Enhancing the efficacy of the 20 m multistage shuttle run test

A D Flouris, G S Metsios, Y Koutedakis

Objective: Maximal oxygen uptake (Vo2max) of 44 ml kg⁻¹ min⁻¹ is an accepted criterion (Vo2CR) below which health and fitness for young male adults may be compromised. New algorithms validated for Vo2CR screening using the 20 m multistage shuttle run test (20mMST) were developed.

Methods: Vo2max 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. Vo2max predicted from the 20mMST in 70 subjects was used for cross validation. Two equations predicting Vo2max during 20mMST (EQMST) and TT (EQTT) were developed.

Results: Significant energy cost variance (ECV) was detected between TT and 20mMST (p<0.001), correlated significantly with subject height, and was a significant predictor of Vo2max differences between TT and 20mMST. The r² of EQMST was 0.92 (p<0.001). Predicted Vo2max values from EQMST correlated with directly measured 20mMST Vo2max 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 Vo2CR was 0.37.

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

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 (Vo2max) of 44 ml kg⁻¹ min⁻¹ for young male adults (18-29 years of age) has been generally accepted as a criterion (Vo2CR) below which both health and fitness may be compromised. 4, 5, 6

The 20 m multistage shuttle run test (20mMST) represents an acceptable field assessment tool for CF, and has been repeatedly employed in different health 7, 8 and fitness 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 Vo2max 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. 12 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. 13 It has been established that severely biased linear relationships can occur owing to sample heterogeneity. 14

From a physiological viewpoint, it could be argued that the curtailed ability of the original 20mMST model to predict treadmill Vo2max 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 Vo2max during the 20mMST is significantly higher compared to a treadmill test. 15, 16 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 Vo2max and increased efficacy in screening for Vo2CR. The objective of the present investigation was to develop a new Vo2max 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 Vo2max and screening for Vo2CR.

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...
Efficacy of the 20mMST

Data collection
Laboratory assessment of $\text{Vo}_{2\max}$ (TT)

A modified Bruce treadmill test (TT) to exhaustion was used.14 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 ($\text{VO}_2$ (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 $\text{Vo}_{2\max}$ 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 $\text{VO}_2$ curve. EC in kcal was calculated for each individual minute/stage as the product of mean $\text{VO}_2$ (1 min$^{-1}$) by the corresponding caloric equivalent.17

Field assessment of $\text{Vo}_{2\max}$ (20mMST)

This test was conducted according to established procedures.6 In the model group a portable gas analyser (K4b2®; 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 $\text{VO}_2$ curve. The EC in kcal was calculated for each individual minute/stage as the product of mean $\text{VO}_2$ (1 min$^{-1}$) by the corresponding caloric equivalent.17 In the validation group, $\text{Vo}_{2\max}$ was predicted from the 20mMST performance according to established procedures.6

The K4b2® 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 $\text{Vo}_{2\max}$ value recorded by the stationary (Vmax 29, SensorMedics) and the portable (K4b2®, Cosmed) gas analyser (48.7 (SD 3.1) $\nu$ 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 (ECV) on the original 20mMST prediction model (EQLEG*) was assessed via a simultaneous general linear model (GLM). This model aimed to predict $\text{Vo}_{2\max}$ differences/errors between TT and EQLEG* using mean ECV as an independent variable. In addition, Pearson’s correlation coefficients were used to detect linearity between ECV and various anthropometrical characteristics.

For the calculation of the novel prediction model, the generalised estimating equations (GEE)14 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 independently distributed response variables.19 A GLM framework

![Figure 1 Energy cost and oxygen uptake during 20mMST and TT. Data obtained from indirect calorimetry.](image-url)

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Dependent Variable</th>
<th>$r^2$</th>
<th>$\chi^2$</th>
<th>SEE</th>
<th>$\text{Vo}_{2\max}$</th>
<th>$\text{AVo}_{2\max}$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQMST</td>
<td>MAS</td>
<td>0.79</td>
<td>236.4*</td>
<td>2.72</td>
<td>48.3 (5.9)†</td>
<td>46.9 (5.7)†</td>
<td>0.91†</td>
</tr>
<tr>
<td>EQT1</td>
<td>EQMST</td>
<td>0.92</td>
<td>456.2*</td>
<td>1.70</td>
<td>49.2 (5.9)</td>
<td>49.2 (5.9)</td>
<td>0.96*</td>
</tr>
</tbody>
</table>

