ORIGINAL ARTICLE

Enhancing the efficacy of the 20 m multistage shuttle run test

A D Flouris, G S Metsios, Y Koutedakis

....

Objective: Maximal oxygen uptake (Vo_{2max}) of 44 ml kg^{-1} min^{-1} is an accepted criterion (Vo_{2CR}) below which health and fitness for young male adults may be compromised. New algorithms validated for Vo_{2CR} screening using the 20 m multistage shuttle run test (20mMST) were developed.

Methods: Vo_{2max} was assessed in 110 males using a stationary gas analyser in a treadmill test (TT) and in 40 of these subjects using a portable gas analyser in the 20mMST. Vo_{2max} predicted from the 20mMST in 70 subjects was used for cross validation. Two equations predicting Vo_{2max} during 20mMST (EQ_{MST}) and TT (EQ_{TT}) were developed.

Results: Significant energy cost variance (EC_V) was detected between TT and 20mMST ($p < 0.001$), correlated significantly with subject height, and was a significant predictor of Vo_{2max} differences between TT and 20mMST. The r^2 of EQ_{MST} was 0.92 ($p < 0.001$). Predicted Vo_{2max} values from EQ_{MST} correlated with directly measured 20mMST Vo_{2max} at $r = 0.96$ ($p < 0.001$). ANOVA detected no mean difference ($p > 0.05$) between predicted and measured values. Prevalence of low fitness based on Vo_{2CR} was 0.37. McNemar χ^2 indicated significant differences in sensitivity ($p < 0.001$) and specificity ($p < 0.05$) between the original 20mMST equation ($EQ_{LÉG}$) and EQ_{TT}, regarding Vo_{2CR} screening. Cohen's κ demonstrated higher agreement with TT Vo_{2max} for EQ_{TT} ($p < 0.001$) than $EQ_{LÉG}$ ($p < 0.05$). TT Vo_{2max} correlated with the end result of both $EQ_{LÉG}$ and EQ_{TT} at $r = 0.75$ ($p < 0.001$). Unlike EQ_{TT} ($p > 0.05$), mean predicted Vo_{2max} from $EQ_{LÉG}$ was significantly higher compared to TT Vo_{2max} ($p < 0.001$).

Conclusion: These algorithms increase the efficacy of 20mMST to accurately evaluate aspects of health and fitness.

See end of article for authors' affiliations
....

Correspondence to:
Yiannis Koutedakis, University of Thessaly, Department of Sport and Exercise Science, Karies, Trikala GR42100, Greece; y.koutedakis@uth.gr

Despite the vast amounts of research focusing on various cardiorespiratory fitness (CF) assessments and the acceptance of specific CF cut offs in national health guidelines,[1 2] statistical screening methodology such as calculating receiver operating characteristics (ROC) curves has not been employed hitherto. The ROC curve analysis is extensively used in epidemiology to provide a graphic means for assessing the accuracy of a diagnostic instrument.[3] The difficulty in adopting ROC curves in sports medicine is mainly attributed to the fact that most outcome measures are in continuous format. However, these biomarkers can be dichotomised using dummy variables according to clinically accepted critical values Q and defined positive or negative if the test outcome measure is greater or lesser than Q. For instance, a maximal oxygen uptake (Vo_{2max}) of 44 ml kg^{-1} min^{-1} for young male adults (18–29 years of age) has been generally accepted as a criterion (Vo_{2CR}) below which both health and fitness may be compromised.[1 4 5]

The 20 m multistage shuttle run test (20mMST)[6] represents an acceptable field assessment tool for CF, and has been repeatedly employed in different health[7 8] and fitness[9] settings. However, the popularity of the 20mMST is mainly attributed to its practical use for simultaneous measurement of large groups of individuals. Studies evaluating its accuracy in predicting laboratory Vo_{2max} have reported contradictory results.[9–11] More importantly, the efficacy (that is, the extent to which a specific procedure produces a valid classification of data in relation to established criteria) of the original 20mMST model in screening for CF remains unknown.

