ARCHIVES of Exercise in Health and Disease

DOI: 10.5628/aehd.v3i1-2.20

Original Research Manuscript

Calculation and validation of models for estimating VO_{2max} from the 20-m shuttle run test in children and adolescents

Gustavo Silva, Nórton Luis Oliveira, Luísa Aires, Jorge Mota, José Oliveira, José Carlos Ribeiro

CIAFEL, Faculty of Sports, University of Porto, Portugal

Objective: The purposes of this study were to calculate and validate two models to estimate maximal oxygen uptake (VO_2max) in Portuguese youths, aged 10-18 years, using a 20-meter shuttle run test (SR). ***Design:*** Subjects (54 girls and 60 boys) were divided into estimation (*n*= 91) and cross-validation (*n*=23) groups, and their VO_{2max} was directly measured by wearing a portable gas analyzer during the SR. The Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) tests were carried out considering sex, age, height, weight, body mass index (BMI) and SR stage as predictors of VO_{2max}. Estimations from MLR and ANN were compared with three other previously published equations. ***Results:*** In summary, the equation estimated by MLR is more appropriate for Portuguese youths than the equation estimate by ANN or the other three previously-published equations (validation coefficient for the MLR model: r=0.84, P<0.001; systematic error=-0.01±5.2, P>0.05; SEE=4.9). ***Conclusion***: For Portuguese youths, the following equation would be recommended: VO_{2max}= 43.313 + 4.567*sex - 0.560*BMI + 2.785*stage. However, findings from this study also warn researchers that the use of equations to estimate VO_{2max} may not be sensitive enough to detect small changes in individuals' cardiorespiratory fitness in longitudinal observations and intervention studies, according to the dispersions of random error for all equations.

Arch Exerc Health Dis 3 (1-2):145-152, 2012

Key Words: physical fitness; oxygen consumption; youths; multiple linear regression; artificial neural network

INTRODUCTION

There is a good deal of evidence suggesting that cardiorespiratory fitness (CRF) is an important marker of cardiovascular health. In adults, low levels of CRF are considered predictors of mortality as a result of cardiovascular diseases (10, 28, 30). In young people, poor CRF is associated with obesity and features of metabolic syndrome (2, 13, 18, 32, 33).

In terms of measurement, the maximum oxygen uptake (VO_{2max}) is the criterion measure of CRF (38). However, protocols to directly measure VO_{2max} usually require a research laboratory with sophisticated equipment, trained staff and increased costs. Alternatively, several field tests and equations have been developed to estimate VO_{2max}, which are crucial for large sample studies and health surveillance policies.

The 20-meter Shuttle Run Test (SR) created by Léger and colleagues (20) is a widely used field test. The SR, or some modified version, is included in a broad number of physical fitness test batteries. One of the most frequently-used versions is the PACER (Progressive Aerobic Cardiovascular Endurance Run), the standard CRF test for the FITNESSGRAM battery (39).

The equation suggested by Léger et al. (20) to estimate VO_{2max} remains the most frequently used, and it is included by default in the FITNESSGRAM software to estimate VO_{2max} (14). On the other hand, several researchers have developed alternative equations to estimate VO_{2max} for different population groups, including children and adolescents (8, 20, 24, 25, 34), attempting to improve the validity and the accuracy of VO_{2max} estimations.

A common issue that arises in the use of several equations is the external validity of estimations. Usually, when estimation models are employed in a different population than that in which the equation originated, the validity and accuracy of estimations

Corresponding author:
Gustavo Silva: Faculty of Sport, University of Porto, Rua Dr. Plácido Costa, 91, 4200-450 Porto, Portugal • Tel: 00351225074785 • Fax: 00351225500689 • Email: gugonsilva@gmail.com

decrease (35). These limitations could be related to the variety of methodological options available to estimate and validate equations, including statistical procedures, error analysis and presence/absence of a cross-validation group. In a previous study, the validity and accuracy of five different equations were tested in Portuguese youths (35), including equations estimated by Multiple Linear Regression (8, 20, 25) and Artificial Neural Networks (34). Findings from that study suggested that more research is needed to calculate a valid and precise indicator of CRF, one that is determined from the SR in Portuguese children and adolescents.

Therefore, this study aimed to estimate and validate two different models (MLR and ANN) to predict VO_{2max} in Portuguese youths aged 10-18 years.

MATERIAL AND METHODS

Participants

For this study, a total of 122 healthy young individuals (57 girls and 65 boys) from three schools in the District of Porto, Portugal, volunteered to participate in the study. The study was carried out following the Declaration of Helsinki guidelines for human research. The study's purpose, nature, benefits and risks were explained to participants, parents/guardians and teachers. Informed written consent was obtained from the participants' parents/guardians. The experimental protocol was approved by the Review Committee of the Institutional Scientific Board, as well as by the Foundation of Science and Technology from Portugal.