Values in parentheses are standard deviations (SD). Significant ANOVA between $\text{Vo}_{2\max}$ and $\text{AVo}_{2\max}$ values: $p<0.001$. $\gamma$ significant at $p<0.001$. $\text{AVo}_{2\max}$, actual values measured during testing; $\text{EQMST}_{TT}$, calculated regression model to predict TT $\text{Vo}_{2\max}$; $\text{EQT1}_{TT}$, calculated regression model to predict 20mMST $\text{Vo}_{2\max}$; Dependent Variable, $\text{Vo}_{2\max}$; Independent variable; $\text{AVo}_{2\max}$, 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; $\text{TTVo}_{2\max}$, $\text{Vo}_{2\max}$ measured during the TT; $\gamma$, $\chi^2$, chi-square.
with GEE estimation was introduced to generate an equation (EQMST) predicting Vo2max 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 EQTT model which aimed to predict the reference standard TT Vo2max (dependent variable) using the end result of EQMST as an independent variable. This procedure was employed to produce a 20mMST Vo2max model that accounts for ECV. In order to ensure that the procedures followed in the calculation of the EQTT model were indeed superior to the traditional approach, a GLM was calculated using TT Vo2max (dependent variable) and MAS (independent variable). ANOVA and Pearson’s correlation coefficients were used to detect possible bias between the mean actual and predicted Vo2max values for the three models.

Data from the remaining 70 subjects (referred to as the validation group) were used to cross validate EQTT and the original EQLEG model. Correlation coefficients, ANOVA, 95% limits of agreement analyses (LIMAG), and percent coefficients of variation (CVv) were adopted to validate the two models according to established procedures. Ninety five percent confidence intervals (CI95%) 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. The demarcation point for Vo2CR was set at 44 ml kg\(^{-2}\) min\(^{-1}\) according to available guidelines. Calculated sensitivity and specificity with corresponding CI95% were used to determine the efficacy of the two equations in screening for Vo2CR. Sensitivity (S\(_E\)) was defined as the proportion of subjects below the Vo2CR who demonstrated a 20mMST predicted value below 44 ml kg\(^{-2}\) min\(^{-1}\). Specificity (S\(_P\)) was defined as the proportion of subjects above the Vo2CR who revealed a 20mMST predicted value above or equal to 44 ml kg\(^{-2}\) min\(^{-1}\). McNemar \(\chi^2\) analysis examined the differences between calculated sensitivity and specificity at the cut off point for both equations. Cohen’s \(k\) 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 EQLEG**

ANOVA detected significant differences in EC and Vo2max between TT and EQLEG (\(p<0.001\); fig 1). Further, GLM results indicated that mean EC was a significant predictor of Vo2max differences between TT and EQLEG (\(r^2 = 0.25, \text{F}_{1, 38} = 28.89, p<0.001\)). A significant linearity was also detected between ECV and subject height (\(r = 0.94, p<0.001\)).

**Prediction of Vo2max achieved via 20mMST and TT**

Table 1 shows relevant statistics for the calculated models (that is, EQMAS, EQMST, and EQTT). 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 Vo2max 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, Vo2max). As significant linearity was detected between ECV and subject height (see previous section), initial calculations for EQMST and EQTT included height as a covariate. Nevertheless, the latter variable was not a significant predictor (\(p>0.05\)) for either model.

\[
\begin{align*}
\text{EQMAS:} & \quad \text{Vo2max} = \text{MAS} \times 6.87 - 39.54 \\
\text{EQMST:} & \quad \text{Vo2max} = \text{MAS} \times 6.65 - 35.8 \\
\text{EQTT:} & \quad \text{Vo2max} = \text{EQMST} \times 0.95 + 0.182
\end{align*}
\]

