

Establishing Age- and Sex-Specific Norms for Pediatric Return-to-Sports Physical Performance Testing

John R. Magill,^{*†‡§} DPT, Heather S. Myers,[‡] DPT, ATC, Trevor A. Lentz,^{||¶} PT, PhD, MPH, Laura Pietrosimone,^{†‡} DPT, PhD, Thomas Risoli,[#] MS, Cindy L. Green,^{||**} PhD, Emily K. Reinke,[¶] PhD, Michael R. Messer,^{††} DPT, and Jonathan C. Riboh,^{†‡§§} MD

Investigation performed at Duke Sports Medicine, Durham, North Carolina, USA

Background: Graft tears and contralateral anterior cruciate ligament (ACL) tears are common in pediatric athletes after ACL reconstruction. Use of objective return-to-sports (RTS) criteria, in particular physical performance tests (PPTs), is believed to reduce the incidence of secondary injury; however, pediatric norms for these tests are unknown.

Purpose: To establish a proof of concept for the creation of age- and sex-based norms for commonly used RTS PPTs in healthy pediatric athletes, allowing the creation of growth curves for clinical referencing.

Study Design: Cross-sectional study; Level of evidence, 3.

Methods: A total of 100 healthy people who were between the ages of 6 and 18 years and involved in organized sports were enrolled, with even distributions of age and sex. All participants underwent 9 common RTS PPTs: stork test, stork test on Bosu, single-leg squat, single-leg squat on Bosu, clockwise and counterclockwise quadrant hops, single-leg hop for distance, 6-m timed hop, and triple crossover hop for distance. Mean performance across limbs was calculated for each individual. Chronological age, height, weight, sex, and self-reported Pubertal Maturational Observational Scale (PMOS) score were recorded. Univariable and multivariable models were created for each PPT, assessing the importance of the recorded descriptive variables. Quantile regression was used to create growth curves for each PPT.

Results: The cohort was 52% female, and the mean \pm standard deviation age was 11.7 ± 3.6 years. PMOS was highly correlated with age ($r = 0.86$) and was excluded from the regressions. In univariable regression, age, height, and weight were strong predictors of performance for all PPTs, whereas sex was a predictor of performance on the single-leg and triple crossover hops for distance (with males outperforming females). Height and weight were excluded from multivariable regression because of multicollinearity with age. Multivariable regression showed predictive patterns for age and sex that were identical to those shown in the univariable analysis. Given ceiling effects, quantile regression for the stork tests was not possible, but quantile regression growth curves were successfully created for the 7 remaining PPTs.

Conclusion: Chronological age and sex accurately predicted performance on common RTS PPTs in pediatric patients. The growth curves presented herein could assist clinicians with benchmarking pediatric patients postoperatively against a healthy athletic cohort.

Keywords: pediatric sports medicine; return to sport testing; knee; ACL

Sports medicine surgeons are frequently faced with the difficult task of deciding when athletes are ready to return to sports (RTS) after anterior cruciate ligament (ACL) reconstruction. This decision is particularly challenging in the pediatric athlete, in whom the rate of second ACL injury after RTS approaches 35%.^{1,3-5,15} The majority of surgeons and physical therapists use some combination of physical performance tests (PPTs), most commonly hop tests, to assess readiness for RTS. These tests are often used in

conjunction with patient-reported outcomes³ and psychological metrics of readiness⁶ to formulate a decision about return to play. Unfortunately, the performance of PPTs in predicting risk of reinjury is still highly contested.^{4,8}

A unique challenge in the pediatric population is that age- and sex-based norms for common PPTs are unknown. As a result, athletes are judged solely based on a limb symmetry index (LSI), which compares their postoperative limb to their unaffected limb, typically requiring LSI >90% for RTS. This is an intrinsically flawed strategy, as recent data have shown that fewer than half of healthy pediatric athletes can achieve LSI >90% on hop testing, even in the absence of injury.⁶ Additionally, it is well

The Orthopaedic Journal of Sports Medicine, 9(8), 23259671211023101
DOI: 10.1177/23259671211023101
© The Author(s) 2021

This open-access article is published and distributed under the Creative Commons Attribution - NonCommercial - No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits the noncommercial use, distribution, and reproduction of the article in any medium, provided the original author and source are credited. You may not alter, transform, or build upon this article without the permission of the Author(s). For article reuse guidelines, please visit SAGE's website at <http://www.sagepub.com/journals-permissions>.

established that deconditioning of the nonoperative limb occurs after ACL reconstruction, thereby inflating LSI at the time of RTS.¹⁰

We therefore believe that benchmarking pediatric athlete performance against age- and sex-based norms, in addition to considering LSI, might improve the predictive accuracy of RTS testing. The present study is aimed at filling a critical gap in the literature by providing a proof of concept to establish age- and sex-specific normative data from healthy pediatric athletes for the most commonly used PPTs in ACL-related RTS decision making. Clinically, this will allow the treating physician to plot an individual patient's performance against a standard growth curve and thereby understand how that patient is faring compared with age- and sex-matched healthy athletes. This provides a more complete picture of recovery than does a simple LSI and may inform clinical decisions about return to play. This is particularly important in the growing athlete, who often returns to sport with a peer group that is developmentally 9 to 12 months ahead of where the athlete was when he or she became injured, making that athlete's own baseline an improper metric of readiness for return to competition. Our findings will also enable future investigators to evaluate the predictive ability of norm-enhanced RTS testing.

METHODS

Cohort Selection

This study included healthy, uninjured volunteers (n = 100) between the ages of 6 and 18 years. Participants were recruited from the local community using email correspondence and public flyers and were required to be actively participating in an organized sport at the time of study participation. Individuals were excluded if they exhibited any of the following: (1) prior ligamentous injury to the knee, hip, or ankle, (2) prior spine or lower extremity surgery, (3) any lower extremity injury within the previous 6 months, (4) inability to participate in neuromuscular testing because of pain or cognitive impairment, (5) ongoing use of any brace or lower extremity orthosis, or (6) any

condition or illness that would cause severe shortness of breath with 1 hour of exercise. We made efforts to ensure an even distribution in chronological age of participants. Before enrollment, written informed consent and assent were obtained from a parent and the participant, respectively, for all participants aged 6 to 17 years, and written informed consent was obtained from all 18-year-old participants. The study protocol was approved by the biomedical institutional review board at Duke University Health System.

