical practice. Following adequate practice with hop testing, a patient 16 weeks after ACL reconstruction scores a limb symmetry index of 80% for the overall combination of hops, and the score improves to 90% following 6 weeks of treatment. Upon initial assessment, the clinician can be 90% confident that the true limb symmetry index value could vary from 75% to 85% simply due to measurement error (ie, 80% ± approximately 5%). When tested 6 weeks later, the clinician can be confident that this patient has truly improved because the observed change of 10% (ie, an increase from 80% to 90%) exceeds the minimal detectable change of approximately 7%. Also note that the minimal detectable change could represent deterioration in performance. For example, if the patient's score dropped to 70% upon reassessment, the clinician can be confident that this patient has truly deteriorated because the observed change of 10% (ie, a decrease from 80% to 70%) also exceeds the minimal detectable change of approximately 7%.

The present findings are consistent with our constructs for change and provide evidence of longitudinal validity. When investigating known-groups validity, each of the hop tests demonstrated significantly greater changes on the operative limb than on the nonoperative limb over the 6-week period. When investigating convergent validity, the observed correlations between the change in limb symmetry index and change in both self-report measures, the single hop test, the crossover hop test, and the overall combination of hops met our criteria for at least moderate evidence of convergent validity. Interestingly, only the correlation between the change in the limb symmetry index for the overall combination of tests and the global rating of change exceeded .5 (Tab. 6). We speculate that this is because the change in combination of tests provided a more global representation of change in motor performance than any one test alone.

We decided to keep the order of the individual hop tests that make up the full test consistent with its original description.[18] In our experience, the 4 hop tests progress logically from less difficult to more difficult, and the initial tests may help to improve performance on the later, more difficult tests. Although reliability would not likely differ from the present findings if a clinician decided to administer just the single hop for distance test (indeed, the present ICC is similar to those reported by Kramer et al[29] on just the single hop test), reliability is more likely to change if a clinician decided to administer just one of the more difficult hop tests without adequate practice. Similarly, our experience with these tests suggests that considerable motor learning is likely when first performing them. It is advisable, therefore, to incorporate considerable practice before stable values can be recorded (eg, we used a "practice day" in the present study to ensure that our subjects' performances were stable). The limitation in the generalizability of the present findings to the described order of testing and the use of a practice session should be recognized.

Although no subject reported pain during a test session, it is important to note that 2 subjects experienced thigh pain after 2 consecutive days of testing and subsequently withdrew from this study. The 2 subjects were the only subjects to report pain following testing. They were reviewed by the operating surgeon 6 months postoperatively and had fully recovered with no adverse effects. Although guidelines for the postoperative rehabilitation protocol used in the care of our subjects suggested that hopping activities should be incorporated by the 16th week following surgery, this was not the case for the 2 subjects who experienced thigh pain. Considering the repeated eccentric muscle contractions required for the landing portions of hop tests, we believe these 2 subjects experienced delayed onset muscle soreness. Clinicians should be aware of this possibility, clearly question patients about activities that they are accustomed to performing before deciding to use the hop tests, and clearly state the risk to patients undergoing testing.

Conclusion

The described series of 4 hop tests provide reliable and valid performance-based outcome measures for patients undergoing rehabilitation after ACL reconstruction. These findings support the use and facilitate the interpretation of hop tests for research and clinical practice.

All authors provided concept/idea/research design. Ms Reid, Dr Birmingham, and Mr Stratford provided writing and data analysis. Ms Reid provided data collection and project management. Ms Reid and Dr Birmingham provided fund procurement. Dr Birmingham, Mr Stratford, and Dr Griffin provided consultation (including review of manuscript before submission).

The authors acknowledge the assistance of Michael Hunt and Jennifer Symmes in the completion of this project.

This study was approved by the University of Western Ontario Research Ethics Board for Healthy Sciences Research Involving Human Subjects, which is organized and operates according to the Tri-Council Policy Statement and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines.

This research was undertaken, in part, thanks to funding from the Canadian Orthopaedic Foundation and the Canada Research Chairs Program.

This article was received May 21, 2006, and was accepted November 6, 2006.

10.2522/ptj.20060143

References

1 Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. *J Rheumatol.* 1997;24:799–802.

2 Jackowski D, Guyatt G. A guide to health measurement. *Clin Orthop Rel Res.* 2003; 413:80–89.

3 Alcock GK, Stratford PW. Validation of the lower extremity functional scale on athletic subjects with ankle sprains. *Physiother Can.* 2002;54:233–240.

4 MacDermid J, Stratford PW. Applying evidence on outcome measures to hand therapy practice. *Journal of Hand Therapy.* 2004;17:165– 173.

5 Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther.* 1997;77: 745–750.

6 Risberg MA, Lewek M, Snyder-Mackler L. A systematic review of evidence for anterior cruciate ligament rehabilitation: how much and what type? *Physical Therapy in Sport.* 2004;5:125–145.

7 Gotlin RS, Huie PA. Anterior cruciate ligament injuries: operative and rehabilitation options. *Phys Med Rehabil Clin North Am.* 2000;11:895–928.

8 Fitzgerald GK, Lephart SM, Hwang JH, Wainner MR. Hop tests as predictors of dynamic knee stability. *J Orthop Sports Phys Ther.* 2001;31:588–597.

9 Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol.* 2006;59:160–167.

10 Stratford PW, Kennedy D, Pagura SMC, Gollish JD. The relationship between self-report and performance-related measures: questioning the content validity of timed tests. *Arthritis Rheum.* 2003;49:535–540.