Anthropometric Measures

Height and weight were measured before testing, with participants wearing shorts and t-shirts only. Height was measured using a Holtain stadiometer (Holtain Ltd., Crymmych, UK) and recorded in centimetres to the nearest millimetre. Weight was measured to the nearest 0.1 kg with a Seca weight scale. Body mass index (BMI) was calculated by the ratio between weight and squared height ($kg.m^{-2}$).

Cardiorespiratory Fitness and Physiological Measurements

Participants performed the SR according to the described by FITNESSGRAM [PACER;(39)]. Briefly, the test consists in running back and forth between two lines 20 meters apart, with running speed determined by audio signals from a pre-recorded music CD. The running speed increases at the end of each one-minute stage. The running speed is 8.0 $km.h^{-1}$ for the first stage, 9.0 $km.h^{-1}$ for the second stage, and thereafter increases by 0.5 $km.h^{-1}$ each minute. The test ends when the subjects twice fail to reach the lines at the time indicated by the audio signals, demonstrating an inability to keep the required pace. All participants were familiar with the test, since the FITNESSGRAM test battery is included in the Portuguese Physical Education curriculum. As a result, most of the students perform the 20m Shuttle Run Test at least three or four times a year. The total number of completed laps was recorded and then transformed into stages. The last completed stage was considering the SR variable and entering it into the equations estimated by this study. All subjects underwent the SR wearing a portable gas analyser (K4*b*2, Cosmed, Rome, Italy) and a heart rate monitor (Polar Electro Oy, Kempele, Finland) to directly measure oxygen consumption (VO_2) and heart rate (HR), respectively. The weight of the Cosmed K4*b*2 was 1.5 kg, including the battery and a specially designed harness. McLaughlin and colleagues (26) reported that it is a valid device when compared with the Douglas bag method. Wearing the portable gas analyser during the 20-m shuttle run test does not significantly alter participants' energy demands (16). Respiratory variables were recorded breath-by-breath, which in turn were averaged over a 10-second period, yielding a "fair" representation of the change in VO_2 during incremental exercise (6). Before each individual test, oxygen and carbon dioxide analysers were calibrated according to the manufacturer's instructions. Directly measured VO_{2max} or VO_{2peak} were the main variables determined using the open-circuit method. Directly measured VO_{2max} was considered when a plateau in the VO_2 curve was detected, defined as an increase in a VO_2 of less than 2 $ml.kg^{-1}.min^{-1}$ with a concomitant increase in speed stage. If a VO_2 plateau was absent (15), the VO_{2peak} was taken and defined as the highest oxygen uptake achieved during the SR at exhaustion (3). For practical reasons, from now on, this paper will refer to the highest VO_2 values achieved in the SR as VO_{2max}. Exhaustion was confirmed when: (1) subjects desired to stop or demonstrated an inability to maintain the required running pace despite strong verbal encouragement; (2) maximal heart rate was greater than 85% of age-predicted maximal heart rate (220-age); (3) the respiratory exchange ratio (RER) was greater than 1.0 at the end of the test; (4) the participants showed symptoms of discomfort and/or signs of high sweating, facial flushing and grimacing (15). Careful control was taken concerning technical and environmental variables that might have had some influence on the results, so that highly reliable metabolic measures could be obtained. The protocol for the SR was carried out in groups of 11 students at time and, for each group of 11 subjects, one student

Table 1. Descriptive values for physical and physiological characteristics and SR performance of study participants.

	Total Sample					Validation Groups	
	All (n=114)			Girls (n=54)	Boys (n=60)	Estimation (n=91)	Cross-Validation (n=23)
Variables	Minimum	Maximum	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
Age (years)	10.0	18.0	14.5±2.0	14.4±1.9	14.5±2.0	14.5±1.9	14.3±2.3
Height (cm)	141.0	187.0	164.0±9.8	160.6±6.5	*167.1±11.2	164.2±9.6	163.2±10.7
Weight (kg)	34.8	96.0	57.6±10.6	55.8±7.6	*59.2±12.6	58.0±10.1	56.1±12.6
BMI ($kg \cdot m^{2}$)	16.0	29.9	21.3±2.7	21.6±2.6	21.0±2.8	21.4±2.7	20.8±2.7
SR laps (number of laps completed)	6.0	135.0	47.9±22.3	34.6±12.5	*60.0±22.4	46.9±20.3	52.0±29.0
SR stage (number of stages completed)	0.0	13.0	5.1±2.2	3.9±1.4	*6.3±2.1	5.1±2.0	5.5±2.7
SR maximal speed ($km \cdot h^{-1}$)	8.0	15.0	11.0±1.1	10.3±0.7	*11.6±1.1	11.0±1.0	11.2±1.4
Measured VO_2max ($ml \cdot kg^{-1} \cdot min^{-1}$)	29.2	70.0	48.1±9.5	42.3±7.5	*53.3±8.1	47.8±9.7	49.4±8.9
Measured VO_2max ($ml \cdot min^{-1}$)	1490.5	4649.4	2765.7±737.3	2344.2±420.2	*3145.0±757.3	2766.5±726.0	2762.6±797.2
Maximal Ventilation ($l \cdot min^{-1}$)	63.5	160.7	102.3±19.1	92.1±11.9	*111.5±19.8	102.6±18.9	101.0±20.3
End-exercise RER ($ml \cdot min^{-1}$)	1.0	2.6	1.3±0.3	1.3±0.3	1.3±0.4	1.3±0.3	1.3±0.3
Predicted Maximal HR ($beats \cdot min^{-1}$)	202.0	210.0	205.5±2.0	205.6±1.9	205.5±2.0	205.5±1.9	205.7±2.3
Measured Maximal HR ($beats \cdot min^{-1}$)	162.0	224.0	199.5 ± 9.9	196.7 ± 10.0	**202.0 ± 9.3	198.9 ± 10.3	202.5 ± 7.9