From a statistical standpoint, the limited accuracy of the 20mMST may be attributed to the repeated measures design used in the original study.[6] It is well known that the inherent dependency of within-subject observations can reduce the power of prediction models.[12] Concurrently, it seems tenable that the theoretical basis of the original 20mMST model may be further compromised by the use of generally large and heterogeneous samples in the validation procedures.[6] It has been established that severely biased linear relationships can occur owing to sample heterogeneity.[13]

From a physiological viewpoint, it could be argued that the curtailed ability of the original 20mMST model to predict treadmill Vo_{2max} values might be attributed to differences in the exercise modes utilised in the validation procedures (that is, shuttle running *v* forward running). Findings from recent investigations suggested that Vo_{2max} during the 20mMST is significantly higher compared to a treadmill test.[14 15] Ergo, a prediction model controlling for differences in energy cost (EC) between the reference standard laboratory assessment and the proxy 20mMST may result in more accurate prediction of Vo_{2max} and increased efficacy in screening for Vo_{2CR}. The objective of the present investigation was to develop a new Vo_{2max} prediction algorithm for the 20mMST using data collected via portable indirect calorimetry and statistical procedures which accounted for within-subject observation dependency. Thereafter, the efficacy of both the original and the novel models was assessed in predicting standard treadmill Vo_{2max} and screening for Vo_{2CR}.

METHODS

Subjects and procedures

A total of 110 healthy males (age: 21.6 (SD 2.5); BMI: 23.6 (2.2)) volunteered. Exclusion criteria included smoking and any muscular or skeletal injuries. Written informed consent was obtained from all participants after full explanation of the procedures involved. The cohort was arbitrarily divided

....

Abbreviations: ANOVA, analysis of variance; CF, cardiorespiratory fitness; CI, confidence interval; CV, coefficient of variation; EC, energy cost; GEE, generalised estimating equations; GLM, general linear model; LIM_{AG}, limits of agreement analysis; MAS, maximal attained speed; ROC, receiver operating characteristics; SD, standard deviation; TT, treadmill test; 20mMST, 20 m multistage shuttle run test
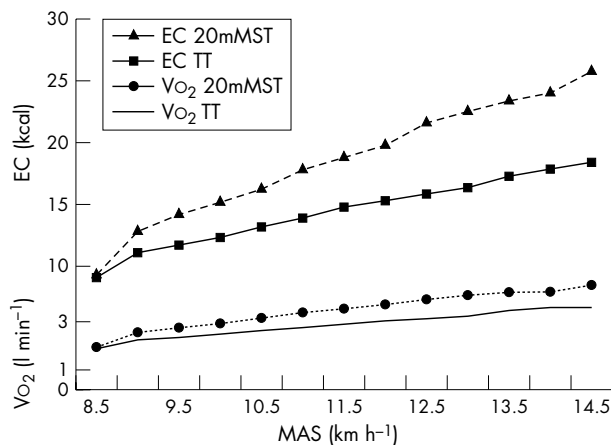
Figure 1 Energy cost and oxygen uptake during 20mMST and TT. Data obtained from indirect calorimetry.

into model (n = 40) and validation (n = 70) groups. Analysis of variance (ANOVA) revealed no significant difference between the two groups in terms of anthropometrical characteristics.

Within a 14 day period, all participants underwent a treadmill Vo_{2max} assessment and performed the 20mMST in an indoor rubber floored gymnasium. Unlike the validation group, participants in the model group were subjected to Vo_{2max} assessment whilst performing the 20mMST using a portable gas analyser. Special care was taken to maintain similar environmental conditions in both measurement sites during assessment. Prior to data collection visits, subjects were familiarised with all assessment protocols. They were also advised to avoid stressful activities 36–48 h prior to the data collection visits. Tests were conducted in a random order, by the same investigators, and at the same time for each subject either between 9:00 and 12:00 h or between 14:00 and 17:00 h. The study was approved by the Research Ethics Board of the University of Wolverhampton.