Data Collection

This study used a prospective, nontherapeutic design. All testing sessions were administered by 1 of 6 licensed physical therapists (including J.R.M., H.S.M., L.P., and M.R.M.). The testing protocol was designed to include a battery of PPTs that could be administered in any clinical or athletic environment to ensure robust external validity. Before administering the testing protocol, investigators recorded descriptive information including age (years), height (centimeters), weight (pounds), leg length (centimeters), and primary sport. Because chronological age alone has been demonstrated to be an insufficient metric of development, participants also completed the Pubertal Maturational Observational Scale (PMOS),⁵ a validated questionnaire assessing pubertal maturation.¹⁶

The performance testing protocol consisted of 9 functional tasks commonly used in rehabilitation and RTS testing, including balance, squatting, and hopping tasks. Detailed descriptions of each of these tasks and procedures for testing are provided in Table 1. All participants completed the study protocol on a firm surface in a physical therapy clinic. All tasks were performed bilaterally and in the same consecutive order. However, the task with which each participant started (tasks 1-9) and the extremity tested first (left or right) were randomized a priori using a random number generator. For each task, the number of errors was recorded according to the error classification system described in Table 1. Participants completed 2 trials on each leg for each test with the exception of the single-leg hop tests, which were performed 3 times. We developed a composite score for each test by averaging trials across

*Address correspondence to John R. Magill, DPT, Doctor of Physical Therapy Division, Duke University Medical Center, 311 Trent Drive, Durham, NC 27710, USA (email: johnryanmagill@gmail.com) (Twitter: @JMagillPT).

[†]Doctor of Physical Therapy Division, Duke University Medical Center, Durham, North Carolina, USA.

[‡]Department of PT/OT, Duke University Health System, Durham, North Carolina, USA.

[§]Department of Physical Therapy Education, Elon University, Elon, North Carolina, USA.

^{||}Duke Clinical Research Institute, Durham, North Carolina, USA.

[¶]Department of Orthopedic Surgery, James Urbaniak Sports Science Institute, Duke University Medical Center, Durham, North Carolina, USA.

[#]Duke Biostatistics, Epidemiology, and Research Design (BERD) Methods Core, Duke University, Durham, North Carolina, USA.

^{**}Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, North Carolina, USA.

^{††}Nashville Soccer Club, Nashville, Tennessee, USA

^{‡‡}OrthoCarolina, Charlotte, North Carolina, USA.

^{§§}Atrium Health Musculoskeletal Institute, Charlotte, North Carolina, USA.

Final revision submitted December 10, 2020; accepted January 25, 2021.

One or more of the authors has declared the following potential conflict of interest or source of funding: J.C.R has received hospitality payments from Stryker Corp and Smith & Nephew. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval for this study was obtained from Duke Medicine (protocol ID: Pro00080805).

TABLE 1
Description of Neuromuscular Testing Protocol^a

Task No.	Task	Description	Scoring Metric	No. of Trials per Limb	Error Classification
1	Single-leg stork	The participant is asked to balance on a single leg.	Time (s) (60-s maximum)	2	Timer stopped if loss of balance and contralateral foot touches ground.
2	Single-leg stork on Bosu	The participant is asked to balance on a single leg on a Bosu ball.	Time (s) (60-s maximum)	2	Timer stopped if loss of balance and contralateral foot touches ground.
3	Single-leg squat	The participant is asked to balance on a single leg while performing a squat deep enough to create a 45° of flexion angle at the knee. The tester predetermines the angle of the knee using a goniometer and then places a target dowel to ensure adequate depth and alignment.	No. of repetitions in 60 s	2	Must achieve 45° of knee flexion to classify as a successful repetition.
4	Single-leg squat on Bosu	A Bosu ball is placed (flat side up) under the stance leg, and the task is performed as described for the single-leg squat. If the participant loses his or her balance, he or she may resume within the 60-s period.	No. of repetitions in 60 s	2	Must achieve 45° of knee flexion to classify as a successful repetition.
5	Quadrant hop—clockwise	The participant is asked to hop on a single leg in a clockwise direction for 30 s. The number of correct hops (at least 50% of the foot surface lands in the correct quadrant) is counted. Each quadrant is 40 cm ² .	No. of repetitions in 30 s	2	Must not allow foot to touch tape to classify as a successful repetition.
6	Quadrant hop—counterclockwise	Same as for the quadrant hop-clockwise, only in a counterclockwise direction.	No. of repetitions in 30 s	2	Must not allow foot to touch tape to classify as a successful repetition.
7	Single-leg hop for distance	The participant is asked to hop as far as possible on 1 foot without losing his or her balance during landing.	Distance (cm)	3	No contralateral foot touchdown or double hop on landing.
8	6-m timed hop	The participant is asked to hop, as quickly as possible, on 1 foot for a distance of 6 m.	Time (s)	3	No contralateral foot touchdown.
9	Triple crossover hop for distance	The participant is asked to hop as far as possible on 1 foot while crossing the tape measure 3 times.	Distance (cm)	3	No contralateral foot touchdown or double hop on landing; entire foot must cross tape line during crossover.

^aTesting protocol originally compiled by Daniel Cooper, MD, and David Surprenant, ATC, of the Carrell Clinic (Dallas, Texas).

sides. This composite score (1 score per participant for each test) was used for analyses.

Statistical Analysis

We used a stepwise approach to the analysis of our data. First, we developed descriptive statistics and plots of the raw data. We then performed univariable regression of all available descriptive data as a screen for meaningful predictors of performance. We used the information from the univariable analysis to build a more robust multivariable model that allowed for estimates of the isolated effects of individual descriptive predictors on athlete performance. Therefore, we based our required sample size on the recommendation that 10 to 20 subjects per variable allows for the optimal assessment of multivariable regression

parameters, and we set out to enroll 100 participants (subjects per variable = 20).² Finally, because ordinary regression allows only an estimation of relationships to the mean, we performed quantile regression. Quantile regression fits models for specific percentiles of a data set, allowing a more thorough investigation of the response variable and allowing for different models based on the percentile. This provides much richer information about the data than does a simple regression, in which the variability of athletes' performance at a given age is reduced to the mean. Additionally, the use of quantile regression allows for the creation of growth curves by allowing modeling of any given percentile of the cohort.

