Estimation of Aerobic Fitness from 20-m Multistage Shuttle Run Test Performance

Matthew T. Mahar, EdD, Ashley M. Guerieri, MS,
Matthew S. Hanna, MS, C. David Kemble, MA

Background: Aerobic fitness (VO_2max) is a key component of youth fitness testing. Criterion-referenced (CR) assessments are used in FITNESSGRAM® to assess health risk.

Purpose: The purpose of this study was to develop and cross-validate regression models to estimate VO_2max from Progressive Aerobic Cardiovascular Endurance Run (PACER) 20-m shuttle run performance in boys and girls aged 10–16 years. Several previously published PACER models were also cross-validated. A secondary purpose was to examine the CR validity of the models.

Methods: PACER performance and VO_2max were assessed in a sample of 244 participants. The sample was randomly split into validation (n=174) and cross-validation (n=70) samples. The validation sample was used to develop the regression models to estimate VO_2max from PACER, BMI, gender, and age. CR validity was evaluated by comparing classification of the prediction models with classification by the criterion of measured VO_2max.

Results: For the Quadratic Model, the multiple correlation between measured and estimated VO_2max was 0.75, and the SE of estimate (*SEE*) was 6.17 mL/kg/min. Similar accuracy was found for Linear Model 2 (R=0.74; *SEE*=6.29 mL/kg/min). Accuracy of these models was confirmed on the cross-validation and total samples. Cross-validation demonstrated that the Quadratic Model and Linear Model 2 were slightly more accurate than previous PACER models. Evidence of CR validity for the newly developed models was of moderate levels.

Conclusions: The Quadratic Model and Linear Model 2 provide valid estimates of VO_2max and compare favorably to previous models. The CR validity evidence for the Quadratic Model and Linear Models developed in this study was slightly better than for the other models examined.
(Am J Prev Med 2011;41(4S2):S117–S123)

Introduction

Aerobic fitness (VO_2max) contributes to health[1–3] and as such is an essential component of youth fitness testing batteries. The FITNESSGRAM® is a comprehensive youth fitness program that includes a variety of fitness assessments, as well as several physical activity assessments. The tests available to assess aerobic fitness include the Progressive Aerobic Cardiovascular Endurance Run (PACER) 20-m multistage shuttle run, the 1-mile run/walk, and the 1-mile walk. The recommended test of aerobic fitness for the FITNESSGRAM is the PACER.[4]

From the Activity Promotion Laboratory, Department of Kinesiology, East Carolina University, Greenville, North Carolina

Address correspondence to: Matthew T. Mahar, EdD, Activity Promotion Laboratory, Department of Kinesiology, 101 Minges Coliseum, East Carolina University, Greenville NC 27858. E-mail: maharm@ecu.edu.
0749-3797/$17.00
doi: 10.1016/j.amepre.2011.07.008

Leger et al.[5] published an equation to estimate aerobic fitness from PACER performance. Their prediction model was developed on a sample of 188 children aged 8–19 years. VO_2max was measured by retroextrapolation immediately after a maximal test and a correlation of 0.71 was reported between measured and estimated VO_2max. Several researchers have examined the accuracy of the Leger et al. model, but few have attempted to develop a more accurate estimation model.[6–8]

Mahar et al.[8] developed a regression model to estimate VO_2max from PACER performance in 135 boys and girls aged 12–14 years and demonstrated that the model was more accurate than the Leger et al.[5] model. Barnett et al.[6] published prediction models that were developed on 55 Chinese students aged 12–17 years. Matsuzaka et al.[7] developed regression models to estimate VO_2max from PACER performance (either laps completed or maximal speed attained), gender, age, and BMI on 132 Japanese youth aged 8–17 years. The small sample sizes, narrow age ranges, and discrete samples make it difficult to employ any of these equations in research or professional practice. The relative validity of the various equations has also not been determined.