Note: BMI=body mass index; SR= 20-meter shuttle run; VO_{2max}=maximal oxygen uptake; RER=respiratory exchange ratio; HR= heart rate; * $P<0.01$ for comparisons between sex; ** $P<0.05$ for comparisons between sex.

was randomly selected to run carrying the portable gas analyser, while the others ran wearing a heart rate monitor (Polar Team System). Finally, a total of 122 subjects had run wearing the Cosmed K4*b*2.

Statistical analysis

Two groups of individuals were separated to estimate and cross-validate the equations. Approximately 80% of participants were randomly assigned in the estimation group. The remaining 20% were defined as the cross-validation group. Descriptive statistics (expressed as mean ± *SD*) were determined to provide anthropometric and physiological characteristics of the participants. Independent t-tests were performed for comparisons between sex and validation subgroups. Two new equations were calculated using the Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) estimation models. Sex, age, weight, height, body mass index (BMI) and the last stage completed (stage) were set up as predictors or inputs (independent) variables. The outcome or the output (dependent) variable was the VO_{2max}, measured directly in the SR by means of the portable gas analyser (measured VO_{2max}). All variables were expressed in their original units, i.e., sex (0=girls and 1=boys), age (years), weight (kg), height (cm), BMI ($kg.m^{-2}$), speed ($km.h^{-1}$) and measured VO_{2max} ($ml.kg^{-1}.min^{-1}$). Before running estimation procedures, Pearson correlation coefficients (*r*) were examined to observe the relationship between measured VO_{2max} and predictors. The MLR model was constructed using the stepwise method. The VO_{2max} estimated by the MLR model was named VO_{2max}MLR. The ANN model was built using the multilayered perceptron method. Predictors and outcome variables were normalized to a 0-1 interval as an ANN step in order to facilitate the model's learning. The resulting multilayered perceptron model consists of the following ANN architecture: 6 inputs (predictors), 4 hidden units (such as a latent dimension for another multivariate analysis) and 1 output (dependent). The best mathematical resolution for the ANN models tested was the logistic activation functions between input variables and hidden units and between hidden units and output. More recently, ANN models have been explored in the context of medicine, health, physical activity and sports sciences (9, 19, 21, 31, 36, 37). For more information about the related techniques used in this study, see Ruiz et al. (34). The VO_{2max} estimated by the ANN model was defined as VO_2maxANN. The normal distribution of the residuals was tested for both the MLR and ANN models. To verify whether the estimation models calculated by this study were better than previously published equations, three other equations were selected for comparison: Léger's

Table 2. Newly developed equations to estimate VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$) from the SR.

Method	*Equation and Inputs*
MLR	VO_2maxMLR= 43.313 + 4.567*sex - 0.560*BMI + 2.785*stage
ANN	VO_2maxANN= (1/(1+EXP(-((1/(1+EXP(-(+((stage)/11)*-5.309 + (sex)*-1.968+((age-10)/8)*4.394 + ((height-141)/46)*1.881 + ((weight-37)/59)*3.078 + ((BMI-16.23)/13.68)*4.429 - 4.302))))*-1.782 + (1/(1+EXP(-(+((stage)/11)*1.790 + (sex)*2.253 + ((age-10)/8)*1.770 + ((height-141)/46)*-1.060 + ((weight-37)/59)*4.978 + ((BMI-16.23)/13.68)*-3.610-2.705))))*9.988 + (1/(1+EXP(-(+((stage)/11)*5.528 + (sex)*-6.357 + ((age-10)/8)*-1.068 + ((height-141)/46)*0.663 + ((weight-37)/59)*1.333 + ((BMI-16.23)/13.68)*0.825-1.608))))*6.384 + (1/(1+EXP(-(+((stage)/11)*8.144 + (sex)*-0.724 + ((age-10)/8)*-0.329 + ((height-141)/46)*6.170 + ((weight-37)/59)*-0.573 + ((BMI-16.23)/13.68)*0.373-4.679))))*-4.278 -3.886))))*39.83 + 29.17