Categorical descriptive variables were summarized using counts and percentages, and continuous descriptive variables were presented as means, standard deviations,

TABLE 2
Characteristics of the Study Population (N = 100)^a

Variable	Value
Age, y	
Mean ± SD (range)	11.7 ± 3.6 (6.0-18.0)
Median (Q1, Q3)	11.5 (9.0, 14.5)
Female sex, n (%)	52 (52.0)
Height, cm	
Mean ± SD (range)	59.3 ± 8.1 (40.0-76.0)
Median (Q1, Q3)	60.0 (53.0, 65.0)
Weight, kg	
Mean ± SD (range)	99.9 ± 38.2 (40.0-195.0)
Median (Q1, Q3)	96.0 (67.0, 132.0)
Pubertal age, PMOS score	
Mean ± SD (range)	4.0 ± 3.3 (0.0-9.0)
Median (Q1, Q3)	3.0 (1.0, 8.0)
Sport played, n (%) ^b	
Basketball	11 (11.1)
Football	1 (1.0)
Gymnastics	5 (5.1)
Running	3 (3.0)
Soccer	52 (52.5)
Swimming	4 (4.0)
Volleyball	2 (2.0)
Wrestling	1 (1.0)
Other	20 (20.2)

^aPMOS, Pubertal Maturational Observational Scale; Q, quartile.

^bn = 99 participants.

medians, 25th and 75th percentiles, and ranges. We constructed univariable linear models to assess the effects of descriptive variables on PPT performance, creating separate models for each PPT variable as well as for each predictor (sex, age, height, weight). Then, multivariable linear models were created to test for adjusted effects of descriptive data (sex and age) for each PPT. Of note, because age, height, and weight were highly correlated with one another, we decided to include only age in the multivariable models. All models were constructed using robust standard errors. For the stork test and stork test on Bosu, the log-normal distribution was assumed for the errors because of an observed ceiling effect at a value of 60. All other models assumed the normal distribution.

To create reference growth curves for clinical use, we performed quantile regression for each PPT as a factor of age, holding other covariates (in this case, sex) constant. This analysis was performed using SAS Version 9.4 (SAS Institute). Age was centered at the mean of 12 years and was modeled by also including a quadratic interaction term. For those PPTs where sex was not predictive of performance in multivariable regression, reference charts for male participants only were created. For those PPTs where sex was predictive of performance in multivariable regression (single-leg hop for distance and triple crossover hop for distance), charts for both male and female participants were created. Predicted norms were reported for the 5th, 10th, 25th, 50th, 75th, 90th, and 95th quantiles.

TABLE 3
Results of Performance Testing (N = 100)^a

Variable	Value
Stork test	
Mean ± SD (range)	50.3 ± 13.8 (6.3-60.0)
Median (Q1, Q3)	59.5 (44.0, 60.0)
Stork test on Bosu	
Mean ± SD (range)	50.3 ± 13.8 (0.3-60.0)
Median (Q1, Q3)	59.5 (5.0, 48.1)
Single-leg squat	
Mean ± SD (range)	48.0 ± 12.4 (16.0-79.8)
Median (Q1, Q3)	46.9 (38.9, 56.0)
Single-leg squat on Bosu	
Mean ± SD (range)	31.9 ± 15.0 (0.0-64.3)
Median (Q1, Q3)	33.0 (23.9, 40.6)
Quadrant hop—clockwise	
Mean ± SD (range)	42.0 ± 16.1 (8.5-73.8)
Median (Q1, Q3)	43.3 (29.3, 53.3)
Quadrant hop—counterclockwise	
Mean ± SD (range)	38.9 ± 15.8 (6.0-73.5)
Median (Q1, Q3)	38.8 (26.0, 50.4)
Single-leg hop for distance	
Mean ± SD (range)	102.1 ± 29.9 (32.0-185.7)
Median (Q1, Q3)	105.0 (78.1, 123.5)
6-m timed hop	
Mean ± SD (range)	2.7 ± 0.8 (1.7-7.0)
Median (Q1, Q3)	2.6 (2.2, 3.0)
Triple crossover hop for distance	
Mean ± SD (range)	287.9 ± 106.2 (97.0-557.3)
Median (Q1, Q3)	294.1 (208.9, 368.7)

^aQ, quartile.

RESULTS

Cohort Characteristics

The characteristics of the cohort are summarized in Table 2, and results of the performance tests are shown in Table 3.

Performance as a Factor of Age and Sex

The raw performance data for all individuals were plotted for each PPT as a factor of age and sex (Figure 1). In our cohort, there was an almost perfect linear correlation between PMOS and chronological age (Pearson $r = 0.86$); therefore, only chronological age was used for all subsequent analyses and regressions. Predicted ceiling effects occurred at 60 seconds for the stork test and stork test on Bosu, as both tests were stopped if the individual reached 60 seconds. Plots of performance as a function of height and weight were not included because age, height, and weight were highly correlated with each other.

