Enhancing the efficacy of the 20 m multistage shuttle run test

A D Flouris, G S Metsios, Y Koutedakis

Objective: Maximal oxygen uptake (\(\text{VO}_2\max\)) of 44 ml kg\(^{-1}\) min\(^{-1}\) is an accepted criterion (\(\text{VO}_2\text{CR}\)) below which health and fitness for young male adults may be compromised. New algorithms validated for \(\text{VO}_2\text{CR}\) screening using the 20 m multistage shuttle run test (20MST) were developed.

Methods: \(\text{VO}_2\max\) 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. \(\text{VO}_2\max\) predicted from the 20MST in 70 subjects was used for cross validation. Two equations predicting \(\text{VO}_2\max\) during 20MST (EQMST) and TT (EQTT) were developed.

Results: Significant energy cost variance (\(\text{EC}_\text{V}\)) was detected between TT and 20MST \((p<0.001)\), correlated significantly with subject height, and was a significant predictor of \(\text{VO}_2\max\) differences between TT and 20MST. The \(r^2\) of EQMST was 0.92 \((p<0.001)\). Predicted \(\text{VO}_2\max\) values from EQMST correlated with directly measured 20MST \(\text{VO}_2\max\) 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 \(\text{VO}_2\text{CR}\) was 0.37. McNemar \(\chi^2\) indicated significant differences in sensitivity \((p<0.001)\) and specificity \((p<0.05)\) between the original 20MST equation (EQLEG) and EQTT, regarding \(\text{VO}_2\text{CR}\) screening. Cohen’s \(\kappa\) demonstrated higher agreement with TT \(\text{VO}_2\max\) for EQTT \((p<0.001)\) than EQLEG \((p<0.05)\). TT \(\text{VO}_2\max\) correlated with the end result of both EQLEG and EQTT at \(r=0.75\) \((p<0.001)\). Unlike EQTT \((p<0.05)\), mean predicted \(\text{VO}_2\max\) from EQLEG was significantly higher compared to TT \(\text{VO}_2\max\) \((p<0.001)\).

Conclusion: These algorithms increase the efficacy of 20MST to accurately evaluate aspects of health and fitness.
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 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.⁴
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 measure-
ment, the gas analyser was calibrated with standard gases.
Exhaustion was confirmed when at least two of the follow-
ing 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 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.⁴

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 VO₂max value recorded by the stationary 
(Ymax 29, SensorMedics) and the portable (K4b², Cosmed) 
gas analyser (48.7 (SD 3.1) vs 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) on the original 20mMST prediction model 
(EQLEG²) was assessed via a simultaneous general linear 
model (GLM). This model aimed to predict VO₂max differ-
ences/errors between TT and EQ LE'G using mean EC 
as an independent variable. In addition, Pearson's correlation 
coefficients were used to detect linearity between EC and 
various anthropometrical characteristics.
For the calculation of the novel partition 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 
dependently distributed response variables.¹⁹ A GLM framework

### Table 1

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>r</th>
<th>χ²</th>
<th>SEE</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\nu} )</th>
<th>( \hat{\sigma} )</th>
<th>r</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>
<td></td>
</tr>
<tr>
<td>EQMST</td>
<td>20mMST</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>
<td></td>
</tr>
<tr>
<td>EQTT</td>
<td>EQMST</td>
<td>0.89</td>
<td>317.3*</td>
<td>1.94</td>
<td>47.0 (0.8)</td>
<td>46.9 (5.7)*</td>
<td>0.94*</td>
<td></td>
</tr>
<tr>
<td>EQTT</td>
<td>20mMST</td>
<td>0.89</td>
<td>317.3*</td>
<td>1.94</td>
<td>47.0 (0.8)</td>
<td>46.9 (5.7)*</td>
<td>0.94*</td>
<td></td>
</tr>
</tbody>
</table>

Values in parentheses are standard deviations (SD). Significant ANOVA between \( \nu_{VO_{2max}} \) and \( \nu_{VO_{2max}} \); \( r \) significant at \( p < 0.001 \). χ² significant at \( p < 0.001 \). 