Inputs: sex (0=girls; 1=boys); age (years); height (cm); weight (kg), BMI ($kg \cdot m^{-2}$) and stage (number of stages completed); MLR=multiple linear regression (stepwise method); ANN=artificial neural network (multilayered perceptron method); BMI= body mass index; equations are expressed with this shape for easy use with an Excel spreadsheet or SPSS syntaxes.

equation (20) [VO_{2max}Léger], Barnett's equation (8) [VO_{2max}Barnett], and Ruiz's equation (34) [VO_{2max}Ruiz]. The Léger's equation was chosen because it was the original equation created with the SR. Findings from a previous study (35) suggest that Barnett and Ruiz equations had better agreement results with measured VO_2max. The equation selected from Barnett et al. (8) used sex, age and speed as predictors. All five equations were tested for validity and error analyses as recommended (7, 17). Simple linear regression was used to calculate the validity correlation (correlation between the criterion measure and the estimation – measured and estimated VO_{2max}) and the standard error of the estimate (SEE). A paired-sample T-test was used to examine the mean differences (systematic error) between measured and estimated VO_{2max} (7). The agreement between measured and estimated $VO2_{max}$ was observed with the Bland-Altman method (11, 12) for VO_{2max}MLR and VO_{2max}ANN. All analyses were completed with SPSS 19.0 (SPSS Inc., Chicago, United States), with a significance level of 0.05.

RESULTS

After the determination of exhaustion parameters, those subjects who did not meet the exhaustion criteria (end-exercise RER superior to 1.0) were excluded from analyses. Thus, all analyses were carried out with a final sample of 114 subjects (54 girls and 60 boys), separated in the estimation (n=91) and cross-validation (n=23) groups. A VO_{2max} plateau was detected in 87.7% of participants. Table 1 shows descriptive statistics for participants' physical characteristics and SR performance. Significant statistical differences were found between sex for height, weight, SR laps, SR stage, SR speed, measured VO_{2max}, maximal ventilation and measured maximal HR. No differences were found between estimation and cross-validation groups (P>0.05).

Sex, age, height, weight, BMI and stage were considered to be predictors for measured VO_{2max}. Correlation analyses indicated significant Pearson-r coefficients between measured VO_{2max} and age (r=0.26; P<0.001), height (r=0.24; P<0.001), BMI (r=-0.34; P<0.001) and stage (r=0.77; P<0.001). When groups were split by sex, the variables age (r=0.26; P<0.001), weight (r=-0.35; P<0.05) and stage (r=0.68; P<0.001) were correlated with measured VO_{2max} in girls and age (r=0.36; P<0.01), BMI (r=-0.38; P<0.01) and stage (r=0.65; P<0.001) were correlated with measured VO_{2max} in boys. Table 2 provides equations estimated for the multiple linear regression and the artificial neural network methods. The residuals for both equations were normally distributed (non-significant Kolmogorov-Smirnov test; P>0.05).

To test the validity and the accuracy of the estimated VO_{2max} (Table 3), the SEE, the validity correlation, the systematic error and the 95% limits of agreement (LOA) were observed for the two newly estimated equations (MLR and ANN) and for the other previous published equations (Léger, Barnett and Ruiz). SEE ranged between 4.9 (estimated MLR equation for the cross-validation group) and 7.1 $ml.kg^{-1}.min^{-1}$ (Léger's equation for the total sample). Validation coefficients (correlation between estimated and measured VO_{2max}) were significant for all equations (0.86>r>0.67; P<0.001). No mean difference (systematic error; P>0.05) was detected in the comparison between measured and estimated VO_{2max} for the newly developed equations (MLR and ANN). However, a significant underestimation of measured VO_{2max} (P<0.001) was found for Léger and Ruiz equations in the total sample and in the cross-validation group.

Table 3. Descriptive values for estimated VO_2max and validation parameters.

Equation	*Estimated VO_2max (Mean ± SD)*	*Mean Difference (95% LOA)*	*r*	*SEE*
Total Sample (n=114)				
MLR	48.1 ± 8.1	0.0 (-11.2; 11.2)	0.80**	5.7
ANN	47.8 ± 8.5	-0.3 (-10.0; 9.4)	0.86**	5.0
Léger	45.2 ± 5.9	-2.9 (-16.8; 11.0)*	0.67**	7.1
Barnett	47.5 ± 5.4	-0.6 (-14.0; 12.9)	0.71**	6.8
Ruiz	44.1 ± 9.6	-4.0 (-16.8; 8.8)*	0.77**	6.2
Cross-validation (n=23)				
MLR	49.3 ± 9.4	-0.1 (-10.2; 10.0)	0.84**	4.9
ANN	47.9 ± 8.8	-1.5 (-12.8; 9.8)	0.79**	5.6
Léger	46.6 ± 7.5	-2.7 (-10.6; 7.3)*	0.82**	5.2
Barnett	48.5 ± 6.5	-0.9 (-11.4; 9.6)	0.80**	5.4
Ruiz	45.2 ± 10.7	-4.2 (-17.0; 8.6)*	0.80**	5.5