Univariable Regression

Univariable regression of the performance data revealed that age, height, and weight were significant predictors of performance for all of the 9 PPTs evaluated in this study (Table 4). In contrast, sex was predictive of performance only for the single-leg hop and triple crossover hop for

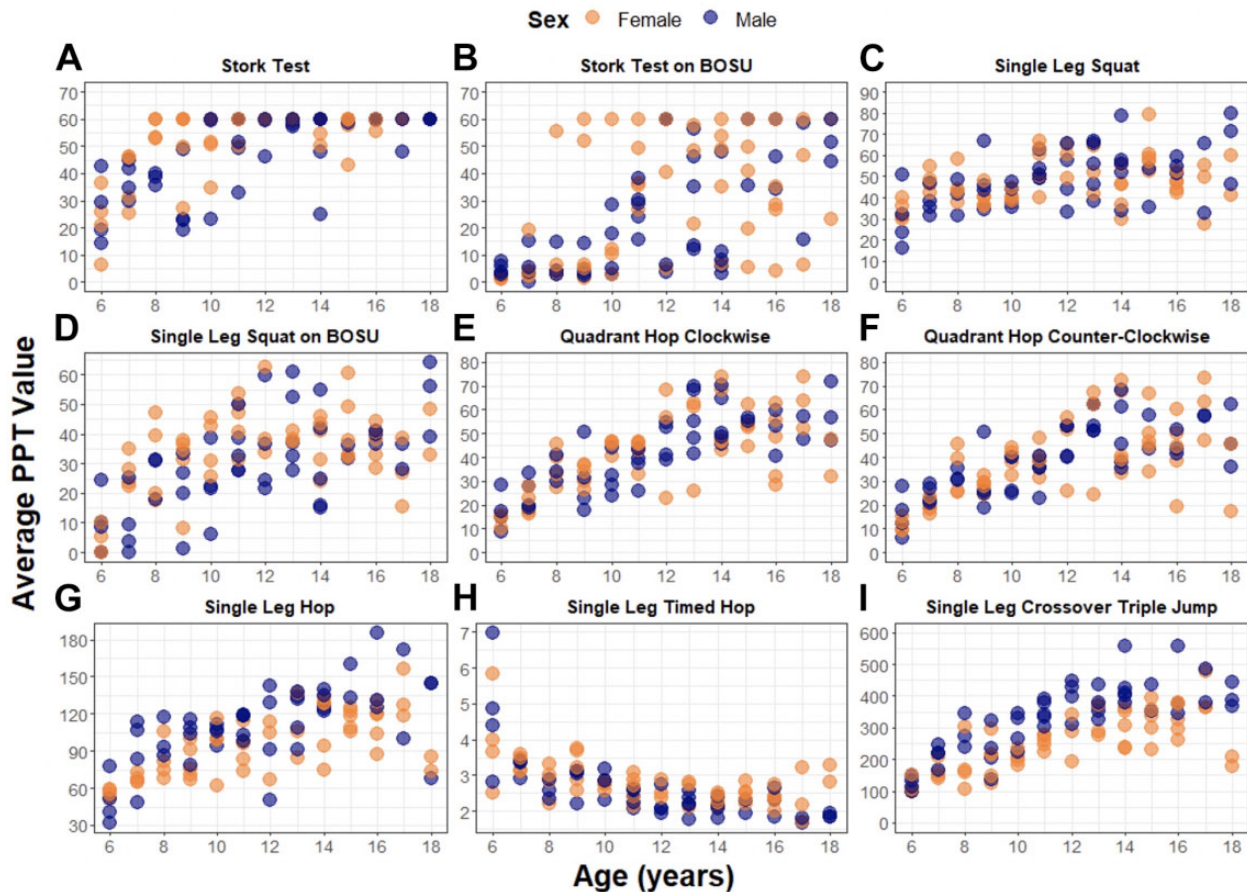


Figure 1. Physical performance test (PPT) results as a factor of age and sex. (A) Stork test; y-axis represents time in seconds. (B) Stork test on Bosu; y-axis represents time in seconds. (C) Single-leg squat test; y-axis represents number of repetitions. (D) Single-leg squat on Bosu; y-axis represents number of repetitions. (E) Quadrant hop—clockwise; y-axis represents number of repetitions. (F) Quadrant hop—counterclockwise; y-axis represents number of repetitions. (G) Single-leg hop for distance; y-axis represents hop distance in centimeters. (H) 6-m timed hop; y-axis represents total hop time in seconds. Of note, a shorter time denotes better performance. (I) Triple crossover hop for distance; y-axis represents total hop distance in cm.

distance, with male participants performing better than their female counterparts performed (Table 4).

Multivariable Regression

Significant multicollinearity existed among age, height, and weight, so height and weight were eliminated from the final multivariable model. In this model, age was predictive of performance for all 9 PPTs, with increasing age predicting improved performance. Consistent with the univariable regression, sex was found to be predictive of performance for only the single-leg hop and triple crossover hop for distance, with male athletes performing better than female athletes performed (Appendix Table A1).

Quantile Regression

Quantile regression was used to create growth curves of predicted performance for each PPT as a factor of age, controlling for sex. Of note, given the strong ceiling effects for

the stork test and stork test on Bosu, these were not appropriate for quantile regression. These growth curves are shown for the squat tests in Figure 2, for the quadrant hop tests in Figure 3, and for the linear hop tests in Figure 4.

DISCUSSION

The principal finding of this study is that chronological age is a strong predictor of performance on common RTS PPTs, allowing for the creation of growth curves using quantile regression that can help clinicians benchmark their patients against a healthy, uninjured pediatric normative group when making return-to-play decisions.

The decision to clear a child for sports participation after ACL reconstruction is one of the most challenging clinical situations we face in pediatric sports medicine. Although historically a period of 6 months was used as a blanket criterion for RTS,³ ample data have shown that in children, the requisite time for successful RTS is significantly longer, often upward of 1 year.^{6,7,9} Although time alone is not a

TABLE 4
Univariable Regression of PPT Results as a Factor of Descriptive Variables^a