Data collection
Laboratory assessment of Vo_{2max} (TT)
A modified Bruce treadmill test (TT) to exhaustion was used.[16] The treadmill running speed was manipulated accordingly in order to bring the subject to exhaustion in 7–10 min. The treadmill inclination was increased by 2.5° every 3 min from an initial 3.5°. Oxygen uptake (Vo_2 (ml kg^{-1} min^{-1})) was measured via open circuit spirometry using an automated gas analyser (Vmax 29, SensorMedics, Yorba Linda, CA) previously calibrated with standard gases. Respiratory parameters were recorded every 20 s during testing, while subjects inspired room air through a low resistance two-way Rudolph valve. To ensure that subjects achieved Vo_{2max}, measurements were considered for further

analysis when at least two of the following criteria were met: (i) maximal heart rate greater than 185 bpm, (ii) respiratory exchange ratio greater than 1.1, and/or (iii) detection of plateau in Vo_2 curve. EC in kcal was calculated for each individual minute/stage as the product of mean Vo_2 (l min^{-1}) by the corresponding caloric equivalent.[17]

Field assessment of Vo_{2max} (20mMST)
This test was conducted according to established procedures.[6] In the model group a portable gas analyser (K4b^2, Cosmed, Rome, Italy) was used to record respiratory parameters every 20 s during testing, while subjects inspired room air through a facemask. Maximal oxygen uptake was the main parameter determined using the open circuit method. Prior to measurement, the gas analyser was calibrated with standard gases. Exhaustion was confirmed when at least two of the following criteria were met: (i) maximal heart rate greater than 185 bpm, (ii) respiratory exchange ratio greater than 1.1, and/or (iii) detection of plateau in Vo_2 curve. The EC in kcal was calculated for each individual minute/stage as the product of mean Vo_2 (l min^{-1}) by the corresponding caloric equivalent.[17] In the validation group, Vo_{2max} was predicted from the 20mMST performance according to established procedures.[6]

The K4b^2 gas analyser weighed 475 g and was not expected to significantly alter the subjects' energy demands. A pilot study using five subjects (age: 21.6 (SD 1.3); BMI: 24.3 (1.5)) was conducted in order to investigate additional energy demands and ensure that significant agreement existed between the two gas analysers employed. The subjects, who did not partake in the main part of the investigation, performed the previously described TT twice using both gas analysers. Results showed no significant difference (p>0.05) between the mean Vo_{2max} value recorded by the stationary (Vmax 29, SensorMedics) and the portable (K4b^2, Cosmed) gas analyser (48.7 (SD 3.1) v 49.1 (3.5) ml kg^{-1} min^{-1}, respectively), with an average absolute error of 0.51 (SD 0.18) ml kg^{-1} min^{-1}.

Statistical analyses
ANOVA was used to compare mean EC between TT and 20mMST. The effect of energy-cost variance between TT and 20mMST (EC_V) on the original 20mMST prediction model ($EQ_{LÉG}$[6]) was assessed via a simultaneous general linear model (GLM). This model aimed to predict Vo_{2max} differences/errors between TT and $EQ_{LÉG}$ using mean EC_V as an independent variable. In addition, Pearson's correlation coefficients were used to detect linearity between EC_V and various anthropometrical characteristics.

For the calculation of the novel prediction model, the generalised estimating equations (GEE)[18] approach was employed to account for subject specific dependency between the repeated observations. The GEE is a powerful approach in fitting generalised linear models to non-normally but *dependently* distributed response variables.[18] A GLM framework

Table 1 Univariate statistics (mean (SD)) and generalised estimated equations analyses for predicting Vo_{2max} during the 20mMST and the TT in the model group (n = 40)

