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

Results: Significant energy cost variance (ECV) was detected between TT and 20MST (p<0.001), correlated significantly with subject height, and was a significant predictor of Vo2max differences between TT and 20MST. The r² of EQMST was 0.92 (p<0.001). Predicted Vo2max values from EQMST correlated with directly measured 20MST 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. McNemar χ² indicated significant differences in sensitivity (p<0.001) and specificity (p<0.05) between the original 20MST equation (EQLEG) and EQTT, regarding Vo2CR screening. Cohen’s κ demonstrated higher agreement with TT Vo2max for EQTT (p<0.001) than EQLEG (p<0.05). TT Vo2max correlated with the end result of both EQLEG and EQTT at r=0.75 (p<0.001). Unlike EQTT (p>0.05), mean predicted Vo2max from EQLEG was significantly higher compared to TT Vo2max (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,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

The 20 m multistage shuttle run test (20MST) represents an acceptable field assessment tool for CF, and has been repeatedly employed in different health4–5 and fitness6 settings. However, the popularity of the 20MST 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.4–11 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 20MST model in screening for CF remains unknown.

From a statistical standpoint, the limited accuracy of the 20MST may be attributed to the repeated measures design used in the original study.4 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 20MST model may be further compromised by the use of generally large and heterogeneous samples in the validation procedures.7 It has been established that severely biased linear relationships can occur owing to sample heterogeneity.7

From a physiological viewpoint, it could be argued that the curtailed ability of the original 20MST model to predict treadmill Vo2max values might be attributed to differences in the exercise modes utilised in the validation procedures (that is, shuttle running or forward running). Findings from recent investigations suggested that Vo2max 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 Vo2max and increased efficacy in screening for Vo2CR. The objective of the present investigation was to develop a new Vo2max 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 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

Field assessment of \( \text{V}O_2\text{max} \) (20mMST)

This test was conducted according to established procedures.\(^6\) In the model group a portable gas analyser (K4\(^b\) 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 \( \text{V}O_2\) curve. The EC in kcal was calculated for each individual minute/stage as the product of mean \( \text{V}O_2 \) (1 min\(^{-1}\)) by the corresponding caloric equivalent.\(^7\)

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 (EQLEG\(^c\)) was assessed via a simultaneous general linear model (GLM). This model aimed to predict \( \text{V}O_2\text{max} \) differences/errors between TT and EQLEG\(^c\) 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)\(^c\) 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.\(^9\)

A GLM framework 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 \( \text{V}O_2\) curve. EC in kcal was calculated for each individual minute/stage as the product of mean \( \text{V}O_2 \) (1 min\(^{-1}\)) by the corresponding caloric equivalent.\(^7\)

Data collection

Laboratory assessment of \( \text{V}O_2\text{max} \) (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\(^\circ\) accordingly in order to bring the subject to exhaustion in Data collection

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.

### Table 1

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>( r^2 )</th>
<th>( \chi^2 )</th>
<th>SEE</th>
<th>( \text{V}O_2\text{max} )</th>
<th>( \Delta \text{V}O_2\text{max} )</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQMAS MAS</td>
<td>( \text{t}V \text{O}_2\text{max} )</td>
<td>0.79</td>
<td>236.4(^*)</td>
<td>2.72</td>
<td>48.3 (5.9)(^\dagger)</td>
<td>46.9 (5.7)(^\dagger)</td>
<td>0.91(^*)</td>
</tr>
<tr>
<td>EQMST MAS</td>
<td>( \text{k}V \text{O}_2\text{max} )</td>
<td>0.92</td>
<td>456.2(^*)</td>
<td>1.70</td>
<td>49.2 (5.9)(^\dagger)</td>
<td>49.2 (5.9)(^\dagger)</td>
<td>0.96(^*)</td>
</tr>
<tr>
<td>EQTT EQMST</td>
<td>( \text{t}V \text{O}_2\text{max} )</td>
<td>0.89</td>
<td>317.3(^\dagger)</td>
<td>1.94</td>
<td>47.0 (0.8)</td>
<td>46.9 (5.7)(^\dagger)</td>
<td>0.94(^\dagger)</td>
</tr>
</tbody>
</table>

Values in parentheses are standard deviations (SD). Significant ANOVA between \( \text{V}O_2\text{max} \) and \( \text{V}O_2\text{max} \) of the model (GLM). This model aimed to predict \( \text{V}O_2\text{max} \) 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)\(^c\) 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.\(^9\) A GLM framework
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 (CVs) were adopted to validate the two models according to established procedures. Ninety-five percent confidence intervals (CIs) of the EQTT 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-1 min-1 according to available guidelines. 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 χ2 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 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 (r2 = 0.25, F1, 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 χ2<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.

\[ \text{[EQMAS]} \text{ Vo2max} = \text{MAS} \times 6.87 - 39.54 \]
\[ \text{[EQMST]} \text{ Vo2max} = \text{MAS} \times 6.65 - 35.8 \]
\[ \text{[EQTT]} \text{ Vo2max} = \text{EQMST} \times 0.95 + 0.182 \]

Thus,

\[ \text{[EQTT]} \text{ Vo2max} = (\text{MAS} \times 6.65 - 35.8) \times 0.95 + 0.182 \]

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

<table>
<thead>
<tr>
<th></th>
<th>Vo2max (SD)</th>
<th>LIMAG (error)</th>
<th>CVs</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></td>
</tr>
<tr>
<td>TT</td>
<td>47.3 (5.5)</td>
<td>-</td>
<td>16.3 (1.5)</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).

*p<0.001, ANOVA different from TT.

CVs, percent coefficient of variation for Vo2max; EQLEG, original 20mMST prediction model; EQTT, calculated regression model to predict TT Vo2max; LIMAG, error of measurement; calculated limits of agreement (error of measurement) for Vo2max; MAS, maximal attained speed (km h-1); Time, exercise time to exhaustion (min); TT, treadmill test; Vo2max, predicted values in the calculated models and assessed value in TT (mean (SD)).

Figure 2 ROC curve for EQLEG and EQTT regression models. The ROC curve is defined as the curve of the results from validation-group variance and EQLEG or EQTT regression models, respectively. Asterisks indicate the designated cutoff point of 44 ml kg-1 min-1.
developed a prediction equation which incorporates indirect
EQMAS (4.4 and 2.7 ml kg
-1
) subjects according to Vo
2
CR. To our knowledge, the present
significantly increased the efficacy of the 20mMST to discern
Results suggested that the developed prediction models
to enhance the efficacy of the 20mMST for CF screening.
The aim of the present investigation was to utilise the most
unreliable as the prediction of EQ TT. ROC curve analysis
be reported as absolute measurements. 21 Finally, unlike EQ TT
such as Vo
2
CR.

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
).

|               | S
ª (CI
º95%
) | S
º (CI
º95%
) | PV (CI
º95%
) | PV (CI
º95%
) | LR (CI
º95%
) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></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>∞</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.

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
2
CR. 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
2
CR.

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
indicated that both EQTT and EQLE´G were highly specific in
discriminating individuals according to Vo
2
CR. However,
sensitivity in the former was significantly increased com-
pared 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
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. 11 This phenomenon has been demon-
strated explicitly by Anderson 25 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 EQMAS prediction model, as compared to EQTT.
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 ECV 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 Vo
2
max during the 20mMST can
be used to calculate the oxygen transport demands of shuttle
running, when such information is required.

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 subjects according to VO2CR.

Authors’ affiliations
A D Flouris, Faculty of Applied Health Sciences, Brock University, St Catherine, 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

Competing interests: none declared

REFERENCES
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/