The purpose of this study was to develop and cross-validate regression models to estimate VO_2max from PACER performance in a large sample of boys and girls aged 10–16 years. The impact of several predictors was examined in an effort to improve prediction accuracy, including age, gender, and BMI. In addition, several prediction models were cross-validated for comparison purposes. A secondary purpose of this study was to examine the accuracy of the newly developed and previously developed prediction models in a criterion-referenced (CR) framework. Specifically, the ability of these models to accurately classify participants into fitness categories using FITNESSGRAM CR standards[4] was examined.

Methods

Participants

Participants included in the analyses were 244 children and adolescents (126 girls, 118 boys) aged 10–16 years. The study was approved by the IRB of East Carolina University. Written informed consent was obtained from parents, and assent was obtained from participants. Physical characteristics of the participants are presented in Table 1. Data were examined for outliers; one standardized residual score greater than 2.5 was found and deleted from the analyses.

Validation and Cross-Validation Samples

The full data set was formed from data collection on three age groups (10–11 years, 12–14 years, and 15–16 years). Data from the participants aged 12–14 years were used in a previous publication.[8] To allow cross-validation of the prediction equation developed in this study, validation and cross-validation samples were formed. Approximately one third of participants were randomly selected for the cross-validation group (n=70). The remaining participants formed the validation group (n=174).

Testing took place in two sessions. During the first visit, all participants were habituated to treadmill exercise and the PACER test. The majority of participants reported having previous experience with both treadmill exercise and the PACER. Also at the first visit, participants had height, body mass, and skinfolds assessed.

Height was assessed with a wall-mounted stadiometer (Perspective Enterprises, Portage, MI). Body mass was measured on a balance beam scale (Healthometer, Boca Raton, FL). BMI was calculated by standard formula. Skinfolds were measured at the triceps and calf sites with Lange calipers, and percent fat was estimated with the equations of Slaughter et al.[9] After body composition assessment, participants completed either the maximal treadmill test or PACER 20-m multistage shuttle run test. The other test of aerobic fitness was administered during the second session, and the order of testing was counterbalanced.

The PACER was administered following standardized procedures.[4] Participants ran from one marker to another marker set 20 m apart, while keeping pace with a prerecorded cadence. The cadence is set to music and increased every minute. Participants were instructed to keep up with the cadence for as long as possible. The test was terminated when a participant failed to reach the appropriate marker in the allotted time twice or could no longer maintain the pace. The number of laps completed was recorded.

Participants underwent a graded exercise test to volitional exhaustion on a Trackmaster (model TMX425C) or Quinton (model Q65) treadmill to determine peak oxygen consumption (VO_2max). For the girls aged 12–16 years, the speed of the treadmill was increased to 5.0 mph within the first minute. This speed was maintained for the remainder of the test. For boys aged 12–16 years, speed was increased to 5.5 mph within the first minute and maintained thereafter. For both girls and boys, at the beginning of the second minute, the treadmill grade was increased to 2%. Every

Table 1. M±SD of physical characteristics and PACER 20-m shuttle run performance

	Validation sample		Cross-validation sample	
Variable	**Girls (n=90)**	**Boys (n=84)**	**Girls (n=36)**	**Boys (n=34)**
Age (years)	13.1±1.8	12.9±1.8	13.2±1.5	12.9±1.8
Height (cm)	158.3±9.3	161.6±13.7	160.8±8.7	161.9±12.2
Body mass (kg)	51.9±13.9	54.2±15.2	56.1±14.4	55.7±15.4
BMI	20.5±4.2	20.4±3.6	21.5±4.5	20.9±3.9
Percent fat (%)	27.1±9.0	22.4±9.7	28.3±9.4	23.4±8.8
PACER (# laps)	30.2±15.4	45.3±22.5	30.1±15.8	46.5±24.2
Max PACER speed (km/hour)	10.1±0.9	10.9±1.2	10.1±0.8	10.9±1.2
VO_2max (mL/kg/min)	40.5±7.6	48.7±9.1	38.6±7.5	49.5±8.2
Max heart rate (beats/min)	200.6±9.7	198.9±8.7	201.7±10.2	199.3±8.1
Maximal RER	1.14±0.10	1.13±0.09	1.14±0.10	1.13±0.08