	$_{Ind}$Variable	$_{Dep}$Variable	r^2	χ^2	SEE	$_p Vo_{2max}$	$_A Vo_{2max}$	r
EQ_{MAS}	MAS	$_{TT}Vo_{2max}$	0.79	236.4*	2.72	48.3 (5.9)†	46.9 (5.7)†	0.91*
EQ_{MST}	MAS	$_{K4}Vo_{2max}$	0.92	456.2*	1.70	49.2 (5.9)	49.2 (5.9)	0.96*
EQ_{TT}	EQ_{MST}	$_{TT}Vo_{2max}$	0.89	317.3*	1.94	47.0 (0.8)	46.9 (5.7)	0.94*

Values in parentheses are standard deviations (SD). Significant ANOVA between $_A Vo_{2max}$ and $_p Vo_{2max}$ values: †p<0.001. χ^2 significant at *p<0.001.
$_A Vo_{2max}$, actual values measured during testing; $EQ_{MAS, TT}$, calculated regression models to predict TT Vo_{2max}; EQ_{MST}, calculated regression model to predict 20mMST Vo_{2max}; $_{Dep}$Variable, dependent variable; $_{Ind}$Variable, independent variable; $_{K4}Vo_{2max}$, Vo_{2max} measured during the 20mMST using the K4b^2 portable analyser; MAS, maximal attained speed (km h^{-1}); $_p Vo_{2max}$, predicted values using the calculated models; r, correlation coefficient between actual and predicted values; r^2, coefficient of determination; SEE, standard error of the estimate; $_{TT}Vo_{2max}$, Vo_{2max} measured during the TT; χ^2, chi-square.

Table 2 Comparisons between the two tests in the validation group (n = 70)

	Vo_{2max} (SD)	LIM_{AG} (error)	$CV_\%$	MAS (SD)	Time (SD)
$EQ_{LéG}$	51.3 (5.0)*	4.0 (8.2)	8.6	12.7 (0.8)*	9:05 (1:27)*
EQ_{TT}	46.7 (4.8)	−0.6 (6.5)	7.0		
TT	47.3 (5.5)	–	–	16.3 (1.5)	13:34 (2:47)

Correlation coefficient with TT for Vo_{2max}, MAS, and time indices: $r = 0.75$ ($p < 0.001$).
*$p < 0.001$, ANOVA different from TT.
$CV_\%$, percent coefficient of variation for Vo_{2max}; $EQ_{LéG}$, original 20mMST prediction model; EQ_{TT}, calculated regression model to predict TT Vo_{2max}; LIM_{AG} (error), calculated limits of agreement (error of measurement) for Vo_{2max}; MAS, maximal attained speed (km h^{-1}); Time, exercise time to exhaustion (min); TT, treadmill test; Vo_{2max}, predicted values in the calculated models and assessed value in TT (mean (SD)).

with GEE estimation was introduced to generate an equation (EQ_{MST}) predicting Vo_{2max} measured during the 20mMST using the model group data (n = 40). For the latter model, the maximal attained speed (MAS) during the 20mMST was set as the independent variable. Thereafter, a second GLM with GEE estimation was performed generating the EQ_{TT} model which aimed to predict the reference standard TT Vo_{2max} (dependent variable) using the end result of EQ_{MST} as an independent variable. This procedure was employed to produce a 20mMST Vo_{2max} model that accounts for EC_V. In order to ensure that the procedures followed in the calculation of the EQ_{TT} model were indeed superior to the traditional approach, a GLM was calculated using TT Vo_{2max} (dependent variable) and MAS (independent variable). ANOVA and Pearson's correlation coefficients were used to detect possible bias between the mean actual and predicted Vo_{2max} values for the three models.