Physical Performance Test	Sex ^b	Age	Height	Weight
Stork test ^c				
exp(β)	1.10	1.07	1.03	1.005
95% CI	0.95 to 1.29	1.04 to 1.09	1.02 to 1.04	1.003 to 1.007
<i>P</i> value	.202	<.001	<.001	<.001
Stork test on Bosu ^c				
exp(β)	1.18	1.24	1.09	1.02
95% CI	0.72 to 1.95	1.18 to 1.31	1.06 to 1.11	1.01 to 1.02
<i>P</i> value	.503	<.001	<.001	<.001
Single-leg squat				
β	-0.640	1.48	0.610	0.109
95% CI	-5.566 to 4.287	0.82 to 2.15	0.306 to 0.913	0.041 to 0.177
<i>P</i> value	.797	<.001	<.001	.002
Single-leg squat on Bosu				
β	4.26	2.24	0.822	0.152
95% CI	-1.65 to 10.17	1.55 to 2.94	0.490 to 1.153	0.084 to 0.221
<i>P</i> value	.156	<.001	<.001	<.001
Quadrant hop-clockwise				
β	0.084	3.23	1.14	0.223
95% CI	-6.253 to 6.420	2.58 to 3.87	0.08 to 1.48	0.153 to 0.293
<i>P</i> value	.979	<.001	<.001	<.001
Quadrant hop-counterclockwise				
β	-1.11	2.93	1.09	0.210
95% CI	-7.30 to 5.08	2.20 to 3.65	0.76 to 1.43	0.143 to 0.277
<i>P</i> value	.723	<.001	<.001	<.001
Single-leg hop for distance				
β	-14.56	5.24	2.32	0.405
95% CI	-26.12 to -3.01	3.65 to 6.80	1.67 to 2.97	0.251 to 0.558
<i>P</i> value	.014	<.001	<.001	<.001
6-m timed hop				
β	0.192	-0.136	-0.058	-0.011
95% CI	-0.123 to 0.506	-0.185 to -0.086	-0.079 to -0.038	-0.015 to -0.007
<i>P</i> value	.230	<.001	<.001	<.001
Triple crossover hop for distance				
β	-77.44	20.76	9.58	1.76
95% CI	-116.80 to -38.07	16.09 to 25.43	7.68 to 11.49	1.32 to 2.19
<i>P</i> value	<.001	<.001	<.001	<.001

^aBolded *P* values indicate statistical significance. PPT, physical performance test.

^bEstimate of female vs male participants.

^cLog-normal distribution for the errors; all other outcomes assumed normal distribution.

sufficient criterion for clearance, there is evidence in this population that each additional month of rehabilitation before RTS decreases the risk of second ACL injury by 13%,^{6,22} a finding corroborated in young adults.^{10,20} This has led to evidence-based guidelines recommending that rehabilitation be pursued for 9 to 12 months in the young, active population.^{11,19}

Although there is agreement that RTS timelines should likely be extended in children, there is no agreement in the literature about what specific criteria define a safe cutoff to end rehabilitation and resume sporting activities.^{4,14} LSI values >90% on hop tests and/or isokinetic quadriceps strength testing are frequently used criteria,^{3,13} although LSIs have intrinsic limitations, including baseline limb asymmetries that exist in children⁸ and deconditioning of the nonoperative leg after ACL reconstruction,^{17,21} which can lead to erroneous

interpretation of LSIs. In this study, we included a population of healthy athletic people between the ages of 6 and 18 years in order to develop growth curves that document predicted performance on 9 common RTS tests based on chronological age. This was supported by the finding that in multivariable regression, age was an independent predictor of performance on all PPTs. Additionally, sex was an independent predictor of performance in only 2 of 9 PPTs (single-leg hop and triple crossover hop for distance), eliminating the need for sex-specific growth curves for the majority of PPTs. Because our results were generated in a healthy population, we cannot draw any conclusions about the predictive value of including norm-based benchmarking in RTS decision making; however, we expect the data presented herein to facilitate future prospective studies that specifically address this question.

