A comparison of methods of predicting maximum oxygen uptake

S Grant, K Corbett, A M Amjad†, J Wilson‡ and T Aitchison†

Department of Physical Education and Sports Science, †Department of Statistics, and ‡Institute of Physiology, University of Glasgow, Scotland, UK

The aim of this study was to compare the results from a Cooper walk run test, a multistage shuttle run test, and a submaximal cycle test with the direct measurement of maximum oxygen uptake on a treadmill. Three predictive tests of maximum oxygen uptake—linear extrapolation of heart rate of VO₂, collected from a submaximal cycle ergometer test (predictedₗₑ), the Cooper 12 min walk, run test, and a multi-stage progressive shuttle run test (MST)—were performed by 22 young healthy males (mean(s.d.) age 22.1(2.4) years; body mass 72.4(8.9 kg)) and the values compared to those obtained by direct measurement on a maximal treadmill test. All of the subjects were regular exercisers. The mean(s.d.) from the various tests in ml·kg⁻¹·min⁻¹ were as follows: treadmill 60.1(8.0), Cooper 60.6(10.3), MST 55.6(8.0), and predictedₗₑ 52.0(8.4).

The Cooper test had a correlation with the treadmill test of 0.92, while the MST and the predictedₗₑ had correlations of 0.86 and 0.76 respectively. Both the MST and predictedₗₑ showed systematic underprediction of the treadmill value. On average, the MST was 4.5 ml·kg⁻¹·min⁻¹ (s.e. 0.9) lower than the treadmill VO₂max while the predictedₗₑ was 7.8 ml·kg⁻¹·min⁻¹ (s.e. 1.4) lower than the treadmill VO₂max.

These findings indicate that, for the population assessed, the Cooper walk run test is the best predictor of VO₂max among the three tests. (Br J Sports Med 1995; 29: 147–152)

Keywords: maximum oxygen uptake; prediction of maximum oxygen uptake; field tests

Aerobic power or maximum oxygen uptake (VO₂max), the maximum rate at which oxygen can be consumed, is a commonly used measure of cardiorespiratory (or aerobic) fitness. Direct measurement of VO₂max is expensive both in terms of time and cost of precise gas analysis. Thus, various predictive tests have been devised to evaluate aerobic fitness. These include performance related measures, for example walking or running for a given time,¹ performing a multistage progressive shuttle test (MST) with increases in speed every minute² or the measurement of heart rate during given workloads and extrapolating to a predicted maximum heart rate to determine an estimate of VO₂max.³ The Cooper and MST can cater for large numbers of subjects simultaneously but have the
disadvantage that they are maximal and thus create a potential health risk. In addition, these tests require a high degree of motivation to elicit a truly representative maximal effort. A submaximal test on a cycle ergometer does not rely on subject motivation but is time consuming.

The validity of the Cooper, MST, and linear extrapolation methods has been assessed separately but no study has compared all three tests within the same population. A limitation of some of the previous studies is that only the correlation between the predicted score and direct measurement has been given. Correlations do not provide the whole picture. At the time of writing this paper, the Physiology Section of the British Association of Sport and Exercise Sciences is compiling a guide for field testing. The results from this study can contribute to the current body of knowledge relating to field testing. A study which compared the Cooper, MST, and linear extrapolation methods in a wide ranging young male population would provide information on the applicability of these tests.

The present study was designed to compare three predictive tests commonly used in the Department of Physical Education and Sports Science at the University of Glasgow—Cooper walk/run test, multistage shuttle run test (MST), and linear extrapolation with the direct measurement of maximum oxygen uptake on a treadmill.

Methods

Subjects

Twenty-two healthy males from the university community volunteered to participate in the study. All subjects gave informed consent and the study was approved by the local ethics committee. The subjects were involved in a variety of sports, mostly of an endurance nature, but at differing levels of ability and training status.