Data from the remaining 70 subjects (referred to as the validation group) were used to cross validate EQ_{TT} and the original $EQ_{LéG}$ model. Correlation coefficients, ANOVA, 95% limits of agreement analyses (LIM_{AG}) and percent coefficients of variation ($CV_\%$) were adopted to validate the two models according to established procedures.[19] Ninety five percent confidence intervals ($CI_{95\%}$) and ROC curve analysis were calculated using statistical software incorporated in SAS/Macro/IML. The latter software is designed specifically to fit ROC curves using dummy variables for data obtained from repeated measures designs. The area under the ROC curve was estimated using the Wilcoxon non-parametric method.[20] The demarcation point for Vo_{2CR} was set at 44 ml kg^{-1} min^{-1} according to available guidelines.[1 4 5]
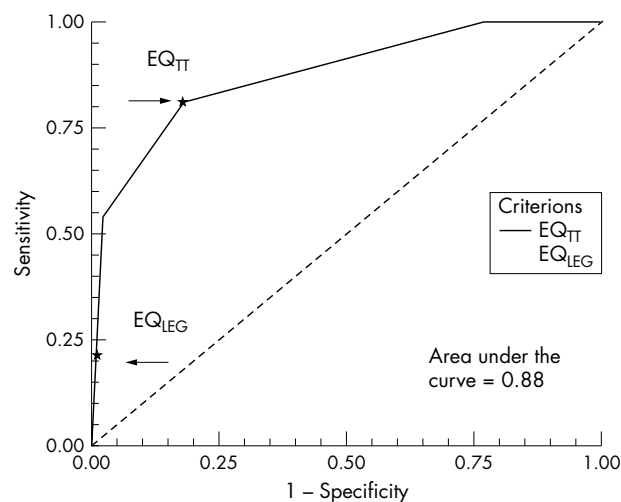

Figure 2 ROC curve for $EQ_{LéG}$ and EQ_{TT} regression models. The ROC curve is defined as the curve of the results from validation-group variance and $EQ_{LéG}$ or EQ_{TT} regression models, respectively. Asterisks indicate the designated cutoff point of 44 ml kg^{-1} min^{-1}.

Calculated sensitivity and specificity with corresponding $CI_{95\%}$ were used to determine the efficacy of the two equations in screening for Vo_{2CR}. Sensitivity (S_E) was defined as the proportion of subjects below the Vo_{2CR} who demonstrated a 20mMST predicted value below 44 ml kg^{-1} min^{-1}. Specificity (S_P) was defined as the proportion of subjects above the Vo_{2CR} who revealed a 20mMST predicted value above or equal to 44 ml kg^{-1} min^{-1}. McNemar χ^2 analysis examined the differences between calculated sensitivity and specificity at the cut off point for both equations. Cohen's κ statistic was used to evaluate the agreement between the prediction models and the reference standard test. Finally, ANOVA and Pearson's correlation coefficients were used to detect possible bias between the mean actual and predicted values. All statistical analyses were carried out with SPSS (version 11.5; SPSS, Chicago, IL) and SAS (version 8.2; SAS Institute, Cary, NC, USA) statistical software packages. The level of significance was set at $p < 0.05$.

RESULTS

Effect of energy-cost variance on $EQ_{LéG}$
ANOVA detected significant differences in EC and Vo_{2max} between TT and $EQ_{LéG}$ ($p < 0.001$; fig 1). Further, GLM results indicated that mean EC_V was a significant predictor of Vo_{2max} differences between TT and $EQ_{LéG}$ ($r^2 = 0.25$, $F_{1, 38} = 28.89$, $p < 0.001$). A significant linearity was also detected between EC_V and subject height ($r = 0.94$, $p < 0.001$).

Prediction of Vo_{2max} achieved via 20mMST and TT
Table 1 shows relevant statistics for the calculated models (that is, EQ_{MAS}, EQ_{MST}, and EQ_{TT}). Routine pre-analysis screening procedures were used to assess whether the data conformed to the assumptions of GLM. Although normally distributed, the variables used in these analyses were not independent of one another. Examination of residuals scatterplots detected no violation of normality, linearity, and homoscedasticity between predicted Vo_{2max} scores and errors of prediction. Mahalanobis distance of each case to the centroid of all cases detected no multivariate outliers for $\chi^2 < 0.001$. As expected the values in the variables utilised were multicollinear, being similar measures of the same parameter (that is, Vo_{2max}). As significant linearity was detected between EC_V and subject height (see previous section), initial calculations for EQ_{MST} and EQ_{TT} included height as a covariate. Nevertheless, the latter variable was not a significant predictor ($p > 0.05$) for either model.