Note: Percent fat is estimated from triceps and calf skinfolds using the equations of Slaughter et al.[9]
min, minute; PACER, Progressive Aerobic Cardiovascular Endurance Run; RER, respiratory exchange ratio

minute thereafter, the treadmill grade was increased by an additional 2% until the participant was no longer able to continue. For participants aged 10 and 11 years, treadmill speed was set at 2.5 mph for the first minute and increased by 0.5 mph each minute until 5.0 mph was reached. Treadmill grade was maintained at 0% until 5.0 mph was reached. If a participant did not achieve a maximal effort before 5.0 mph, speed was then maintained and grade was increased by 3% each minute until the participant was no longer able to continue.

VO_2 was assessed using a COSMED K4b^2 portable metabolic system for participants aged 10, 11, 15, and 16 years. For participants aged 12–14 years, VO_2 was assessed using a Consentius Technologies-ParvoMedics TrueMax 2400 metabolic measurement system. Prior to testing, the systems were calibrated using known concentration sample gases. VO_2max was accepted as a maximal index if two of the following three conditions were satisfied: the participant was habituated to the test procedures and environment, and showed signs of intense effort (e.g., hyperpnea, facial flushing and grimacing, unsteady gait, sweating)[10]; peak heart rate reached a value at least 90% of age-predicted maximal heart rate; and respiratory exchange ratio (RER) $\geq$1.0.[11] Heart rate was monitored throughout the test with a Polar heart rate monitor (Polar Electro, Inc.).

Estimation of VO_2max

Multiple regression was used in the validation sample to predict VO_2max from the number of laps completed on the PACER, gender, age, and BMI. After examination of scatterplots, a quadratic term for PACER laps was tested. In addition, two interaction terms (i.e., age × gender; gender × PACER laps) were entered into the model to examine whether they contributed significantly to the prediction. The equations developed on the validation sample were applied to the cross-validation sample and to the total sample. The correlation coefficients between measured VO_2max and VO_2max predicted from the equations developed on the validation sample were calculated. Prediction error was assessed with two equations. The SE of estimate (*SEE*) was calculated as: $SEE = S_Y\sqrt{1 - R^2_{YY'}}$. The cross-validation SE of estimate (referred to as total error [*TE*]) was calculated as: $TE = \sqrt{\Sigma(Y - Y')^2/N}$. For these equations, Y is measured VO_2max and Y′ is VO_2max estimated from the equation developed on the validation sample. Comparison of these two error estimates quantified the overestimation or underestimation on prediction accuracy.

For comparison, cross-validation was conducted on several previously published prediction models. The Leger et al.[5] model is:

$VO_2max' = 31.025 + (3.238 \times \text{speed in km h}^{-1}) - (3.248\times age) + (0.1536\times\text{speed}\times\text{age})$,

where speed is maximal speed attained on the PACER test and age is in years. The Mahar et al.[8] model is:

$VO_2max' = 50.945 + (0.126\times\text{PACER laps}) + (4.946\times\text{gender}) - (0.655\times\text{BMI})$,

where gender = 1 if boy or 0 if girl.

Two models from Barnett et al.[6] were tested. The Barnett et al. A model is:

$VO_2max' = 25.8 - (6.6\times\text{gender}) - (0.2\times\text{mass in kg}) + (3.2\times\text{speed in km h}^{-1})$

where gender = 0 if boy or 1 if girl.

The Barnett et al. B model is:

$VO_2max' = 24.2 - (5.0\times\text{gender}) - (0.8\times\text{age}) + (3.4\times\text{speed in km h}^{-1})$

where gender = 0 if boy or 1 if girl.

Two models from Matsuzaka et al.[7] were also examined. The Matsuzaka et al. A model is:

$VO_2max' = 25.9 - (2.21\times\text{gender}) - (0.449\times\text{age}) - (0.831\times\text{BMI}) + (4.12\times\text{speed in km } h^{-1})$

where gender = 0 if boy or 1 if girl.