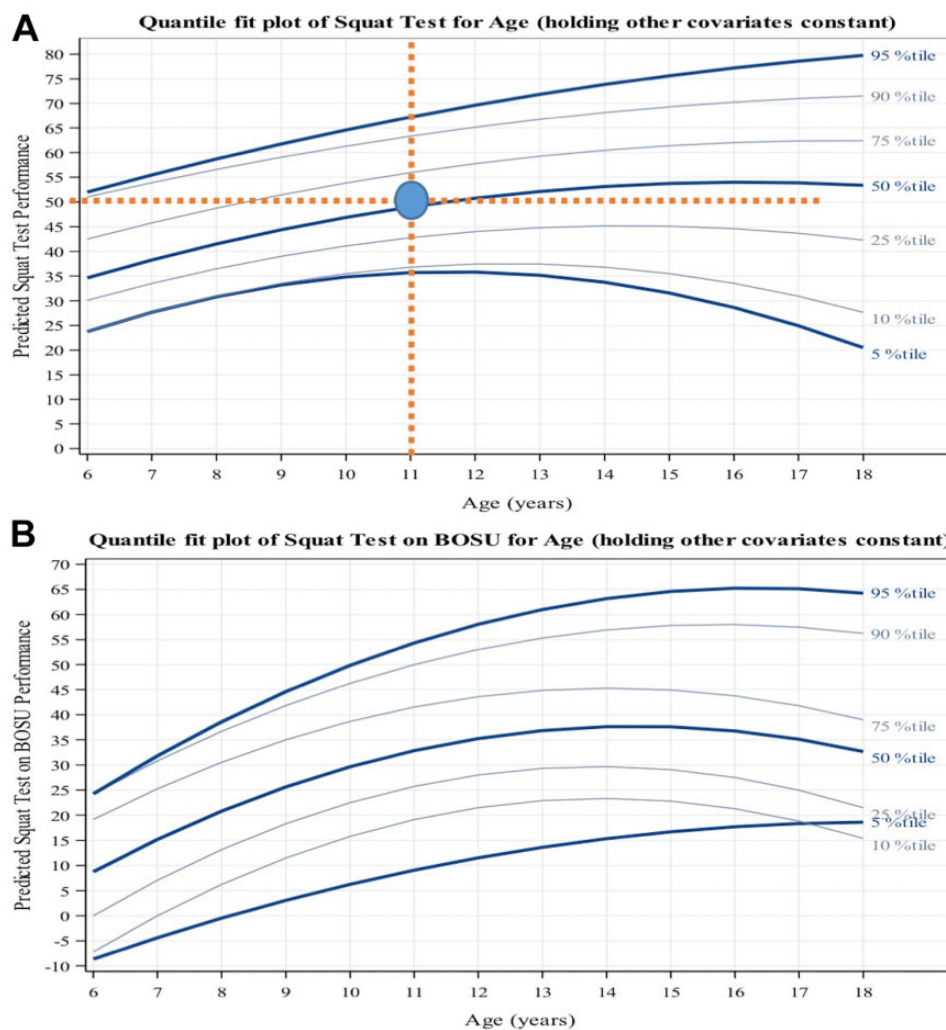


Figure 2. Quantile regression plots for the squat tests: (A) single-leg squat and (B) single-leg squat on Bosu. The x-axis represents participant age, and the y-axis represents predicted performance (number of repetitions). Regression curves are shown for the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles. Physical performance test results were modeled using age (the linear term, centered at the mean age of 12 years, and a quadratic interaction term) and sex as predictors. (A) Clinical use of the growth curve is demonstrated for male participants. Assume a hypothetical 11-year-old male athlete who performed 50 single-leg squat repetitions during testing. A vertical line is drawn from the “11-year” mark on the x-axis, and a horizontal line is drawn from the “50” mark on the y-axis. The intersection of these lines falls at the 50th percentile growth curve, informing the clinician that the male participant performed at the expected 50th percentile for his age.

The sex-specific differences in our data set are interesting. Sex was predictive of performance for only the single-leg and triple crossover hop for distance. This is concordant with previously published studies showing that male athletes outperform female athletes in measures of lower extremity power,¹¹ including improved performance on single-leg hop testing in high school-aged male athletes.¹³ Interestingly, we saw no sex-based differences in tests that focused on balance (Bosu-based tests and stork tests), agility (quadrant hops), or muscular endurance (squat tests). Therefore, during ACL rehabilitation, young female athletes should be held to the same standards as male athletes in terms of absolute performance on balance, agility, and muscular endurance tests. It will be interesting

to see whether these sex-based patterns hold true in a post-ACL reconstruction clinical cohort, as female sex is a known predictor of second ACL injury.¹³

We were somewhat surprised to find that the PMOS was so strongly correlated with chronological age ($r = 0.86$) that chronological age alone could be used for the predictive regression and quantile analyses. This could be because of inherent limitations in self-reporting of pubertal maturation that have been previously described.¹⁸ Unfortunately, Tanner staging for large epidemiological studies is both inconvenient and known to be inaccurate when performed by orthopaedic surgeons.¹⁸ In light of these limitations, we believe that our findings of strong correlation between self-reported pubertal

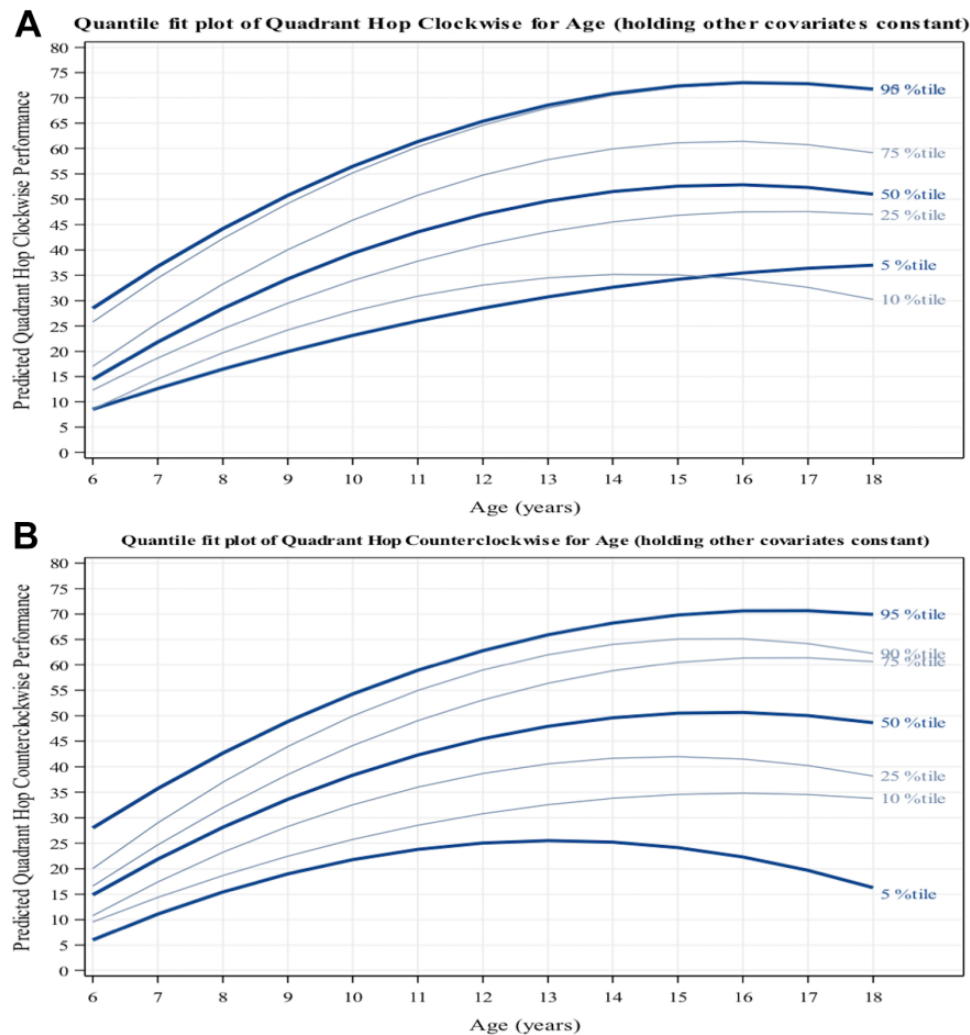


Figure 3. Quantile regression plots for the quadrant hop tests: (A) quadrant hop—clockwise and (B) quadrant hop—counterclockwise. The x-axis represents participant age, and the y-axis represents predicted performance (number of repetitions). Regression curves are shown for the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles. Physical performance test results were modeled using age (the linear term, centered at the mean age of 12 years, and a quadratic interaction term) and sex as predictors. These curves represent growth curves for male participants.

maturity and chronological age provide justification for using chronological age alone in pediatric and adolescent benchmarking, greatly simplifying the process for future prospective studies.