Thus,

\[
\text{EQTT:} \quad \text{Vo2max} = (\text{MAS} \times 6.65 - 35.8) \times 0.95 + 0.182
\]
Efficacy of the 20mMST

Table 3  Results for ROC curve and McNemar χ² analyses in the validation group (n=70) for the designated cut off point (44 ml kg⁻¹ min⁻¹)

<table>
<thead>
<tr>
<th></th>
<th>S₀ (CI95%)</th>
<th>S₁ (CI95%)</th>
<th>PV (CI95%)</th>
<th>PV (CI95%)</th>
<th>LR (CI95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQLEG</td>
<td>0.23 (0.16)</td>
<td>1.00 (0.00)*</td>
<td>1.00 (0.16)</td>
<td>0.69 (0.10)*</td>
<td></td>
</tr>
<tr>
<td>EQTT</td>
<td>0.81 (0.15)**</td>
<td>0.82 (0.11)</td>
<td>0.72 (0.16)</td>
<td>0.88 (0.10)*</td>
<td>4.44 (0.19)</td>
</tr>
</tbody>
</table>

McNemar χ² increased at: *p<0.05; **p<0.001.

The Theoretical basis of the EQTT 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. This phenomenon has been demonstrated explicitly by Anderson 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 EQmax prediction model, as compared to EQTT.

On another note, the present results are in line with previous studies suggesting increased energy demands during shuttle running compared to treadmill running. 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 ECV is exacerbated with increased body stature. It is tenable that various biomechanical complexities of shuttle running may account for this. The EQMST model developed herein to predict VO2max during the 20mMST can be used to calculate the oxygen transport demands of shuttle running when such information is required.

DISCUSSION

Sedentary lifestyle is a common phenomenon in modern societies, representing a major risk factor for numerous pathologies. 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 reliable and valid prediction model for VO2max during the 20mMST, as well as for VO2max and CVAV, which appear in Table 2. Preliminary analyses for LIMxG and CVxG were inconclusive and only two subjects (2.9%) were diagnosed below the VO2CR using the reference measurements (given by either (the mean of EQLEG and TT) or (mean of EQTT and TT)), respectively. Thus, the LIMxG can be reported as absolute measurements. Finally, unlike EQTT and TT (t = 1.46, p = 0.05), the mean difference (error) was within the limits of agreement in EQTT and TT (t = -8.86, p<0.001) instead of being 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 Figure 2. Twenty six subjects (37.1%; CI95%; 9.9%) were diagnosed below the VO2CR using the reference standard TT. In contrast, EQLEG and EQTT identified six and 29 subjects below the VO2CR, respectively. Cohen's k statistic demonstrated significant agreement with the TT measurement for both the EQLEG (p<0.05) and the EQTT (p<0.001).

Table 3 Results for ROC curve and McNemar χ² analyses in the validation group (n=70) for the designated cut off point (44 ml kg⁻¹ min⁻¹).

<table>
<thead>
<tr>
<th></th>
<th>S₀ (CI95%)</th>
<th>S₁ (CI95%)</th>
<th>PV (CI95%)</th>
<th>PV (CI95%)</th>
<th>LR (CI95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQLEG</td>
<td>0.23 (0.16)</td>
<td>1.00 (0.00)*</td>
<td>1.00 (0.16)</td>
<td>0.69 (0.10)*</td>
<td></td>
</tr>
<tr>
<td>EQTT</td>
<td>0.81 (0.15)**</td>
<td>0.82 (0.11)</td>
<td>0.72 (0.16)</td>
<td>0.88 (0.10)*</td>
<td>4.44 (0.19)</td>
</tr>
</tbody>
</table>

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.44 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 VO2CR.

Authors' affiliations
A D Flouris, Faculty of Applied Health Sciences, Brock University, St Catharines, Ontario, Canada L2S 3A1
G S Metsios, School of Sports, University of Wolverhampton, Wolverhampton, UK
Y Koutedakis, Department of Sport and Exercise Science, University of Thessaly, Trikala, Greece

Competing interests: none declared

REFERENCES