The Matsuzaka et al. B model is:

$VO_2max' = 61.1 - (2.20\times\text{gender}) - (0.462\times\text{age}) - (0.862\times\text{BMI}) + (0.192\times PACER \text{ laps})$

where gender = 0 if boy or 1 if girl.

The final analyses were conducted in a CR framework. Values of measured VO_2max and VO_2max predicted from each of the previously developed models were categorized using the new FITNESSGRAM standards.[4] For the first analysis, the following categories were used: healthy fitness zone (HFZ); needs improvement zone, some risk (NIZ–some risk); and needs improvement zone, higher risk (NIZ– higher risk). For the second analysis, only two categories were used: HFZ and needs improvement zone (NIZ). For these categoric variables, the proportion of agreement, modified kappa, and phi coefficient statistics between measured and predicted VO_2max were calculated.

Results

Descriptive statistics for the validation and cross-validation samples are presented in Table 1. Validation and cross-validation samples did not differ significantly ($p>0.05$) on any variable. In the validation sample, the correlation between measured VO_2max and PACER performance was 0.66. The first model tested (Linear Model 1) included the predictors of PACER laps, gender, and age. In this model, PACER performance and gender were significant predictors, but age did not contribute significantly to the prediction of measured VO_2max. The multiple *R* for Linear Model 1 was 0.68 (see Table 2). Linear Model 2 tested the inclusion of BMI to the predictors used in Linear Model 1. All predictors in Linear Model 2 contributed significantly to the prediction of measured VO_2max and the multiple *R* was increased to 0.74.

When entered into the model, the quadratic term for PACER laps was significant. The gender X age interaction term made a significant contribution to the prediction of VO_2max, but the gender X PACER laps interaction term was not significant. The Quadratic Model developed on the validation sample was:

$VO_2max' = 41.76799 + (0.49261\times\text{PACER}) - (0.00290\times\text{PACER}^2) - (0.61613\times\text{BMI}) + (0.34787\times\text{gender}\times\text{age})$

$[R = 0.75, R^2 = 0.56, \text{SEE} = 6.17 \text{ ml/kg/min}]$,

where PACER is the number of laps completed; for gender, 1 = boy and 0 = girl; and age is in years. Standardized regression coefficients demonstrated that PACER perfor-

Table 2. Multiple regression models to estimate VO_2max (mL/kg/min): validation sample (n=174)

Variable	Linear Model 1	Linear Model 2	PACER Quadratic Model
Intercept	32.56941 (25.104, 40.035)[a]	40.34533 (32.921, 47.770)	41.76799 (38.804, 48.732)
PACER (# laps)	0.27297 (0.213, 0.333)	0.21426 (0.155, 0.273)	0.49261 (0.302, 0.683)
PACER (squared)	—	—	–0.00290 (–0.005, –0.001)
BMI	—	–0.79472 (–1.078, –0.512)	–0.61613 (–0.871, –0.362)
Gender	3.25225 (1.007, 5.497)	4.27293 (2.170, 6.376)	—
Age (years)	0.02961 (–0.571, 0.630)	0.79444 (0.177, 1.412)	—
Gender × age	—	—	0.34787 (0.189, 0.507)
R	0.68	0.74	0.75
R^2	0.46	0.54	0.56
SEE (mL/kg/min)	6.84	6.29	6.17

Note: Gender is categorized as 1 = boy and 0 = girl.
[a]95% CI of the regression coefficients are shown within parentheses.
min, minute; PACER, Progressive Aerobic Cardiovascular Endurance Run; *SEE*, SE of estimate

mance contributed more to the prediction than other variables in the model.

Table 3 shows the accuracy of the equations developed on the validation group when applied to the cross-validation group and to the total sample. Although some shrinkage occurred, the accuracy of the models was confirmed. The mean differences between measured and estimated VO_2max were less than 0.2 mL/kg/min for the Quadratic Model and Linear Model 2 for both the cross-validation and total samples. Mean differences were slightly larger for Linear Model 1. The correlations between measured and estimated VO_2max were 0.69 and 0.73 for the Quadratic Model for the cross-validation and total samples, respectively, as compared to 0.75 on the validation sample. Corresponding correlations were slightly lower for Linear Model 1 and Linear Model 2.