Note: MLR=multiple linear regression model; ANN=artificial neural network model; VO_2max=maximal oxygen uptake; Mean Difference = VO_{2max} estimated by the equation - VO_{2max} measured directly (expressed in $ml \cdot kg^{-1} \cdot min^{-1}$); LOA=limits of agreement (mean difference ± 1.96*SD*); *r* = Pearson correlations between estimated and measured VO_2max (validation coefficient); SEE = standard error of the estimate expressed in $ml \cdot kg^{-1} \cdot min^{-1}$; **P*<0.001 for comparison between estimated and measured VO_{2max}; ***P*<0.001 for correlations between estimated and measured VO_{2max}.

By analysing the 95%LOA, it would be expected that errors of estimation by these five equations would lie between –17.0 (for Ruiz's equation) and +12.9 $ml.kg^{-1}.min^{-1}$ (for Barnett's equation). The Bland-Altman plots were explored in the cross-validation group for the two newly estimated equations (MLR and ANN), as represented in Figure 1. For the observed plots, the systematic error and the dispersion of random error (95% LOA) were reduced for VO_{2max}MLR in comparison to VO_{2max}ANN.

DISCUSSION

In the present study, two new models to predict the maximal oxygen uptake (VO_{2max} – the standard measure for cardiorespiratory fitness) were estimated and validated for Portuguese children and adolescents, based on their performance in the 20-meter shuttle run test. Also, the validity of three other previously-published equations [Léger's, Barnett's and Ruiz's equations (8, 20, 34)] was verified for Portuguese youths. In summary, the new equation estimated from the multiple linear regression could be considered the most satisfactory to estimate VO_2max for Portuguese children and adolescents. The strengths of the study lie in the strategy to estimate and validate the two new equations, based on several assumptions: (a) the objective and direct measure of oxygen uptake while children and adolescents performed the SR; (b) the employment of two different statistical methods for estimating equations; (c) the analysis of data on three group levels (total sample, estimation group and cross-validation group); (d) the analysis of equation accuracies in a cross-validation group; and (e) the use of robust statistics for validation analysis (7, 11, 12, 17).

The findings of the current study are of interest and are timely, since recent studies (1, 23, 33) have reported that reference-based standards for cardiorespiratory fitness expressed in units of VO_{2max} ($ml.kg^{-1}.min^{-1}$) are valid for targeting young people at risk for metabolic syndrome and cardiovascular diseases. Also, the findings are opportune because there are a variety of equations for the prediction of VO_{2max} that are based in SR performance, and there is no agreement in the literature concerning the appropriate equation for different populations. Moreover, the SR is widely used in Portugal. The FITNESSGRAM battery is the standard set of physical fitness tests implemented by the Portuguese Physical Education Curriculum.

The SR is a logically valid test, since its pacing simulates the criterion test to which it is most often compared: a maximal graded exercise test (14). In the present study, the validation model used by Ruiz and colleagues (34, 35), where subjects ran during the SR wearing a portable gas analyser, was used as an alternative to the usual laboratory treadmill reference test. In this case, the SR was assumed to be a maximally graded test, since physiological variables were measured directly and exhaustion criteria were satisfactorily achieved. Metabolic variables (VO_2 and VCO_2) were measured breath-by-breath and averaged across 10-second periods, which allowed a trustworthy depiction of the metabolism response to incremental effort. This sampling interval is more precise than 30 or 60 seconds periods to determine the VO_2 plateau and/or the VO_{2max}. On the other hand, raw breath-by-breath acquisition is subject to variability due to fluctuations in breathing frequency and tidal volume (6). This approach yields a more sensitive method than the backward VO_{2max} extrapolation technique used by Léger et al. (20). The backward extrapolation is