Design

The subjects performed four tests on three separate days. On day one subjects performed a submaximal cycle ergometer test followed 30 min later by a direct measurement on a treadmill. These tests were carried out first as they were laboratory based. Thus, it was possible to monitor the ECG of each subject before he undertook the other maximum tests. The Cooper test and the MST were carried out on separate days with the order of testing randomized. An interval of at least two days was allowed between test days.

Submaximal cycle ergometer test

This predicted protocol was a modification of the YMCAs of American protocols. The initial workload selected was based on the responses to the warm-up, body mass and training status of the athlete, to elicit a steady state heart rate around 100 beats min⁻¹. The other two workloads were selected to elicit steady state heart rates between 120–140 beats min⁻¹ and 150–170 beats min⁻¹.

After a 3 min warm-up the test consisted of three continuous submaximal exercise bouts of 5 min duration on a Monark cycle ergometer (818 Monark Crescent, Varberg, Sweden). Cycling was performed at 60 revs per minute. Heart rates were measured during the fourth and fifth minute of each stage using a three lead ECG S & W Medical cardio aid (type CB600M) to measure heart rate. Expired air was collected in the final minute of each stage. Subjects were connected to a Douglas bag by a rubber mouthpiece, Hans Rudolph 2700 valve, and standard tubing. Gases were analysed using a PK Morgan CO₂ analyser, type 801A, and a Taylor Servomex O₂ analyser, type 570A. Volume was measured using a Parkinson Cowan volume meter.

All analysers were calibrated before each test with known gas concentrations.

To allow for comparison of various methods of extrapolation, VO₂max was predicted using the linear extrapolation method as follows: (a) the measured submaximal VO₂ and heart rate were extrapolated to the measured maximum heart rate; (b) the measured submaximal VO₂ and heart rate were extrapolated to the predicted maximum heart rate (220 minus age); and (c) the estimated VO₂ and measured heart rate were extrapolated to the predicted maximum heart rate (220 minus age).

Maximal treadmill test

Thirty minutes elapsed between the submaximal test and the maximal treadmill test. During this time the subjects were given a warm-up on the treadmill and familiarization with the treadmill, nose clip, and mouthpiece was provided. The maximal treadmill test protocol was based upon that recommended by the British Association of Sports Science. Heart rate and oxygen consumption were measured as for the submaximal cycle ergometer test.

Multistage progressive shuttle run test (MST)

This test was performed on a hard synthetic surface located in an indoor sports arena. Subjects ran between two lines 20 m apart in time with a sound signal which was emitted from an audio cassette. The frequency of the sound signals increased every minute. The test was terminated when the subject was no longer able to follow the set pace and did not reach the targeted line on three consecutive occasions. The level attained and the number of shuttles at that level allowed a prediction of VO₂max to be made according to Ramsbottom et al. When possible, subjects ran in pairs in order to provide a source of motivation.

Cooper 12 min walk/run test

This test was also performed in the indoor arena, on a 234 m jogging track. The warm-up included both aerobic activities and flexibility exercises. The subjects ran in pairs in order to provide a source of motivation. A prediction of VO₂max from the distance covered at the end of the 12 min period was obtained using the table.
provided by Cooper. The subjects who ran more than 2 miles had VO_{2\text{max}} predicted by applying the distance run to the Cooper regression equation.

**Statistical analysis**

The basis of the statistical analyses were regressions of each test (that is, Cooper, MST, and predicted_{\text{LE}}) separately on the treadmill value (that is, the predictive test was the response variable and the treadmill value the explanatory variable). Separate standard I test of the slopes being equal to 1 and the intercepts being equal to 0 (that is, no bias) were carried out for each of these regressions.

In the illustrations of the effect of calibration, a measurement value of 60 ml·kg\(^{-1}\)·min\(^{-1}\) was selected for each test separately to 'predict' the corresponding treadmill value. The simple approximate formula for the 95% confidence interval was used:

\[
\frac{60-A}{B} \pm 2S
\]

Where A, B and S are the estimated intercept, slope and s.d. of the regression of the test on the treadmill value.