11 Brosky JA, Nitz AJ, Malone TR, et al. Intrarater reliability of selected clinical outcome measures following anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther.* 1999;29:39–48.

12 Kennedy D, Stratford PW, Pagura SMC, et al. Comparison of gender and group differences in self-report and physical performance measures in total hip and knee arthroplasty candidates. *J Arthroplasty.* 2002;17:70–77.

13 Brach JS, VanSwearingen JM, Newman AB, Kriska AM. Identifying early decline of physical function in community-dwelling older women: performance-based and self-report measures. *Phys Ther.* 2002;82:320–328.

14 Borsa PA, Lephart SM, Irrgang JJ. Comparison of performance-based and patient-reported measures of function in anterior-cruciate-ligament-deficient individuals. *J Orthop Sports Phys Ther.* 1998;28:392–399.

15 Daniel DM, Malcom L, Stone ML, et al. Quantification of knee stability and function. *Contemporary Orthopaedics.* 1982; 5:83–91.

16 DeCarlo MS, Sell KE. The effects of the number and frequency of physical therapy treatments on selected outcomes of treatment in patients with anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther.* 1997;26:332–339.

17 Petschnig R, Baron R, Albrecht M. The relationship between isokinetic quadriceps strength test and hop tests for distance and one-legged vertical jump test following anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther.* 1998;28:23–31.

18 Noyes FR, Barber SD, Mangine RE. Abnormal lower limb symmetry determined by function hop tests after anterior cruciate ligament rupture. *Am J Sports Med.* 1991; 19:513–518.

19 Fischer DA, Tewes DP, Boyd JL, et al. Home-based rehabilitation for anterior cruciate ligament reconstruction. *Clin Orthop Rel Res.* 1998;1:194–199.

20 Heckman TP, Noyes FR, Barber-Westin BS. Autogenic and allogenic anterior cruciate ligament rehabilitation. In: Ellenbecker TS, ed. *Knee Ligament Rehabilitation.* New York, NY: Churchill Livingstone Inc; 2000: 132–150.

21 Manal TJ, Snyder-Mackler L. Practice guidelines for anterior cruciate ligament rehabilitation: a criterion-based rehabilitation progression. *Operative Techniques in Orthopaedics.* 1996;6:190–196.

22 Williams GN, Chmielewski T, Rudolph K, et al. Dynamic knee stability: current theory and implications for clinicians and scientists. *J Orthop Sports Phys Ther.* 2001; 31:546–566.

23 Eastlack ME, Axe MJ, Snyder-Mackler L. Laxity, instability, and functional outcome after ACL injury: copers versus noncopers. *Med Sci Sports Exerc.* 1999;31:210–215.

24 Fitzgerald GK, Axe MJ, Snyder-Mackler L. A decision-making scheme for returning patients to high-level activity with nonoperative treatment after anterior cruciate ligament rupture. *Knee Surg Sports Traumatol Arthrosc.* 2000;8:76– 82.

25 Fitzgerald GK, Axe MJ, Snyder-Mackler L. The efficacy of perturbation training in nonoperative anterior cruciate ligament rehabilitation programs for physically active individuals. *Phys Ther.* 2000;80:128–140.

26 Rudolph KS, Axe JM, Snyder-Mackler L. Dynamic stability after ACL injury: who can hop? *Knee Surg Sports Traumatol Arthrosc.* 2000;8:262–269.

27 Rudolph KS, Axe JM, Buchanan TS, et al. Dynamic stability in the anterior cruciate ligament deficient knee. *Knee Surg Sports Traumatol Arthrosc.* 2001;9:62–71.

28 Bolgla LA, Keskula DR. Reliability of lower extremity functional performance tests. *J Orthop Sports Phys Ther.* 1997;26:138–142.

29 Kramer JF, Nusca D, Fowler P, Webster-Bogaert S. Test-retest reliability of the one-leg hop test following ACL reconstruction. *Clin J Sport Med.* 1992;2:240–243.

30 Paterno MV, Greenberger HB. The test-retest reliability of a one legged hop for distance in young adults with and without ACL reconstruction. *Isokinet Exerc Sci.* 1996;6:1–6.

31 Hopper DM, Goh SC, Wentworth LA, et al. Test-retest reliability of knee rating scales and functional hop tests one year following anterior cruciate ligament reconstruction. *Physical Therapy in Sport.* 2002;3:10–18.

32 Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis.* 1957;16:494–502.

33 Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability. *Statistics in Medicine.* 1998;17:101–110.

34 Barber SD, Noyes FR, Mangine RE, et al. Quantitative assessment of functional limitations in normal and anterior cruciate ligament-deficient knees. *Clin Orthop Rel Res.* 1990;255:204–214.

35 Daniel KM, Stone ML, Riehl B, Moore MR. A measurement of lower limb function: the one leg hop for distance. *Am J Knee Surg.* 1982;1:212–214.

36 Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther.* 1999;79:371–383.

37 Stratford PW, Hart DL, Binkley JM, et al. Interpreting lower extremity functional status scores. *Physiother Can.* 2005;57:154–162.

38 Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–415.

39 Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol.* 2002;55:900–908.

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

41 Streiner DL, Norman JG. *Health Measurement Scales: A Practical Guide to Their Development and Use.* 2nd ed. Oxford, United Kingdom: Oxford Medical Publications; 1995.

42 Stratford PW, Binkley JM, Solomon P, et al. Defining the minimum level of detectable change for the Roland-Morris Questionnaire. *Phys Ther.* 1996;76:359–368.

43 Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York, NY: John Wiley & Sons Inc; 1986.