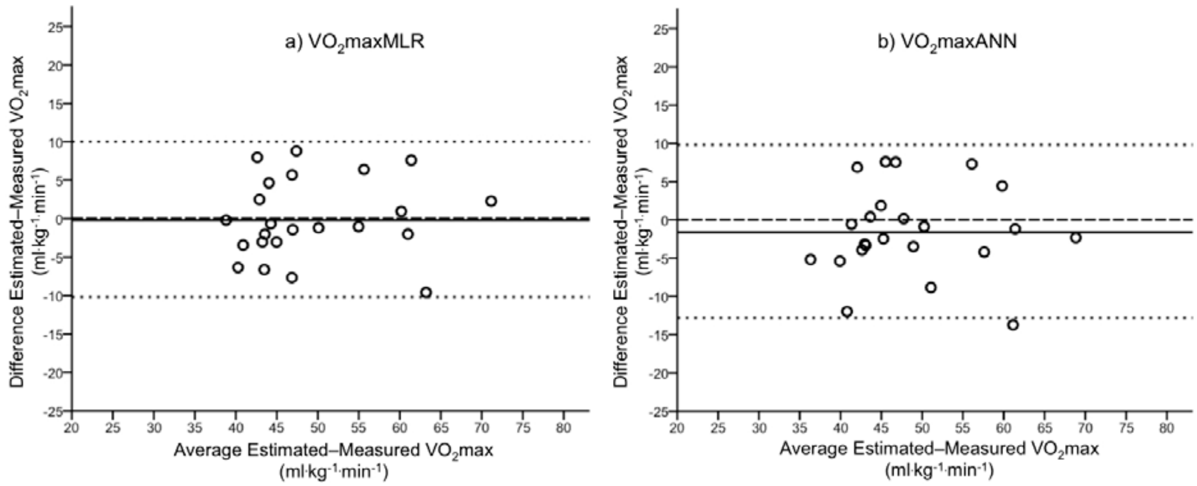

Figure 1. Bland-Altman plots for directly measured VO_{2max} and estimated VO_{2max}. Figure 1a and 1b respectively represents results for MLR and ANN equations in Bland-Altman plots. Solid lines represent mean differences (systematic error) between estimated and measured VO_{2max}. Upper and lower broken lines represent 95% limits of agreement (mean difference±1.96 standard deviations of differences). Centered broken lines represent the reference line for 0 (zero) mean differences.

considered an outdated method (16, 34, 35), since it does not allow the detection of the VO_2 plateau, and provides only an estimation of actual VO_2. The use of new techniques, such as a field measure using a portable gas analyser, has enhanced the understanding of metabolic kinetics during maximal effort measurement, or any other habitual daily task.

A possible discrepancy factor for the comparison of equations is based on the fact that other studies determine VO_{2max} values as directly measured from treadmill-based protocols, using this as the reference (standard) measure (8, 16, 24, 25). The exercise-mode differences in protocol to determine the reference measure of VO_{2max} seem to indicate some discrepancy in energy demands. Metsios et al. (27) and Flouris et al. (16) suggested that VO_{2max} for shuttle running appears to be higher than for treadmill running, but is still strongly correlated (r=0.96). This can be attributed to differences in exercise mode and technique, since shuttle runs incorporate pivot and turning movements characterizing another muscular and, consequently, metabolic involvement (16, 27). Previous studies (8, 22, 34, 35) showed that Léger's equation (20) has an acceptable but not exceptionally strong concurrent validity, although Léger's equation (20) is still used by the FITNESSGRAM software (14). Therefore, some effort should be made to improve VO_2max prediction. Those efforts should consider methodological/statistical alternatives, as well as the selection/inclusion of CRF predictors.

The selection of variables assumed to be VO_{2max} predictors were based on gender differences detected in height, weight, SR laps, SR stages, SR speed, measured VO_{2max}, maximal ventilation and maximal heart rate. Also, by exploring the relationship between the predictors and the outcome, age, height, weight, BMI and stage were found to be correlated with measured VO_{2max}, which is in accordance with previous research (8, 24, 25, 34). Furthermore, it is expected that the inclusion of sex, age and anthropometrical measures would improve the percentage of the VO_2max variance explained by prediction models (8, 24, 25, 34). Despite this, age, height and weight, for example, are indicators of growth and maturation, expressing changes in body size and composition, which can influence the expression of VO_{2max} (4, 5, 29).

Regarding the analysis of errors and validation coefficients, no significant systematic error was detected for the models estimated from the Portuguese population (MLR and ANN). However, a significant underestimation of VO_2max was found for the Léger and Ruiz equations. The results for validation correlations are in accordance with previous studies, varying between 0.79 and 0.86 for MLR and ANN. For Barnett et al. (8), correlations ranged between 0.82 and 0.85 for Hong Kong Chinese students. Matsuzaka et al. (25) reported correlations of 0.76 and 0.75 for Japanese children and adolescents. Mahar et al. (24) showed correlations between 0.65 and 0.67 for North American youths. Ruiz et al. (34) reported a correlation of 0.96 between measured and estimated $VO2_{max}$. However, it should be highlighted that higher validation coefficients were described for the estimation samples. Actually, few studies tested the estimated equations in a cross-validation group (24, 34). When the accuracy of a certain equation is tested in a cross-validation group, generally, validation coefficients drop, even when estimation and cross-validation groups are similar. This was observed when the developed ANN model was employed in the cross-validation group and when the previous published equations were tested in the total sample of the present study, corroborating with findings from a previous study (35). The only exception for this behaviour was observed in the present findings for the developed MLR equation, which obtained improved validation values in the cross-validation group.