**Results**

The mean (s.d.) age of the subjects was 22.1(2.4) (19–29 years) and the means(s.d.) body mass was 72.4(8.0) kg (56.4–92.6 kg).

The mean(s.d.) from the various tests in ml·kg\(^{-1}\)·min\(^{-1}\) are given in Table 1. The linear extrapolation method used the measured maximum heart rate.

<table>
<thead>
<tr>
<th>Treadmill (\text{VO}_{2\text{max}})</th>
<th>Cooper (\text{VO}_{2\text{max}})</th>
<th>MST (\text{VO}_{2\text{max}})</th>
<th>Predicted_{\text{LE}} (\text{VO}_{2\text{max}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>60.1(8.0) ml·kg(^{-1})·min(^{-1})</td>
<td>60.6(10.3) ml·kg(^{-1})·min(^{-1})</td>
<td>56.68(0.0) ml·kg(^{-1})·min(^{-1})</td>
<td>52.08(4.0) ml·kg(^{-1})·min(^{-1})</td>
</tr>
</tbody>
</table>

**Table 1. Summary statistics of VO_{2\text{max}} scores derived from various tests (means(s.d.))**

As none of the slopes of the regression of each of the various predicted values on measured treadmill \(\text{VO}_{2\text{max}}\) was significantly different from 1, then any bias on the predicted values can be assumed constant across the range of measured treadmill \(\text{VO}_{2\text{max}}\) measured (that is, 45–75 ml·kg\(^{-1}\)·min\(^{-1}\)). Accordingly, estimates of such systematic biases, as well as correlation coefficients, for each predicted value are given in Table 2. The bias is the respective test \(\text{VO}_{2\text{max}}\) minus the treadmill \(\text{VO}_{2\text{max}}\) and both the MST and predicted_{\text{LE}} tests showed a significant systemic underestimation. In addition, all of the predicted values correlated reasonably well with the treadmill \(\text{VO}_{2\text{max}}\) and of these, the Cooper test gave the highest correlation.

**Figure 1. The relation between the Cooper and treadmill tests**

**Figure 2. The relation between the MST and treadmill tests**

**Figure 3. The relation between the predicted_{\text{LE}} and treadmill tests**
Prediction of maximum oxygen uptake: S Grant et al.

Table 2. Summary of the performance of each predicted value with the measured treadmill VO$_{2\text{max}}$ (ml·kg$^{-1}$·min$^{-1}$). (a) Heart rate extrapolated to measured maximum heart rate in treadmill test using measured submaximal steady state VO$_{2}$; (b) Heart rate extrapolated to predicted maximum heart rate (220 minus age) using measured submaximal steady state VO$_{2}$; (c) Heart rate extrapolated to predicted maximum heart rate (220 minus age) using estimated VO$_{2}$ for given workloads

<table>
<thead>
<tr>
<th>Test</th>
<th>Correlation with treadmill VO$_{2\text{max}}$</th>
<th>Bias (s.e.) † to VO$_{2\text{max}}$</th>
<th>Significance of bias (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooper</td>
<td>0.92</td>
<td>+0.5(0.9)</td>
<td>0.6</td>
</tr>
<tr>
<td>MST</td>
<td>0.86</td>
<td>-4.5(0.9)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Predicted L/E (a)</td>
<td>0.76</td>
<td>-7.8(1.4)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Predicted L/E (b)</td>
<td>0.74</td>
<td>-8.2(1.3)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Predicted L/E (c)</td>
<td>0.76</td>
<td>-8.4(1.6)</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

†This is the estimated average difference of the test and the treadmill VO$_{2\text{max}}$ (i.e. test minus treadmill VO$_{2\text{max}}$)

Table 3. Illustration of the predictive performance for each test at a target VO$_{2\text{max}}$ of 60 ml·kg$^{-1}$·min$^{-1}$