$[EQ_{MAS}]$ $Vo_{2max} = MAS \times 6.87 - 39.54$
$[EQ_{MST}]$ $Vo_{2max} = MAS \times 6.65 - 35.8$
$[EQ_{TT}]$ $Vo_{2max} = EQ_{MST} \times 0.95 + 0.182$

Thus,

$[EQ_{TT}]$ $Vo_{2max} = (MAS \times 6.65 - 35.8) \times 0.95 + 0.182$

Table 3 Results for ROC curve and McNemar χ^2 analyses in the validation group (n = 70) for the designated cut off point (44 ml kg^{-1} min^{-1})

	S_E (CI$_{95\%}$)	S_P (CI$_{95\%}$)	$^+$PV (CI$_{95\%}$)	$^-$PV (CI$_{95\%}$)	LR (CI$_{95\%}$)
EQ$_{LÉG}$	0.23 (0.16)	1.00 (0.00)*	1.00 (0.16)	0.69 (0.10)	∞
EQ$_{TT}$	0.81 (0.15)**	0.82 (0.11)	0.72 (0.16)	0.88 (0.10)*	4.44 (0.19)

McNemar χ^2 increased at: *p<0.05; **p<0.001.
CI$_{95\%}$, 95% confidence interval; EQ$_{LÉG}$, original 20mMST prediction model; EQ$_{TT}$, calculated regression model to predict TT Vo$_{2max}$; LR, likelihood ratio; $^-$PV, negative predicted value; $^+$PV, positive predicted value; S_E, sensitivity; S_P, specificity; ∞, mathematical infinity.

Model cross validation

Means (SD) and comparisons of various performance indices from the TT and the 20mMST, as well as results for LIM$_{AG}$ and CV$_\%$ appear in table 2. Preliminary analyses for LIM$_{AG}$ revealed no positive relationship between the differences/ errors (either (EQ$_{LÉG}$–TT) or (EQ$_{TT}$–TT)) and the size of measurements (given by either (the mean of EQ$_{LÉG}$ and TT) or (mean of EQ$_{TT}$ and TT)), respectively. Thus, the LIM$_{AG}$ can be reported as absolute measurements.[21] Finally, unlike EQ$_{TT}$ and TT ($t = 1.46$, p>0.05), the mean difference (error) between estimates from EQ$_{LÉG}$ and TT ($t = -8.86$, p<0.001) was biased.

Relevant univariate statistics and ROC curve analyses for the designated cut off point (that is, 44 ml kg^{-1} min^{-1}) appear in table 3 and fig 2. Twenty six subjects (37.1%; CI$_{95\%}$: 0.9%) were diagnosed below the Vo$_{2CR}$ using the reference standard TT. In contrast, EQ$_{LÉG}$ and EQ$_{TT}$ identified six and 29 subjects below the Vo$_{2CR}$, respectively. Cohen's κ statistic demonstrated significant agreement with the TT measurement for both the EQ$_{LÉG}$ (p<0.05) and the EQ$_{TT}$ (p<0.001).

DISCUSSION

Sedentary lifestyle is a common phenomenon in modern societies, representing a major risk factor for numerous pathologies.[22] Consequently, screening for, and evaluation of, CF has become important for both health and fitness. The aim of the present investigation was to utilise the most salient physiological and epidemiological procedures in order to enhance the efficacy of the 20mMST for CF screening. Results suggested that the developed prediction models significantly increased the efficacy of the 20mMST to discern subjects according to Vo$_{2CR}$. To our knowledge, the present study represents the first direct clinical appraisal of the 20mMST as a screening tool for specific CF cut off points such as Vo$_{2CR}$.

