Enhancing the efficacy of the 20 m multistage shuttle run test

A D Flouris, G S Metsios, Y Koutedakis

Objective: Maximal oxygen uptake (V_{O2max}) of 44 ml kg^{-1} min^{-1} is an accepted criterion (V_{O2CR}) below which health and fitness for young male adults may be compromised. New algorithms validated for V_{O2CR} screening using the 20 m multistage shuttle run test (20MST) were developed.

Methods: V_{O2max} 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 20MST. V_{O2max} predicted from the 20MST in 70 subjects was used for cross validation. Two equations predicting V_{O2max} during 20MST (EQ_{MST}) and TT (EQ_{TT}) were developed.

Results: Significant energy cost variance (EC_v) was detected between TT and 20MST (p<0.001), correlated significantly with subject height, and was a significant predictor of V_{O2max} differences between TT and 20MST. The r^2 of EQ_{MST} was 0.92 (p<0.001). Predicted V_{O2max} values from EQ_{MST} correlated directly measured 20MST V_{O2max} 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 V_{O2CR} was 0.37. McNemar \chi^2 indicated significant differences in sensitivity (p<0.001) and specificity (p<0.05) between the original 20MST equation (EQ_{LEG}) and EQ_{TT}, regarding V_{O2CR} screening. Cohen’s k demonstrated higher agreement with TT V_{O2max} for EQ_{TT} (p<0.001) than EQ_{LEG} (p<0.05). TT V_{O2max} correlated with the end result of both EQ_{LEG} and EQ_{TT} at r=0.75 (p<0.001). Unlike EQ_{TT} (p<0.05), mean predicted V_{O2max} from EQ_{LEG} was significantly higher compared to TT V_{O2max} (p<0.001).

Conclusion: These algorithms increase the efficacy of 20MST 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,

From a physiological viewpoint, it could be argued that the curtailed ability of the original 20MST model to predict treadmill V_{O2max} values might be attributed to differences in the exercise modes utilised in the validation procedures (that is, shuttle running vs forward running). Findings from recent investigations suggested that V_{O2max} during the 20MST 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 20MST may result in more accurate prediction of V_{O2max} and increased efficacy in screening for V_{O2CR}. The objective of the present investigation was to develop a new V_{O2max} prediction algorithm for the 20MST 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 V_{O2max} and screening for V_{O2CR}.

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 musculoskeletal injuries. Written informed consent was obtained from all participants after full explanation of the procedures involved. The cohort was arbitrarily divided into two groups:-board.
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 VO2max assessment and performed the 20mMST in an indoor rubber floored gymnasium. Unlike the validation group, participants in the model group were subjected to VO2max 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 VO2max (TT)

A modified Bruce treadmill test (TT) to exhaustion was used. The treadmill inclination was increased by 2.5° accordingly in order to bring the subject to exhaustion in 2–10 min. The treadmill inclination was increased by 2.5° every 3 min from an initial 3.5°. Oxygen uptake (VO2) was measured 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 ppm, (ii) respiratory exchange ratio greater than 1.1, and/or (iii) detection of plateau in VO2 curve. EC in kcal was calculated for each individual minute/stage as the product of mean VO2 (1 min−1) by the corresponding caloric equivalent.17

Field assessment of VO2max (20mMST)

This test was conducted according to established procedures. 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 ppm, (ii) respiratory exchange ratio greater than 1.1, and/or (iii) detection of plateau in VO2 curve. The EC in kcal was calculated for each individual minute/stage as the product of mean VO2 (1 min−1) by the corresponding caloric equivalent.17 In the validation group, VO2max was predicted from the 20mMST performance according to established procedures.

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 investigations, performed the previously described TT twice using both gas analysers. Results showed no significant difference (p>0.05) between the mean VO2max value recorded by the stationary (Vmax 29, SensorMedics) and the portable (K4b2, 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 (ECV) on the original 20mMST prediction model (EQLEG) was assessed via a simultaneous general linear model (GLM). This model aimed to predict VO2max 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) 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 for the calculation of the novel prediction model, the generalised estimating equations (GEE) 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](https://example.com/figure1.png)

**Figure 1** Energy cost and oxygen uptake during 20mMST and TT. Data obtained from indirect calorimetry.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Univariate statistics (mean (SD)) and generalised estimated equations analyses for predicting VO2max during the 20mMST and the TT in the model group (n = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IndVariable</td>
<td>DepVariable</td>
</tr>
<tr>
<td>EQLEG</td>
<td>MAS</td>
</tr>
<tr>
<td>EQLEG</td>
<td>V20mMST</td>
</tr>
<tr>
<td>EQTT</td>
<td>EQMST</td>
</tr>
</tbody>
</table>

Values in parentheses are standard deviations (SD). Significant ANOVA between aVO2max and aΔVO2max values: tp<0.001. *X2 significant at p<0.001.
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 was introduced to generate the EQTT set as the independent variable. Calculated sensitivity and specificity with corresponding CIs% were used to determine the efficacy of the two equations in screening for VO2CR. Sensitivity (SE) was defined as the proportion of subjects below the VO2CR who demonstrated a 20mMST predicted value below 44 ml kg\(^{-1}\) min\(^{-1}\). Specificity (SP) was defined as the proportion of subjects above the VO2CR who revealed a 20mMST predicted value above or equal to 44 ml kg\(^{-1}\) min\(^{-1}\). McNemar \(\chi^2\) analysis examined the differences between calculated sensitivity and specificity at the cut off point for both equations. Cohen’s \(\kappa\) 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 VO2max values for the three models.