<table>
<thead>
<tr>
<th>Workload</th>
<th>Percent difference measured</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mean ± s.d.)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5.2(18.0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.3(11.1)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.9(5.7)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Percent* differences between measured and predicted submaximal VO$_{2}$

<table>
<thead>
<tr>
<th>Workload</th>
<th>Percent difference measured (mean ± s.d.)</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mean ± s.d.)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5.2(18.0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.3(11.1)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.9(5.7)</td>
<td></td>
</tr>
</tbody>
</table>

*Expressed as [(measured−predicted)/measured] × 100

For each of the three tests a value of 60 ml·kg$^{-1}$·min$^{-1}$ was assumed for that particular test and the corresponding treadmill value calculated from the appropriate regression (see Table 3).

Both MST and predicted, L/E predictions of the treadmill VO$_{2\text{max}}$ are shifted significantly to the right of the target value of 60 ml·kg$^{-1}$·min$^{-1}$, and both have a large range of prediction. The Cooper test is best as it at least straddles the target value of 60 ml·kg$^{-1}$·min$^{-1}$.

Across the full range of measurement, 45–75 ml·kg$^{-1}$·min$^{-1}$, there was consistent bias and large variability for the MST and predicted L/E tests but markedly less so for the Cooper test.

A comparison of the measured and predicted submaximal VO$_{2}$ for the cycle ergometer was made. The percentage differences between measured and predicted submaximal VO$_{2}$ are given in Table 4. There was no significant difference between measured and predicted VO$_{2}$.

Discussion

The direct treadmill based measurement of VO$_{2\text{max}}$ of 60.1(8.0) ml·kg$^{-1}$·min$^{-1}$ represents a relatively fit group of individuals, similar to that reported by McCutcheon et al., who found values of 61.5(7.1) ml·kg$^{-1}$·min$^{-1}$ in a group of recreational sportsmen. However, it should be noted that the sample group in this study was not a homogeneous group. For example, VO$_{2\text{max}}$ ranged from 44.9 ml·kg$^{-1}$·min$^{-1}$ to 75.3 ml·kg$^{-1}$·min$^{-1}$.

Cooper 12 min walk/run test

For the Cooper test the correlation coefficient of 0.92 relative to the treadmill VO$_{2\text{max}}$ in this study compares favourably with the corresponding values of 0.90 and 0.84 in the studies of Cooper and McCutcheon et al., respectively. Other studies have reported a range of correlation coefficients (0.7–0.9) in other categories of subjects including 14- to 15-year-old boys, young female athletes, and 12-year-old boys and girls.

The age variation, status and number of subjects, and the homogeneity or otherwise of the group must be considered when comparison with other groups is made. In the original study, Cooper assessed 115 US Air Force personnel with a mean age of 22(range 17–54 years). The VO$_{2\text{max}}$ of the group ranged from 31–59 ml·kg$^{-1}$·min$^{-1}$, none of the Cooper subjects could complete more than two miles within 12 min compared to 50% in this study. Anderson stressed that high correlations are to be expected in groups with a large age variation as in the Cooper study. The wide range of VO$_{2\text{max}}$ values in this study could contribute to the high correlation. McDade highlighted the potential problem of comparing predictive tests with direct measurement of VO$_{2\text{max}}$ in a homogeneous group. They stated that when the original Cooper data were confined to college men the correlation was lowered from 0.90 to 0.59.

The correlations between this study and the Cooper study are similar. However, correlations do not give the whole picture. A formal hypothesis test shows that there was no systematic bias for the Cooper test in this study. However, McCutcheon et al. found that the Cooper test gave a systematic underestimation of 4 ml·kg$^{-1}$·min$^{-1}$ in their male and female subjects (mean age of 25 years).

Some of the subjects ran over two miles. For these subjects it was assumed that the Cooper regression equation applied. This assumption was considered to be acceptable, as the basis of the Cooper test is that the average running speed is dependent on oxygen uptake.