Table 3 also shows the results for the cross-validation of other PACER prediction models. The newly developed PACER Quadratic Model was slightly more accurate (i.e., smaller mean differences, higher correlations, and lower SEs of estimate) than all other models examined.

The performance of the prediction equations in a CR framework on the total sample was also examined. Table 4 shows the proportion of agreement, modified kappa, and phi coefficient for how participants would be categorized (HFZ, NIZ-some risk, or NIZ-higher risk) using measured VO_2max versus VO_2max estimated from the prediction models. The newly developed Quadratic Model had similar or better classification agreement compared to all other prediction models. Using estimates of VO_2max from the Quadratic Model resulted in correct classification of 70% of participants into one of the three categories. Fifteen percent of participants were incorrectly classified into the HFZ, and 9% of participants who had adequate levels of measured VO_2max were incorrectly classified into one of the NIZs. Another 6% of participants had inadequate levels of measured VO_2max, but were incorrectly classified into one of the NIZs. Linear Model 1, Linear Model 2, and the Leger et al.[5] model misclassified fewer than 18% of participants into the HFZ, which is comparable to the accuracy achieved with the Quadratic Model. Although the other models had relatively similar levels of proportion of agreement and modified kappa, the Barnett et al.[6] B model incorrectly classified nearly 30% of participants into the HFZ.

Discussion

The PACER 20-m multistage shuttle run is a widely used field test of aerobic fitness[12] and is the recommended (default) test for the FITNESSGRAM. The FITNESSGRAM converts PACER performance to estimated VO_2max to evaluate fitness levels of participants. In this paper, linear and quadratic models to estimate VO_2max were developed on a large sample of boys and girls aged 10–16 years. The accuracy of these models was then compared to previously developed prediction models for estimation of VO_2max and for classification into FITNESSGRAM fitness categories (i.e., HFZ, NIZ–some risk, and NIZ–higher risk).

Regression results demonstrated that the newly developed Quadratic Model and Linear Model 2 were slightly more accurate than previous models for estimation of VO_2max. The overall correlations for these models between measured and estimated VO_2max (R=0.75 and 0.74) were similar to the corresponding correlation re-

Table 3. Cross-validation analysis of regression models on cross-validation sample and total sample, mL/kg/min

Model	M±SD	$R_{YY'}$	*SEE*	*TE*
CROSS-VALIDATION SAMPLE (*n*=70)				
Measured VO_2max	43.9±9.5			
Models from current study				
PACER Quadratic Model	44.1±6.9	0.69	6.91	6.87
PACER Linear Model 1	44.9±6.8	0.64	7.34	7.39
PACER Linear Model 2	44.1±7.1	0.64	7.32	7.34
Leger[5]	43.7±5.9	0.60	7.63	7.58
Mahar[8]	44.3±5.4	0.66	7.18	7.19
Barnett A[6]	44.8±6.4	0.66	7.16	7.17
Barnett B[6]	46.9±5.3*	0.69	6.86	7.53
Matsuzaka A[7]	44.5±6.7	0.61	7.53	7.55
Matsuzaka B[7]	43.0±6.4	0.59	7.70	7.74
TOTAL SAMPLE (*N*=244)				
Measured VO_2max	44.5±9.3			
Models from current study				
PACER Quadratic Model	44.5±6.9	0.73	6.39	6.37
PACER Linear Model 1	44.8±6.4	0.66	6.99	6.99
PACER Linear Model 2	44.5±6.9	0.71	6.61	6.59
Leger[5]	43.7±5.6	0.58	7.63	7.65
Mahar[8]	44.7±5.1	0.69	6.81	6.92
Barnett A[6]	45.3±6.1	0.66	7.06	7.10
Barnett B[6]	46.8±5.1*	0.64	7.20	7.61
Matsuzaka A[7]	45.1±6.4	0.66	7.02	7.03
Matsuzaka B[7]	43.5±6.0*	0.65	7.14	7.19