To account for the increased energy requirements of shuttle running compared to forward treadmill running,[14 15] we developed a prediction equation which incorporates indirect calorimetry data collected while the subjects performed the 20mMST. Results from the newly developed model demonstrated increased accuracy in predicting Vo$_{2max}$ and a minimised standard error of the estimate (1.9 ml kg^{-1} min^{-1}) compared to the original EQ$_{LÉG}$ and EQ$_{MAS}$ (4.4 and 2.7 ml kg^{-1} min^{-1}, respectively).[6] Although the limits of agreement in EQ$_{TT}$ are still relatively wide, this range is more likely to be acceptable compared to EQ$_{LÉG}$ and EQ$_{MAS}$. Further, as illustrated by the present CV$_\%$ indices, the traditional Vo$_{2max}$ prediction can be up to 1.2 times as unreliable as the prediction of EQ$_{TT}$. ROC curve analysis indicated that both EQ$_{TT}$ and EQ$_{LÉG}$ were highly specific in discriminating individuals according to Vo$_{2CR}$. However, sensitivity in the former was significantly increased compared to the latter model (81% v 23%).

The theoretical basis of the EQ$_{TT}$ model is advantageous in that it seeks to parallel the energy utilisation of the human body during the 20mMST and the TT, rather than relying on statistical inference from a generally large and heterogeneous sample. The cohort consisted entirely of males to avoid the well known phenomenon of severely biased (that is, nonsense or spurious) linear relationships attributed to sample heterogeneity.[13] This phenomenon has been demonstrated explicitly by Anderson[23] who examined various factors associated with prediction power in the original 20mMST model. Anderson concluded that research utilising large heterogeneous samples in the validation process of predictive tests of aerobic capacity must be suspect. It seems reasonable to suggest that the prediction models developed using these procedures are rather generalised, representing merely vague indicators of the true values. These hypotheses are verified in the present study by the reduced accuracy of the EQ$_{MAS}$ prediction model, as compared to EQ$_{TT}$.

On another note, the present results are in line with previous studies suggesting increased energy demands during shuttle running compared to treadmill running.[14 15] This may well be attributed to differences in factors such as intensity, exercise mode, technique, and musculature employed between the two conditions. These factors should be considered in the design of physical training programmes that incorporate shuttle running elements. This information should also be taken into account when designing the physical training for sports incorporating shuttle running (for example, football, basketball, rugby). In addition, the present results suggest that EC$_V$ is exacerbated with increased body stature. It is tenable that various biomechanical complexities of shuttle running may account for this. The EQ$_{MST}$ model developed herein to predict Vo$_{2max}$ during the 20mMST can be used to calculate the oxygen transport demands of shuttle running, when such information is required.

It is important to acknowledge, however, that the 20mMST is a test requiring maximal effort. Therefore, it may not be suitable for populations with specific diseases. In addition, the novel EQ$_{TT}$ model represents a strict means of assessing CF. Three subjects with CF above the Vo$_{2CR}$ in our cross validation sample were mis-screened as performing below the Vo$_{2CR}$. Practicing such strict screening techniques may be beneficial in circumstances where adequate levels of CF are crucial (for example, military training). The applications from the present investigation would be further increased by

---

What is already known on this topic

The 20 m multistage shuttle run test (20mMST) is an acceptable field assessment tool for cardiorespiratory fitness but its original prediction model is subject to significant bias.

What this study adds

The prediction models introduced in the present study increase the efficacy of 20mMST thus providing increased accuracy in evaluating aspects of health and fitness.

calculating additional prediction models for both males and females of various age groups. In addition, it is worth mentioning that the present results are subject to some variability among different models of metabolic carts.[24] Within the limits of the present investigation, it is concluded that the developed models can be valuable tools that explicitly increase the efficacy of the 20mMST to discern subjects according to Vo_{2CR}.

. . . .