Multistage shuttle run test

It might have been expected that the MST and Cooper tests would provide similar predictions as they both involve running. The MST showed a slightly lower correlation of 0.86 between the directly determined VO$_{2\text{max}}$ and the predicted VO$_{2\text{max}}$ from the MST. These results are similar to the correlations between the MST and direct measurement of VO$_{2\text{max}}$ reported by Ramsbottom et al. (0.83), Paliczka et al. (0.93), and Leger and Lambert (0.91). However, the systematic underestim-
Prediction of maximum oxygen uptake: S Grant et al.

consumption relationship is linear; and (c) the mechanical efficiency of all subjects is the same.

The heart rates from the three submaximal points were extrapolated to the measured maximum heart rate in the treadmill test. There was no significant difference on average between the measured maximum heart rate and predicted maximum heart rate (220 minus age). Thus it is not surprising that extrapolation to the age predicted heart rate (220 minus age) does not make any meaningful difference to the prediction (a correlation of 0.76 based on the measured maximum heart rate against 0.74 based on predicted maximum heart rate and the bias remains similar).

Individual subject correlations between heart rate and VO₂ on the three submaximal workloads were calculated and had a median value of 0.99, with a range of 0.91–1.00. When such correlations were calculated on the basis of the three submaximal and the maximum values, the median correlation was again 0.99, with a range of 0.93–1.00, and indicated a very strong linear relationship across all workloads for heart rate of VO₂.

There was no evidence of any significant difference between the predicted and measured VO₂ at the three submaximal workloads. There was a strong indication, particularly at the first and second workloads, of a wide range in oxygen cost.

Further, the correlations based on the measured and predicted submaximal VO₂ respectively with the direct measurement of VO₂max were very similar (0.74 and 0.76 respectively). Overall, therefore, there was no indication that the measurement of submaximal VO₂ or indeed maximum heart rate, made any significant contribution to VO₂max prediction.

The underprediction found in this study has been recorded by other researchers who have estimated VO₂max from an extrapolation to maximum heart rate using measured VO₂ and heart rate.15-17 Zwirner et al18 and Louhevaara et al19 both reported an overestimate using heart rate/workload points from two and three submaximal loads respectively to predicted maximum heart rates.

It is plausible that the reason for the low predictive capacity of the predictedLE method can be partly explained by the fact that a running VO₂max is known to be higher than a VO₂max measurement using cycling (except if cyclists are used. A cycling VO₂max is around 10% lower than a running VO₂max.20-22 Thus, it would be expected that the treadmill VO₂max value of 60 ml·kg⁻¹·min⁻¹ would be around 54 ml·kg⁻¹·min⁻¹ on a cycle. This value of 54 ml·kg⁻¹·min⁻¹ is close to the predictedLE value of 52.0 ml·kg⁻¹·min⁻¹. It should be pointed out that other studies have reported an underprediction of VO₂max using a cycle test for prediction and a treadmill test for direct measurement of VO₂max.15-17 19 23 24 Studies which have used a cycle test for prediction and a treadmill test for direct measurement have also found an underprediction of VO₂max16-25

Summary

The predictive tests used in this study yielded different VO₂max scores. Thus it is inappropriate to compare...
Prediction of maximum oxygen uptake: S Grant et al.

$\text{VO}_{2\max}$ scores obtained from different predictive tests. The findings from this study indicate that, among the three tests for the population assessed, the Cooper walk/run test is the best predictor of $\text{VO}_{2\max}$. Predictions from the MST and the submaximal ergometer test provided a less accurate measure of maximum oxygen uptake compared to the Cooper walk/run test and gave a systematic underestimate of $\text{VO}_{2\max}$.

References
15 Davies CTM. Limitation to the prediction of maximum oxygen intake from cardiac frequency measurements. J Appl Physiology 1968; 24: 700-6.