Note: PACER Quadratic Model [from validation sample]: $VO_2max = 41.76799 + (0.49261 \times \text{PACER laps}) - (0.00290 \times \text{PACER squared}) - (0.61613 \times \text{BMI}) + (0.34787 \times \text{gender} \times \text{age})$; PACER Linear Model 1 [from validation sample]: $VO_2max = 32.56941 + (0.27297 \times \text{PACER laps}) + (3.25225 \times \text{gender}) + (0.02961 \times \text{age})$; PACER Linear Model 2 [from validation sample]: $VO_2max = 40.34533 + (0.21426 \times \text{PACER laps}) - (0.79472 \times \text{BMI}) + (4.27293 \times \text{gender}) + (0.79444 \times \text{age})$; $R_{YY'}$ is the correlation between measured and estimated VO_2max.
*<0.05, significantly different from measured VO_2max
min, minute; PACER, Progressive Aerobic Cardiovascular Endurance Run; *SEE*, SE of estimate; *TE*, total error; VO_2max, maximal oxygen consumption

ported by Cureton et al.[13] for the 1-mile run/walk test (R=0.71). Other PACER prediction models have been developed on either small[6] or geographically distinct[6,7] samples. The Leger et al.[5] model that has been used in previous editions of the FITNESSGRAM program was noticeably less accurate than the new Quadratic Model and Linear Model 2. The new Quadratic Model and Linear Model 2 improve slightly on the accuracy of the Mahar et al.[8] equation and expand the age group of youth for whom the model is appropriate.

Comparison of SEs of estimate from different studies is difficult because the SD of the criterion variable (VO_2max, in this case) may differ from study to study, and the SD is used in calculation of the *SEE*. However, within the same study, comparison of the SEs of estimate provides a good indication of comparative accuracy. That is, the models with the lowest *SEE* and *TE* can be considered most accurate. Table 3 demonstrates the comparative accuracy of several prediction models. For the total sample, the new PACER Quadratic Model and Linear Model 2 produced fairly similar levels of accuracy (i.e., *TEs* of 6.37 and 6.59 mL/kg/min). The Barnett et al.[6] B model and the Leger et al.[5] model had relatively higher *TEs* compared to the other models.

The Matsuzaka et al.[7] models, developed on children aged 8–17 years, and the Barnett et al.[6] B model, developed on children aged 12–17 years, have a negative regression coefficient associated with the age predictor. This does not agree with the expected increase in VO_2max for boys and the expected decrease in VO_2max for girls aged 12–19 years.[14] In the newly developed Quadratic Model, the interaction between gender and age contributed significantly to the prediction of VO_2max and accounts for age-related differences in aerobic fitness between boys and girls. In the equation to estimate VO_2max from 1-mile run/walk performance, which was developed on 753 participants, Cureton et al.[13] also reported a significant gender X age interaction term. In smaller samples or samples with a more limited age range, the interaction term, which appears to be important to prediction accuracy, may be masked.

Others have reported correlation coefficients between measured VO_2max and number of laps completed on the PACER,[15–17] although the sample sizes, and thus the confidence that can be placed in the findings, varied substantially. Boreham et al.[15] reported a correlation coefficient of 0.87 (0.64 for 23 boys and 0.90 for 18 girls) between measured VO_2max and the number of laps completed on the PACER for adolescents aged 14–16 years. Suminski et al.[16] reported correlation coefficients between laps completed and measured VO_2max of 0.63 for boys (n=58) and 0.58 for girls (n=67) aged 10–12 years. Liu et al.,[17] in participants aged 12–15 years, found a correlation coefficient of 0.69 (.65 for 22 boys and 0.51 for 26 girls) between measured VO_2max and laps completed.