Authors' affiliations
A D Flouris, Faculty of Applied Health Sciences, Brock University, St Catherines, Ontario, Canada L2S 3AI
G S Metsios, School of Sports, University of Wolverhampton, Wolverhampton, UK
Y Koutedakis, Department of Sport and Exercise Science, University of Thessaly, Trikala, Greece

REFERENCES
1 **Åstrand P**, Rodahl K. *Textbook of work physiology. Physiological bases of exercise.* New York: McGraw-Hill, 1986.
2 **De Backer G**, Ambrosioni E, Borch-Johnsen K, *et al.* European guidelines on cardiovascular disease prevention in clinical practise. *Eur Heart J* 2003;**24**:1601–10.
3 **Knapp RG**, Miller MC. *Clinical epidemiology and biostatistics.* Philadelphia, PA: Williams and Wilkins, 1992.
4 **Wilmore J**, Costill D. *Physiology of sport and exercise.* Champaign, IL: Human Kinetics, 2004.
5 **Shvartz E**, Reibold R. Aerobic fitness norms for males and females aged 6–75: a review. *Aviat Space Environ Med* 1990;**61**:3–11.
6 **Léger L**, Gadoury C. Validity of the 20 m shuttle run test with 1 min stages to predict VO2max in adults. *Can J Sports Sci* 1989;**14**:21–6.
7 **Bouziotas C**, Koutedakis Y, Shiner R, *et al.* The prevalence of selected modifiable coronary heart disease risk factors in 12-year-old Greek children. *Pediatr Exerc Sci* 2001;**13**:173–84.
8 **Flouris AD**, Faught BE, Hay J, *et al.* Modeling risk factors for coronary heart disease in children with developmental coordination disorder. *Ann Epidemiol* 2003;**13**:591.
9 **Grant S**, Corbett K, Amjad A, *et al.* A comparison of methods of predicting maximum oxygen uptake. *Br J Sports Med* 1995;**29**:147–52.
10 **Stickland M**, Petersen S, Bouffard M. Prediction of maximal aerobic power from the 20-m multi-stage shuttle run test. *Can J Appl Physiol* 2003;**28**:272–82.
11 **Boreham C**, Paliczka V, Nichols A. A comparison of the PWC170 and 20-MST tests of aerobic fitness in adolescent schoolchildren. *J Sports Med Phys Fitness* 1990;**30**:19–23.
12 **Dunson D**, Watson M, Taylor J. Bayesian latent variable models for median regression on multiple outcomes. *Biometrics* 2003;**59**:296–304.
13 **Hassler U**, Thadewald T. Nonsensical and biased correlation due to pooling heterogeneous samples. *J R Stat Soc Ser D Stat* 2003;**52**:367–79.
14 **Flouris A**, Carrillo A, Metsios G, *et al.* Probing the energy equilibrium approach for enhanced proxy design in bioenergetics. *Can J Appl Physiol* 2003;**28**:S53.
15 **Flouris A**, Metsios G, Koutedakis Y. Attaining energy equilibrium in proxies assessing human bioenergetics. *Eur Heart J* 2003;**24**:183.
16 **Ramsbottom R**, Brewer J, Williams C. A progressive shuttle run test to estimate maximal oxygen uptake. *Br J Sports Med* 1988;**22**:141–4.
17 **Krogh A**, Lindhard J. The relative value of fat and carbohydrate as sources of muscular energy. *Biochem J* 1920;**14**:290.
18 **Liang K**, Zeger S. Longitudinal data analysis using general linear models. *Biometrica* 1986;**73**:13–22.
19 **Bland J**, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**8**:307–10.
20 **Bamber D**. The area above the ordinal dominance graph and the area below the receiver operating graph. *J Math Psychol* 1975;**12**:387–415.
21 **Nevill A**, Atkinson G. Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. *Br J Sports Med* 1997;**31**:314–8.
22 **Tokmakidis S**, Volaklis K. Training and detraining effects of a combined-strength and aerobic exercise program on blood lipids in patients with coronary artery disease. *J Cardiopulm Rehabil* 2003;**23**:193–200.
23 **Anderson GS**. A comparison of predictive tests of aerobic capacity. *Can J Sports Sci* 1992;**17**:304–8.
24 **Babineau C**, Léger L, Long A, *et al.* Variability of maximal oxygen consumption measurements in various metabolic systems. *J Strength Cond Res* 1999;**13**:318–24.