In addition to potentially helping guide ACL RTS decisions, our data also have value for the more general assessment of functional capacity or injury risk in the pediatric population. Indeed, hop testing and functional movement screens have been assessed in numerous studies for their injury prediction abilities, with mixed success.¹² In fact, the predictive value of these tests was worse in younger than in older athletes.¹² It is possible that by accounting for age-specific norms described in this study, PPTs might provide meaningful risk stratification tools for the pediatric athlete, perhaps even warranting inclusion in preparticipation physicals.¹⁴

Our study has several limitations, including a comparatively small cohort size, a predominance of soccer athletes (approximately 50% of the cohort), and a narrow geographic distribution of participants, all of which can affect the external validity of our findings. All efforts were made to optimize the internal validity via a single testing surface, a small and consistent group of assessors, rigorous statistical modeling, and strict inclusion and exclusion criteria. However, given the small sample, the results should be interpreted with some caution. A robust population sample is the foundational element in the establishment of normative values. Therefore, this study should be viewed as a successful proof of concept for the feasibility of creating normative growth curves. Given our analysis, we recommend future large-scale studies be done to continue the creation of this normative data.

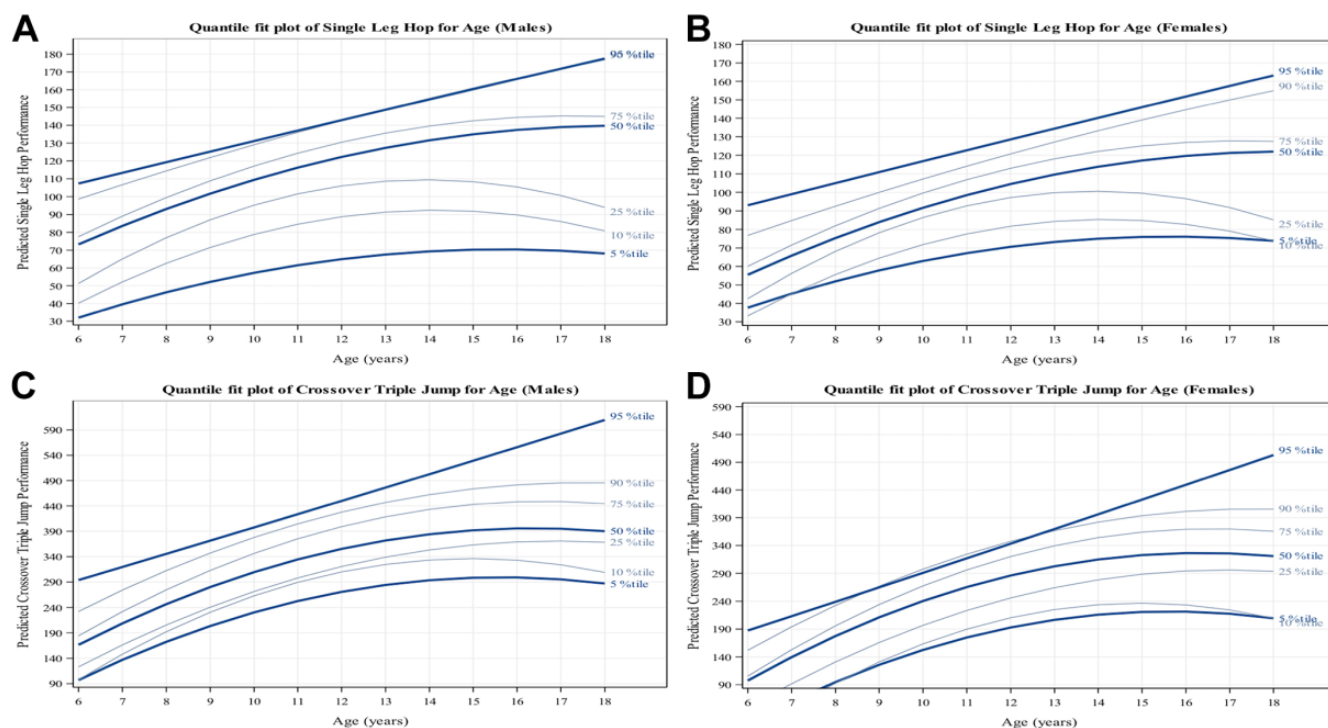


Figure 4. Quantile regression plots for the linear hop tests: (A) single-leg hop for distance (male participants), (B) single-leg hop for distance (female participants), (C) triple crossover hop for distance (male participants), and (D) triple crossover hop for distance (female participants). In each plot, the x-axis represents participant age, and the y-axis represents predicted performance (distance in centimeters). Regression curves are shown for the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles. Physical performance test results were modeled using age (the linear term, centered at the mean age of 12 years, and a quadratic interaction term) and sex as predictors.

CONCLUSION

In this study, we provided potential benchmarks for clinicians treating pediatric patients (in particular, those recovering from ACL reconstruction) to interpret the results of common PPTs within the context of age- and sex-specific norms. We hope that these results will serve as a springboard for ongoing prospective research assessing the value of PPT benchmarking in children and adolescents for risk stratification in various clinical settings, including ACL-related return to play.