Table 4. Classification agreement with measured VO_2max (mL/kg/min; N=244)

Model	Three categories[a]			Two categories[b]		
	Pa	Kq	Phi	Pa	Kq	Phi
New Quadratic Model	0.70	0.55	0.49	0.76	0.52	0.44
New Linear Model 1	0.67	0.50	0.48	0.76	0.52	0.45
New Linear Model 2	0.69	0.54	0.48	0.75	0.50	0.41
Leger[5]	0.64	0.47	0.42	0.73	0.45	0.35
Mahar[8]	0.70	0.55	0.44	0.74	0.48	0.38
Barnett A[6]	0.68	0.51	0.41	0.73	0.47	0.36
Barnett B[6]	0.68	0.51	0.24	0.70	0.41	0.27
Matsuzaka A[7]	0.70	0.56	0.45	0.74	0.48	0.37
Matsuzaka B[7]	0.67	0.50	0.39	0.71	0.42	0.31

[a]The categories are healthy fitness zone; needs improvement zone, some risk; and needs improvement zone, higher risk.
[b]The categories are healthy fitness zone and needs improvement zone.
Pa, proportion of agreement; *Kq*, modified Kappa

The corresponding correlation coefficient in the present study was 0.66 (0.64 for boys and 0.54 for girls). The correlation coefficient reported by Cureton et al.[13] between 1-mile run/walk time and measured VO_2max was −0.54. Thus, although the correlation coefficients between measured VO_2max and the primary outcome variable of a field test (e.g., 1-mile time or laps completed) are significant, they are fairly modest and need to be combined with other variables to provide an accurate estimate of VO_2max.

Perhaps the most important aspect of the accuracy of field tests of youth fitness, such as the PACER, is the accuracy of classification into fitness categories. The FITNESSGRAM report contains not only the estimated VO_2max, but also where a participant's fitness level suggests he or she should be categorized. For the FITNESSGRAM aerobic fitness component, participants are categorized in the HFZ if their estimated values of VO_2max fall above the CR standard or in the NIZ if their estimated values of VO_2max fall below the CR standard. The NIZ is further categorized as NIZ-some risk, and NIZ-higher risk. CR validity can be examined by classifying participants into fitness categories based on their estimated values of VO_2max and then comparing this with the classification based on an appropriate criterion measure, in this case, the classification based on measured VO_2max.

Criterion-referenced validity evidence for the newly developed Quadratic Model and Linear Models 1 and 2 was only slightly better than the other models examined, and overall was only moderate. Compared to all other prediction models, the Quadratic Model and Linear Models 1 and 2 misclassified a lower percentage of participants into the HFZ (<17%). The nature of the quadratic curve makes it more difficult for participants with a low number of laps on the PACER to be categorized into the HFZ. However, the Quadratic Model was not substantially more accurate than the newly developed Linear Models 1 and 2 in terms of classification accuracy. For most of the other models, more than 20% (range: 21% to 29%) of participants were inaccurately classified into the HFZ.

The nature of the quadratic term does, however, cause a problem in estimation of VO_2max at the higher performance levels. After a certain number of laps completed on the PACER, because the regression coefficient for the PACER squared term is negative, estimated VO_2max starts to decrease. This may cause an underestimation of measured VO_2max. Simulations with various levels of BMI and age, for both genders, indicates that estimated VO_2max starts to level off between 70 and 80 laps and starts to decrease slightly between 90 and 100 laps. In the sample used in this study, 14 participants (5.7%) completed more than 75 laps, and four participants (1.6%) completed more than 90 laps. Slightly underestimated VO_2max for these highly fit participants would not, however, affect classification accuracy because their predicted values for VO_2max are much higher than the CR standard for the HFZ.

Results of the CR analyses suggest that use of the newly developed Quadratic Model and Linear Models 1 and 2 would result in accurate classification of approximately 70% to 75% of participants. Of the misclassified participants, relatively similar numbers of participants would be incorrectly classified into the HFZ and NIZ. Because the other models examined produced relatively similar levels of classification agreement, practitioners can probably expect this level of classification accuracy from field tests of aerobic fitness.