* \( \nu_{VO_{2max}} \) actual values measured during testing; EQMST, calculated regression model to predict TT VO₂max; EQMST, calculated regression model to predict 20mMST VO₂max; DepVariable, dependent variable; \( \nu \) Variable, independent variable; \( \nu_{VO_{2max}} \), \( \nu_{VO_{2max}} \) measured during the 20mMST using the K4b² portable analyser; MAS, maximal attained speed (km h⁻¹); \( \nu_{VO_{2max}} \), 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; \( \hat{\nu}_{VO_{2max}} \), \( \hat{\nu}_{VO_{2max}} \) measured during the TT; χ², chi-square.
with GEE estimation was introduced to generate an equation (EQMST) predicting $\text{Vo}_{2\text{max}}$ 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 $\text{Vo}_{2\text{max}}$ (dependent variable) using the end result of EQMST as an independent variable. This procedure was employed to produce a 20mMST $\text{Vo}_{2\text{max}}$ 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 $\text{Vo}_{2\text{max}}$ (dependent variable) and MAS (independent variable). ANOVA and Pearson's correlation coefficients were used to detect possible bias between the mean actual and predicted $\text{Vo}_{2\text{max}}$ 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 LE\'G model. Correlation coefficients, ANOVA, 95% limits of agreement analyses ($\text{LIM}_{\text{AG}}$), and percent coefficients of variation ($\text{CV}_{V}$) were adopted to validate the two models according to established procedures. Ninety five percent confidence intervals ($\text{CI}_{95\%}$) 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. The demarcation point for $\text{Vo}_{2\text{CR}}$ was set at 44 ml kg$^{-1}$ min$^{-1}$ according to available guidelines.

Calculated sensitivity and specificity with corresponding $\text{CI}_{95\%}$ were used to determine the efficacy of the two equations in screening for $\text{Vo}_{2\text{CR}}$. Sensitivity ($S_{E}$) was defined as the proportion of subjects below the $\text{Vo}_{2\text{CR}}$ 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 $\text{Vo}_{2\text{CR}}$ 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 values. All statistical analyses were carried out with SPSS (version 11.5; SPSS, Chicago, IL) and 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 EQLE\'G

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

### Prediction of $\text{Vo}_{2\text{max}}$ 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 $\text{Vo}_{2\text{max}}$ 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, $\text{Vo}_{2\text{max}}$). 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.

Thus,

\[
\text{EQTT} \text{Vo}_{2\text{max}} = (\text{MAS} \times 6.65 - 35.8) \times 0.95 + 0.182
\]
Model cross validation

Means (SD) and comparisons of various performance indices from the TT and the 20mMST, as well as results for LIM_MG and CV_v appear in table 2. Preliminary analyses for LIM_MG revealed no positive relationship between the differences/errors (either (EQ_LEG-TT) or (EQ_TT-TT)) and the size of measurements (given by either (the mean of EQ_LEG and TT) or (mean of EQ_TT and TT)), respectively. Thus, the LIM_MG 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_LEG 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%; CI95%: 0.9%) were diagnosed below the Vo_2CR using the reference standard TT. In contrast, EQ_LEG and EQ_TT identified six and 29 subjects below the Vo_2CR, respectively. Cohen’s k statistic demonstrated significant agreement with the TT measurement for both the EQ_LEG (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 reliable tool available to assess cardiorespiratory fitness. 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.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>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})</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ_LEG</td>
<td>Sₚ (CI₉₅% )</td>
</tr>
<tr>
<td>EQ_TT</td>
<td>0.81 (0.15)**</td>
</tr>
</tbody>
</table>

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

CI₉₅, 95% confidence interval; EQ_LEG, original 20mMST prediction model; EQ_TT, calculated regression model to predict TT Vo_2max; LR, likelihood ratio; PV, negative predicted value; Sₚ, positive predicted value; S, sensitivity; S, specificity; \(\infty\), mathematical infinity.

indicated that both EQ_TT and EQ_LEG 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 15 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_MAX 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 EQ_MAX is exacerbated with increased body stature. It is tenable that various biomechanical complexities of shuttle running may account for this. The EQ_MAX 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.

www.bjsportmed.com
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 VO2CR.

---

References


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

www.bjsportmed.com
Enhancing the efficacy of the 20 m multistage shuttle run test

A D Flouris, G S Metsios and Y Koutedakis

doi: 10.1136/bjsm.2004.012500

Updated information and services can be found at:
http://bjsm.bmj.com/content/39/3/166

These include:

References
This article cites 19 articles, 5 of which you can access for free at:
http://bjsm.bmj.com/content/39/3/166#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/