In the search for an estimation model that better predicts VO_{2max} for Portuguese youths, MLR and ANN were calculated. While MLR models are common and easier to understand, ANN models are more complex and characterized by a "black box" nature (34). By this complexity, it would be expected that the explained variance and the SEE for ANN models would be more satisfactory. Indeed, this is only for the estimation group. When ANN models are utilized in a different group of individuals, like the cross-validation sample, validation parameters are more favourable to the use of the MLR model, as suggested by these results. Also, the stepwise method used in the MLR model reduces the presence of some co-linearity between predictors, which is not an option for ANN models. Based on the mathematical law of parsimony, since the more complex model achieves no substantial improvement, it is more appropriate to choose the simplest one. When comparing the estimations between MLR and ANN models, no statistical differences were detected.

The findings from the present study should be considered together with some limitations. The equations estimated by this study were calculated and validated for the Portuguese population. Further research should be done to explore the accuracy of these equations in other populations. Moreover, it should be considered that other variables could be included and analysed as potential predictors. The maturation status, for example, could eventually contribute to the understanding of the variation of $VO2_{max}$, as well as the predictors, over the growth and development process. Eventually, physical activity levels could also be included in the analyses. Unfortunately, the lack of this information does not allow exploration of whether the estimation models would behave in a different way if maturation status and physical activity levels were accounted for. The results also point to a substantial dispersion of random error (95%LOA), existing in all equations. Researchers should consider the possibility of biased results when the analysis requires more sensitive differential data, as required in longitudinal and intervention studies, for example, where SR assessed CRF. If that was the case, maybe a raw result from the SR, such as the number of completed laps or stages, could be more sensitive for detecting small but significant changes in CRF. However, for prospective studies, the clinical value of reference standards for of VO_2max standards for the discrimination of children SR laps is not known. On the other hand, the accuracy and adolescents at high and low cardio-metabolic risk had already been described (1, 23, 33). Also, FITNESSGRAM Reference Standards for the number of laps were arbitrarily determined by inverting the equation from Léger for VO_{2max} reference standards (14). From the perspective of future research, it would be useful to verify whether the raw value of the SR performance could determine a cut-off that could indicate subjects with a poor cardio-metabolic profile.

In conclusion, the model estimated from the multiple linear regression in the current study would be recommended for estimating the maximal oxygen uptake in Portuguese children and adolescents, as suggested by analyses of systematic error, validation coefficients and the standard error of the estimates, especially in the cross-validation group. However, users should be aware that the equations to estimate VO_{2max} could not be sensitive enough to detect small changes in individuals' CRF in longitudinal observations and intervention studies, according to the dispersions of random error.

ACKNOWLEDGEMENTS

The Foundation of Science and Technology (FCT) supported this study with a PhD grant (SFRH/BD/45090/2008) and a research project (PTDC/DES/72424/2006).

REFERENCES

1. Adegboye AR, Anderssen SA, Froberg K, Sardinha LB, Heitmann BL, Steene-Johannessen J, Kolle E, and Andersen LB. Recommended aerobic fitness level for metabolic health in children and adolescents: a study of diagnostic accuracy. *Br J Sports Med* 2010;45:722-728.
2. Andersen LB, Hasselstrøm H, Grønfeldt V, Hansen SE, and Karsten F. The relationship between physical fitness and clustered risk, and tracking of clustered risk from adolescence to young adulthood: eight years follow-up in the Danish Youth and Sport Study. *Int J Behav Nutr Phys Act* 2004;1: 6.
3. Armstrong N and Welsman J. Aerobic fitness: What are we measuring? *Med Sports Sci* 2007;50: 5-25.
4. Armstrong N and Welsman JR. Development of aerobic fitness during childhood and adolescence. *Pediatr Exerc Sci* 2000;12: 128-149.
5. Armstrong N and Welsman JR. Peak oxygen uptake in relation to growth and maturation in 11- to 17-year-old humans. *Eur J Appl Physiol* 2001;85: 546-551.
6. Astorino TA. Alterations in VO2max and the VO2 plateau with manipulation of sampling interval. *Clin Physiol Funct Imaging* 2009;29:60-67.
7. Atkinson G and Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998;26:217-238.
8. Barnett A, Chan LYS, and Bruce IC. A preliminary study of the 20-m multistage shuttle run as a predictor of peak VO2 in Hong Kong Chinese students. *Pediatr Exerc Sci* 1993;5: 42-50.
9. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995;346:1135-1138.
10. Blair SN, Kohl HW, 3rd, Paffenbarger RS, Jr., Clark DG, Cooper KH, and Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* 1989;262:2395-2401.
11. Bland JM and Altman DG. Comparing methods of measurement: Why plotting difference against standard method is misleading. *Lancet* 1995;346:1085-1087.

12. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-310.
13. Brage S, Wedderkopp N, Ekelund U, Franks PW, Wareham NJ, Andersen LB, and Froberg K. 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;27:2141-2148.
14. Cureton KJ and Plowman SA. *Aerobic capacity assessments. In: FITNESSGRAM Reference Guide* (3rd ed.), edited by Welk GJ and Meredith MD. Dallas, TX: The Cooper Institute, 2008, p. 96-120.
15. Docherty D. *Measurement in pediatric exercise science.* Champaign, IL: Human Kinetics, 1996.
16. Flouris AD, Metsios GS, and Koutedakis Y. Enhancing the efficacy of the 20 m multistage shuttle run test. *Br J Sports Med* 2005;39: 66-170.
17. Hopkins WG, Marshall SW, Batterham AM, and Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc* 2009;41:3-13.
18. Hürtig-Wennlöf A, Ruiz JR, Harro M, and Sjöström M. Cardiorespiratory fitness relates more strongly than physical activity to cardiovascular disease risk factors in healthy children and adolescents: The European Youth Heart Study. *Eur J Cardiovasc Prev Rehabil* 2007;14:575-581.
19. Itchhaporia D, Snow PB, Almassy RJ, and Oetgen WJ. Artificial neural networks: Current status in cardiovascular medicine. *J Am Coll Cardiol* 1996;28: 515-521.
20. Léger LA, Mercier D, Gadoury C, and Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 1988;6:93-101.
21. Linder R, Mohamed EI, De Lorenzo A, and Pöppl SJ. The capabilities of artificial neural networks in body composition research. *Acta Diabetologica* 2003;40: S9-S14.
22. Liu NY, Plowman SA, and 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;63:360-365.
23. Lobelo F, Pate RR, Dowda M, Liese AD, and Ruiz JR. Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. *Med Sci Sports Exerc* 2009;41: 1222-1229.
24. Mahar MT, Welk GJ, Rowe DA, Crotts DJ, and McIver KL. Development and Validation of a Regression Model to Estimate VO2peak From PACER 20-m Shuttle Run Performance. *J Phys Act Health* 2006; S34-S46.
25. Matsuzaka A, Takahashi Y, Yamazoe M, Kumakura N, Ikeda A, Wilk B, and Bar-Or O. Validity of the multistage 20-m shuttle-run test for Japanese children, adolescents, and adults. *Pediatr Exerc Sci* 2004;16:113-125.
26. McLaughlin JE, King GA, Howley ET, Bassett Jr DR, and Ainsworth BE. Validation of the COSMED K4 b2 portable metabolic system. *Int J Sports Med* 2001;22:280-284.
27. Metsios GS, Flouris AD, Koutedakis Y, and Nevill A. Criterion-related validity and test-retest reliability of the 20m square shuttle test. *J Sci Med Sport* 2008;11:214-217.
28. Mora S, Redberg RF, Cui Y, Whiteman MK, Flaws JA, Sharrett AR, and Blumenthal RS. Ability of exercise testing to predict cardiovascular and all-cause death in asymptomatic women: a 20-year follow-up of the lipid research clinics prevalence study. *JAMA* 2003;290:1600-1607.
29. Mota J, Guerra S, Leandro C, Pinto A, Ribeiro JC, and Duarte JA. Association of maturation, sex, and body fat in cardiorespiratory fitness. *Am J Hum Biol* 2002;14: 707-712.
30. Myers J, Prakash M, Froelicher V, Do D, Partington S, and Atwood JE. Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* 2002;346:793-801.
31. Ravé JMG and Prieto JAA. Potential challenges in physical education through the knowledge of Artificial Neural Networks. *Journal of Human Movement Studies* 2003;45:81-96.
32. Ribeiro JC, Guerra S, Oliveira J, Teixeira-Pinto A, Twisk JWR, Duarte JA, and Mota J. Physical activity and biological risk factors clustering in pediatric population. *Prev Med* 2004;39:596-60.
33. Ruiz JR, Ortega FB, Rizzo NS, Villa I, Hürtig-Wennlöf A, Oja L, and Sjöström M. High cardiovascular fitness is associated with low metabolic risk score in children: The European Youth Heart Study. *Pediatr Res* 2007;61:350-355.
34. Ruiz JR, Ramirez-Lechuga J, Ortega F, Castro-Pinero J, Benitez JM, Arauzo-Azofra A, Sanchez C, Sjöström M, Castillo M, Gutierrez A, and Zabala M. Artificial neural network-based equation for estimating VO2max from the 20 m shuttle run test in adolescents. *Artif Intell Med* 2008;44:233-245.
35. Ruiz JR, Silva G, Oliveira N, Ribeiro J, Oliveira J, and Mota J. Criterion-related validity of the 20-m shuttle run test in youths aged 13-19 years. *J Sports Sci* 2009;27:899-906.
36. Ryguła I. Artificial neural networks as a tool of modeling of training loads. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, Shanghai, 2005, p. 2985-2988.
37. Staudenmayer J, Pober D, Crouter S, Bassett D, and Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol* 2009;107:1300-1307.
38. Taylor HL, Buskirk E, and Henschel A. Maximal oxygen intake as an objective measure of cardiorespiratory performance. *J Appl Physiol* 1955;8:73-80.
39. The Cooper Institute for Aerobics Research. *Fitnessgram & Activitygram Test Administration Manual*. Champaign, Illinois, USA: Human Kinetics, 2010.