In summary, regression equations to estimate VO_2max from PACER 20-m multistage shuttle run performance and demographic variables in a heterogeneous sample of children aged 10–16 years were developed and cross-validated. The newly developed Quadratic Model and Linear Model 2 provided more accurate estimates of VO_2max than other PACER prediction models and the accuracy of these models appears to be similar to that of the 1-mile run/walk. The CR validity evidence for the Quadratic and Linear Models developed in this study was slightly better than comparable evidence for the other models examined and resulted in fewer participants incorrectly classified into the HFZ.

This work was supported by a grant from the Research Consortium of the American Alliance for Health, Physical Education, Recreation and Dance.

Publication of this article was supported by The Cooper Institute through a philanthropic gift from Lyda Hill.

No other financial disclosures were reported by the authors of this paper.

References

1. Blair SN, Kohl HW 3rd, Paffenbarger RS Jr., Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. JAMA 1989;262(17):2395–401.
2. Brage S, Wedderkopp N, Ekelund U, et al. Features of the metabolic syndrome are associated with objectively measured physical activity and fitness in Danish children: the European Youth Heart Study (EYHS). Diabetes Care 2004;27:2141–48.
3. Ortega FB, Ruiz JR, Castillo MJ, Sjostrom M. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes 2008;32:1–11.
4. Meredith MD, Welk GJ, eds. FITNESSGRAM®/ACTIVITYGRAM: test administration manual. Updated 4th ed. Champaign IL: Human Kinetics, 2010.
5. Leger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. J Sports Sci 1988;6(2):93–101.
6. Barnett A, Chan LYS, Bruce IC. A preliminary study of the 20-m multistage shuttle run as a predictor of a peak VO_2 in Hong Kong Chinese students. Ped Ex Sci 1993;5:42–50.
7. Matsuzaka A, Takahashi Y, Yamazoe M, et al. Validity of the multistage 20-m shuttle-run test for Japanese children, adolescents, and adults. Ped Ex Sci 2004;16:113–25.
8. Mahar MT, Welk GJ, Rowe DA, Crotts DJ, McIver KL. Development and validation of a regression model to estimate VO_2peak from PACER 20-m shuttle run performance. J Phys Act & Hlth 2006; 3(S2):S34–46.
9. Slaughter MH, Lohman TG, Boileau RA, et al. Skinfold equations for estimation of body fatness in children and youth. Hum Biol 1988; 60(5):709–23.
10. Rowland TW. Does peak VO_2 reflect VO_2max in children?: evidence from supramaximal testing. Med Sci Sports Exerc 1993;25(6):689–93.
11. Armstrong N, Welsman JR. Assessment and interpretation of aerobic fitness in children and adolescents. Exerc Sport Sci Rev 1994;22: 435–76.
12. Tomkinson GR, Leger LA, Olds TS, Cazorla G. Secular trends in the performance of children and adolescents (1980–2000): an analysis of 55 studies of the 20m shuttle run test in 11 countries. Sports Med 2003;33(4):285–300.
13. Cureton KJ, Sloniger MA, O'Bannon JP, Black DM, McCormack WP. A generalized equation for prediction of VO_2peak from 1-mile run/walk performance. Med Sci Sports Exerc 1995;27(3):445–51.
14. Pate RR, Wang CY, Dowda M, Farrell SW, O'Neill JR. Cardiorespiratory fitness levels among U.S. youth 12 to 19 years of age: findings from the 1999–2002 National Health and Nutrition Examination Survey. Arch Pediatr Adolesc Med 2006;160(10):1005–12.
15. Boreham CA, Paliczka VJ, Nichols AK. A comparison of the PWC170 and 20-MST tests of aerobic fitness in adolescent schoolchildren. J Sports Med Phys Fitness 1990;30(1):19–23.
16. Suminski RR, Ryan ND, Poston CS, Jackson AS. Measuring aerobic fitness of Hispanic youth 10 to 12 years of age. Int J Sports Med 2004;25(1):61–7.
17. Liu NY, Plowman SA, Looney MA. The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. Res Q Exerc Sport 1992;63(4):360–5.