Table 2 shows relevant statistics for the calculated models and assessed value in TT (mean (SD)).

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

<table>
<thead>
<tr>
<th>Test</th>
<th>VO2max (SD)</th>
<th>LIMAG (error)</th>
<th>CV(v)</th>
<th>MAS (SD)</th>
<th>Time (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQLEG</td>
<td>51.3 (5.0)*</td>
<td>4.0 (8.2)</td>
<td>8.6</td>
<td>12.7 (0.8)*</td>
<td>9.05 (1.27)*</td>
</tr>
<tr>
<td>EQTT</td>
<td>46.7 (4.8)</td>
<td>-0.6 (6.5)</td>
<td>7.0</td>
<td>-</td>
<td>16.3 (1.5)</td>
</tr>
<tr>
<td>TT</td>
<td>47.3 (5.5)</td>
<td>-</td>
<td>-</td>
<td>13.34 (2.47)</td>
<td></td>
</tr>
</tbody>
</table>

Correlation coefficient with TT for VO2max, MAS, and time indices: \(r=0.75\) (\(p<0.001\)). ANOVA different from TT.

**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 ECv was a significant predictor of VO2max differences between TT and EQLEG \(r^2 = 0.25, F_{1, 38} = 28.89, p<0.001\). 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 variables was not a significant predictor (\(p>0.05\)) for either model.

\[
\begin{align*}
\text{[EQMAS]} VO2max &= \text{MAS} \times 6.87 - 39.54 \\
\text{[EQMST]} VO2max &= \text{MAS} \times 6.65 - 35.8 \\
\text{[EQTT]} VO2max &= \text{EQMST} \times 0.95 + 0.182
\end{align*}
\]

Thus,

\[
\begin{align*}
\text{[EQTT]} VO2max &= (\text{MAS} \times 6.65 - 35.8) \times 0.95 + 0.182
\end{align*}
\]
Model cross validation

Means (SD) and comparisons of various performance indices from the TT and the 20mMST, as well as results for LIMxG and CVx, appear in Table 2. Preliminary analyses for LIMxG revealed no positive relationship between the differences/errors (either (EQLE’G−TT) or (EQTT−TT)) and the size of measurements (given by either (the mean of EQLE’G 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) between estimates from EQLE’G 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. Six subjects (37.1%; CI\(_{95}%\): 0.9%) were diagnosed below the \(V_{O_2CR}\) using the reference TT. In contrast, EQLE’G and EQTT identified six and 29 subjects below the VO\(_{2CR}\), respectively. Cohen’s \(k\) statistic demonstrated significant agreement with the TT measurement for both the EQLE’G (p < 0.05) and the EQTT (p < 0.001).

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 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 \(V_{O_2CR}\). 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 \(V_{O_2CR}\).

To account for the increased energy requirements of shuttle running compared to forward treadmill running, 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 \(V_{O_2max}\) and a minimised standard error of the estimate (1.9 ml kg\(^{-1}\) min\(^{-1}\)) compared to the original EQLE’G and EQMAS (4.4 and 2.7 ml kg\(^{-1}\) min\(^{-1}\), respectively). Although the limits of agreement in EQTT are still relatively wide, this range is more likely to be acceptable compared to EQLE’G and EQMAS. Further, as illustrated by the present CVx indices, the traditional \(V_{O_2max}\) prediction can be up to 1.2 times as unreliable as the prediction of EQTT. ROC curve analysis indicated that both EQTT and EQLE’G were highly specific in discriminating individuals according to \(V_{O_2CR}\). However, sensitivity in the former was significantly increased compared to the latter model (81% v 23%).

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 \(V_{O_2max}\) 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 EC\(_V\) is exacerbated with increased body stature. It is tenable that various biomechanical complexities of shuttle running may account for this. The EQMAS model developed herein to predict \(V_{O_2max}\) 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 EQTT model represents a strict means of assessing CF. Three subjects with CF above the \(V_{O_2CR}\) in our cross validation sample were mis-screened as performing below the \(V_{O_2CR}\). 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

---

**Table 3** Results for ROC curve and McNemar $\chi^2$ 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>(S_p) (CI(_{95}%))</th>
<th>(S_n) (CI(_{95}%))</th>
<th>(PV) (CI(_{95}%))</th>
<th>(PV) (CI(_{95}%))</th>
<th>LR (CI(_{95}%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQLE’G</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>$\infty$</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 $\chi^2$ increased at: $*p<0.05; ^{**}p<0.001$.

---

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. 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 variability among different models of metabolic carts. 24

References