REFERENCES

- Astur DC, Novaretti JV, Cavalcante ELB, et al. Pediatric anterior cruciate ligament ruptures are related to lower functional scores at the time of return to activity: a prospective, midterm follow-up study. *Orthop J Sports Med.* 2019;7(12):2325967119888888.
- Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. *J Clin Epidemiol.* 2015;68(6):627-636.
- Burgi CR, Peters S, Ardern CL, et al. Which criteria are used to clear patients to return to sport after primary ACL reconstruction? A scoping review. *Br J Sports Med.* 2019;53(18):1154-1161.
- Capin JJ, Snyder-Mackler L, Risberg MA, Grindem H. Keep calm and carry on testing: a substantive reanalysis and critique of "what is the evidence for and validity of return-to-sport testing after anterior cruciate ligament reconstruction surgery? A systematic review and meta-analysis." *Br J Sports Med.* 2019;53(23):1444-1446.
- Davies PL, Rose JD. Motor skills of typically developing adolescents: awkwardness or improvement? *Phys Occup Ther Pediatr.* 2000;20(1):19-42.
- Dekker TJ, Godin JA, Dale KM, Garrett WE, Taylor DC, Riboh JC. Return to sport after pediatric anterior cruciate ligament reconstruction and its effect on subsequent anterior cruciate ligament injury. *J Bone Joint Surg Am.* 2017;99(11):897-904.
- Geffroy L, Lefevre N, Thevenin-Lemoine C, et al. Return to sport and re-tears after anterior cruciate ligament reconstruction in children and adolescents. *Orthop Traumatol Surg Res.* 2018;104(8):S183-S188.
- Greenberg E, Dabbous M, Leung A, et al. Less than half of youth athletes can achieve 90% limb symmetry on a battery of single leg hop tests. *Orthop J Sports Med.* 2019;7(3)(suppl):2325967119S00052.
- Greenberg EM, Greenberg ET, Ganley TJ, Lawrence JTR. Strength and functional performance recovery after anterior cruciate ligament reconstruction in preadolescent athletes. *Sports Health.* 2014;6(4):309-312.
- Grindem H, Snyder-Mackler L, Moksnes H, Engebretsen L, Risberg MA. Simple decision rules can reduce reinjury risk by 84% after ACL reconstruction: the Delaware-Oslo ACL cohort study. *Br J Sports Med.* 2016;50(13):804-808.
- Lopes TJA, Simic M, Alves DdS, et al. Physical performance measures of flexibility, hip strength, lower limb power and trunk endurance in healthy Navy cadets. *J Strength Cond Res.* 2021;35(2):458-464.
- Moore E, Chalmers S, Milanese S, Fuller JT. Factors influencing the relationship between the functional movement screen and injury risk in sporting populations: a systematic review and meta-analysis. *Sports Med.* 2019;49(9):1449-1463.
- Onate JA, Starkel C, Clifton DR, et al. Normative functional performance values in high school athletes: the functional pre-participation evaluation project. *J Athl Train.* 2018;53(1):35-42.
- Paterno MV, Huang B, Thomas S, Hewett TE, Schmitt LC. Clinical factors that predict a second ACL injury after ACL reconstruction and return to sport: preliminary development of a clinical decision algorithm. *Orthop J Sports Med.* 2017;5(12):2325967117745279.

15. Paterno MV, Rauh MJ, Schmitt LC, Ford KR, Hewett TE. Incidence of second ACL injuries 2 years after primary ACL reconstruction and return to sport. *Am J Sports Med.* 2014;42(7):1567-1573.
16. Rasmussen AR, Wohlfahrt-Veje C, de Renzy-Martin KT, et al. Validity of self-assessment of pubertal maturation. *Pediatrics.* 2015;135(1):86-93.
17. Sadeqi M, Klouche S, Bohu Y, Herman S, Lefevre N, Gerometta A. Progression of the psychological ACL-RSI score and return to sport after anterior cruciate ligament reconstruction: a prospective 2-year follow-up study from the French Prospective Anterior Cruciate Ligament Reconstruction Cohort Study (FAST). *Orthop J Sports Med.* 2018;6(12):2325967118812819.
18. Slough JM, Hennrikus W, Chang Y. Reliability of Tanner staging performed by orthopedic sports medicine surgeons. *Med Sci Sports Exerc.* 2013;45(7):1229-1234.
19. Van Melick N, van Cingel REH, Brooijmans F, et al. Evidence-based clinical practice update: practice guidelines for anterior cruciate ligament rehabilitation based on a systematic review and multidisciplinary consensus. *Br J Sports Med.* 2016;50(24):1506-1515.
20. Webster KE, Hewett TE. What is the evidence for and validity of return-to-sport testing after anterior cruciate ligament reconstruction surgery? A systematic review and meta-analysis. *Sports Med.* 2019;49(6):917-929.
21. Wellsandt E, Failla MJ, Snyder-Mackler L. Limb symmetry indexes can overestimate knee function after anterior cruciate ligament injury. *J Orthop Sports Phys Ther.* 2017;47(5):334-338.
22. Wiggins AJ, Grandhi RK, Schneider DK, Stanfield D, Webster KE, Myer GD. Risk of secondary injury in younger athletes after anterior cruciate ligament reconstruction. *Am J Sports Med.* 2016;44(7):1861-1876.

APPENDIX

TABLE A1
Multivariable Regression of Physical Performance Test Results as a Factor of Sex and Age^a

Physical Performance Test	Sex ^b	Age
Stork test ^c		
exp(β)	1.09	1.07
95% CI	0.96 to 1.24	1.04 to 1.09
P value	.165	<.001
Stork test on Bosu ^c		
exp(β)	1.14	1.24
95% CI	0.77 to 1.70	1.18 to 1.31
P value	.507	<.001
Single-leg squat		
β	0.888	1.49
95% CI	-5.314 to 3.539	0.82 to 2.15
P value	.692	<.001
Single-leg squat on Bosu		
β	3.89	2.23
95% CI	-1.08 to 8.86	1.52 to 2.94
P value	.124	<.001
Quadrant hop—clockwise		
β	-0.454	3.23
95% CI	-4.824 to 3.916	2.59 to 3.87
P value	.837	<.001
Quadrant hop—counterclockwise		
β	-1.60	2.93
95% CI	-6.17 to 2.98	2.21 to 3.65
P value	.490	<.001
Single-leg hop for distance		
β	-15.45	5.30
95% CI	-24.22 to -6.67	3.79 to 6.80
P value	.001	<.001
6-m timed hop		
β	0.214	-0.136
95% CI	-0.029 to 0.458	-0.185 to -0.087
P value	.083	<.001
Triple crossover hop for distance		
β	-80.94	21.03
95% CI	-106.35 to -55.53	16.76 to 25.29
P value	<.001	<.001

^aBolded P values indicate statistical significance.

^bEstimate of females vs males.

^cLog-normal distribution for the errors; all other outcomes assumed normal distribution.